In silico validation of the Autoinflammatory disease damage index

Authors

Nienke M ter Haar,[1,2] Amber LJ van Delft,[3] Henk van Stel[3] Sulaiman M Al-Mayouf,[4] Gayane Amaryan,[5]
Jordi Anton,[6] Karyl S Barron,[7] Susanne M Benseler,[8] Paul A Brogan,[9]
Luca Cantarini,[10] Marco Cattalini,[11] Alexis-Virgil Cochino,[12] Fabrizio De Benedetti,[13]
Fatma Dedeoglu,[14] Adriana A De Jesus,[15]
Erkan Demirkaya,[16] Pavla Dolezalova,[17] Karen L Durrant,[18] Giovanna Fabio,[19]
Romina Gallizzi,[20] Raphaela Goldbach-Mansky,[15] Eric Hachulla,[21]
Veronique Hentgen,[22] Troels Herlin,[23] Michaël Hofer,[24,25] Hal M Hoffman,[26]
Antonella Insalaco,[27] Annette F Jansson,[29] Tilmann Kallinich,[29] Isabelle Koné-Paut,[30]
Anna Kozlova,[31] Jasmin B Kuemmerle-Deschner,[32] Helen J Lachmann,[33]
Ronald M Laxer,[34] Alberto Martini,[35] Susan Nielsen,[36] Irina Nikishina,[37]
Amanda K Ombrello,[38] Seza Ozen,[39] Efimia Papadopoulou-Alataki,[40]
Pierre Quartier,[41] Donato Rigante,[42] Ricardo Russo,[43] Anna Simon,[44]
Maria Trachana,[45] Yosef Uziel,[46] Angelo Ravelli,[47] Grant Schulert [48] Marco Gattorno,[47] Joost Frenkel[3]

Author affiliations
1. Laboratory for Translational Immunology, University Medical Centre Utrecht, Utrecht, The Netherlands
2. Department of Paediatric Immunology, University Medical Centre Utrecht, Utrecht, The Netherlands
3. Department of Paediatrics, University Medical Centre Utrecht, Utrecht, The Netherlands
4. Department of Paediatrics, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia
5. National Paediatric Centre for Familial Mediterranean Fever and Gastroenterology Service, Arabkir Medical Centre-Institute of Child & Adolescent Health, Yerevan, Armenia
6. Paediatric Rheumatology Unit, Hospital Sant Joan de Déu, Barcelona, Spain
7. ivision of Intramural Research and National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, USA
8. Department of Paediatrics and Department of Rheumatology, Alberta Children's Hospital, Calgary, Canada
9. Department of Infection, Inflammation and Rheumatology, University College London Institute of Child Health, London, UK
10. Department of Medical Sciences, Surgery and Neurosciences, Rheumatology Unit, University of Siena, Siena, Italy
11. Paediatric Clinic, University of Brescia and Spedali Civili di Brescia, Brescia, Italy
12. Paediatrics Department, National Institute for Mother and Child Health Alessandrescu-Rusescu, Bucharest, Romania
13. Division of Rheumatology, Ospedale Pediatrico Bambino Gesù, Rome, Italy
14. Division of Immunology, Rheumatology Program, Boston Children's Hospital, Harvard Medical School, Boston, USA
15.Translational Autoinflammatory Disease Section, NIAID, National Institutes of Health, Bethesda, USA
16. Department of Paediatric Rheumatology, Gulhane Military Medical Faculty, Ankara, Turkey
17. Department of Paediatrics and Adolescent Medicine, Charles University, General University Hospital, Prague, Czech Republic
18. Autoinflammatory Alliance, San Fransisco, USA
19. Department of Internal Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy
20. Department of Paediatrics, Rheumatology, AOU G Martino, Messina, Italy
21. Département de Médecine Interne et Immunologie Clinique, Université de Lille, Lille, France
22. Reference centre for autoinflammatory diseases (CEREMAI), Versailles Hospital, Le Chesnay, France

23. Department of Paediatrics, Aarhus University Hospital, Aarhus, Denmark
24. Department of Paediatric Rheumatology, University of Lausanne, Lausanne, Switzerland
25. Department of Paediatric Rheumatology, University Hospital of Geneva, Geneva, Switzerland
26. Department of Paediatrics, University of California, San Diego, USA
27. Dipartimento di Medicina Pediatrica, IRCCS Ospedale Pediatrico Bambino Gesù, Rome, Italy
28. Department of Rheumatology&Immunology, Dr. von Hauner Childrens Hospital, Ludwig-Maximilians-University, Munich, Germany
29. Paediatric Pneumology and Immunology and Interdisciplinary Centre for Social Paediatrics, Charité University Medicine Berlin, Berlin, Germany
30. Department of Paediatric Rheumatology and CEREMAI, Bicêtre Hospital, APHP, University of Paris Sud, Paris, France
31. Department of Immunology, Federal Research and Clinical Centre for Paediatric Haematology, Oncology and Immunology, Moscow, Russia
32. Division of Paediatric Rheumatology, Department of Paediatrics, University Hospital Tuebingen, Tuebingen, Germany
33. Division of Medicine, University College London, London, UK
34. Department of Paediatrics and Medicine, University of Toronto and the Hospital for Sick Children, Toronto, Canada
35. Direzione Scientifica, G Gaslini Institute, Genova, Italy
36. Paediatric Rheumatology unit 4272, Rigshospitalet, Copenhagen, Denmark
37. Department of Paediatric Rheumatic diseases, V.A. Nasonova Research Institute of Rheumatology, Moscow, Russia
38. Inflammatory Disease Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, USA
39. Department of Paediatric Rheumatology, Hacettepe University, Ankara, Turkey
40. Fourth Department of Paediatrics, Aristotle University of Thessaloniki, Thessaloniki, Greece
41. Department of Paediatric Immunology-Hematology and Rheumatology Unit and IMAGINE Institute, Institution Necker-Enfants Malades Hospital and Paris-Descartes University, Paris, France
42. Institute of Paediatrics, Fondazione Policlinico Universitario A. Gemelli, Università Cattolica Sacro Cuore, Rome, Italy
43. Servicio de Inmunología y Reumatología, Hospital de Pediatría Garrahan, Buenos Aires, Argentina
44. Internal Medicine, Radboud Expertise Centre for Immunodeficiency and Autoinflammation, Radboud University Medical Centre, Nijmegen, The Netherlands
45. Paediatric Immunology and Rheumatology Referral Centre, first Paediatric clinic, Aristotle University of Thessaloniki, Thessaloniki, Greece
46. Department of Paediatrics, Meir Medical Centre, Kfar Saba, Tel Aviv University, Sackler School of Medicine, Tel Aviv, Israel
47. Institution Università degli Studi di Genova and G. Gaslini Institute, Genova, Italy
48. Pediatric rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, United States of America

