Development and validation of psoriatic arthritis switch quality assessment tool (PASQAL) – an outcomes measurement tool to assess the quality of biologic switch decisions in psoriatic arthritis

Laires PA[1], Carrilho M[2], Tavares-Costa J[3], Lucas P[2], Machado P[4], Cunha-Miranda L[5], Pimentel-Santos F[6], Santos H[5], Vieira-Sousa E[7], Santos MJ[8]

ACTA REUMATOL PORT. 2020;45:46-57

ABSTRACT

Background: Switching between biologic therapies is a recommended strategy for Psoriatic Arthritis (PsA) patients that show an insufficient response or adverse events. Although the choice of the subsequent biologic may be dependent on many factors, assessing the quality of the switch decision is of utmost relevance.
Objectives: To develop and validate two outcomes measurement tools (for patients with peripheral and axial PsA phenotypes) that address the quality of treatment decisions in PsA regarding the switch of biologic therapies in clinical practice.
Methods: A Task Force and an Expert Panel were specifically assembled for this purpose. The Psoriatic Arthritis Switch Quality Assessment tool (PASQAL) development comprised a modified-Delphi method in a four-step procedure: 1) literature search and experts' opinion collection about quality indicators for PsA management; 2) Delphi design to address the development of the measurement tool; 3) three Delphi questionnaire rounds; 4) final consensus meeting. This phase resulted in the definition of two measurement tools, one to evaluate the quality of biologic switch in peripheral (pPASQAL) and another one in axial PsA (axPASQAL). For the validation of PASQAL, 12 experienced rheumatologists were asked to evaluate and classify the biologic switch of 80 clinical cases (40 with predominant peripheral and 40 with predominant axial PsA). Clinical judgement was defined to be the "gold standard" against which the performance of PASQAL was assessed. The results were used to assess tools' performance (sensitivity/specificity analysis) and the agreement between the tools and the gold standard (Cohen's kappa).
Results: PASQAL consists of 6 domains (joint disease activity, dactylitis, enthesis, physical function, quality of life, and skin and nail manifestations), respective instruments and thresholds. The classification of the biologic switch was divided into three quality levels: "Good", based on treat-to-target thresholds; "Moderate", based on improvement from baseline; and the remaining as "Insufficient". pPASQAL was found to be highly sensitive (92%) with the "Good" quality level and specific (97%) with the "Insufficient" quality level. Whilst axPASQAL showed overall higher sensitivity and specificity for all quality levels, as well as a higher level of agreement between the tool and the gold standard than pPASQAL (k=0.87 vs k=0.71).
Conclusion: PASQAL was developed and showed good criterion validity for the evaluation of the quality of switch in both peripheral and axial PsA phenotypes.

1. Health Economics & Outcomes Research - Novartis Farma, Oeiras, Portugal; Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade NOVA de Lisboa
2. Consulting & Services, IQVIA, Oeiras, Portugal
3. Rheumatology Department, ULSAM, Ponte de Lima, Portugal
4. Centre for Rheumatology and MRC Centre for Neuromuscular Diseases, University College London, London, UK; Department of Rheumatology, University College London Hospitals NHS Foundation Trust, London, UK; Department of Rheumatology, Northwick Park Hospital
5. Rheumatology Department, Portuguese Institute of Rheumatology, Lisbon, Portugal
6. Chronic Diseases Research Centre, NOVA Medical School | Faculdade de Ciências Médicas – NOVA University of Lisbon; Rheumatology Department, Centro Hospitalar Lisboa Ocidental, Hospital de Egas Moniz, Lisbon, Portugal
7. Serviço de Reumatologia e Doenças Ósseas Metabólicas, Hospital de Santa Maria, CHLN; Unidade de Investigação em Reumatologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Centro Académico de Medicina de Lisboa, Portugal;
8. Serviço de Reumatologia, Hospital Garcia de Orta, Almada; Unidade de Investigação em Reumatologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Portugal

These tools may be used in research as well as in clinical practice, to support rheumatologists in making more informed therapeutic decisions.

Keywords: Psoriatic Arthritis; Biologic switch; Consensus; Outcomes measurement tools; Therapeutic decisions

INTRODUCTION

Psoriatic arthritis (PsA) is a chronic inflammatory musculoskeletal disease preferentially managed by rheumatologists. PsA can affect the peripheral joints, the axial skeleton, entheses, skin and nails, and it can also have systemic involvement. According to Moll and Wright's 1973 criteria, PsA can be divided into five subgroups: asymmetrical oligoarthritis, symmetrical polyarthritis, distal interphalangeal disease, arthritis *mutilans,* and spondylitis[1]. However, this classification can be resumed in two main PsA phenotypes: peripheral (predominantly peripheral involvement) and axial (predominantly spinal involvement).

Treatment options have expanded and are increasingly more effective, enabling the simultaneous control of multiple PsA manifestations. Given the heterogeneity of disease manifestations, the assessment of disease activity and consequent response to treatment is more complex in PsA than in other inflammatory rheumatic diseases.

The Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group has defined a core set of domains to be measured to study treatment outcomes, aiming to guide therapeutic decisions in PsA. These domains include peripheral joint activity, skin activity, pain, patient global assessment, physical function, and health-related quality of life[2]. Consequently, outcome measures to evaluate these PsA domains were defined, with the vast majority being adapted from related diseases (e.g. 28-joint Disease Activity Score [DAS28], initially developed for rheumatoid arthritis and the Ankylosing Spondylitis Disease Activity Score [ASDAS], originally developed for ankylosing spondylitis) and only some of them validated for PsA[3]. Composite measures, which combine different domains into a single score, have also been created. One example is the Minimal Disease Activity (MDA), a composite remission score created specifically for PsA. This composite measure comprises a core set of domains (tender joint count ≤1; swollen joint count ≤1; Psoriasis Activity and Severity Index ≤1 or body surface area ≤3; patient pain visual analogue score (VAS) ≤15; patient global disease activity VAS ≤20; health assessment questionnaire ≤0.5; tender entheseal points ≤1). These domains are then used to define a "state" of MDA (high versus low disease activity), which in turn can be used as a target for treatment in clinical practice. If a patient does not achieve MDA criteria (at least 5 out of 7 criteria), the treatment may be modified either by increasing therapeutic doses or by adding new drugs or even by switching therapies[4]. However, MDA, as other outcome measures, has limitations, as it only defines a disease state and does not allow assessment of different levels of disease activity[5]. This is different from the ASDAS, a continuous composite index for which cut-offs for disease activity states and improvement scores have been published. Four disease activity states have been defined for ASDAS: inactive disease (remission-like state[6]), low, high and very high disease activity[7]. However, the ASDAS was not validated in PsA.

