The glucocorticoid toxicity index: measuring change in glucocorticoid toxicity over time

Stone, J. H., McDowell, P. J., Jayne, D. R. W., Merkel, P. A., Robson, J., Patel, N. J., Zhang, Y., Yue, H., Bekker, P., & Heaney, L. G. (2022). The glucocorticoid toxicity index: measuring change in glucocorticoid toxicity over time. *Seminars in Arthritis and Rheumatism*, *55*, Article 152010. https://doi.org/10.1016/j.semarthrit.2022.152010

Published in:
Seminars in Arthritis and Rheumatism

Document Version:
Publisher's PDF, also known as Version of record

Seminars in Arthritis and Rheumatism 55 (2022) 152010

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

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

The glucocorticoid toxicity index: Measuring change in glucocorticoid toxicity over time

John H. Stone*, P. Jane McDowell, David R.W. Jayne, Peter A. Merkel, Joanna Robson, Naomi J. Patel, Yuqing Zhang, Huibin Yue, Pirow Bekker, Liam G. Heaney

Harvard Medical School, Massachusetts General Hospital, Boston, MA, United States

ARTICLE INFO

Keywords:
Glucocorticoid toxicity index
Minimum clinically important difference
ANCA-associated vasculitis
Systemic lupus erythematosus
Asthma
Lupus nephritis
Giant cell arteritis
Polymyalgia rheumatica
Pemphigus vulgaris
Bullous pemphigoid
IgG4-related disease
Sarcoidosis
Behçet's disease
Myasthenia gravis
Idiopathic inflammatory myopathies
Kawasaki's disease
Cumulative worsening score
Aggregate improvement score
Bone mineral density
Glucocorticoid myopathy

ABSTRACT

Glucocorticoids (GCs) have been the cornerstone of treating dozens of inflammatory conditions for more than seven decades. GC toxicity is ubiquitous in both clinical trials and clinical practice, and toxicities associated with GC use are central to the experience of most patients being treated for immune-mediated conditions. These conditions span the full range of medical specialties, including rheumatology, nephrology, gastroenterology, neurology, pulmonology, ophthalmology, and others. One of the goals of novel therapies for inflammatory disease must be to diminish the effects of GC toxicity in clinically important ways, thereby differentiating these new treatments from existing approaches. Despite the importance of glucocorticoids in the treatment of inflammatory disease for more than 70 years, no reliable means of calculating the degree to which GC toxicity has worsened or improved over the course of treatment has been available. The Glucocorticoid Toxicity Index (GTI), developed by an international group of subspecialty physician experts as a clinician-facing clinical trials outcome measure, is a standardized, validated measure of the phenomenon known as GC toxicity. The purpose of the instrument is to measure change in GC toxicity between two points in time: for example, between the baseline visit and the time of the primary efficacy outcome assessment. The instrument is designed to quantify both worsening and improvement in GC toxicity. The GTI has been validated in both real-world experiences and clinical trials, including a phase 3, label-enabling trial in ANCA associated vasculitis. This article reviews the history and rationale for the development of the GTI, describes key data from validation studies, considers the minimum clinically important difference, and provides instructions for use of the instrument.

Introduction

Glucocorticoids (GCs) have been the cornerstone of treating dozens of inflammatory conditions for more than seven decades. Although these medications suppress many types of inflammation swiftly, the toxicities associated with GC use are central to the experience of patients with most inflammatory diseases. GC toxicity is ubiquitous in both clinical trials and clinical practice [1,2]. Despite the phenomenon of GC toxicity, there has been no practical, reliable means of calculating the degree to which GC toxicity worsened or improved. The potential impact of a new medication on reducing GC toxicity could be evaluated only crudely (e.g., by counting numbers of adverse events presumed to be related to GCs).

One of the goals of novel therapies for inflammatory disease is to diminish the effects of GC toxicity in clinically important ways, thereby differentiating new treatment approaches from current standards of care and justifying their value to society. A simple search of Clinicaltrials.gov on the terms "glucocorticoids" yields 6,689 clinical trials between the years 2001 and 2021. Not a single one employed a validated instrument to assess the impact of their treatment approach on GC toxicity – if indeed the demonstration of GC toxicity reduction was a stated priority at all. In short, GC toxicity has hung heavily over the treatment landscape and been an accepted part of the patient experience for those with inflammatory disease for too long. It is time to acknowledge it, to measure it, and to change it.

The Glucocorticoid Toxicity Index (GTI), developed by an international group of subspecialty physician experts as a clinical outcome assessment, is the first standardized, validated measure of GC

Abbreviations: GC, Glucocorticoid; BMI, Body mass index; GTI, Glucocorticoid Toxicity Index; CWS, Cumulative worsening score; AIS, Aggregate Improvement Score; MCID, Minimum clinically important difference; ANCA, anti-neutrophil cytoplasmic antibody

* Corresponding author.
E-mail address: jhstone@mgh.harvard.edu (J.H. Stone).

https://doi.org/10.1016/j.semarthrit.2022.152010

toxicity. The instrument's purpose is to measure **change** in glucocorticoid (GC) toxicity between two points in time: for example, between the baseline visit and the time of the primary efficacy outcome assessment. The instrument is designed to quantify both worsening and improvement in GC toxicity. This article reviews the history and rationale for the development of the GTI, outlines the ways in which the instrument has been validated and used in both clinical trials and real-world investigations, and reviews the general instructions for use. Readers are encouraged to explore the on-line **Appendix**, which provides additional details the GTI.

Challenges in the assessment of GC toxicity

Depending upon how one classifies and counts the individual toxicities, nearly 70 toxicities caused by GCs are commonly recognized (**Supplement** Table 1). The complexity of GC toxicity posed a high hurdle to the development of an instrument to measure it. Other challenges to the measurement of GC toxicity are summarized in Table 1.

Table 1
Challenges in the Measurement of Glucocorticoid Toxicity.

