Seminars in Arthritis and Rheumatism 56 (2022) 152068

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

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

The pediatric glucocorticoid toxicity index

Paul Brogan [a], Ray Naden [b], Stacy P. Ardoin [c], Jennifer C. Cooper [d], Fabrizio De Benedetti [e], Jean-Francois Dicaire [f], Despina Eleftheriou [a], Brian Feldman [g], Jon Goldin [a], Seth E. Karol [h], Fiona Price-Kuehne [i], David Skuse [a], Constantine A. Stratakis [j], Nicholas Webb [k,l], John H. Stone [a,*]

[a] Great Ormond Street Hospital NHS Foundation Trust, University College London Great Ormond Street Institute of Child Health, London UK
[b] McMasterUniversity, Hamilton, Ontario, Canada
[c] Ohio State University; Columbus, OH, USA
[d] University of Colorado Anschutz Medical Alifornia, San Francisco, CA, USA
[e] Hopital Bambino Gesu, Rome, Italy
[f] Pinnacle, Montreal, Quebec, Canada
[g] Hospital for Sick Children, Toronto, Ontario, Canada
[h] St. Jude Children's Research Hospital, Memphis, TN, USA
[i] Cambridge University Hospitals, Cambridge, UK
[j] Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA
[k] Royal Manchester Children's Hospital, Manchester, UK
[l] Rheumatology Clinic, Bulfinch 165, Massachusetts General Hospital, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA

ARTICLE INFO

Keywords:
Glucocorticoid
Toxicity
Adverse effects
Clinical outcome assessment
Body mass index
Growth
Hemoglobin A1c
Myopathy
Bone mineral density
Hypertension
Acne
Hirsutism
Sleep
Mood
Infection
Osteonecrosis
Damage
Asthma
Kawasaki's disease
Juvenile idiopathic arthritis

ABSTRACT

Objectives: To develop a Pediatric glucocorticoid toxicity index (pGTI), a standardized, weighted clinical outcome assessment that measures change in glucocorticoid (GC) toxicity over time.

Methods: Fourteen physician experts from 7 subspecialties participated. The physician experts represented multiple subspecialties in which GCs play a major role in the treatment of inflammatory disease: nephrology, rheumatology, oncology, endocrinology, genetics, psychiatry, and maternal-fetal medicine. Nine investigators were from Canada, Europe, or New Zealand, and 5 were from the United States. Group consensus methods and multi-criteria decision analysis were used. The pGTI is an aggregate assessment of GC toxicities that are common, important, and dynamic. These toxicities are organized into health domains graded as minor, moderate, or major and are weighted according to severity. The relative weights were derived by group consensus and multi-criteria decision analysis using the 1000Minds^{TM} software platform. Two quantitative scores comprise the overall toxicity profile derived from pGTI data: (1) the Cumulative Worsening Score; and (2) the Aggregate Improvement Score. The pGTI also includes a qualitative, unweighted record of GC side-effects known as the Damage Checklist, which documents less common toxicities that, although potentially severe, are unlikely to change with varying GC dosing.

Results: One hundred and seven (107) toxicity items were included in the pGTI and thirty-two (32) in the Damage Checklist. To assess the degree to which the pGTI corresponds to expert clinical judgement, the investigators ranked 15 cases by clinical judgement from highest to lowest GC toxicity. Expert rankings were then compared to case ranking by the pGTI, yielding excellent agreement (weighted kappa 0.86). The pGTI was migrated to a digital environment following its development and initial validation. The digital platform is designed to ensure ease-of-use in the clinic, rigor in application, and accuracy of scoring. Clinic staff enter vital signs, laboratory results, and medication changes relevant to pGTI scoring. Clinicians record findings for GC myopathy, skin toxicity, mood dysfunction, and infection. The pGTI algorithms then apply the weights to these raw data and calculate scores. Embedded logic accounts for the impact of age- and sex-related reference ranges on several health domains: blood pressure, lipid metabolism, and bone mineral density. Other algorithms account for

Abbreviations: pGTI, Pediatric Glucocorticoid Toxicity Index; CWS, Cumulative Worsening Score; AIS, Aggregate Improvement Score; MCID, Minimum clinically important difference.

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

https://doi.org/10.1016/j.semarthrit.2022.152068

Available online 14 July 2022
0049-0172/© 2022 Massachusetts General Hospital. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

anticipated changes in the height Z-scores used in the growth domain, thereby addressing a concern unique to GC toxicity in children. The Damage Checklist ensures comprehensive measurement of GC toxicity but does not contribute to pGTI scoring, because the scored domains emphasize manifestations of GC toxicity that are likely to change over the course of a trial.

Conclusions: We describe the development and initial evaluation of a weighted, composite toxicity index for the assessment of morbidity related to GC use in children and adolescents. Developing the pGTI digital platform was essential for performing the nuanced calculations necessary to ensure rigor, accuracy, and ease-of-use in both clinic and research settings.

Introduction

Glucocorticoids have been a cornerstone of treatment for dozens of pediatric diseases since their introduction over seventy years ago [1]. Glucocorticoids are effective in many settings but their use is associated with substantial treatment-related morbidity and the potential for life-long complications in children and adolescents [2,3]. Improved therapeutic options, particularly the development of effective immunomodulatory agents, have provided new treatment approaches for pediatricians, who can now reduce their patients' GC use and diminish the well-known adverse effects of these agents by selecting alternative treatment approaches [4–7].