While there are no agreed definitions about the quality of biologics' switch in PsA and which outcome measure should be used for this quality assessment, the choice of subsequent biologic continues to be empirical, which makes the clinical evaluation of PsA patients a challenge to rheumatologists. Additionally, there is no composite measure that addresses the quality of the switch in the two different phenotypes of PsA. Therefore, the objective of this study was to develop and validate two outcomes measurement tools (for patients with peripheral and axial PsA phenotypes) that address the quality of treatment decisions in PsA regarding the switch of biologic therapies in clinical practice.

MATERIALS AND METHODS

This study consisted of two phases: 1) the development of two outcomes measurement tools, named PASQAL (Psoriatic Arthritis Switch Quality Assessment tooL), one for peripheral (pPASQAL) and another for axPASQAL, 2) the subsequent validation of these tools.

Research activities were developed by a Task Force and an Expert Panel specifically created for this study.

TASK FORCE AND EXPERT PANEL SELECTION

The Task Force comprised six members from different backgrounds, including health economics and outcomes research, and was chaired by a rheumatolo-

gist. The Expert Panel comprised seven experienced rheumatologists, all but one practice in Portugal. These rheumatologists were chosen from the Spondyloarthritis Study Group of the Portuguese Society of Rheumatology, recognized by their strong know-how in the PsA area. Both groups had significant roles during this study: Task Force was responsible for structuring the discussion process, while the Experts' Panel was responsible for decision making.

PHASE 1: TOOL DEVELOPMENT

To develop both PASQAL tools, a modified-Delphi method was applied in a four-step procedure: 1) literature search and experts' opinion collection about quality indicators for PsA management; 2) Delphi design to address the development of the measurement tool; 3) three Delphi questionnaire rounds; 4) consensus meeting to discuss the results and to decide about the outcomes measurement tools' components. During all phases, the Task Force and the Experts' Panel interacted regularly and in different formats (questionnaires, interviews, meetings) to ensure both tools comprised all relevant components and to promote an effective design of the modified-Delphi process.

Step 1: A literature search was performed in the PubMed database up until October 2017 to identify and collect relevant data to support the identification of potential quality indicators in PsA valued by all relevant stakeholders in this area (patients, rheumatologists, dermatologists, and others). The selection of the information was in English language and had, as a starting point, the OMERACT core set of domains agreed for PsA[8]. In PubMed, search terms reporting PsA management, switch recommendations, patient-reported outcomes (PROs) and clinical outcome indicators and measures were used. Search limits included links to full text only, humans, English language articles, males and females, and all adult ages. Additional information was collected through specific questionnaires and interviews administered to experts to understand the best practices on PsA management. Experts were asked about the different PsA phenotypes commonly managed in clinical practice, which indicators are commonly evaluated in PsA patients (from clinical to patient-related), and how they are measured. Additionally, questions about which indicators do patients value the most, when should indicators be evaluated, how indicators should be organized and combined, what constitutes an "appropriate" versus "inappropriate" biologic switch, and how many indicators should be included in a tool, were asked. With the information of both literature search and physicians' questionnaires/interviews, a long-list of potential quality indicators was gathered, and a list of domains was elaborated and organized according to three perspectives: physician, patient, and society. The categorization of perspectives was created together with the Experts to facilitate the systematization of information (e.g. elimination of repeated domains; grouping domains) and to structure the next steps.

Step 2: From the interactions between the Task Force and the Experts' Panel, it was agreed that the biologic switch outcome should be classified into three quality levels: "Good", based on treat-to-target thresholds; "Moderate", based on improvement from baseline status; and "Insufficient". The Delphi questionnaire rounds were discussed and validated by the Task Force, considering the Expert Panel input.

Step 3: A three-round Delphi session was then applied to start the decision-making process needed to define two outcomes measurement tools. In the first round, experts were asked which domains to include in each perspective and to rank these domains by relevance. In the second and third rounds, each Expert received a personalized questionnaire, which included previous answers, the Experts' Panel response interval, and the average response. The second and third rounds included additional questions related to the prioritization of domains, selection of instruments, the definition of quality thresholds, and definition of tool's output (i.e. quality of switch) given the possible combinations on the domains. In each subsequent round, experts reanswered the questions from the previous round and answered new questions. The three Delphi rounds resulted in a pre-selection of domains and respective instruments and quality thresholds that served as the discussion starting point for the consensus meeting[9].

Step 4: Tools' definition was concluded at the consensus meeting, to which all members from the Experts Panel and the Task Force attended. The Delphi results were presented and discussed. A decision-making process was promoted to reach conclusions regarding the definition of both outcomes measurement tools. This process involved structured discussion and democratic voting whenever necessary (e.g. distribution of points, individually or in groups, by the domains to be included in both tools; hands raised to reach consensus on the instruments cut-offs).

The modified-Delphi consensus resulted in the definition of two outcomes measurement tools to evaluate

the quality of biologic switch in peripheral and axial PsA patients.

PHASE 2: TOOLS' VALIDATION

The Task Force agreed to focus the analysis on the quantification of the performance of the tools for each quality level and calculation of the level of agreement between both tools and the gold standard (criterion validity).