Challenge	Comment
Measurement-associated challenges	
Capturing all GC toxicities in a concise instrument	Many GC-related adverse events have been described. The development of a concise and accurate instrument that can be used in a clinical trial setting represents a significant challenge.
Variability in time course	Some GC toxicities occur acutely (within hours to days), others subacutely (weeks to months), and others chronically (many months to years). Chronic toxicity may not be captured during a typical clinical trial.
GC use before period of interest	Many patients in clinical trials have used GCs before study entry. Therefore, GC-associated morbidity occurring during the trial may reflect pre-trial GC treatment.
Background rate and attribution	Some toxicities can occur in the absence of GC treatment (e.g., hypertension). Accounting for the background rate of such adverse events and correctly attributing the toxicity to GC use can introduce bias into measurement of GC toxicity.
Degrees of toxicity	GC toxicities range from mild and short-lived to severe and life-altering. Capturing the severity of adverse events included in a measurement tool is a major challenge.
Disease-associated challenges	
Variability in patient populations	GCs are used broadly in medical practice, but patient populations vary significantly according to disease, with variation in age and sex that impact susceptibility to GC toxicity.
Co-morbidities	Patients may have co-morbidities that precede GC treatment (e.g., obesity, hypertension, etc.), which vary among patient populations and disease states.
GC utilization	The doses of GCs required to achieve disease control vary widely according to disease. Furthermore, practice patterns of GC use vary across medical specialties and from country to country.
Disease-associated complications	Disease features can mimic GC complications or have a synergistic effect with GCs in causing the toxicity. As examples, inflammation in systemic lupus erythematosus and rheumatoid arthritis contributes to atherogenesis leading to cardiovascular and cerebrovascular disease.
Concomitant therapies	Medications used frequently with GCs may have a synergistic effect with certain GC toxicities. As an example, immunomodulating agents may heighten the risk of infection, an effect that can be difficult to distinguish from that of GCs.

History of GTI development

The GTI was developed by a group of 17 international physician-investigators. The group represented eleven subspecialties (**Supplemental Table 2**). The full Scientific Committee was comprised of participants from rheumatology, pediatric rheumatology, pulmonology, nephrology, neurology, ophthalmology, dermatology, infectious disease, and psychiatry. Ten investigators were from the United States, and 9 from Europe, Canada, Australia, or New Zealand. The investigators were invited to participate because of their expertise within areas that represent a large portion of the inflammatory disease spectrum affecting adults. Two investigators had also participated in the development of European League Against Rheumatism (EULAR) recommendations for GC toxicity monitoring [3].

Work began in 2015, with investigator recruitment and the first conference calls. The overall approach involved well-established group consensus methods and multi-criteria decision analysis. The effort capitalized on the 1000Minds software platform (Dunedin, New Zealand), which facilitates multi-criteria decision analysis among groups. This methodology of group consensus has been utilized widely in rheumatology for the development of weighted classification criteria for systemic sclerosis, gout, systemic lupus erythematosus, calcium pyrophosphate dihydrate deposition disease, IgG4-related disease, and other conditions [4–10].

Intermediate aims on the path to the development of a fully-validated GTI were: 1) to define the optimal setting(s) for GTI use; 2) to define toxicity items for inclusion in the GTI; 3) to assign relative weights to those items; and, 4) to begin the process of validating the GTI using paper cases. The investigators established milestones for development of the GTI at the start of the process (**Supplemental Table 3**). Before the full, weighted GTI prototype was complete, the development process included ten one-hour conference calls conducted over a period of three months (July-September, 2015); work between the calls by the investigators who had special expertise within individual domains of GC toxicity; and one day-long, face-to-face meeting (October, 2015).

The GTI investigators agreed that a principal use of the GTI is in the context of prospective, randomized, controlled clinical trials in which GCs are part of the treatment protocol. The instrument can also be used in a variety of other types of clinical studies and in practice, but randomization and blinding serve the critical purposes of controlling for the background rate of adverse events and prior GC treatment. The procedure of randomization also eliminates the requirement for study investigators to attribute an adverse event observed to either GCs or to a confounding factor – a task that is often difficult and prone to bias. Subsequent studies have also confirmed the utility of the GTI in non-randomized clinical trials and real-world studies [11,12].

The initial validation phase was completed in early 2017. A paper describing the GTI development process and the investigators' initial validation efforts was published [13]. Improvements in this instrument clarified the approaches to scoring the GTI [11]. The GTI was

adopted as the key secondary outcome measure of treatment efficacy in a randomized, double-blind, placebo-controlled trial of avacopan, a first-in-class inhibitor of the C5A receptor, in ANCA-associated vasculitis [14,15]. Between 2017 and 2020, the GTI digital platform was developed. This web-based app that can be used directly for data entry or integrated into a larger clinical trial database [11]. The digital platform is 21 CFR Part 11-enabled, addressing certain regulatory standards of the U.S. Food & Drug Administration.

More than 45 studies have now licensed the GTI to date, including 12 phase 3 clinical trials. The indications under study with this instrument now include ANCA-associated vasculitis, asthma, systemic lupus erythematosus, lupus nephritis, giant cell arteritis, polymyalgia rheumatica, pemphigus vulgaris, bullous pemphigoid, congenital adrenal hyperplasia, IgG4-related disease, sarcoidosis, Behçet's disease, myasthenia gravis, idiopathic inflammatory myopathies, Kawasaki's disease, and juvenile idiopathic arthritis (the last two employing the pediatric version of the instrument, known as the pGTI) (**Supplemental Table 4**).

Selection of the GTI domains

Candidate items for inclusion in the Core GTI were divided into clinical domains (e.g., glucose metabolism, infection, and neuropsychiatric effects). Experts were assigned to consider each domain based on their area of expertise. The experts' initial goal was to create domains that were independent of each other and consisted of three to five items of increasing toxicity. Each domain was then presented to the full Scientific Committee. Items were selected for inclusion by nominal group technique.

Item inclusion for the Core GTI was based on four main principles: 1) likelihood of occurrence greater than 5% over the course of 6 months to 3 years (i.e., the toxicity is common); 2) importance to both providers and patients [16]; 3) item independence; and, 4) the dynamic nature of the specific toxicity (i.e., the likelihood of change over time – either worsening or improvement – with varying GC dosing). Toxicities were included in the GTI only if they were more likely to be due to the effect of GC therapy than to the disease itself or to the background rate of that particular event. In addition, the investigators excluded toxicities that were difficult to separate from either concurrent co-morbidities or effects of the underlying disease. For example, toxicities such as atherosclerosis, myocardial infarction, and stroke were not included in the GTI because all of those toxicities are confounded frequently by either co-morbid conditions (e.g., smoking) or the effects of the disease under treatment (e.g., systemic lupus erythematosus). Toxicities that could not be evaluated objectively without a requirement for invasive testing, e.g., gastritis, were also excluded.