The authors thank Dr Nicolino Ruperto and the PRINTO's staff for their precious collaboration.

Contributors NMtH and ALJvDA are joint first authors. MG and JF are joint last authors. NMtH, ALJvD and JF designed the study and wrote the manuscript. All other authors contributed to the online surveys and/or the consensus meeting, and attributed to and approved the manuscript.

Funding The project was supported by ERANET-PRIOMEDCHILD RaDiCEA Project No. 40-41800-98-007. The Eurofever Registry was funded by the Executive Agency for Health and Consumers (EAHC, Project No. 2007332). The work was supported by an unrestricted grant by Novartis Pharma AG.

Competing interests Novartis Pharma AG financially supported meetings with the methodologist. They did not have any influence on the selection of participants or on the content of the ADDI/consensus meeting or the reporting of the findings. PB: Consultancy/speaking fees Novartis, Roche, SOBI, UCB. FdB: Novartis, Novimmune, Hoffmann-La Roche, SOBI, AbbVie. LC: speaker's fee for Novartis and SOBI. MC: consultancy fees for Novartis, SOBI and Abbvie. KLD: consultancy work for SOBI and Novartis, donations, honorariums and unrestricted grants have been received by the Autoinflammatory Alliance from SOBI, Novartis, and Regeneron. RG: consultant for Abbvie. RGM: study support from SOBI, Novartis, Regeneron. VH: honorariums and educational grants from Novartis, honorariums from SOBI. MH: consultant for Novartis. HMH: consultant for Novartis and SOBI, and speaker for Novartis. TK: research grant by Novartis, speaker's bureau by Roche, BMS, Novartis and SOBI. JKD: consultant/speaker for Novartis and SOBI and has received grant support from SOBI and Novartis. RML: ad board and consultant for Abbvie and Novartis. PQ: investigator, consultant and speaker's bureau for Novartis and SOBI. MG: consultant for and unrestricted grants to Eurofever and speaker's fee from SOBI and Novartis. YU: Y. Uziel Grant/Research Support from Novartis, Consultant for Novartis, Speaker Bureau of Abbvie, Neopharm, Novartis, Roche. JF: consultant for Novartis.

Ethics approval The Medical Ethical Committee of the University Medical Centre Utrecht.

Word count abstract

240

Word count article without figures/tables and supplementary

2879

Abstract

Introduction Autoinflammatory diseases can cause irreversible tissue damage due to systemic inflammation. Recently, the Autoinflammatory Disease Damage Index (ADDI) was developed. This is the first instrument to quantify damage in Familial Mediterranean Fever, Cryopyrin Associated Periodic Syndromes, Mevalonate Kinase Deficiency and Tumor Necrosis Factor Receptor Associated Periodic Syndrome. The aim of this study was to validate this tool for its intended use in a clinical/research setting.

Methods The ADDI was scored on 110 semi-fictional cases by at least 3 physicians per case, independently of each other. Face and content validity were assessed by requesting comments on the ADDI. Reliability was tested by calculating the intra-class correlation coefficient (ICC) using an 'observer-nested-within-the-subject' design. Construct validity was determined by correlating the ADDI-score to the physician's global assessment (PGA) of damage and disease activity. Redundancy of individual items was determined with Cronbach's alpha.

Results The ADDI was validated on a total of 110 paper clinical cases by 37 experts in autoinflammatory disease. This yielded an ICC of 0.838 (95% CI 0.782-0.888). The ADDI score correlated strongly with PGA-damage (R=0.92, 95% CI 0.88-0.95), but was not strongly influenced by disease activity (R=0.395, 95% CI 0.209-0.553). After comments from the experts, some item definitions were refined. The inter-item correlation in all different subcategories was lower than 0.7 indicating that there was no redundancy between individual damage items.

Conclusion The ADDI is a reliable and valid instrument to quantify damage in individual patients, and can be used to compare disease outcomes in clinical studies.