Assembly of clinical cases: A total of 80 clinical cases who switched biologic treatment, 40 with predominantly peripheral PsA and 40 with predominantly axial PsA, were assembled, based on real-world patients, retrieved from the Reuma.pt registry. Reuma.pt, the Rheumatic Diseases Portuguese Register from the Portuguese Society of Rheumatology (SPR), is a nationwide web-based platform where Portuguese rheumatologists collect data on a routine basis regarding rheumatic patients (rheumatoid arthritis, axial spondyloarthritis, psoriatic arthritis, and juvenile idiopathic arthritis) receiving biological therapies or synthetic disease-modifying anti-rheumatic drugs (DMARDs)[10]. The clinical cases included information about age, gender, symptoms, disease activity and other parameters related to the current patient's disease state (e.g. physical function and quality of life), relevant to make a judgment about the quality of the biologic switch. In order to have a significant sample of each of the tools' switch levels for the analyses, 12 cases classified as "Good", 16 as "Moderate" and 12 as "Insufficient" were considered for both tools' validations.

Clinical judgment: A total of 12 experienced rheumatologists were recruited to evaluate clinical cases before and three months after switching to a second biologic therapy and were asked to classify the quality of that biologic switch as "Good", "Moderate" or "Insufficient". Rheumatologists did not receive any information about the tool's components or quality levels before clinical judgement to avoid a bias in the results. Experts' clinical judgment was defined as the "gold standard", against which the tools' output was compared. Three experts evaluated each clinical case, and only those with consensual clinical judgment (same biologic switch classification of at least 2 out of 3 rheumatologists) were included in the validation analysis.

Tools' performance: The performance was assessed through a sensitivity/specificity analysis, which evaluates the level of performance of the tool in classifying the biologic switch correctly, comparing the tool's outcome with the clinical judgment result. Sensitivity is the tool's ability to correctly classify the biologic switch as "Good", "Moderate" or "Insufficient" (true positive rate). Specificity is the ability of the tool to correctly classify the biologic switch as "Not Good", "Not Moderate" or "Not Insufficient" (true negative rate)[11,12].

Agreement between the tool and the gold standard: The agreement between the output of each tool and the gold standard was quantified using the Cohen's kappa. Generally, Cohen's kappa ranges from 0 to 1, whereas larger numbers (i.e. above 0.6) represent a good level of agreement. For the calculation of Cohen's kappa, a 95% confidence interval (CI) was also reported[13].

RESULTS

TOOL DEVELOPMENT

From the literature search and experts' opinion, a total of forty-five domains relevant for PsA management were identified (Annex - Table I) and organized according to three perspectives: physician (n=19), patient (n=20) and society (n=6).

During the Delphi questionnaire rounds, experts pre-selected eleven domains (Annex - Table II) to be included in both outcomes' measurement tools (physician=5, patient=4, society=2), and listed instruments and the respective quality thresholds for nine domains. For the remaining two (therapy costs and indirect costs from society's perspective), no consensus was achieved regarding the instruments and respective quality thresholds.

At the consensus meeting, experts agreed that both tools should be feasible to implement in daily clinical practice. As such, some domains were clustered (e.g. disease activity, swollen and tender joints, inflammatory parameters, joint pain and patient global disease evaluation into a composite "Joint Disease Activity" domain) and experts agreed on a final selection of six domains to be included in each tool (physician=4, patient=2, society=0). Experts concluded that none of the domains included in the preliminary list of society perspective should be included in the final tool due to the inexistence of internationally accepted and easy to use instruments to quantify these domains, which consequently would generate difficulties on the measurement and results' interpretation. Considering these domains, experts defined the most appropriate instruments and thresholds for both peripheral and axial phenotypes, concluding the core set criteria of the two outcomes measurement tools (Table I and Table II).

TABLE I. DOMAINS, INSTRUMENTS AND THRESHOLD DEFINITIONS FOR pPASQAL

Domains	Instrument	"Good" threshold[1]	"Moderate" threshold[2]
1: Disease Activity	DAPSA[a]	≤ 4	≥ 50% reduction
2: Dactylitis	Number of fingers	≤ 1 finger	Clinical improvement (Yes/No scale)
3: Enthesis	SPARCC[b]	≤ 1 enthesis	≥ 50% reduction
4: Physical Function	HAQ-DI[c]	≤ 0.5	≥ 0.35 reduction
5: Quality of Life	EQ-5D[d]	≥ 0.65[e]	≥ 0.18[f] improvement
6: Skin and Nail Manifestations	sPGA[g]	≤ 1	≥ 50% reduction

1: based on treat-to target thresholds; 2: based on improvement from the baseline
a: Disease Activity in Psoriatic Arthritis is a composite instrument that covers 5 domains of the initial pool; b: Spondyloarthritis Research Consortium of Canada index; c: Health Assessment Questionnaire Disability Index; d: EuroQol five-dimensional questionnaire; e: Based on the thresholds for Patient Acceptable Symptom State (PASS) in line with the average Portuguese Population above 50 years old = [0.60–0.69] and equivalent to HAD-DI ~1; f: Based on the Minimal Clinically Important Improvement (MCII); g: Visual analogue scale performed by physicians based on the GRAPPA 0-5 scale

TABLE II. DOMAINS, INSTRUMENTS AND THRESHOLD DEFINITIONS FOR axPASQAL

Domains	Instrument	"Good" threshold[1]	"Moderate" threshold[2]
1: Disease Activity	ASDAS[a]	< 1.3	≥ 1.1 reduction
2: Dactylitis	Number of fingers	≤ 1 finger	Clinical improvement (Yes/No scale)
3: Enthesis	MASES[b]	≤ 1 finger	≥ 50% reduction
4: Physical Function	BASFI[c]	≤ 2 points	≥ 2 points reduction
5: Quality of Life	EQ-5D[d]	≥ 0.65[e]	≥ 0.18[f] improvement
6: Skin and Nail Manifestations	sPGA[g]	≤ 1	≥ 50% reduction