Derivation of the weights of the GTI items

After Composite GTI toxicity items had been selected, organized by domain, assessed for content and face validity, and ranked in order of increasing toxicity within each domain, relative weights for each item were derived at the face-to-face meeting [13].

Multi-criteria decision analysis was used in the following approach. The investigators were asked to assess the relative toxicity of items by selecting the higher toxicity level from a paired patient scenario differing in two toxicity items. **Supplemental Figure 1** shows examples of the types of comparisons the Scientific Committee considered in deciding which combinations of clinical complications of GC use constituted the higher degree of GC toxicity. Based on the number of domains and toxicity items, there were 62,208 possible paired patient scenarios differing in two toxicity items. The investigators completed 103 scenarios, reaching agreement on all combinations. The remaining 62,105 scenarios were then resolved implicitly by the 1000Minds software using the transitivity principle, and relative weights for each of the GTI toxicity items were derived [11,14].

Orientation of the GTI: a clinician-facing instrument reliant upon patient input

The GTI is designed as a clinician-facing instrument that relies on patient input. It is therefore a clinician-reported outcome as opposed to a patient-reported outcome. Most GTI domains are based upon simple clinical measurements (e.g., body mass index, blood pressure) or laboratory tests (e.g., hemoglobin A1c) (**Table 2**). Several domains require direct patient interaction and careful consideration of the impact of GC toxicity on their lives (see Appendix Sections 5 and 6 as examples). Even for these domains, however, data capture is designed to be simple. Examples of the direct clinician-patient interaction incorporated into the GTI are shown in **Table 3**.

Scoring approach to the GTI

The analytic approach to the GTI derives two scores calculated from the same data: the **Cumulative Worsening Score (CWS)** and the **Aggregate Improvement Score (AIS)**. The two scores are outlined below. Additional information about weighting, scoring, and analyzing the GTI is found in **Appendix Section 3**.

Cumulative worsening score

The CWS is designed to assess the total GC toxicity that has occurred since baseline, regardless of whether the toxicity is permanent or transient. New toxicities that occur are added. Toxicities that resolve on follow-up are not removed from the CWS. Thus, the CWS serves as a lasting record of any GC toxicity that has occurred following entry of the patient into the trial, study, or practice. An investigational agent that is effective in lowering GC toxicity over time in a clinical trial will lead to a lower CWS compared to standard treatment.

Aggregate improvement score

Many patients entering a clinical trial or starting a new "steroid-sparing" drug in clinical practice have received GC before. Therefore, they often have substantial baseline GC toxicity, even before treatment with the new agent begins. The AIS is important in establishing that the new therapy is effective at lowering any baseline GC toxicity over the course of follow-up. As noted, the GTI domains were selected in part because they represent GC toxicities that are anticipated to change over time: to increase if the underlying disease requires additional GC treatment, but also to decrease if a new therapy permits less GC exposure. With the AIS, toxicities such as insomnia and acne are removed if they resolve during follow-up. If an

Table 2
GTI Domains Assessed Through Simple Clinical Measures or Laboratory Tests.

Domain	Description
Body Mass Index	Height, weight
Blood Pressure*	Systolic and diastolic blood pressures
Glucose Metabolism*	Hemoglobin A1c
Lipid Metabolism*	Low-density lipoprotein
Bone Mineral Density	Dual X-ray absorptiometry (DEXA scan)
Glucocorticoid Myopathy	Physical examination testing for proximal muscle weakness
Skin Toxicity	Physical examination
Neuropsychiatric Effects	Patient interview
Infection	Adverse events reporting

* Increases and decreases in medications for hypertension, glucose metabolism, and hyperlipidemia are considered in the GTI scoring algorithm.

Table 3
Direct Clinician-Patient Interaction Is Required For Three Domains.

Domain	Description
Myopathy	Assessment of patients' muscle strength and its impact on day-to-day function. This is accomplished by simple tests of strength in the shoulders and legs.
Skin	Examination for cutaneous findings of glucocorticoid toxicity (acne, hirsutism, easy bruising, and striae), with assessments that consider impact on activities of daily living.
Neuropsychiatric	Determination through discussion with the patient of the degree to which patients' lives and day-to-day functioning are impacted by, for example, insomnia, depression, steroid-induced violence.

investigational agent is effective at lowering GC toxicity over time, the AIS will be lower over the course of the trial in the investigational treatment arm.

Worsening and improvement in GC toxicity are weighted equally

The AIS scores improvement in GC toxicity using the same weights for worsening of GC toxicity. The principle is that an improvement in GC toxicity is assigned the same absolute weight as a worsening of GC toxicity of the same magnitude. As an example, an increase in the body mass index (BMI) more than 5 BMI units to a BMI of greater than 25 is associated with an increase in the GTI score of +36 points. In contrast, a decrease in BMI of more than 5 BMI units towards a normal BMI is associated with an improvement in the score of -36 points.

In summary, for each patient, the GTI measures the change in GC toxicity in a nuanced manner for each GTI interval. The totals of the individual patient scores are calculated within treatment groups to provide summary scores for each group within a trial. Accurate GTI scores are calculated easily with the GTI digital platform, but investigators can also calculate scores themselves.

Validation

The process of validating the GTI extended over several years. Following item inclusion and definition development in the initial development phase from 2015 to 2017, the GTI was evaluated by the Scientific Committee for clarity, format, visual design, organization, and navigation. Revisions of each domain were performed in an iterative manner. Each expert was asked to submit four cases of real patients who had experienced changes in GC toxicity during treatment. From those 68 cases, 15 were selected because they reflected the spectrum of potential GC toxicity. Each expert was then asked to score the 15 cases using the draft GTI. Discussion of the exercise results by the Scientific Committee led to refinement of some item definitions as well as to modification of the GTI design. The performance of the GTI on paper cases was evaluated among the participating experts to assess the consistency with which the GTI was applied and the degree to which the GTI score reflected the experts' clinical judgment of relative GC toxicities [13].