To evaluate the potential clinical and health economic benefits of new treatment approaches, investigators must be able to assess the ability of new drugs to prevent or reverse GC-related toxicity. An adult GTI is now employed in more than 40 studies, including phase 3 clinical trials in asthma, vasculitis, systemic lupus erythematosus, lupus nephritis, inflammatory myopathies, sarcoidosis, polymyalgia rheumatica, pemphigus vulgaris, and congenital adrenal hyperplasia [7,8]. Recently, the adult GTI was used as the most important secondary endpoint of efficacy in a trial of a novel complement inhibitor in anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis [6]. In that trial, the new medication demonstrated unequivocal superiority over standard care in terms of GC toxicity while showing non-inferiority for the primary efficacy outcome. The between-group differences were highly statistically significant, achieved the GTI threshold for minimum clinically important difference (MCID), and separated the two treatment groups across several GTI thresholds of GC toxicity [6–8]. The drug, avacopan, was awarded regulatory approval in the United States, Europe, Japan, and other countries in 2021.

The assessment of GC toxicity in children is more complex than in adults, because evaluations must consider the impact of age reference ranges on certain toxicity domains as well as the effect of GCs on growth. No reliable instrument has been designed and developed to measure GC-related toxicity in pediatric patients. We aimed to develop a pediatric glucocorticoid toxicity index (pGTI) for use across pediatric disciplines to assess GC-associated morbidity within an age range of 2–18 years. To manage the complexity of GC toxicity assessment in children, we planned to migrate the developed instrument to a digital environment. The pGTI user interface is organized in a manner consistent with the flow of a typical clinic visit, to facilitate ease of use. The logic and algorithms within the application ensure rigor of application and accuracy in scoring.

Methods

Overview of approach and aims

We assembled a multi-specialty group of physician experts and employed established group consensus methods and multi-criteria decision analysis. The multi-criteria decision analysis stage of pGTI development, essential in the development of relative weights for the various levels of toxicities within domains, relied upon use of the 1000Minds^{TM} software platform (www.1000Minds.com). Intermediate aims on the path to the development of the pGTI were: (1) to define the optimal settings for pGTI use; (2) to define toxicity items for inclusion in the pGTI; (3) to assign relative weights to the pGTI items; and (4) to begin the process of assessing the instrument's reliability and validity. Once the analog version of the pGTI was complete, we developed a cloud-based platform for the instrument.

Participants

The participating experts, recruited from multiple specialties, had extensive experience in the use of GCs and the management of GC toxicity in children. The Steering Committee was comprised of a patient representative (FPK), two pediatric rheumatologists (PAB, SPA), a methodologist with expertise in group consensus methods and classification criteria development (RPN), a logistics expert (JFD), and an adult rheumatologist (JHS) who led the development of the adult GTI. The Steering Committee invited 14 physician experts to participate in the pGTI development, representing seven subspecialties. The Scientific Committee was comprised of participants from pediatric and adult rheumatology, pediatric endocrinology and genetics, pediatric nephrology, pediatric oncology, pediatric psychiatry, and maternal-fetal medicine. Nine investigators were from Canada, Europe, or New Zealand, and 5 were from the United States.

Study procedures

The Steering Committee created milestones for pGTI development (Fig. 1).

The development process included eight one-hour conference calls conducted over a period of three months; substantial work between the calls; and one day-long, face-to-face meeting at which the weights for the pGTI were derived. This was followed by additional validation work over two months and the development of the digital platform.

Broad characteristics of the pGTI

The pGTI is a clinician-facing instrument that is weighted and quantitative. The weighted core includes GC toxicities that are common, important, and dynamic. Each of the toxicity domains included in the pGTI is estimated to occur in at least 5% of patients treated with a substantial GC course and, although the pGTI is a clinician-facing instrument, all of the domains represent toxicities that matter to patients as well as to physicians. The pGTI domains are dynamic, in the sense that they have the potential to change with varying GC doses, and are therefore quantifiable from one clinic visit to the next.

The instrument also includes a qualitative, unweighted Damage Checklist. The Checklist is designed to capture permanent GC-related toxicities that are unlikely to be reversed by lower exposure to GCs (examples: osteonecrosis, cataracts, tendon rupture). The pGTI with the accompanying weights of each item and the Damage Checklist are shown in Table 1 and Table 2, respectively.

Candidate domains and items

Candidate GC-related toxicities were generated based on a literature review and on consensus exercises. Each toxicity was grouped into a

specific health domain (e.g., body mass index, glucose metabolism, or infection). The domains were then subdivided further into levels of severity: minor, moderate, or major.

pGTI domains

The goal was to create domains of GC toxicity that are independent of each other (e.g., hypertension, skin toxicity, growth impact, and infection). The Scientific Committee then selected the items for inclusion using a nominal group technique. The principles guiding selection of the pGTI items are shown in Fig. 2.

Definitions for each pGTI item were developed by the experts from the relevant clinical specialty, presented to the full Scientific Committee, and revised according to consensus. Domains were refined such that only one item within the domain could be assigned to a given patient (i.e., the items are mutually exclusive). Items not included in the pGTI were considered for incorporation into the Damage Checklist.

Accounting for increasing age in pediatric populations

Age-related reference ranges apply to certain pGTI domains. For example, normative blood pressure ranges, growth (change in height Z-score), low-density lipoprotein levels, and bone mineral density are all affected by age [9,10]. The need to apply age-related references to certain domains constitutes a major difference between the pGTI and the adult GTI, reflecting the additional complexity of the assessment of GC toxicity in children.