1: based on treat-to target thresholds; 2: based on improvement from the baseline
a: Ankylosing Spondylitis Disease Activity Score – measures of Axial Spondyloarthritis disease activity; b: Maastricht Ankylosing Spondylitis Enthesitis Score; c: Bath Ankylosing Spondylitis Functional Index – measures disease activity and function of axial spondyloarthritis;
d: EuroQol five-dimensional questionnaire; e: Based on the thresholds for Patient Acceptable Symptom State (PASS) in line with the average Portuguese Population above 50 years old = [0.60 – 0.69] and equivalent to HAD-DI ~1; f: Based on the Minimal Clinically Important Improvement (MCII); g: Visual analogue scale performed by physicians based on the GRAPPA 0-5 scale

Despite skin clearance being one of the most valued symptoms in patients with PsA and Psoriasis, this domain shows limitations since the agreed measures (e.g. Psoriasis Area Severity Index - PASI, Body Surface Area - BSA) are difficult to evaluate and to use in clinical practice. Therefore, these parameters are rarely completed by rheumatologists during patient assessment. Given the importance of this assessment, the Task Force and the Expert Panel agreed to choose physician global assessment applied to skin and nails (sPGA) as the instrument of "Skin & Nails Manifestations" domain[14]. sPGA is commonly measured using a visual analogue scale (VAS) based on the GRAPPA 0-5 scale with a good correlation with the PASI index, i.e. a "Good" quality level is a patient with a clean or almost clean skin corresponding to a Physician VAS of 0/1. Currently, the "Skin and Nails Manifestations" domain is not routinely assessed by all rheumatologists, hence this domain was agreed to be optional. "Productivity and Activity Impairment" was discussed as a possible seventh domain but was discarded due to feasibility concerns. Due to the need for further literature search, the thresholds for the EQ-5D instrument were defined after the consensus meeting by the Task Force. Two possible options regarding the threshold definition for EQ-5D were discussed. According to Carreño *et al.* 2011, the EQ-5D thresholds can be defined using HAQ-DI scores.[15] The conversion of the "Good" state of HAQ-DI=0.50 results in a value of EQ-5D=0.80, which is too ambitious, given the average results of the

EQ-5D obtained in the overall Portuguese population.[16] Furthermore, the EQ-5D has a sigmoid distribution and decreases sharply at upper values (i.e. better health states) - the perfect health EQ-5D score is 1 and, in the Portuguese population, the closest result (i.e. near-perfect health state) is 0.77. The other option available for the threshold definition of EQ-5D was based on the work of Kvamme *et al*. 2010 which aimed to identifying the thresholds for Patient Acceptable Symptom State (PASS) and Minimal Clinically Important Improvement (MCII) of the Norwegian population with rheumatoid arthritis, ankylosing spondylitis, and PsA17. Thus, the Task Force decided to choose the second option. The "Good" threshold was defined as EQ-5D ≥ 0.65 using PASS as an anchor; this is in line with the average EQ-5D score of the Portuguese population above 50 years old = [0.60 – 0.69] and equivalent to HAQ-DI ~1 according to Lara Ferreira et al. 2014. The "Moderate" threshold was defined as EQ-5D equal or above 0.18, using EQ-5D MCII as the reference value. The research and the Task Force discussion on the definition of the EQ-5D thresholds led to the revision of the "Good" and "Moderate" thresholds of HAQ-DI (function domain). As a result, the "Good" threshold was kept because it was already an accepted value by the medical community; the "Moderate" threshold was changed from 0.33 to 0.35, as the latter seems to be often accepted in the recent literature as the "minimally important change". Although, the practical impact is null since HAQ-DI comprises intervals of 0.125; 0.250; 0.37518.

Regarding overall switch quality, a "Good" biologic switch was defined when comprising most of the domains with "Good" outcome and up to 1 or 2 domains with a "Moderate" outcome, out of 5 or 6 evaluated domains, respectively. A "Moderate" switch outcome was defined when comprising most of the domains with "Good" or "Moderate" outcomes and up to 1 or 2 domains with an "Insufficient" outcome, out of 5 or 6 evaluated domains, respectively. Besides, experts agreed that "Joint Disease Activity" domain outcome is a critical prerequisite of overall switch quality, meaning that if a given biologic switch is to be classified as "Good", then the "Joint Disease Activity" domain also needs to have a "Good" outcome; and if instead the overall quality of switch is classified as "Moderate", then the "Joint Disease Activity" needs to have at least a "Moderate" outcome. When "Joint Disease Activity" had an "Insufficient" outcome the switch outcome should be classified as "Insufficient", regardless of the result obtained in the other domains. (Figure 1)

Tools' validation: From the 80 clinical cases (40 with peripheral PsA and 40 with axial PsA phenotypes) assessed, 30 cases for the peripheral PsA and 25 cases for the axial PsA tools' validation were used. These were the cases where at least 2 out of 3 rheumatologists gave the same biologic switch classification ("Good", "Moderate" or "Insufficient") to the clinical scenario. The overall results leading to the cases used for both tools' validation are presented in the Annex – Table III).

Regarding the performance of the tools, the peripheral PsA tool was found to be more sensitive (92%) for the "Good" quality level and more specific (97%) for the "Insufficient" quality level. The axial PsA tool was found to be more sensitive (100%) for the "Good" quality level and more specific (100%) for the "Insufficient" quality level. Both tools showed to be less sensitive for the "Moderate" quality level. Most cases of discordance between the opinion of the rheumatologists and PASQAL were due to the switch being classified as "Moderate" by the tool and as "Good" by the rheumatologists (83% of the discordant cases in pPASQAL and 70% of the discordant cases of axPASQAL). Additional performance parameters regarding sensitivity/specificity analysis are also shown in the Annex – Table V.