Following the 2015 face-to-face meeting, participants completed an on-line exercise composed of two tasks. The first was to score another set of 15 patient cases with specific clinical scenarios of GC exposure and GC-related adverse events using the GTI. The second was to rank those 15 cases in order of greatest to least GC-toxicity. The investigators were asked to use their overall clinical judgment and were blinded to the weights and the GTI scores of the cases. The experts' rankings were then compared to the ranking assigned by the weighted GTI.

Statistical analysis in the first phase of validation

Interrater reliability among 17 raters on toxicity items in the Composite GTI of the 15 paper cases was assessed using the Kappa statistic [13]. Similarly, agreements between rankings assigned by the weighted Composite GTI and expert clinical judgment rankings were assessed by the Kappa statistic. The overall interrater reliability of the ranking agreements was then calculated by averaging pairwise Kappa values. All statistical analyses were performed on SAS Version 9.3 (SAS Institute, Cary, NC, USA).

Evaluation of the GTI

To assess reliability in the evaluation phase of GTI development, the Scientific Committee assigned toxicity items on the Composite GTI to 15 paper cases. Raters reached a high degree of agreement, with a kappa of 0.88 ($p<0.01$). To assess Composite GTI validity, participants ranked 15 cases in order of highest to lowest toxicity. Expert rankings were compared with rankings according to the GTI score (from highest to lowest), also yielding excellent agreement with a weighted kappa of 0.87 ($p<0.01$).

Validation in a real-world asthma cohort

The GTI was then validated further in a real-world cohort of 101 patients with severe asthma who were beginning treatment with mepolizumab, an anti-interleukin-5 receptor antagonist [11,12]. The patients were receiving high-dose inhaled GC in combination with other medications to control their asthma. More than four fifths (82.2%) were also receiving daily prednisolone to maintain control of their asthma, with a median daily dose of 11.7 mg.

In the cross-sectional component of the study, a measure of baseline GC toxicity based on GTI weights at study entry had a modest correlation with maintenance prednisolone dose (rho=0.26, p=0.01), cumulative exposure in the prior year (rho=0.38, $p<0.001$), and number of GC boosts for asthma flares (rho=0.25, p=0.01) [11]. The GTI also demonstrated strong associations with asthma-related quality-of-life (the Mini-Asthma Quality of Life Questionnaire; mini-AQLQ) and the St. George's Respiratory Questionnaire, both important patient-reported outcome measures in asthma studies [11].

Consideration of the minimum clinically important difference

We used a distribution-based approach to estimating the minimum clinically important difference (MCID) for the GTI. Methods for estimating the MCID for clinical instruments have been divided into distribution-based approaches and anchor-based approaches [17, 18]. Distribution-based approaches are based on statistical characteristics of clinical data.

In the case of the GTI, these data for the initial MCID estimate consisted of evaluations of toxicity change in a total of 510 patient assessments, performed by 34 clinicians. These assessments were performed on summaries of a group of patients with a variety of inflammatory diseases, including ANCA-associated vasculitis, systemic lupus erythematosus, giant cell arteritis, and others. The summaries were evaluated by the 17 original subspecialty investigators and a second group of 17 independent evaluators. This initial estimate of the MCID derived from this effort was then evaluated further and substantiated in both a longitudinal cohort of 101 patients with asthma and a randomized, double-blind, placebo-controlled trial that included 331 patients, as described below.

The basis for estimating the MCID was the standard error of measurement (SEM), defined as the smallest change likely to reflect a true difference as opposed to a measurement error of the effect. AIS data were used to calculate the MCID because the AIS has a normal distribution [11]. In the MCID derivation, the intra-class coefficient

was calculated for the investigators' scores and the SEM was calculated from the standard deviation of those scores. The intra-class correlation coefficient for the AIS in these evaluations was 0.97. The standard deviation of AIS at the follow-up was 57.42. The minimum intra-patient difference in GTI scores between two timepoints that cand be regarded as a true change in GC toxicity, therefore, is $57.42 * \sqrt{1 - 0.971} = 9.8$, or approximately 10 points. Thus, any GTI score > 10 points may be considered to represent a true change in GC toxicity. This estimate, calculated on data related to toxicity change in real patients, can be regarded as an initial estimate of the MCID for the GTI, subject to further evaluation [17,18]. This initial estimate was then studied further in two additional settings, one a real-world longitudinal experience in asthma and the other a phase 3 clinical trial.

Examination of the MCID in a real-world experience

The MCID was assessed in in a longitudinal study of the cohort of patients with severe asthma who were starting mepolizumab therapy [12]. This cohort of 101 patients was followed prospectively for one year following the institution of treatment with mepolizumab and consequent GC reduction. The patients were able to reduce their daily GC requirement following the start of mepolizumab from a median daily prednisolone dose of 11.7 mg at baseline to 6.7 mg at one year. The patients also required fewer GC bursts ("rescue" courses) to treat asthma flares and had fewer Emergency Room visits and hospitalizations compared to the year before starting mepolizumab.

The GTI data confirmed substantial reductions in GC toxicity during this time [12]. The mean AIS for the cohort overall was -35.7, far exceeding the MCID of 10. Of the 101 patients, 62 met the AIS MCID of ≤-10. The mean AIS among patients meeting the MCID was -71.4 (range -165 to -10). Patients who met the MCID had fewer GC rescue courses for exacerbations in the year following the start of mepolizumab (odds ratio 0.73; 95% CI 0.55,0.98; P=0.04). They also had a greater percentage reduction in prednisolone exposure between baseline and one year (48.6% versus 40.8%; OR 0.80; 95% CI 0.66,0.96; P=0.02). In contrast, 39 (39%) of the patients did not meet the MCID. Among those who did not achieve the MCID, the mean AIS was 21.1 (range -9 to +130).

The relationship between the percentage decrease in annual oral GC dosing and the probability of achieving the MCID was examined [**Figure 1**]. The wide confidence intervals of the curve illustrate that measuring the overall percentage reduction of systemic GC use is an inadequate indicator of GC toxicity. The degree of toxicity change varied widely at the individual patient level (AIS range -165 to +130). A reduction in toxicity (AIS <0) was seen in 70% (71/101) of the cohort, while 3% (3/101) had no change (AIS = 0) and 27% (27/101) an increase (AIS ≥1). The GC toxicity change did not have a significant linear correlation with oral GC reduction, indicating that if an important goal of a new therapy is to decrease GC toxicity, then GC toxicity needs to be measured directly rather than relying upon GC reduction – either absolute reduction (mg) or proportional reduction (%) – as a surrogate for toxicity.