Accounting for growth in pediatric populations

A major concern in pediatric patients is the impact of GCs on growth [11–13]. Thus, a domain of GC toxicity focusing on growth was created for the pGTI. The growth domain, which relies upon changes in height Z-score, was modelled specifically on data obtained from a clinical trial in juvenile idiopathic arthritis [14].

Accounting for the impact of medication on GC toxicity

The Scientific Committee also sought to account for the impact of medications on GC-related toxicity, because some pGTI Domains (e.g., Blood Pressure) may improve through medical management rather than a decrease in GC toxicity *per se*. Therefore, item definitions and the scoring algorithms account for the impact of treatment. For example, if blood pressure was lower following an increase in the anti-hypertensive regimen, then the Blood Pressure domain is scored "no significant change" rather than "improvement." Lower blood pressure in the absence of new anti-hypertensive treatment is, however, scored as improvement and weighted accordingly.

Initial evaluation of precision, consistency and usability of the pGTI

Following item inclusion and definition development, the pGTI 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 then asked to submit four cases of real patients who had experienced GC toxicity ranging from major worsening to major improvement during treatment. Fifteen cases, selected to reflect the spectrum of potential GC toxicity, formed the basis of the exercise. Each expert was asked to score the 15 cases (i.e., allocate the appropriate level within the Domains) by reference to a draft version of the pGTI in which the items remained unweighted. Consistency in the scoring of each case was evaluated.

Weight derivation

After the pGTI toxicity items had been selected, organized by domain, and ranked within each domain in order of increasing toxicity, relative weights for each item were derived at a face-to-face meeting lasting one day. This meeting involved facilitated group consensus combined with multi-criteria decision analysis, utilizing using the 1000Minds software. The multi-criteria decision analysis exercise is based on the PAPRIKA method (**P**otentially **A**ll **P**airwise **R**an**K**ings of all possible **A**lternatives) [15]. This methodology of group consensus and

1. Agreement upon the broad characteristics of the pGTI.
2. Determination of the optimal study setting(s) for use of the pGTI.
3. Establishment of criteria for including toxicities in the pGTI.
4. Identification of toxicities for inclusion in the pGTI.
5. Development of specific definitions for the toxicities included.
6. Assignment of toxicity items to clinical domains.
7. Division of items into pGTI or the Damage Checklist.
8. Ranking of pGTI items in order of severity.
9. Assessment of the precision, consistency, and usability of the draft instrument.
10. Derivation of weights for each toxicity item utilizing multi-criteria decision analysis.
11. Evaluation of the pGTI using recorded cases.
12. Development of the digital platform.

Fig. 1. Milestones for development of the pGTI established by the steering committee.

Table 1
Relative weights of items in the pediatric glucocorticoid toxicity index (pGTI).

Domain	Weight	Domain	Weight
Change in Body Weight		**Skin (cont.)**	
Decrease by >/= 5 BMI units	−52	***Hirsutism***	
Decrease by >2 but <5 BMI units	−21	Moderate hirsutism to none	−16
No change (+/−2 BMI units)	0	Moderate to Minor hirsutism	−11
Increase of >2 to <5 BMI units	21	Minor hirsutism to none	−5
Increase of 5 or more BMI units	52	No change	0
		None to Minor hirsutism	5
Growth Velocity		Minor to Moderate hirsutism	11
Increase in z-score more than 0.5	−49	None to Moderate hirsutism	16
No change in z-score (+/−0.5)	0		
Decrease in z-score more than 0.5	49	***Striae/Atrophy***	
		Moderate striae/atrophy to none	−33
Glucose Metabolism		Moderate to Minor striae/atrophy	−23
Improvement in HbA1c AND decrease in medication	−40	Minor striae/atrophy to none	−10
Decrease in medication without improvement in HbA1c	−29	No change	0
Improvement in HbA1c without decrease in medication	−17	No striae/atrophy to Minor	10
No significant change	0	Minor striae/atrophy to Moderate	23
Increase in HbA1c without increase in medication	17	No striae/atrophy to Moderate	33
Increase in medication without increase in HbA1c	29		
Increase in HbA1c AND increase in medication	40	**Neuropsychiatric Toxicity**	
		Sleep Disturbance	
Blood Pressure		Major sleep disturbance to none	−35
Improvement in BP AND decrease in medication	−38	Major sleep disturbance to Minor	−27
Decrease in medication without improvement in BP	−26	Major sleep disturbance to Moderate	−22
Improvement in BP without decrease in medication	−15	Moderate sleep disturbance to none	−13
No significant change	0	Moderate to Minor sleep disturbance	−5
Increase in BP without increase in medication	15	Minor sleep disturbance to none	−8
Increase in medication without increase in BP	26	No change	0
Increase in BP AND increase in medication	38	No sleep disturbance to Minor	8
		Minor sleep disturbance to Moderate	5
Hyperlipidemia		No sleep disturbance to Moderate	13
Decrease in LDL AND decrease in medication	−33	Moderate sleep disturbance to Major	22
Decrease in medication without decrease in LDL	−22	Minor sleep disturbance to Major	27
Decrease in LDL without decrease in medication	−11	No sleep disturbance to Major	35
No significant change	0		
	11	***Mood***	

Table 1 (*continued*)