An additional analysis was conducted for the cases where rheumatologists disagreed between themselves regarding the quality of the switch (without the agreement of at least 2 out of 3 rheumatologists) to understand the main reasons for rheumatologists to evaluate the biologic switch differently from the tools. The example in the Annex – Figure 1 depicts the tool's output of the 40 peripheral PsA cases (coloured balls corresponding to the "Good", "Moderate" and "Insufficient" tool's outcomes) and the cases where the rheumatologists gave a different classification (red crosses). Notably, cases considered "Moderate" by the tool (considering the outcomes of all domains) are classified as "Good" by the rheumatologists mainly because they accepted some residual "Disease Activity" if the patient has a considerable improvement from the baseline (e.g. patient with a DAPSA=5.6 but with an improvement from the baseline of 64%). Similar findings were obtained for axial PsA cases (Annex – Figure 2).

Good levels of agreement of both tools with the gold standard were achieved. The peripheral PsA tool presented a kappa=0.87, whereas the axial PsA tool had a kappa=0.71. The summary of the validation results for both tools is shown in Table III.

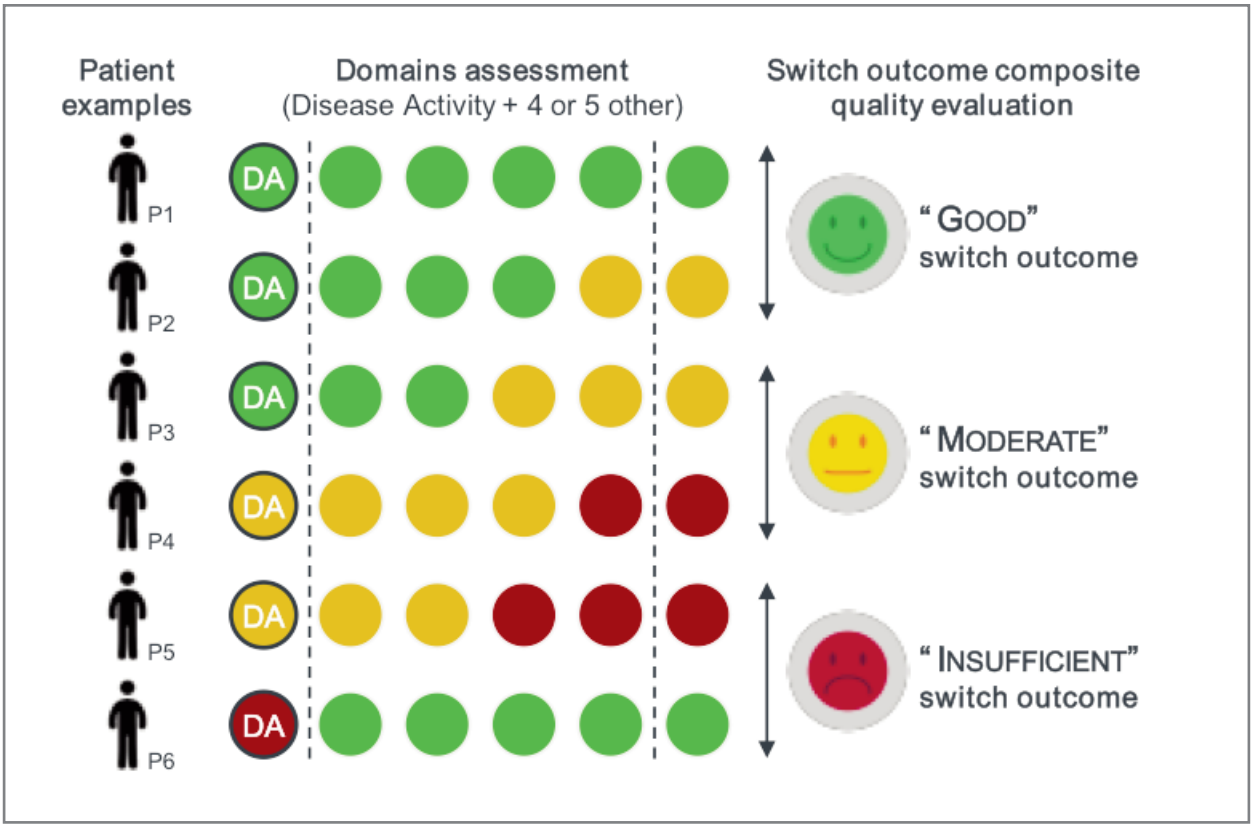


FIGURE 1. Definition of "Good", "Moderate" and "Insufficient" biologic switch outcomes

TABLE III. VALIDATION RESULTS OF pPASQAL AND aPASQAL

Outcome Measurement Tool	Biologic Switch Classification[1]	"Sensitivity"[2]	"Specificity"[3]	Agreement between the tool and rheumatologists[4]
pPASQAL	"Good"	92%	89%	k=0.71 (95% CI = 0.59-0.83)
	"Moderate"	73%	85%	
	"Insufficient"	75%	97%	
aPASQAL	"Good"	100%	90%	k=0.87 (95% CI = 0.78-0.96)
	"Moderate"	83%	97%	
	"Insufficient"	86%	100%	

1. All patient profiles (40 with Peripheral PsA and 40 with Axial PsA) were evaluated by 3 experts and the biologic switch was classified into three quality levels: "Good", "Moderate" or "Insufficient"
2. "Sensitivity": Only of those consensual patient profiles, proportion (%) correctly classified as "Good"; "Moderate" or "Insufficient"
3. "Specificity": Only of those consensual patient profiles, proportion (%) correctly classified as not "Good"; "Moderate" or "Insufficient"
4. The agreement between the tool and the "gold standard" was calculated by Cohen's kappa (k); values above 0.61 are considered a good level of agreement

These validation analyses were also performed considering all cases without discarding the non-consensual cases, i.e. without considering at least 2 out of 3 rheumatologists giving the same biologic switch classification. The answers of the experts to all 80 clinical cases were analyzed and compared with the tool output through the metrics mentioned above (Annex – Table 4). The results reflect a higher divergence in the opinion of experts, i.e. less coherence with the tool. As expected, the sensitivity and specificity for all quality

levels as well as the reliability were lower for both tools.

DISCUSSION

A tool to assess the quality of biologic switch decisions in PsA was developed and shown to have good content validity. The fact that there are no agreed definitions about the quality of biologics' switch in PsA and which outcome measure should be used for this quality assessment makes this project the first attempt to create a new dedicated overall measure to evaluate the biologic switch in PsA patients overcoming these current limitations.