Examination of GTI thresholds in the ADVOCATE trial

ADVOCATE, a phase 3, randomized, double-blind, controlled trial in 330 patients, evaluated the efficacy and safety of avacopan [14, 15]. Patients with ANCA-associated vasculitis (new-onset or relapsing disease) were randomized to receive avacopan 30 mg twice daily, or a standard 20-week oral prednisone taper, plus matching placebos, in addition to background immunosuppressant therapy (rituximab or cyclophosphamide/azathioprine). GCs were allowed as pre-medication for rituximab to reduce hypersensitivity reactions, taper after GCs given during the screening period, treatment of ANCA-associated vasculitis (persistent vasculitis, worsening of vasculitis, or relapses), as well as for non-vasculitis reasons such as adrenal insufficiency. The median (mean) total GC dose from all sources over the 52-week study was 400 mg (1349 mg) in the avacopan group versus 2939 mg (3655 mg) in the prednisone group.

The investigators examined the percentages of patients in the avacopan versus prednisone groups that exceeded selected GTI

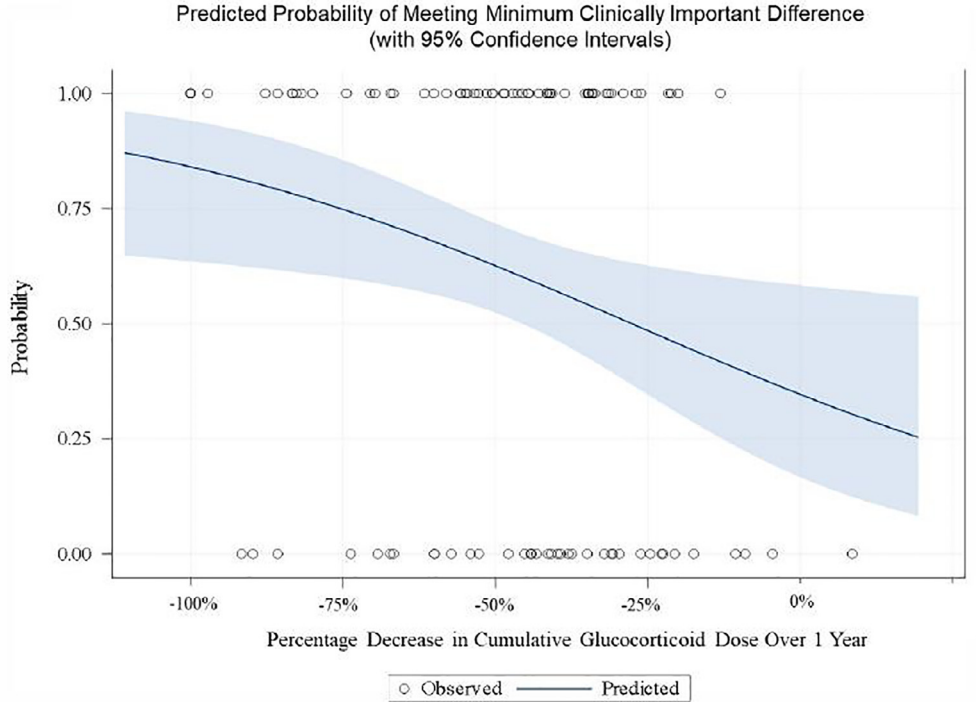

Fig. 1. Relationship between the percentage decrease in annual oral glucocorticoid (GC) dosing (X-axis) and the probability of achieving the minimal detectable difference for the GTI (Y-axis). The wide confidence intervals of the curve illustrate that measurement of the overall percentage reduction of systemic GC use is an inadequate indicator of GC toxicity.

threshold values, beginning with 10 points (the estimated MCID) and also examining the 20- and 30-point thresholds [19]. The percentages of patients exceeding the specified AIS differentiated the two treatment groups in GC toxicity at all three thresholds. In the avacopan group, 48% of the patients compared with 60% of those in the prednisone group exceeded the 10-point (MCID) threshold (P=0.02). At 20-point comparison, 30% versus 45% of the patients exceeded this threshold, respectively (P=0.003); and at the 30-point threshold the percentages for the two groups were 18% and 34%, respectively (P=0.001) (**Table 4**). The CWS differentiated the avacopan group from the prednisone group at the 20- and 30-point thresholds (58% vs 73%, respectively, at the 20-point threshold [P=0.002]; 41% vs 56% at the 30-point threshold [P=0.007]). **Figure 2** shows the CWS and AIS comparisons between groups at both 13 and 26 weeks. At both timepoints, the avacopan had substantially lower GTI scores.

These data provide further support for the concept that 10 points represents not only the minimum detectable difference for the GTI, it is also a reasonable estimate of the MCID. Moreover, the ability of the GTI to demonstrate a difference in GC toxicity as early as 13 weeks even when both treatment groups received some GC therapy during this time period underscores the ability of the instrument to measure changes in GC toxicity over relatively short time intervals.

Consistency of scoring across GTI domains

Examination of the individual GTI domain scores from the ADVOCATE trial is also instructive. The GTI scores demonstrated a consistent reduction in GC toxicity across domains in the avacopan group. As shown in **Figure 3**, 7 of the 8 domains differentiated the two treatment groups well with regard to the development of new GC toxicities after baseline. The only domain that did not differentiate the two treatment groups was the blood pressure domain. This finding was not surprising because 81% of the patients in the trial overall had glomerulonephritis, a disease feature often associated with hypertension and the need for anti-hypertensive therapies.

Similar findings were observed with regard to the AIS (**Figure 4**). Three of the domains (glucose tolerance, skin toxicity, and neuropsychiatric toxicity) actually improved over time in the avacopan group, with bigger improvements occurring at the 26-week timepoint compared to the 13-week timepoint. This observation corresponded to the pattern of GC use in the avacopan group. Glucose tolerance and neuropsychiatric toxicity also approved in the standard care group, consistent with the lower mean daily GC doses over time. Comparisons within the body mass index, lipid, skin toxicity, and infection domains all strongly favored the avacopan group.