List of abbreviations

ADDI	Autoinflammatory Disease Damage Index
AID	Autoinflammatory Disease
AIDAI	AutoInflammatory Disease Activity Index
CAPS	Cryopyrin Associated Periodic Syndromes
CI	Confidence interval
FCAS	Familial Cold Autoinflammatory Syndrome
FMF	Familial Mediterranean Fever
ICC	Intra-class correlation coefficient
MA	Mevalonic Aciduria
MKD	Mevalonate Kinase Deficiency
MWS	Muckle Wells Syndrome
NOMID	Neonatal-Onset Multisystemic Inflammatory Disease
PGA	Physician Global Assessment
TRAPS	Tumor Necrosis Factor Receptor Associated Periodic Syndrome

INTRODUCTION

Autoinflammatory diseases (AIDs) are characterized by seemingly unprovoked, recurrent episodes of inflammation caused by activation of the innate immune system. The four main monogenic AIDs are cryopyrin-associated periodic syndromes (CAPS), tumor necrosis factor-associated periodic fever (TRAPS), mevalonate kinase deficiency (MKD) and familial Mediterranean fever (FMF). (1),(2). Chronic inflammation in AID's may cause irreversible damage in multiple organ systems, e.g. visual loss, deafness, joint restriction and amyloidosis(3).

Even though targeted therapy for these AIDs has become available (4-6), permanent damage may still accumulate pre-diagnostically or pre-therapeutically in these patients. Furthermore, the majority of studies on new biological therapies for AIDs are recent, with limited follow-up, hence the potency of these drugs to prevent or stop the development of damage is not yet completely known. (3, 7, 8). In order to study the effect of different (new) therapies on the long term burden of AIDs, it is important to have a systematic index to quantify damage. The autoinflammatory disease damage index (ADDI) is such an index. The main purpose of the ADDI is to serve as a comprehensive tool to determine damage in AID patients.(5)

To properly validate a damage index such as the ADDI, several aspects are important: reliability, content validity, face validity, criterion validity and construct validity.(9) A reliable index means that for a given patient, different observers will give the same score; this can be assessed by calculating the inter-observer variability (intra-class correlation coefficient, ICC). Content validity tests whether the content of the index is important for the subject the index applies to. Face validity is an indication whether the index 'seems' valid in practical terms. Criterion validity tests whether an index is as good as the gold standard. Construct validity consists of convergent and divergent validity: convergent validity determines whether an index correlates to a similar index (e.g. whether the ADDI correlates to other indices of damage or impairments in daily living), whereas discriminant validity determines whether the index is different from a dissimilar index (e.g. the ADDI should not correlate to indices of disease activity).

Continuously during development and validation of the ADDI, content validity, face validity and adherence to the OMERACT (The OMERACT Filter for Outcome Measures in Rheumatology, Boers et al.) principles (truth, discrimination and feasibility) were assessed (10). As a gold standard for disease damage in AIDs is lacking, criterion validity cannot be determined. PGA-damage can be considered the best alternative for a gold standard, but as it is not a validated measurement we decided it was more applicable to test construct validity instead of criterion validity. Therefore, in this study we aimed to investigate reliability and construct validity, using semi-fictional paper cases of patients with FMF, CAPS, TRAPS and MKD, designed to ensure that all the damage items were adequately covered.

METHODS

Development of the validation

Together with an experienced methodologist (HS), a validation plan was developed. Semi-fictional cases, which were based on real patient data but modified to protect patient privacy and to ensure that all damage items would be sufficiently present and different degrees of damage could be tested. Using a pilot with a limited number of cases and expert participants, a preliminary ICC was calculated, on which the final number of cases for the full validation exercise was calculated. All experts that participated in the development of the ADDI (top-40 enrollers in the Eurofever registry and 9 experts from the Americas) were invited to participate in the validation of the ADDI. One expert (JF) did not take part in the scoring of the cases as he facilitated development of all the cases.

Development of the cases

The cases for validation of the ADDI were derived from anonymized clinical data from patients with confirmed FMF, CAPS, TRAPS and MKD acquired from the European based online Eurofever registry. (11, 12). All physicians involved in the Eurofever project (reference) as disease experts were asked to complete follow up data on patients they had entered in the Eurofever registry. To also retrieve non-European cases, experts from America were asked to send cases using a pre-formed case template. The patient information retrieved from the Eurofever registry and American cases was rewritten into semi-fictional case scenarios and modified to ensure that all items of the ADDI were represented at least 4 times in the sample. Precautions were made to provide a similar amount of cases for each disease, and to have cases with different grades of disease activity and damage. One expert (JF) checked all cases to verify that the case summaries were realistic and comprehensible.

Case distribution

The case summaries were distributed via a web-based survey, in which experts completed the ADDI, estimated the degree of disease damage and disease activity using a 10-point physician's global assessment (PGA-damage and PGA-activity, respectively) and could give comments. The distribution of cases followed the 'observer-nested-within-subject' design, meaning that a large group of experts all scored a subset of the cases.(9) Each group of 4 experts scored 10 cases, a minimum of 3 doctors was needed per group to calculate the ICC; additional experts were asked to complete the survey when necessary. An equal division of adult and pediatric physicians and distribution among different countries and centers was ensured in each participant group.

Statistical analysis

Statistical analyses were performed in IBM SPSS Statistics version 21. The ICC was determined to assess the reliability of the damage index as a whole, as well as for the eight subcategories and all individual items. It was also assessed for the PGA-damage and the PGA-activity, in order to determine whether these measurements would be sufficiently reliable to test construct validity. An ICC of 0.8 or higher was considered indicative for excellent reliability (9)(13). Cronbach's alpha was used to determine possible redundancy of different items (e.g. whether two items would score the same damage). An inter-item correlation of more than 0.7 was considered to indicate redundancy.(14) A spearman rank test was used to assess discriminant and convergent validity, correlating the ADDI to PGA-activity and PGA-damage, respectively. A spearman rank test with an r from 0.1-0.3 was considered weak, r 0.3-0.5 was considered moderate and r >0.5 was considered strong. (15).