Domain	Weight	Domain	Weight
Increase in LDL without increase in medication			
Increase in medication without increase in LDL	22	Major mood disturbance to none	−57
Increase in LDL AND increase in medication	33	Major mood disturbance to Minor	−46
		Major mood disturbance to Moderate	−31
Bone Mineral Density		Moderate mood disturbance to none	−26
Increase in BMD (gain of more than 0.5)	−32	Moderate to Minor mood disturbance	−15
No change in BMD (+/−0.5)	0	Minor mood disturbance to none	−11
Decrease in BMD (loss of more than 0.5)	32	No change	0
		No mood disturbance to Minor	11
Steroid Myopathy		Minor mood disturbance to Moderate	15
Moderate weakness to none	−41	No mood disturbance to Moderate	26
Moderate to Minor weakness	−29	Moderate mood disturbance to Major	31
Minor weakness to none	−12	Minor mood disturbance to Major	46
No significant change	0	No mood disturbance to Major	57
None to Minor weakness (without functional limitation)	12		
Minor to Moderate weakness	29	***Cognitive impairment***	
None to Moderate weakness (with functional limitation)	41	Moderate cognitive impairment to none	−20
		Moderate to Minor cognitive impairment	−11
Skin Glucocorticoid Toxicity		Minor cognitive impairment to none	−9
Acne		No change	0
Moderate acne to none	−12	No cognitive impairment to Minor	9
Moderate to Minor acne	−8	Minor cognitive impairment to Moderate	11
Minor acne to none	−4	No cognitive impairment to Moderate	20
No change	0		
None to Minor acne	4	**Infections**	
Minor to Moderate acne	8	Decrease in Infections – Grade 3 or 4 infection to None	−48
None to Moderate acne	12	Decrease in Infections – Grade 3 or 4 infection to Specific infection (<Grade3)	−34
		Decrease in Infections – Specific infection (<Grade3) to None	−14
Easy Bruising		No significant change	0
Moderate easy bruising to none	−11	Increase in Infections – None to Specific infection (<Grade3)	14
Moderate to Minor easy bruising	−9	Increase in Infections – Specific infection (<Grade3) to Grade 3 or 4 infection	34
Minor easy bruising to none	−2	Increase in Infections – None to Grade 3 or 4 infection	48
No change	0		
None to Minor easy bruising	2		
Minor to Moderate easy bruising	9		
None to Moderate easy bruising	11		

multi-criteria decision analysis, employed in the development of the adult GTI [7,8,16], has also been used in the development of weighted classification criteria for rheumatoid arthritis, systemic sclerosis, gout, IgG4-related disease, lupus, and other conditions [17–20]. Experts voted

Table 2
Damage checklist.

	At Baseline or Before	NEW since Baseline
Body Mass Index		
- An absolute increase in BMI of more than 8 units (and >24.9 kg/m^2)[BMI z-score]		
Growth Velocity		
- Major decrease in growth velocity, defined as a decrease in height SDS by more than 2.0 standard deviations		
Pubertal Delay / Sex Hormone Axis Interruption		
- If post-pubertal, maintenance of the same Tanner stage for more than one year		
- Delayed start of puberty		
- New-onset secondary amenorrhea or oligomenorrhea since the start of glucocorticoid therapy		
Glucose Tolerance		
- Diabetic nephropathy		
- Diabetic neuropathy		
- Diabetic retinopathy		
Endocrine		
- Symptomatic adrenal insufficiency		
Blood Pressure / Vascular		
- Hypertensive emergency		
- Posterior reversible encephalopathy syndrome		
Bone Health		
- Osteonecrosis of one site		
- Osteonecrosis of more than site		
- Bone mineral density decrease $\geq$ 1 standard deviation (z-score)		
- Insufficiency fracture		
- Insufficiency fracture in more than one bone		
Muscle & Tendon		
- Major glucocorticoid myopathy		
- Tendon rupture		
- More than one tendon rupture		
Skin		
- Major skin toxicity		
Neuropsychiatric		
- Sleep disturbance: Severe sleep disturbance with latency >60min, >4 night awakenings, total sleep <8 h (3–5yrs) <7 hrs (6+ yrs)		
- Mood regulation: Persistent irritability, depression with loss of activity, suicidal, or elevated mood, irrational ambitions		
- Cognitive impairment: Substantial learning difficulties leading to impaired educational progress		
- Psychosis: Clear psychotic features, with persistent hallucinations, thought disorder, hypomania (grandiosity)		
- Glucocorticoid-induced violence toward self or others		
Infection		
- Grade 5 infection (death from infection)		
Eye		
- Central serous retinopathy		
- New-onset or worsened elevation of intra-ocular pressure requiring treatment or change in treatment		
- Posterior subcapsular cataracts (or history of same)		
Gastrointestinal Tract		
- Gastrointestinal perforation (occurring in the absence of regular nonsteroidal anti-inflammatory drug use)		
- Peptic ulcer disease confirmed by endoscopy (excluding *H. pylori*) or severe dyspeptic symptoms despite treatment		

on a series of pairwise decisions about hypothetical cases, each defined by two criteria from two domains. An example is shown in Supplement Fig. 1 (Appendix).

Evaluation process

The performance of the pGTI on recorded cases was evaluated among the participating experts to assess: (1) the consistency of application (reliability); (2) the degree to which the instrument reflected experts' clinical judgment of relative GC toxicity (face validity).