The tendency across domains with regard to GC toxicity – likely more pronounced within individual patients as opposed to treatment cohorts as a whole – underscores the importance of a composite measure that accounts for the occurrence of GC toxicities in multiple domains.

Consistency of GTI scores with reported glucocorticoid adverse events and patient-reported outcome measures

Findings with regard to GC toxicity as measured by the GTI were highly consistent with the incidence of individual GC toxicities as outlined in the EULAR guidelines [3]. They were also consistent with adverse events in the trial overall [15], suggesting that GC-related adverse events accounted for a substantial proportion of the treatment-related morbidity in the trial. Finally, the GTI scores were consistent with data observed in patient-reported outcomes such as the SF-36 and the EuroQoL 5D-5L [15]. The SF-36 Physical Component Domain scores are shown in **Supplemental Figure 2**.

Potential weaknesses and strengths

The GTI has both potential weaknesses as well as strengths. Ease of application is important with any clinical assessment tool. It might be considered that an assessment tool consisting of multiple domains is too difficult and time-consuming for investigators to use easily. This point was considered in the development of the GTI, however. Many of the domains of GC toxicity represented in the GTI consist of data that are already collected in the context of trials. For example, body mass index, blood pressure, hemoglobin A1c, low-density lipoprotein concentrations, medication changes, and occurrences of infections are all captured routinely and rigorously in clinical trials. Assessments of muscle strength, skin toxicity, and the neuropsychiatric effects of glucocorticoids are similarly simple to collect. The ease of use of the GTI has now been demonstrated in multiple real-world and clinical trial experiences. Moreover, data collection can be streamlined using carefully-designed case report forms, electronic data entry, and the digital platform of the GTI. The digital platform can be linked to clinical trial data bases through application programming interfaces, if desired, though this is not necessary. All of these measures, singly or combined, streamline data collection with the GTI.

Another potential weakness is that the GTI, even though a composite instrument, may capture only a small spectrum of GC toxicities, of which there are multiple dozen (**Supplemental Table 1**). As noted, however, the GTI domains were selected to emphasize those that are common, easy to measure, have substantial clinical impact, and are dynamic – i.e., are susceptible to change with varying glucocorticoid doses. GC toxicities such as avascular necrosis and cataracts, for example, which represent irreversible damage, are captured as other items of GC toxicity in the GTI but do not impact scoring.

There may be some concern that the GTI scores are difficult to calculate. This potential challenge is addressed by having the GTI users simply record the data elements for the GTI. The scoring can be performed either by hand and through the digital platform, either in real time as data are entered or following a data upload at the time of data analysis.

Table 4
Percentages of Patients in ADVOCATE Exceeding Selected GTI Thresholds at Week 26.

GTI threshold/study group	n (%) exceeding CWS threshold	P-value	n (%) exceeding AIS threshold	P-value
GTI worsening > 10 points				
Avacopan (N = 164)	138 (83%)	0.147	80 (48%)	0.022
Prednisone (N = 166)	144 (88%)		99 (60%)	
GTI worsening > 20 points				
Avacopan (N = 164)	96 (58%)	0.002	49 (30%)	0.003
Prednisone (N = 166)	120 (73%)		74 (45%)	
GTI worsening > 30 points				
Avacopan (N = 164)	68 (41%)	0.007	30 (18%)	0.001
Prednisone (N = 166)	91 (56%)		55 (34%)	

AIS, Aggregate Improvement Score; CWS, Cumulative Worsening Score; GTI, Glucocorticoid Toxicity Index

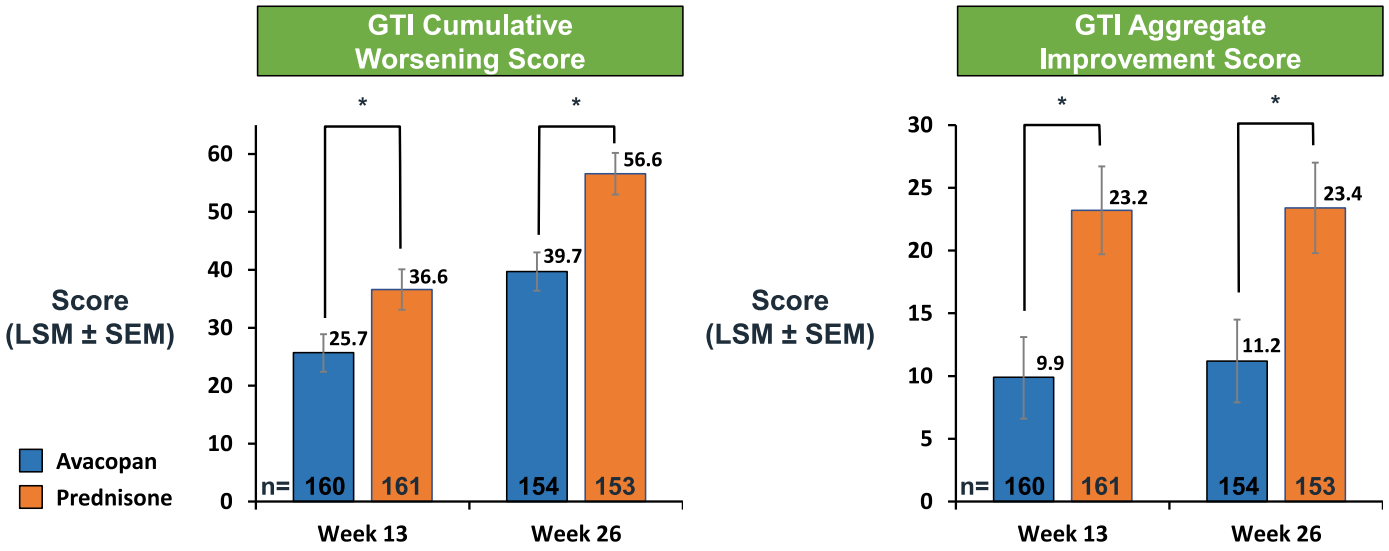

Fig. 2. GTI scores in the ADVOCATE trial at 13 and 26 weeks. Left: Cumulative Worsening Scores (CWS). Right: Aggregate Improvement Scores (AIS). Glucocorticoid (GC) toxicity in the avacopan group was lower at both 13 and 26 weeks for both GTI scores (CWS: P = 0.01 at 13 weeks; P = 0.0002 at 26 weeks. AIS: P = 0.003 at 13 weeks; P = 0.008 at 26 weeks). The data shown are least square means (LSM) and the standard error of measurement (SEM), estimates based on mixed-model repeated measures of the longitudinal GTI data (from Day1 to Week 26).