The methodology used for the validation analysis was carefully chosen and needed to ensure that both tools are valid, also having clinical application and feasibility in mind. The validation results showed that aPASQAL is more sensitive classifying a biologic switch as "Good", "Moderate" or "Insufficient" and more specific classifying a biologic switch as not "Good", "Moderate" and "Insufficient". Also, the axial PsA clinical cases classified as "Good" by both tools were 100% correctly classified by the rheumatologists, enabling the statement that a negligible error will be involved when classifying these cases in the future. The "Moderate" quality level (in both tools) was the one showing higher discrepancies between tool's output and clinical judgement, where rheumatologists seem to be more "optimistic" than the tool, given that, in most of the discordant cases, they classified the biologic switch as "Good" instead. Also, the axial PsA tool has shown to have a higher agreement when compared with the gold standard than the peripheral PsA tool. However, if all cases were considered, including the ones without an agreement, a higher divergence in the opinions of experts would have been observed, resulting in lower sensitivity, specificity, and level of agreement for both tools.

PASQAL's development comprised several interactions between the Task Force and the Expert's Panel, which were essential to arrive at the final tools' definition. The main objective of this phase was to develop a tool that should be feasible in daily clinical practice. Experts reached consensus on having the "Joint Disease Activity" domain as a prerequisite in both tools as they considered essential for the patient to meet this "requirement" to have a "Good" or "Moderate" biologic switch, otherwise, it is an "Insufficient" switch. The special importance given to this domain was then validated during clinical judgement namely in cases without an agreement, as most rheumatologists focused mainly on the joints to give their opinion about the quality of the biologic switch.

The analysis of the cases without an agreement for both tools led to the conclusion that often rheumatologists take into consideration only one or two key parameters when evaluating the biologic switch. From all domains involved, the "Joint Disease activity" proved to be the main driver of rheumatologists' opinion. Overall, the discordant cases showed that rheumatologists are less demanding when classifying the biologic switch than the tool. For instance, they accepted a "Good" outcome for "Articular Disease Activity" even when the patient was not in remission/inactive disease. This also suggests that PASQAL may be more "demanding" that the rheumatologists, related to the fact that this tool considers a broader number and type of domains.

This study has some limitations. First, it is a subjective process and dependent on the opinion of the involved Experts, which may not necessarily reflect the view of the entire community of rheumatologists. Second, although the patient perspective was considered in the literature search, the lack of direct patient involvement in this study is a limitation. Third, the validation results are based on a small sample, mainly due to lack of patients' information (some hospitals do not report measures that are outputs of the tools) available at the time of the study. Thus, PASQAL needs to be tested in a larger sample to further analyze its performance, usability and feasibility. Lastly, both tools were created to evaluate the outcome three months after switching to a second biologic therapy. Therefore, it would be interesting to test PASQAL in other stages of the disease and management (e.g. first biologic switch and DMARD switch).

In conclusion, two outcome measurement tools were developed to address the quality of treatment decisions regarding biologics' switch in PsA management. Both outcomes measurement tools define states of the disease and levels of improvement and their use can be extended to other therapeutic interventions in PsA, such as the introduction of first biologic DMARD or to subsequent switches, and consequently gain a broader scope and utility. Besides, the developed tools are more demanding than the rheumatologists' opinion since they take into consideration relevant parameters that are not always considered by rheumatologists daily, and include the patients' perspective (through the

quality of life assessment measured by the EQ-5D).

In the future, it would be interesting to evaluate the implications of PASQAL output on long-term "hard" outcomes (e.g. persistence, quality of life, remission, structural damage), namely to test if a "Good" switch can predict significant positive outcomes, and therefore enhancing the therapeutic decisions of rheumatologists and PsA management. Finally, this project can also contribute to the implementation of PASQAL in the Reuma.pt registry and be used to evaluate the quality of treatment decisions done by rheumatologists in PsA, as well as to evaluate how well controlled are biologic-persistent patients according to a core set of quality indicators. Hence directly supporting clinical practice and real-world research.

CORRESPONDENCE TO

Pedro A. Laires
Centro de Investigação em Saúde Pública
Escola Nacional de Saúde Pública
Universidade NOVA de Lisboa
E-mail: pedro.laires@ensp.unl.pt