Discussion on the items and definitions

A small team discussed all items with an ICC below 0.7 (NtH, AvD, JF). This discussion encompassed possible explanations for a low score (e.g. unclear definition of an item, or the lack of a growth chart hampering easy scoring of growth failure). Further, based on experts' comments and suggestions during the scoring, possibilities to improve the item and/or definition were discussed. The old and refined items were proposed to all experts via a web-based survey and subsequently discussed in an open face to face meeting at the Pediatric rheumatology congress in Athens (PReS 2017). Consensus was considered achieved if at least 70% of the experts agreed.

RESULTS

Pilot

A pilot study with 15 paper cases was completed by 4 experts. This yielded a preliminary ICC of 0.85 (95% CI 0.70-0.94), which implied that a minimum of 90 cases would be needed for the validation of the ADDI. We therefore decided to assign 110 cases to the experts.

Collection of cases

A total of 120 patients with follow up were identified in the Eurofever registry, and an additional 20 cases were submitted by non-European experts. By selecting and combining case information, a total of 110 cases were compiled from these 140 cases. The final semi-fictional cases included 29 CAPS, 27 TRAPS, 29 FMF and 25 MKD patients.

Validation

A total of 37 of 44 participants responded. In 10 groups at least 3 participants responded, which led to 100 cases that could be used for the analyses; due to insufficient response in one group, 10 cases could not be used. Every item was scored at least 18 times.

ICC

The ICC of the ADDI was 0.84 (95% confidence interval (CI) 0.78-0.89). This indicated good interrater-reliability. The ICC per disease, for different organ systems and the individual damage items are shown in table 1. The highest ICC was found for the item 'hearing loss' (0.861 95%CI 0.812-0.901) exceeding the overall ICC; the lowest ICC was found for the item 'puberty delay' (0.292 95% CI 0.164-0.425).

Table 1 – ICC

	ICC (95% CI)
Overall	0.84 (0.78-0.89)
Per disease	
CAPS	0.82 (0.71-0.91)
TRAPS	0.62 (0.39-0.80)
FMF	0.84 (0,72-0,92)
MKD	0.73 (0.55-0.86)
Per category	
Reproductive	0.67 (0.59-0.76)
Sub/infertility	0.72 (0.63-0.79)
Amenorrhea	0.57 (0.46-0.67)
Renal/amyloidosis	0.88 (0.84- 0.92)
Amyloidosis	0.76 (0.69-0.82)
Proteinuria	0.80 (0.73-0.85)
Renal insufficiency	0.84 (0.78-0.88)
Developmental	0,54 (0,43-0,64)
Growth failure	0.57 (0.46-0.67)
Puberty delay	0.30 (0.17-0.42)
Serosal	
Serosal scarring	0.64 (0.54-0.72)
Neurological	0,75 (0,67-0,81)
Developmental delay	0.48 (0.37-0.60)
Cognitive impairment	0.54 (0.43-0.65)
Elevated ICP	0.65 (0.56-0.74)
CNS involvement	0.67 (0.58-0.75)
Ears	
Hearing loss	0.86 (0.82-0.90)
Ocular	
Ocular	0.74 (0.66-0.80)
Musculoskeletal	0,73 (0,64 – 0,80)
Joint restriction	0.52 (0.41-0.63)
Bone deformity	0.74 (0.66-0.81)
Osteoporosis	0.86 (0.81-0.90)
Musculoskeletal pain	0.47 (0.35-0.58)

Construct validity

The ICC of PGA-damage (0.75, 95% CI: 0.67-0.81) and PGA-activity (0.62, 95% CI: 0.52-0.71) were considered sufficiently reliable to determine construct validity. A strong relation was found between the score of the ADDI and the estimated damage (PGA-damage), with a correlation coefficient higher than 0.9 (Spearman's rho 0.92, 95% CI 0.88-0.95, $p<0.001$, figure 2). This correlation coefficient indicated that an increase in the ADDI score is indeed an increase in the total damage according to the expert physicians' overall opinion. The relation between disease activity (PGA-activity) and the ADDI score was much weaker (Spearman's rho 0.395, 95% CI 0.209-0.553 $p<0.001$, figure 3), indicating that the ADDI was not strongly influenced by disease activity.

Inter item correlation

In order to assess whether items had too much overlap, inter item correlation was determined using Cronbach's alpha. Of specific interest was the inter-item correlation between cognitive impairment (mainly relating to adult patients, or adolescents) and developmental delay (mainly relating to pediatric patients), as the experts worried that these might have too much overlap. The inter-item correlation between cognitive impairment and developmental delay was 0.66 indicating that there was minimal redundancy. All inter-item correlation matrixes can be found in supplementary table 2.