Following the face-to-face meeting, investigators completed an online exercise with two tasks. First, the investigators scored a second set of 15 patient cases with a specific clinical scenario of GC exposure and GC-related toxicities using the pGTI. Inter-rater reliability was assessed in this context. Second, the investigators ranked the 15 patient cases in order of greatest to least change in GC-toxicity. Each case was assigned a unique non-alphanumeric identifier to reduce potential bias that can be present when alphanumeric labeling is utilized. The investigators were asked to use their overall clinical judgment and remained blinded to the domain weight scores of the cases. The investigators' rankings were then compared to the ranking assigned by the weighted pGTI.

Development of the pGTI digital application

The Scientific Committee recognized that assessments of GC toxicity in pediatric patients have the potential to become either: (1) impractically complex, thereby rendering them unusable; or, (2) overly simplistic and incomplete, such that crucial GC toxicities are not addressed. To address the challenge of complexity, the investigators planned for development of a digital platform for the pGTI that would ensure ease-of-use in data capture and facilitate the calculations critical to understanding of changes in GC toxicity over time. The recording of the data for each pGTI domain is simple: the domains consist of toxicities that clinicians are familiar with and can grade easily. The application of weights to each toxicity and the algorithms that yield the pGTI scores are, however, complex: they require consideration of the impact of medication use on the overall toxicity yielded by the Blood Pressure domain, for example, as well as the impact of changes related to age, sex, and growth on other domains. The creation of a digital platform was the logical solution to these complexities.

A cloud-based digital interface was developed in partnership with the ADK Group (Boston, Massachusetts). With the digital pGTI, investigators are required to record only raw data (e.g., height, weight, hemoglobin A1c value) for each domain. The platform tabulates, analyzes, and scores the pGTI in real time (see video demonstration in Appendix). An example of the investigator inputs and the resulting pGTI outputs is shown in Fig. 3. The pGTI algorithms assign weights to each item of toxicity so that investigators need to enter those values.

Data from the patient visit can be recorded quickly, efficiently, and directly – either in the application or on case-report forms that can be uploaded later for analysis. The pGTI information from a large data set can also be bulk-uploaded into the platform from another database (e.g.,

Item selection for the pGTI was based on four main principles:

1. Likelihood of occurrence of at least 5% in a patient over the course of a clinical trial (typically considered to range between 6 months and 3 years in duration);
2. Independence of the items;
3. Item equivalence (i.e., several measures of glucocorticoid toxicity could be included within a single item, provided they were within the same clinical domain and were equivalent in their degree of toxicity); and,
4. The toxicity was more likely to be due to the effect of glucocorticoid therapy than to the disease itself or to the background rate of that particular morbidity.

Fig. 2. Principles guiding selection of the pGTI Items.

pGTI	Clinical Inputs		pGTI Domain Scores	
Module	**Baseline Visit 1**	**Follow-up Visit 2**	**Worsening Scores**	**Improvement Scores**
Height (m)	1.53	1.59	0	-21
Weight (kg)	78	74.7	0	0
Z Score for Growth Velocity	-1.12	-0.21	0	-49
Blood pressure (mmHG)	144/92	135/88	0	-29
Hemoglobin HbA1c (%)	6.4	6.1	0	-26
LDL (mmo/L)	3.2	3.3	0	0
BP medication (mg)	20	10	0	0
Glucose medication (mg)	10	0	0	0
Lipid medication (mg)	N/A	N/A	0	0
Bone mineral density (g/cm^2)	N/A	N/A	N/A	N/A
Skin	Moderate	Minor	0	-12
Neuropsychiatric	Moderate	Minor	0	-13
Infection	Grade 1	None	0	0
Glucocorticoid Myopathy	Moderate	Minor	0	-29
pGTI Toxicity Profile			0	-179
			CWS	AIS

Fig. 3. Example of data entered by the clinician and the resulting pGTI output.

an overarching clinical trial data base), a statistical software package, or a spreadsheet. Integration with other databases (e.g., electronic data capture or electronic medical records systems) can further increase speed and efficiency and eliminate user error by auto-populating data routinely gathered at the patient visit. As examples, the patients' vital signs can be pulled into the pGTI from existing electronic data capture systems. Final scores and individual domain data can be downloaded for analysis and data reconciliation.

Statistical analysis

Interrater reliability among 14 raters on toxicity items in the pGTI of

the 15 recorded cases was assessed using the Kappa statistic. Agreements between rankings assigned by the weighted pGTI and expert clinical judgment rankings were also 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).

Results

Item inclusion, definition, and grouping

Ten pGTI domains, including Skin and Neuropsychiatry subdomains, incorporate 107 toxicity items (Table 1). Thirteen domains and 32 items comprise the Damage Checklist (Table 2). Every pGTI domain addresses the possibility that either improvement or worsening in GC-related toxicity can occur between visits. Improvement and worsening of toxicity within a given domain are accorded equal weights in the pGTI. Specifically, improvement within a domain between two pGTI time points is weighted equivalently to worsening of the same severity. As an example, an improvement of GC myopathy from moderate to minor has the same absolute weight as worsening of GC myopathy from minor to moderate: −29 points in the setting of improvement, and +29 points in the setting of worsening. Higher pGTI scores represent higher levels of GC toxicity (worsening). Lower pGTI scores represent lower levels of GC toxicity (improvement).