REFERENCES

1. Moll JM, Wright V. Psoriatic arthritis. Semin. Arthritis Rheum. 3, 55–78 (1973).
2. Gladman DD et al. Consensus on a core set of domains for psoriatic arthritis. J. Rheumatol. 34, 1167–1170 (2007).
3. Coates LC, Fransen J, Helliwell PS. Defining minimal disease activity in psoriatic arthritis: a proposed objective target for treatment. Ann. Rheum. Dis. 69, 48 LP-53 (2010).
4. Smolen JS et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann. Rheum. Dis. 77, 3–17 (2018).
5. Tucker LJ, Coates LC, Helliwell PS. Assessing Disease Activity in Psoriatic Arthritis: A Literature Review. Rheumatol. Ther. 6, 23–32 (2018).
6. Machado P, Landewé R. Is it time to replace BASDAI with ASDAS? Nat. Rev. Rheumatol. 9, 388 (2013).
7. Machado PM, Landewé R, van der Heijde D Assessment of SpondyloArthritis international Society (ASAS)Ankylosing Spondylitis Disease Activity Score (ASDAS): 2018 update of the nomenclature for disease activity statesAnnals of the Rheumatic Diseases 2018;77:1539-1540
8. Orbai AM et al. Updating the Psoriatic Arthritis (PsA) Core Domain Set: A Report from the PsA Workshop at OMERACT 2016. J. Rheumatol. 44, 1522 LP-1528 (2017).
9. K, F., SJ, B., MS, A. & Al., E. RAND/UCLA Appropriateness method user's manual. (2001).
10. Canhao H, Faustino A, Martins F, Fonseca JE. Reuma.pt - the rheumatic diseases portuguese register. Acta Reumatol. Port. 36, 45–56 (2011).
11. Petri M et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 64, 2677–2686 (2012).
12. Fawcett T. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874 (2006).
13. Kwiecien R, Kopp-Schneider A, Blettner M. Concordance analysis: part 16 of a series on evaluation of scientific publications. Dtsch. Arztebl. Int. 108, 515–521 (2011).
14. Cauli A et al. Physician's Global Assessment in Psoriatic Arthritis: A Multicenter GRAPPA Study. J. Rheumatol. 45, 1256 LP-1262 (2018).
15. Carreño A, Fernández I, Badia X, Varela C, Roset M. Using HAQ-DI to estimate HUI-3 and EQ-5D utility values for patients with rheumatoid arthritis in Spain. Value Heal. 14, 192–200 (2011).
16. Ferreira PL, Ferreira LN, Pereira LN. Contribution for the Validation of the Portuguese Version of EQ-5D. Acta Med. Port. 26, 664–675 (2013).
17. Kvamme M, Kristiansen I, Lie E, Kristian Kvien T. Identification of Cutpoints for Acceptable Health Status and Important Improvement in Patient-Reported Outcomes, in Rheumatoid Arthritis, Psoriatic Arthritis, and Ankylosing Spondylitis. The Journal of rheumatology 37, (2009).
18. Mease PJ et al. Minimally Important Difference of Health Assessment Questionnaire in Psoriatic Arthritis: Relating Thresholds of Improvement in Functional Ability to Patient-rated Importance and Satisfaction. J. Rheumatol. 38, 2461 LP-2465 (2011).

ANNEX

ANNEX – TABLE I. PRELIMINARY LIST OF THE 45 DOMAINS RELEVANT FOR PSA SELECTED PRIOR TO DELPHI PROCESS

Perspective	Domain
Physician (N=19)	Joint disease activity
	Swollen joints
	Dactylitis
	Inflammatory parameters
	Enthesis
	Side-effects, safety and tolerability
	Global disease evaluation (physician)
	Skin manifestations
	Tender joints
	Ophthalmologic manifestations
	Imaging evaluation of the inflammatory activity
	Joint pain (related to Peripheral PsA)
	Gastrointestinal manifestations
	Axial pain (related to Axial PsA)
	Nail manifestations
	Persistence
	Structural damage
	Axial mobility
	Cardiometabolic events
Patient (N=20)	Global patient satisfaction
	Global disease evaluation (patient)
	Physical function
	Joint pain (related to Peripheral Joint PsA)
	Quality of life
	Side-effects, safety and tolerability
	Treatment satisfaction
	Work productivity
	Axial pain (related to Axial PsA)
	Fatigue
	Emotional well-being
	Social life
	Anxiety and depression
	Sleep disturbance
	Impact on family and/or caregiver
	Axial manifestations
	Convenience
	Sexual life
	Rejection, discrimination and shame
	Fear and uncertainty
Society (N=6)	Therapy costs
	Indirect costs
	Direct medical costs
	Social life
	Impact on family and/or caregiver
	Persistence

ANNEX – TABLE II. PRE-SELECTED 11 DOMAINS AFTER DELPHI PROCESS

Perspective	Domain
Physician (N=5)	Joint disease activity
	Swollen joints
	Dactylitis
	Inflammatory parameters
	Enthesis
Patient (N=4)	Global patient satisfaction
	Global disease evaluation (patient)
	Physical function
	Joint pain (related to Peripheral Joint PsA)
Society (N=2)	Therapy costs
	Indirect costs

ANNEX – TABLE III. RESULTS OF THE TWO OUTCOMES MEASUREMENT TOOLS, COMPARED WITH THE "GOLD STANDARD" (CLINICAL JUDGEMENT) USING ALL CASES (WITHOUT CONSIDERING AT LEAST 2 OUT OF 3 RHEUMATOLOGISTS GIVING THE SAME BIOLOGIC SWITCH CLASSIFICATION)

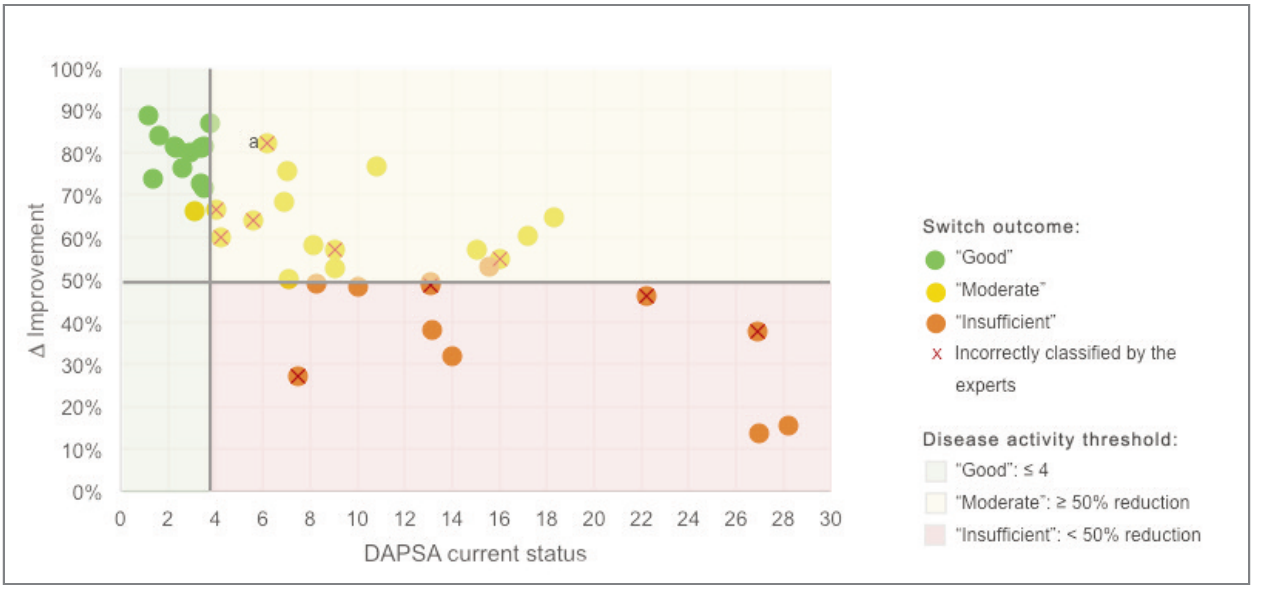
Outcome Measurement Tool	Biologic Switch Classification[1]	"Sensitivity"[2]	"Specificity"[3]	Agreement between the tool and rheumatologists[4]
Peripheral PsA Tool	"Good"	92%	74%	k=0.49 (95% CI = 0.34-0.63)
	"Moderate"	52%	78%	
	"Insufficient"	58%	97%	
Axial PsA Tool	"Good"	100%	70%	k=0.45 (95% CI = 0.30-0.60)
	"Moderate"	44%	79%	
	"Insufficient"	53%	95%	