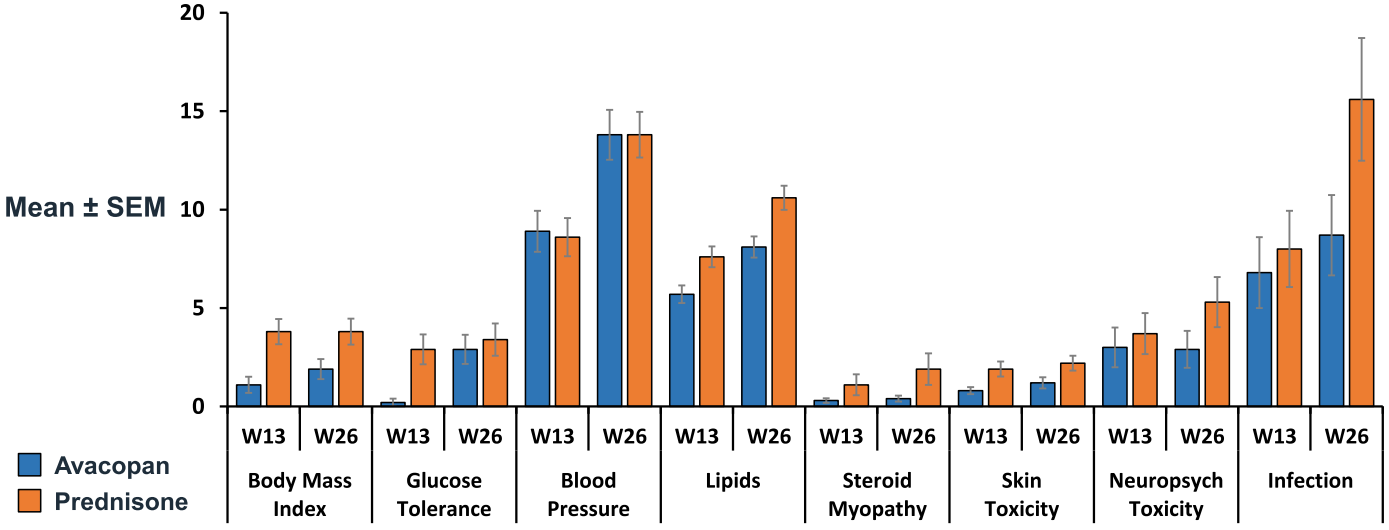

Fig. 3. Cumulative Worsening Scores (CWS) by GTI domain in the ADVOCATE trial. Seven of 8 GTI domains of GC toxicity differentiated the two treatments, with all favoring the avacopan group in a statistically significant manner. Only the hypertension domain showed no significant difference between the treatment groups.

Finally, establishing a reliable MCID for clinical instruments is often challenging. In contrast to the distribution-based approach that we employed, anchor-based approaches that compare the change in a clinical outcome to a second, external measure of change (that is, one that is clearly understood, such as a global assessment) may also be considered. No such anchor-based method has been established for the assessment of change in glucocorticoid toxicity. Nevertheless, further considerations of the MCID based on other potential anchors, e.g., patient-reported quality of life outcome measures, are an important part of the future research agenda for the GTI.

Among the most important strengths of the GTI are its development by an international panel of subspecialty experts; its derivation of weights for each item of GC toxicity; its deployment of a composite of relevant GC domains; its validation in real-world clinical experiences and a phase 3 clinical trial; its correspondence with patient-reported quality-of-life outcomes; and the availability of digital platform that streamlines data collection and scoring.

Ongoing efforts to understand the optimal ways of using the GTI to understand the phenomenon GC toxicity are important. The GTI was developed explicitly as a clinician-facing instrument as opposed to a patient-reported outcome measure. There is ample precedent in clinical investigation for this approach. A patient-reported outcome measure would also be a valuable addition to the evaluation of GC toxicity, undoubtedly providing insights into different facets of this issue. Other GTI analyses of high priority include repeated evaluations of GTI assessments in cohorts of patients over time, pre-specified comparisons of changes in the GTI among patients perceived to have high GC toxicity as opposed to low toxicity, and understanding of the relationships between change in GC toxicity and healthcare resource utilization.

Conclusion

GC toxicity has framed a major portion of the treatment experience for patients with inflammatory disease for more than 70 years.