Assigning weights to domains

Thirteen experts attended the face-to-face meeting at which the weighting exercise was conducted. Based on the number of domains and toxicity items in those ten domains combined, there were 46,848 possible paired patient scenarios in which the hypothetical patient could differ in two toxicity domains. The investigators completed 158 scenarios, reaching agreement on all combinations. The remaining 46,690 scenarios were resolved implicitly by the transitivity principle within the 1000Minds software. The relative weights, generated by 1000Minds, are presented in Table 1.

The development of a major mood disturbance is accorded the highest weight of any GC toxicity (+57 points), followed by an increase in BMI by 5 or more units (+52 points), and the development of a Grade 3 or 4 infection (+48 points). Among the lower-weighted items are minor easy bruising (+2 points), minor acne (+4 points), and minor GC myopathy (+12 points).

Evaluation of the pGTI

To assess reliability in the evaluation phase of pGTI development, participants assigned toxicity items on the pGTI to 15 recorded cases. Raters reached a high degree of agreement, with a kappa of 0.89 ($p < 0.01$). To assess pGTI validity, participants ranked 15 cases in order of highest to lowest toxicity. Expert rankings were compared with rankings according to the pGTI score (from highest to lowest), also yielding excellent agreement with a weighted kappa of 0.86 ($p < 0.01$).

Scoring the pGTI: the cumulative worsening score and the aggregate improvement score

The pGTI measures the change in GC toxicity rather than static or absolute GC toxicity. Thus, evaluation at two time points is required for the scoring. Two complementary scoring algorithms provide information about the ability of a treatment approach to reduce GC toxicity. These scores optimize the granularity with which the pGTI captures GC toxicity. The analytic approach derives two scores that are calculated from the same data: the Cumulative Worsening Score (CWS) and the Aggregate Improvement Score (AIS).

Cumulative worsening score

For trials in some diseases, it may be most important to document any cumulative GC toxicity that occurs over the course of a trial or clinic treatment, even if some toxicities are transient. The CWS is designed to assess cumulative GC toxicity, regardless of whether the toxicity is permanent or transient. New toxicities that occur are added, and even toxicities that resolve remain in the score. Thus, the CWS serves as a lasting record of new GC toxicity that occurs after the baseline evaluation. The CWS can only remain the same or increase over time. In a clinical trial, if an investigational agent is effective at lowering GC toxicity, the CWS will be lower in the investigational treatment arm.

Aggregate improvement score

The AIS is important in establishing that the new therapy is effective at lowering any baseline GC toxicity. This is because many patients may already have evidence of GC toxicity at the time of starting a new therapy, as a result of previous GC treatment. With the AIS, toxicities that resolve are removed. In a clinical trial, if an 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.

Discussion

We describe a multi-specialty effort to develop the pGTI, a systematic, standardized, and weighted outcome measure of GC toxicity over time. A valid measure of GC-related toxicity is crucial to assessing the true utility of immunomodulatory agents that have emerged across pediatric specialties [21]. The ability to measure the impact of novel medications designed to be “steroid-sparing” will permit investigators, clinical trial sponsors, and regulatory authorities to judge the medical utility and the economic value of such agents. Furthermore, the ability to measure improvement and worsening GC toxicity between patient visits in the clinic setting will guide pediatricians in tracking side-effects and tapering GCs to the lowest effective level.

Several challenges in measuring GC toxicity were addressed by development of an instrument that includes two components – the pGTI (a quantitative, weighted, core component) and the Damage Checklist (a qualitative, unweighted ancillary component). The pGTI is designed to differentiate between high and low GC exposure over the periods typical of most clinical trials. The pGTI domains were chosen because they represent common, important GC toxicities and because they are dynamic, worsening or improving with variations in GC doses. The Damage Checklist captures GC toxicities that are often severe and highly impactful but which generally represent irreversible damage. Both the pGTI and Damage Checklist, therefore, capture elements of GC toxicity that are essential to understanding the full spectrum of adverse effects related to GC.

The separation of Damage Checklist items from the pGTI has several distinct advantages. First, many Damage Checklist items reflect toxicity from longstanding GC use. They therefore represent GC-related damage from previous treatment and do not improve or worsen with varying GC doses. Separating these toxicities from the weighted pGTI core avoids confusion that may arise from the inclusion of toxicities that are not affected by new treatments. Second, some items in the Damage Checklist require imaging or invasive procedures for measurement, which would have made the administration of the pGTI in an unbiased and systematic manner challenging.

The assignment of relative weights to each toxicity item in the pGTI is a major strength of the instrument. These weights were assigned through a process that utilized multi-criteria decision analysis [21]. Using the pair-wise, “forced-choice” methods within the 1000Minds platform, we were able to rank all possible combinations in their order of their toxicity severity, thereby deriving a point system that reflects the relative weight of each toxicity. This systematic method carries a lower

cognitive burden than group consensus methods such as Delphi exercises, in which investigators are asked to order all possible toxicities or assign relative weights. Another strength of the pGTI is that it uses data that are easy to collect and, in fact, are often already collected in the context of clinic visits.

The results of our initial evaluation exercise suggest excellent reliability. The pGTI item definitions provide sufficient specificity to allow different clinicians to assign toxicity levels accurately. The pGTI digital platform augments the reliability of the instrument further, because definitions of the individual toxicities are built into the digital instrument and are therefore easily accessible to the clinician or clinical investigator. The clinicians examine the patient and record raw data for each domain. In fact, the data for several pGTI domains can be recorded by either research nurses or research coordinators. The digital application assigns the correct toxicity items and weights based on the data input. Such a digital interface, integrated within a larger clinical data base, can also capitalize on application programming interfaces (APIs) that automate data capture and storage within a centralized database. A simpler approach is to collect data on case-report forms and then to input or upload data later into the pGTI.