1. All patient profiles (40 with Peripheral PsA and 40 with Axial PsA) were evaluated by 3 experts and the biologic switch was classified into three quality levels: "Good", "Moderate" or "Insufficient"
2. "Sensitivity": Only of those consensual patient profiles, proportion (%) correctly classified as "Good"; "Moderate" or "Insufficient"
3. "Specificity": Only of those consensual patient profiles, proportion (%) correctly classified as not "Good"; "Moderate" or "Insufficient"
4. The agreement between the tool and the "gold standard" was calculated by Cohen's kappa (k); values above 0.61 are considered a good level of agreement

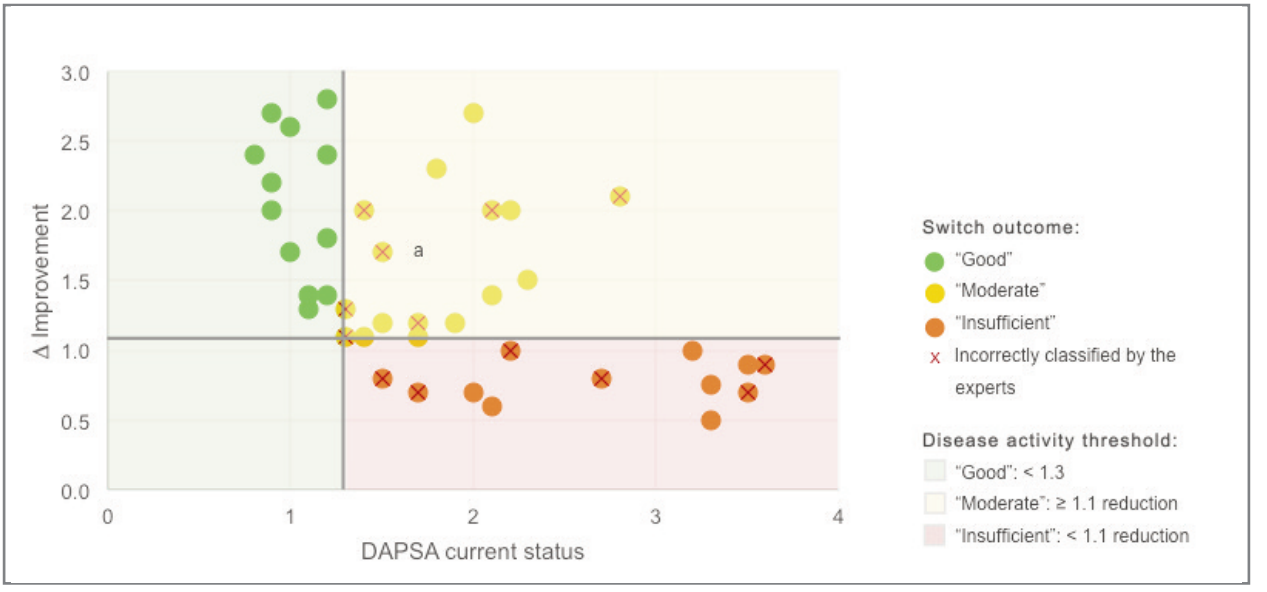
ANNEX – TABLE III. COMMON PERFORMANCE METRICS CALCULATED FROM SENSITIVITY/SPECIFICITY ANALYSIS

Outcome Measurement Tool	Biologic Switch Classification[1]	PPV[1]	NPV[2]	LR+[3]	LR-[4]
pPASQAL	"Good"	0.85	0.94	1.04	0.71
	"Moderate"	0.71	0.86	0.87	1.75
	"Insufficient"	0.90	0.91	0.77	8.75
aPASQAL	"Good"	0.90	1.00	1.13	0
	"Moderate"	0.88	0.95	0.86	4.83
	"Insufficient"	1.00	0.95	0.86	-5

1. Precision or positive predictive value (PPV). PPV = True Positives / (True Positives + False Positives)
2. Negative predictive value (NPV). NPV = True Negatives / (True Negatives + False Negatives)
3. Positive likelihood ration (LR+). LR+ = True Positive Rate / (1-False Positive Rate)
4. Positive likelihood ration (LR-). LR- = (1-True Positive Rate)/False Positive Rate
5. Not applicable as no False Positives were founded in the "Insufficient" switch classification



Annex - Figure 1. Additional analysis in peripheral PsA – Agreement between tools' output and clinical judgement using DAPSA thresholds
a – Example of a case without agreement between clinical judgement and the tool: Experts have evaluated this case as "Good" and the tool has considered it "Moderate" due to DAPSA = 5.60 and DAPSA improvement above 50%.



Annex - Figure 2. Additional analysis in axial PsA – Agreement between tools' output and clinical judgement using ASDAS thresholds
a – Example of a case without agreement between clinical judgement and the tool: Experts evaluate this case as "Good" and the tool has considered it "Moderate" due to ASDAS = 1.7 and ASDAS improvement above 1.1.