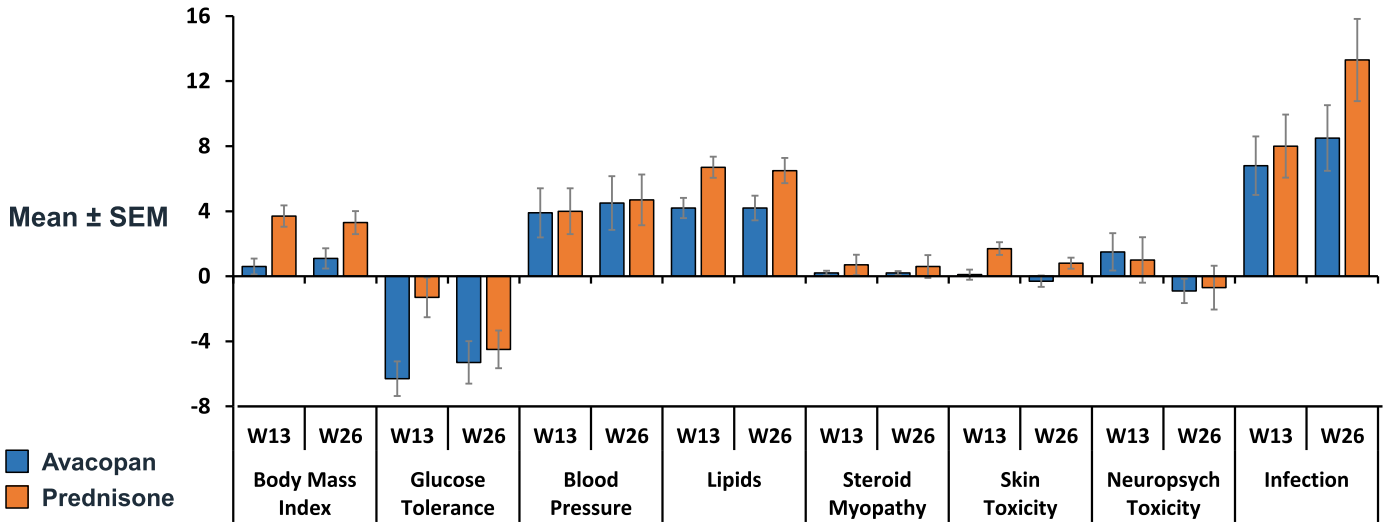

Fig. 4. Aggregate Improvement Scores (AIS) by GTI domain in the ADVOCATE trial. Four of 8 domains strongly favored the avacopan group. A fifth domain (glucocorticoid myopathy) also suggested a larger impact in the avacopan group but was not statistically significant.

The development of novel immunomodulatory agents has given clinicians new potential options for treatment, opening the possibility of reducing GC toxicity while preserving or even improving upon therapeutic efficacy. To assess the full benefits of new medications, investigators must be able to assess the ability of new drugs to prevent or reverse GC toxicity and use this measurement in the definition of treatment response. The GTI has now been employed both in real-world experience and clinical trials and may play a key role in assessing the true value of new "steroid-sparing" medications.

Declaration of Competing Interest

Dr. Stone's hospital, the Massachusetts General Hospital, owns the intellectual property of the Glucocorticoid Toxicity Index (GTI). He is Chairman of the Scientific Advisory Board of STERITAS, which owns the sole licensing rights to the GTI. Dr. Stone has consulted for ChemoCentryx on the topic of glucocorticoid toxicity. Drs. Jayne, Merkel, Robson, and Bekker were investigators in the ADVOCATE trial of avacopan for ANCA-associated vasculitis. Dr. Bekker is an employee of ChemoCentryx.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.semarthrit.2022.152010.

References

[1] Wilson JC, Sarsour K, Collinson N, et al. Incidence of outcomes potentially associated with corticosteroid therapy in patients with giant cell arteritis. Semin Arthr Rheum 2017;46(5):650–6.
[2] Wilson JC, Sarsour K, Collinson N, et al. Serious adverse effects associated with glucocorticoid therapy in patients with giant cell arteritis (GCA): A nested case-control analysis. Semin Arthr Rheum 2017;46(6):819–27.
[3] Van der Goes MC, Jacobs JWG, Boers M, et al. Monitoring adverse events of low-dose glucocorticoid therapy: EULAR recommendations for clinical trials and daily practice. Ann Rheum Dis 2010;69(11):1913–9.
[4] Johnson SR, Naden RP, Fransen J, et al. Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. J Clin Epidemiol 2014;67(6):706–14.
[5] Neogi T, Jansen TLTA, Dalbeth N, et al. 2015 Gout classification criteria: an American College of Rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis 2015;74(10):1789–98.
[6] Neogi T, Aletaha D, Silman AJ, et al. The 2010 American College of Rheumatology/European league against rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. Arthritis Rheum 2010;62(9):2582–91.
[7] Taylor WJ. Pros and cons of conjoint analysis of discrete choice experiments to define classification and response criteria in rheumatology. Curr Opin Rheumatol 2016;28:117–21.
[8] Tedeschi SK, Johnson SR, Boumpas DT, et al. Multicriteria decision analysis process to develop new classification criteria for systemic lupus erythematosus. Ann Rheum Dis 2019;78:634–40.
[9] Tedeschi SK, Pascart T, Latourte A, et al. Identifying potential classification criteria for calcium pyrophosphate deposition disease (CPPD): Item generation and item reduction. Arthritis Care Res 2021 in press (PMID: 33973414).
[10] Wallace ZS, Naden RP, Choi H, et al. The 2019 ACR/EULAR classification criteria for IgG4-related disease. Arthritis Rheum 2020;72(1):7–19.
[11] McDowell PJ, Stone JH, Zhang Y, et al. Quantification of glucocorticoid-associated morbidity in severe asthma using the glucocorticoid toxicity index. J Allergy Clin Immunol Pract 2021;9(1):365–72.
[12] McDowell PJ, Stone JH, Zhang Y, et al. Glucocorticoid toxicity reduction with mepolizumab using the glucocorticoid toxicity index. Eur Resp J 2021; Jul 1:2100160 PMID: 34210787. doi: 10.1183/13993003.00160-2021.
[13] Miloslavsky EM, Naden RP, Bijlsma JW, et al. Development of a glucocorticoid toxicity index (GTI) using multi-criteria decision analysis. Ann Rheum Dis 2017;76:543–6.
[14] Merkel PA, Jayne DRW, Wang C, et al. Evaluation of the safety and efficacy of avacopan, a C5a receptor inhibitor, in patients with antineutrophil cytoplasmic antibody-associated vasculitis treated concomitantly with rituximab or cyclophosphamide/azathioprine: protocol for a randomized, double-blind, active-controlled, phase 3 trial. JMIR Res Protoc 2020;9(4):e16664.
[15] Jayne DRW, Merkel PA, Schall TJ, Bekker P. For the ADVOCATE study group. Avacopan for the treatment of ANCA-associated vasculitis. N Engl J Med 2021;384(7):599–609.
[16] Van der Goes MC, Jacobs JWG, Boers M, et al. Patient and rheumatologist perspectives on glucocorticoids: an exercise to improve the implementation of the European League Against Rheumatism (EULAR recommendations on the management of systemic glucocorticoid therapy in rheumatic disease. Ann Rheum Dis 2010;69:1015–21.
[17] Rai SK, Yazdany J, Fortin PR, Avina-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. Arthritis Research & Therapy 2015;17:143. doi: 10.1186/s13075-015-0658-6.
[18] Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7:541–6.
[19] Patel NJ, Zhang Y, Jayne DRW, Merkel PA, Yue H, Bekker P, Stone JH. Differences between avacopan and prednisone for treatment of ANCA-associated vasculitis at different thresholds of glucocorticoid toxicity. (Abstract accepted for presentation at European League Against Rheumatism 2022).