The assessment of GC toxicity in children is more complex than in adults, for several reasons. Assessments in children must consider the impact of both age and sex reference ranges on certain toxicity domains. In addition, one of the most worrisome effects of GC on children is their impact on children's growth. GCs can also lead to adverse effects on pubertal development. The challenges of measuring GC toxicity in children led to several important differences between the pGTI and the adult GTI. First, the pGTI includes an entirely new domain: growth. Second, the pGTI algorithms take into account pediatric height and sex-related reference ranges for blood pressure and age-related references for other parameters. Third, the more nuanced neuropsychiatric domain of the pGTI encompasses aspects of neurodevelopment that are unique to the pediatric population. Finally, there are slight differences in the numeric value of weights for comparable domains across the pGTI and adult GTI. This reflects differences in the relative importance of specific GC-related toxicities across the two populations, but overall the trends in weights are quite similar.

Our study has limitations. The performance of the pGTI has been evaluated only in the context of recorded cases to date. For the pGTI to be truly valid, it must be able to discriminate between high and low GC use when assessed in prospective clinical studies. Such investigations have been conducted with the adult GTI [6–8] and are now under way in a multi-center clinical trial in Kawasaki disease [22]. Such additional studies will also be required to understand the MCID for the pGTI and the correlation with patient-reported outcome measures. Analyses of the adult GTI have compared treatment groups over a range of adult GTI thresholds, demonstrating consistent and convincing differences between groups [8].

The pGTI is currently being employed as an important secondary outcome in the Kawasaki Disease Coronary Artery Aneurysm Prevention (KD-CAAP) trial (ISRCTN7198747), a multi-center randomized clinical trial now under way at 58 planned sites in the UK and Europe. KD-CAAP is exploring the efficacy and safety of prednisolone and intravenous globulin (IVIG) versus IVIG alone for the prevention of coronary artery aneurysms. The investigators collect data required for the pGTI domains on paper case report forms (available upon request from PAB). These data will be bulk-uploaded from the clinical trial data base into the pGTI digital platform after the database has been locked. Twenty-four patients have been enrolled in the trial to date. Trial investigators thus far have reported that the tool was conceptually simple to grasp following a 15 min training on-line presentation. The case report forms have been simple to complete, facilitated by the fact that item definitions are embedded directly into the forms rather than being maintained in a separate reference appendix.

In conclusion, we describe the development and initial evaluation of the pGTI, a GC toxicity assessment instrument intended for use in prospective, randomized clinical trials in pediatrics and in pediatric practice. This instrument can be employed across clinical disciplines to assess the clinical and economic value of GC-sparing therapies, as well as to measure the impact of GC toxicity. Given the widespread use of GCs and the accelerating pace of immunological drug discovery, this instrument may represent a substantial advance in our ability to assess the utility of new pharmacologic agents.

Financial disclosure statement

The authors have no financial relationships relevant to this article to disclose.

Note: Dr. Stratakis, an employee of Elpen, Inc., and Dr. Webb, an employee of Novartis Pharma AG, held academic positions during the time of this work.

Funding

Development of the pGTI was supported entirely by Dr. Stone's fund at the Massachusetts General Hospital. The intellectual property of the pGTI is owned by the General Hospital Corporation (Massachusetts General Hospital, Boston, Massachusetts).

Definitions

Hypertensive emergency: The blood pressure has reached levels that are damaging organs, or are at sufficiently high levels that risk of organ injury is imminent if untreated. Hypertensive emergencies generally occur at blood pressure levels significantly in excess of the 99th percentile for age, sex and height percentile but can occur at even lower levels in patients whose blood pressure have not been elevated before. Complications can include: seizure, confusion, stroke, loss of consciousness, memory loss, myocardial infarction, hypertensive retinopathy or nephropathy, or pulmonary edema.

Posterior reversible leukoencephalopathy syndrome (PRES): A clinico-radiological entity characterized by headaches, altered mental status, seizures, and visual loss. PRES is associated with white matter vasogenic edema that predominantly affects the posterior occipital and parietal lobes of the brain, but other brain regions may also be affected. Confirmation by MRI is required.

Severe glucocorticoid myopathy: Grade 3 weakness or respiratory accessory muscle use caused by glucocorticoid myopathy.

Central serous retinopathy: a fluid detachment of macula layers from their supporting tissue. Requires formal ophthalmology examination, typically accompanied by optical coherence tomography and/or fluorescein angiography for diagnostic confirmation.

Grade 4 infection: Life-threatening consequences (e.g., septic shock, hypotension, acidosis, necrosis).

Diabetic nephropathy: Macroalbuminuria; i.e., a urinary albumin excretion $>$ 300 mg in a 24 h collection or a urinary protein: creatinine ratio $>$ 300mg/g.

Diabetic neuropathy: Any of four types of peripheral neuropathy occurring in the setting of diabetes mellitus, namely: 1) a distal sensory polyneuropathy; 2) autonomic neuropathy (hypoglycemia unawareness, bladder or bowel problems, erectile dysfunction, and other autonomic nervous system issues); 3) diabetic amyotrophy (muscle infarction); or 4) mononeuritis (e.g., a sixth cranial nerve palsy or foot-drop attributed to diabetes).