Figure 2 Correlation of PGA-damage to ADDI-score

**Every line represents a case scored by at least 3 experts*

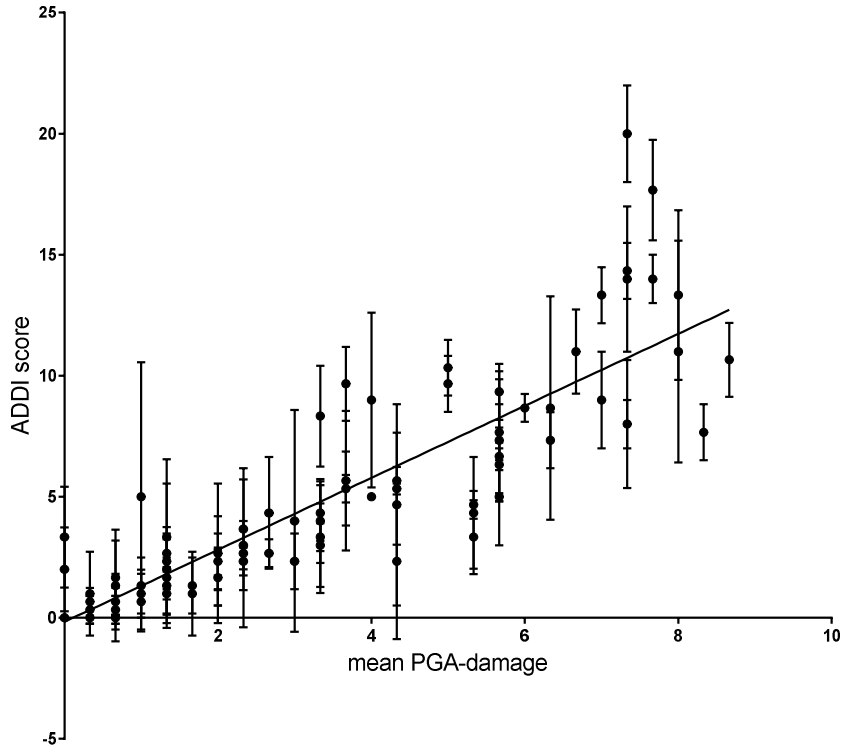
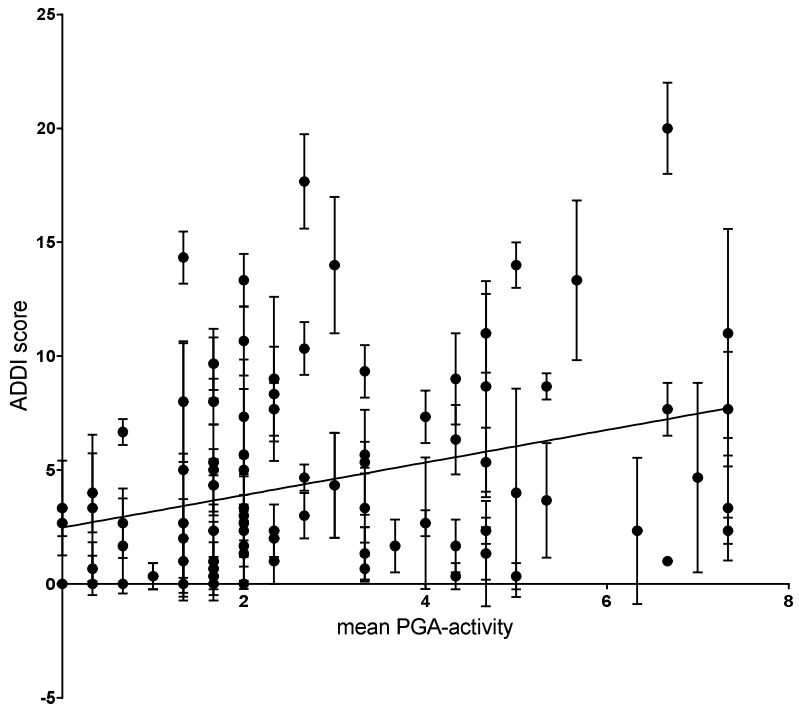


Figure 3 Correlation of PGA-activity to ADDI-score

**Every line represents a case scored by at least 3 experts*



Comments from the experts

The ADDI was considered overall a simple and easily applicable tool. The most important comments given during the survey considered comments and uncertainties about scoring, e.g. due to insufficient information in the case (e.g. the lack of growth charts to completely assess growth failure), unclear definitions in the ADDI (e.g. whether psychiatric comorbidities are part of the item CNS involvement) or doubts about the severity of organ involvement (e.g. severity of visual loss). A full overview of these comments can be found in supplementary 1.

Other important comments considered ADDI content such as scoring of the items (suggesting a higher/lower score weighting), or suggestions to improve item definitions. These suggestions were presented to all expert participants using an online survey, in which experts could consider and respond to these comments. The results of this survey were subsequently discussed in a face to face meeting. Following this meeting, the scoring for the item "reproductive" was changed from 3 to 2. Furthermore, there were slight changes in the definitions for growth failure, central nervous system involvement, joint restriction, puberty delay, and serosal scarring. The revised and now definitive ADDI can be found in figure 4.

All items were considered truthful, discriminative and feasible, however doubts were raised about the reliability and feasibility of the scoring of musculoskeletal pain as there is no easy objective test to assess this. Despite that, it was considered that this particular item was sufficiently valid and important for patients, therefore was kept as part of the ADDI.

Figure 4 Definitive ADDI including glossary of terms.

Preliminary ADDI

Definition of damage: Damage is defined as persistent or irreversible change in structure or function, which is present for at least 6 months. Damage items should not be scored if they are attributed to ongoing disease activity. Damage may be the result of prior disease activity, complications of therapy or co-morbid conditions that developed after the onset of autoinflammatory disease signs and symptoms. If damage has been present for longer than 6 months, but later resolves, it should still be scored in order to capture the damage that was present in the individual for that time period.

Damage item	**Grading**	**Points**
Reproductive		***Max. 2***
Sub/infertility		2
Amenorrhea		1
Renal/amyloidosis		***Max. 6***
Amyloidosis	Limited amyloidosis	2
	Extensive amyloidosis	3
Proteinuria		1
Renal insufficiency	Moderate renal insufficiency	2
	Severe renal insufficiency	3
Developmental		***Max. 3***
Growth failure		2
Puberty delay		1
Serosal		***Max 1***
Serosal scarring		1
Neurological		***Max. 6***
Developmental delay		2
Cognitive impairment		3
Elevated intracranial pressure		2
Central nervous system involvement		3
Ears		***Max. 2***
Hearing loss	Moderate hearing loss of better ear	1
	Severe hearing loss of better ear	2
Ocular		***Max. 3***
Ocular involvement	Mild ocular involvement of better eye	1
	Moderate ocular involvement of better eye	2
	Severe ocular involvement of better eye	3
Musculoskeletal		***Max. 4***
Joint restriction		2
Bone deformity		2
Osteoporosis		1
Musculoskeletal pain		1

Glossary of terms

Amenorrhea: Primary amenorrhea: absence of menarche at the age of 16 years or absence of menarche 5 years after thelarche in a female. Secondary amenorrhea: absence of the menses for six consecutive months or more, in a female who previously had menstrual cycles.

Amyloidosis, Extensive Symptomatic amyloidosis affecting more than one organ and confirmed by examination of tissue sections by Congo red dye or SAP scintigraphy.

Amyloidosis, Limited Symptomatic amyloidosis affecting one organ and confirmed by examination of tissue sections by Congo red dye or SAP scintigraphy.
Bone deformity Bone deformation or overgrowth on clinical examination and/or imaging studies.
Central nervous system involvement[1] Focal deficits (gross and/or fine sensorimotor), diffuse deficits (e.g. memory, behaviour), seizures, and spinal cord symptoms.
Cognitive impairment Requirement of special education because of cognitive impairment or IQ below 70 as defined by neuropsychological assessment (e.g. WISC) or other age-appropriate equivalents.
Developmental delay[2] Failure to reach age-appropriate developmental milestones, including language/speech, motor, social/emotional, and cognitive milestones.
Elevated intracranial pressure[3] Signs and/or symptoms of elevated intracranial pressure supported by appropriate techniques.
Growth failure Defined as the presence of at least two of the three features:
- lower than the 3rd percentile or -2SD height for age
- growth velocity over 6 months lower than the 3rd percentile or -2SD for age
- crossing at least 2 centiles (5%, 10%, 25%, 50%, 75%, 90%, 95%) on growth chart
For patients older than 18 years: Pathological short stature (e.g. below 3rd percentile or -2SD for normal ethnic population)
Hearing loss, moderate Sensorineural hearing impairment confirmed by audiometry or another age appropriate technique without requirement of hearing aids or a cochlear implant
Hearing loss, severe Sensorineural hearing impairment confirmed by audiometry or another age appropriate technique requiring hearing aids or a cochlear implant.
Infertility A disease of the reproductive system defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse, not due to known disorders in the unaffected partner.
Joint restriction Fixed limitation in the normal range of motion of joints affecting function, with or without destructive arthropathy or avascular necrosis.
Musculoskeletal pain Non-inflammatory musculoskeletal pain impairing activities of daily living.
Ocular involvement, mild Ocular damage (e.g. optic nerve atrophy, elevated intraocular pressure or cataract) documented by an ophthalmologist, without visual impairment.
Ocular involvement, moderate Ocular damage (e.g. optic nerve atrophy, elevated intraocular pressure or cataract) documented by an ophthalmologist, resulting in visual impairment.
Ocular involvement, severe Ocular damage (e.g. optic nerve atrophy, elevated intraocular pressure or cataract) documented by an ophthalmologist, resulting in legal blindness.
Osteoporosis Reduced bone mineral density with vertebral collapse and/or pathological fractures confirmed with imaging, which may include bone densitometry. Requires both evidence of decreased bone density and fracture, 'low bone density' by itself is insufficient
Proteinuria Persistent urinary protein to creatinine ratio of >20mg/mmol in the first morning void; and/or a daily protein excretion of > 0.3 g/24 hours, or urine albumin to creatinine ratio of > 15 mg/mmol.
Puberty delay A Tanner stage below minus two standard deviations for age or below the 3rd percentile for age or any tanner stage after pharmacological induction of puberty.
Renal insufficiency, moderate Glomerular filtration rate (GFR) between 15-60 ml/min/1,73m^2.
Renal insufficiency, severe GFR <15 ml/min/1,73m^2, dialysis or transplantation.
Serosal scarring Symptomatic adhesions or fibrosis affecting pericardium, pleura, peritoneum and/or retroperitoneum, supported by imaging techniques, endoscopy or surgery.

[1] Neuropsychiatric disorders unrelated to the disease should not be scored
[2] Only for pediatric patients.
[3] Such as fundoscopy, neuroimaging or lumbar CSF pressure measurement.

DISCUSSION

This validation study indicates that the ADDI is a reliable index to measure damage in the four main monogenic AIDs, with an overall ICC of 0.838. Most items were considered clearly defined and easy to score. An ICC of 0.838 is comparable to other damage indices for rheumatoid diseases such as the Juvenile Arthritis Damage Index (JADI, ICC 0.85-0.97) (16), the Localized Scleroderma Skin Damage Index (LoSDI, ICC 0.99) (17), Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI, ICC 0.86)(18), Vasculitis damage index (VDI, ICC 0.94)(19) and the Combined Damage Assessment (CDA, ICC 0.78)(19).

This is the first validation of a disease damage index for AIDs. A key strength is the high correlation between the ADDI score and the PGA-damage. As this is the closest approximation of a gold standard for damage, this indicates that the ADDI is indeed measuring damage. Another strength of this study is the development of cases, which were based on actual patient data to be as realistic as possible, while modifications were made to ensure a sufficient representation of all damage items. The total of 110 cases is a large number for validation, given the rarity of these diseases. Further, the participation of adult and pediatric experts worldwide who all provided patient cases and scored the ADDI, is an important strength of study. This makes it plausible that the ADDI can be used in clinical settings involving pediatric or adult patients with FMF, CAPS, TRAPS or MKD.