Diabetic retinopathy: Any form of retinopathy associated with diabetes mellitus, including both non-proliferative and proliferative forms of diabetic retinopathy as well as diabetic macular edema. These complications must be confirmed by an ophthalmologist.

Major skin toxicity: Any of the three following manifestations:

Grade 4 acneiform lesions - Papules and/or pustules covering any % BSA, which may or may not be associated with symptoms of pruritus or tenderness and are associated with extensive superinfection with IV

antibiotics indicated or life- threatening consequences

Grade 3 striae - Covering >30% BSA or associated with ulceration

Grade 3 ulcers – Combined area of ulcers >2 cm or full-thickness skin loss involving damage to or necrosis of subcutaneous tissue that may extend down to fascia

CRediT authorship contribution statement

Paul Brogan: Visualization, Resources, Writing – review & editing. **Ray Naden:** Methodology, Resources, Writing – review & editing. **Stacy P. Ardoin:** Visualization, Resources, Writing – review & editing. **Jennifer C. Cooper:** Visualization, Resources, Writing – review & editing. **Fabrizio De Benedetti:** Visualization, Resources, Writing – review & editing. **Jean-Francois Dicaire:** Methodology, Writing – review & editing. **Despina Eleftheriou:** Visualization, Resources, Writing – review & editing. **Brian Feldman:** Visualization, Resources, Writing – review & editing. **Jon Goldin:** Visualization, Resources, Writing – review & editing. **Seth E. Karol:** Visualization, Resources, Writing – review & editing. **Fiona Price-Kuehne:** Visualization, Resources, Writing – review & editing. **David Skuse:** Visualization, Resources, Writing – review & editing. **Constantine A. Stratakis:** Visualization, Resources, Writing – review & editing. **Nicholas Webb:** Visualization, Resources, Writing – review & editing. **John H. Stone:** Conceptualization, Visualization, Funding acquisition, Resources, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors have no conflicts of interest relevant to this article to disclose.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2022.152068.

References

[1] Hench PS, Kendall EC, Slocumb CH, et al. The effect of a hormone of the adrenal cortex (17-hydroxy-11-dehydrocorticosterone; compound E) and of pituitary adreenocorticotropic hormone on rheumatoid arthritis. Preliminary report. Proc Staff Mayo Clin 1949;24:181–97.
[2] McDonough AK, Curtis JR, Saag KG. The epidemiology of glucocorticoid-associated adverse events. Curr Opin Rheumatol 2008;20(2):131–7.
[3] Sarnes E, Crofford L, Watson M, et al. Incidence and US costs of corticosteroid-associated adverse events: a systematic literature review. Clin Ther 2011;33(10):1413–32.
[4] Bel EH, Wenzel SE, Thompson PJ, et al. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. N Engl J Med 2014;371(13):1189–97.
[5] Fortunet C, Pers Y-M, Lambert J, et al. Tocilizumab induces corticosteroid sparing in rheumatoid arthritis patients in clinical practice. Rheumatology 2014;54(4):672–7 (Oxford).
[6] Jayne DRW, Merkel PA, Schall TJ, Bekker P, Study Group A. Avacopan for the treatment of ANCA-associated vasculitis. N Engl J Med 2021;384:599–609.
[7] 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 2021:365–372.e5.
[8] Stone JH, McDowell PJ, Jayne DRW, et al. The glucocorticoid toxicity index: measuring change in glucocorticoid toxicity over time. Semin Arthritis Rheum 2022;55:152010.
[9] Flynn JT, Kaelber DC, Baker-Smith CM, et al. Clinical practice guideline for screening and management of high blood pressure in children and adolescents. Pediatrics 2017;140(3):e20171904.
[10] Weinstein RS. Clinical practice. Glucocorticoid-induced bone disease. N Engl J Med 2011;365(1):62–70.
[11] Mushtaq T, Ahmed SF. The impact of corticosteroids on growth and bone health. Arch Dis Child 2002;87:93–6.
[12] Ole D, et al. the timing of administration of exogenous glucocorticoid affects 24 hour growth hormone secretion in children. Growth Horm IGF Res 2017;35:40–4.
[13] David H, Aupiais C, Louveau B, et al. Growth outcomes after growth hormone therapy of patients given long-term corticosteroids for juvenile idiopathic arthritis. J Clin Endocrinol Metab 2017;102:4578–87.
[14] De Benedetti F, Brunner H, Ruperto N, et al. Catch up growth during Tocilizumab therapy for systemic juvenile idiopathic arthritis. Arthritis Rheumatol 2015;67(3):840–8.
[15] Hansen P, Ombler F. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. J Multi Criteria Decis Anal 2008;15:87–107.
[16] Miloslavsky EM, Naden RP, Bijlsma JW, et al. Development of a glucocorticoid toxicity index (GTI) using multicriteria decision analysis. Ann Rheum Dis 2017;76(3):543–6.
[17] Aringer M, Costenbader K, Daikh D, et al. European league against rheumatism/American college of rheumatology classification criteria for systemic lupus erythematosus. Ann Rheum Dis 2019;78:1151–9. 2019.
[18] 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.
[19] Neogi T, Jansen T, Dalbeth N, et al. Gout classification criteria: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis 2015;74(10):1789–98. 2015.
[20] 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.
[21] Food and Drug Administration. Guidance for industry systemic lupus erythematosus — developing medical products for treatment, Food and Drug Administration.
[22] Kawasaki disease coronary artery aneurysm prevention (KD-CAAP) trial (ISRCTN7198747).