Some weaknesses should also be mentioned. Firstly, the modification of the cases could have resulted in less realistic cases. Furthermore, scoring damage using paper cases is different from use in daily clinical practice, since the ADDI might be easier to score in these paper cases as all the information is summarized and presented in a uniform way. In addition to this the calculated ICC might be too positive, as the experts scoring the cases had also been involved in developing the ADDI. Physicians with less knowledge on the development of the ADDI or AIDs in general might encounter more difficulties interpreting the damage items and scoring the ADDI. On the other hand, due to the nature of cases (semi-fictional instead of real patients) participants may interpret data more ambiguously than they would in real life. Scoring anonymous cases, without knowing the patients or being able to ask additional questions, could be more difficult than in daily practice. The comments of the participants indeed showed that they sometimes experienced difficulties interpreting the data in the cases.

Although the overall ICC was >0.8, the ICC of some individual items was less than 0.6. This could be explained by insufficient information provided in the paper cases (e.g. no growth charts to more carefully assess growth failure, or joint examination not detailed enough to definitively score joint restriction), less experience of adult rheumatologists with pediatric measurements (e.g. scoring of puberty delay) or the more subjective nature of some items (e.g. the lack of an objective measure of musculoskeletal pain). Indeed, more objective items as hearing loss, renal insufficiency

and osteoporosis all had an individual ICC of >0.8. As the overall ICC was good and the nature of the cases may be an important reason for a low ICC, items scoring less than 0.6 were deemed acceptable, albeit sometimes with small alterations in the definition. A study testing the ADDI in real life patients and, if possible, with another group of physicians, would be needed to overcome the abovementioned issues.

Besides the strong correlation between the ADDI score and PGA-damage, the ADDI also moderately correlated to the PGA of disease activity. Ideally there would be no correlation between the ADDI and an activity score, however some degree of correlation is acceptable, since patients with more disease activity over the years generally accrue more damage. Therefore, disease activity had limited (but not zero) influence on the ADDI score, as should be expected from a damage index.

Interestingly, we found a relatively high ICC for the PGA-damage among the experts. One could therefore argue that a detailed damage index is not necessary to accurately score damage when the PGA is also reliable. However, we would still recommend the use of a damage index since the physicians scoring the ADDI were considered experts in AIDs, therefore their estimation of damage might be more accurate than physicians with less experience. Secondly, even though the estimates of PGA-damage might be reliable, an estimate of damage on a numerical scale does not give transparent information on why a certain amount of damage was estimated for a patient. The ADDI thus provides insight to the reasons why a certain amount of damage is scored for a patient. Thirdly, the ADDI provides a useful aide memoir and systematic means of collecting and quantifying damage, which is crucial to enable future comparisons between different studies.

During the face to face meeting, it was suggested to omit musculoskeletal pain from the ADDI, as it was deemed more subjective than the other items. Musculoskeletal pain, and other less objectively scored items such as fatigue and headache, might better be captured by a Patient Reported Outcome Measurement (PROMs) in addition to the ADDI. However, since no such tool exists yet for AID, and because musculoskeletal pain was emphasised by the patient representatives working within this this group as an important long-term disease burden, it was decided to maintain this item. When PROMs for AIDs are developed, the ADDI may need revision to exclude overlapping items.

Due to the design of this validation study, some important issues could not yet be addressed. The responsiveness to change, i.e. whether accrued damage over time is also reflected in an increasing score of the ADDI in an individual patient, could not be determined. Real clinical data and responsiveness to change are needed to assess the minimally clinically important difference of the ADDI. Another important correlation would be between the ADDI score and measures of quality of life (QoL). As the damage items in the ADDI are selected for their influence of patients'

lives, we hypothesize that patients with more damage have a lower QoL. This was impossible to assess because adding a fictional QoL index was neither realistic nor feasible. Lastly, ideally the ADDI should be correlated to a validated activity index, such as the AutoInflammatory Disease Activity index (AIDAI).(20, 21). As we could not derive the AIDAI scores from the patient data, we used PGA-activity as a surrogate marker. However, the ICC of PGA-activity was low with a broad CI, meaning that this estimate for activity as made by the participants is not a very reliable measure. Lastly, the ADDI was designed and validated for FMF, CAPS, TRAPS, and MKD. Its use outside these for diseases is not validated, and therefore cannot be recommended. That said, many of the items might have relevance for damage associated with many of the ever expanding list of AIDs, an area worthy of future study.

In conclusion, the ADDI can be considered a reliable tool to assess disease damage in clinical trials and routine clinical practice for the four most commonly encountered monogenic AIDs.

SUPPLEMENTARY

Index

1. Overview of comments explaining discrepancies and comments discussed in final survey and face-to-face meeting.

Comment	Amount of times suggested	Led to change of scoring yes/no	Led to change of item definition yes/no
Unclear case description/insufficient information/different interpretation of information between observers	73	na	na
Difficulties scoring puberty delay due to pharmacological induction of puberty	3	no	yes
Difficulties scoring serosal scarring	7	no	yes
Difficulties differentiating developmental delay/cognitive impairment from CNS and or on how to interpret psychiatric disorders	6	no	yes
Unclarities considering definitions of joint restriction	2	no	yes
Suggestion: more points for end stage renal disease	1	no	no
Suggestion for higher score for growth failure	1	no	no
Maximum score of the ADDI is always lower in males than in females because of the item amenorrhea	1	yes	no

2. Cronbach's alpha

Not applicable if a category consists of less than 2 items

a. Reproductive (Cronbach's alpha 0.271)

	Amenorrhea	**Infertility**	**Cronbach's alpha if item deleted**
Amenorrhea	-	0.225	*not applicable*
Infertility	-	-	*not applicable*

b. Renal/amyloidosis (Cronbach's alpha 0.777)

	Amyloidosis	**Proteinuria**	**Renal insufficiency**	**Cronbach's alpha if item deleted**
Amyloidosis	-	0.59	0.60	0.71
Proteinuria	-	-	0.73	0.75
Renal insufficiency	-	-	-	0.60

c. Developmental (Cronbach's alpha 0.225)

	Growth failure	**Puberty delay**	**Cronbach's alpha if item deleted**
Growth failure	-	0.25	*not applicable*
Puberty delay	-	-	*not applicable*

d. Neurological (Cronbach's alpha 0.754)

	Developmental delay	**Elevated intracranial pressure**	**Cognitive impairment**	**Central nervous system involvement**	**Cronbach's alpha if item deleted**
Developmental delay	-	0.34	0.66	0.35	0.70
Elevated intracranial pressure	-	-	0.43	0.48	0.73

Cognitive impairment	-	-	-	0.49	0.63
Central nervous system involvement	-	-	-	-	0.72

e. Musculoskeletal (Cronbach's alpha 0.509)

	Musculoskeletal	**Joint restriction**	**Osteoporosis**	**Bone deformity**	**Cronbach's alpha if item deleted**
Musculoskeletal	-	0.25	-0.002	0.15	0.50
Joint restriction	-	-	0.14	0.39	0.29
Osteoporosis	-	-	-	0.22	0.53
Bone deformity	-	-	-	-	0.32

References

1. Federici S, Sormani MP, Ozen S, Lachmann HJ, Amaryan G, Woo P, et al. Evidence-based provisional clinical classification criteria for autoinflammatory periodic fevers. Ann Rheum Dis. 2015 BMJ Publishing Group Ltd and European League Against Rheumatism;74(5):799-805.

2. Masters S, Simon A, Aksentijevich I, Kastner D. Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (*). Annu Rev Immunol. 2009;27:621-68.

3. Savic S, Dickie L, Wittmann M, McDermott M. Autoinflammatory syndromes and cellular responses to stress: pathophysiology, diagnosis and new treatment perspectives. Best Pract Res Clin Rheumatol. 2012;26(4):505-33.

4. ter Haar N, Oswald M, Jeyaratnam J, Anton J, Barron K, Brogan P, et al. Recommendations for the management of autoinflammatory diseases. Ann Rheum Dis. 2015;74(9):1636-44.

5. Ter Haar N, Lachmann H, Özen S, Woo P, Uziel Y, Modesto C, et al. Treatment of autoinflammatory diseases: results from the Eurofever Registry and a literature review. Ann Rheum Dis. 2013;72(5):678-85.

6. Broderick L, Tourangeau L, Kavanaugh A, Wasserman S. Biologic modulators in allergic and autoinflammatory diseases. Curr Opin Allergy Clin Immunol. 2011;11(4):355-60.

7. Vitale A, Rigante D, Lucherini O, Caso F, Muscari I, Magnotti F, et al. Biological treatments: new weapons in the management of monogenic autoinflammatory disorders. Mediators Inflamm. 2013;2013:939847-.

8. Holzinger D, Kessel C, Omenetti A, Gattorno M. From bench to bedside and back again: translational research in autoinflammation. Nat Rev Rheumatol. 2015;11(10):573-85.

9. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2008.

10. ter Haar NM, Annink KV, Al-Mayouf SM, Amaryan G, Anton J, Barron KS, et al. Development of the autoinflammatory disease damage index (ADDI). Ann Rheum Dis. 2017 BMJ Publishing Group Ltd and European League Against Rheumatism;76(5):821-30.

11. Ozen S, Frenkel J, Ruperto N, Gattorno M. The Eurofever Project: towards better care for autoinflammatory diseases. Eur J Pediatr. 2011;170(4):445-52.

12. Toplak N, Frenkel J, Ozen S, Lachmann H, Woo P, Koné Paut I, et al. An international registry on autoinflammatory diseases: the Eurofever experience. Ann Rheum Dis. 2012;71(7):1177-82.

13. Field AP. Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll. Los Angeles: SAGE; 2013.

14. Boyle GJ. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? Personality and Individual Differences. 1991;12(3):291.

15. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hoboken: Taylor and Francis; 1988.

16. Viola S, Felici E, Magni Manzoni S, Pistorio A, Buoncompagni A, Ruperto N, et al. Development and validation of a clinical index for assessment of long-term damage in juvenile idiopathic arthritis. Arthritis Rheum. 2005;52(7):2092-102.

17. Arkachaisri T, Vilaiyuk S, Torok K, Medsger T. Development and initial validation of the localized scleroderma skin damage index and physician global assessment of disease damage: a proof-of-concept study. Rheumatology (Oxford). 2010;49(2):373-81.

18. Albrecht J, Taylor L, Berlin J, Dulay S, Ang G, Fakharzadeh S, et al. The CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. J Invest Dermatol. 2005;125(5):889-94.

19. Suppiah R, Flossman O, Mukhtyar C, Alberici F, Baslund B, Brown D, et al. Measurement of damage in systemic vasculitis: a comparison of the Vasculitis Damage Index with the Combined Damage Assessment Index. Ann Rheum Dis. 2011;70(1):80-5.

20. Piram M, Frenkel J, Gattorno M, Ozen S, Lachmann H, Goldbach Mansky R, et al. A preliminary score for the assessment of disease activity in hereditary recurrent fevers: results from the AIDAI (Auto-Inflammatory Diseases Activity Index) Consensus Conference. Ann Rheum Dis. 2011;70(2):309-14.

21. Piram M, Koné Paut I, Lachmann H, Frenkel J, Ozen S, Kuemmerle Deschner J, et al. Validation of the auto-inflammatory diseases activity index (AIDAI) for hereditary recurrent fever syndromes. Ann Rheum Dis. 2014;73(12):2168-73.