Version

This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 http://www.niso.org/publications/rp/

Suggested Reference


Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

This is the peer reviewed version of the article above which has been published in final form at 10.1002/acr.22875

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for self-archiving.

For more information, see General copyright, Publisher copyright, SHERPA/RoMEO.
What should be the cut point for classification criteria for studies in gout? A conjoint analysis.

1Jaap Fransen, PhD; 2Wietske Kievit PhD; 3Tuhina Neogi MD, PhD; 4Prof. Ralph Schumacher, MD, PhD; 5Tim Jansen, MD, PhD; 6Prof. Nicola Dalbeth, MD, PhD; 7William J Taylor, MD, PhD.

1Department of Rheumatology and 2Department for Health Evidence, Radboud University Medical Centre, Nijmegen, The Netherlands; 3Boston University Schools of Medicine and of Public Health, Boston, USA; 4University of Pennsylvania and VA Medical Center, Philadelphia, USA; 5Department of Rheumatology, Vie Curi Medical Centre, Venlo, and IQ Healthcare, Radboud University Medical Centre, Nijmegen, The Netherlands; 6Department of Medicine, University of Auckland, Auckland. New Zealand; 7Department of Medicine, University of Otago, Wellington. New Zealand.

This study was not supported by commercial or non-commercial funding.

Word count: 2476
Tables: 2
Figures: 1

Adress for correspondence:
Jaap Fransen, PhD.
Department of Rheumatology, Radboudumc.
PO Box 9101, 6500HB Nijmegen, The Netherlands.
Phone: +31-24-36 16530
E-mail: Jaap.Fransen@Radboudumc.nl

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an ‘Accepted Article’, doi: 10.1002/acr.22875
© 2016 American College of Rheumatology
Received: Aug 10, 2015; Revised: Jan 18, 2016; Accepted: Feb 23, 2016
Summary

Objective To determine the acceptable level of positive predictive value (PPV) and negative predictive value (NPV) for classification criteria for gout, given the type of study.

Methods We conducted an international web-based survey with 91 general practitioners and rheumatologists, experienced in gout. Conjoint analysis was used as the framework for designing and analyzing pairs of two profiles each describing a study type, a PPV and NPV. There were 5 study types presented: a phase 3 RCT of an NSAID versus prednisone for acute gout flares; a phase 3 RCT of a biologic agent for acute gout flares; a phase 2 RCT of a novel uricosuric drug of unknown efficacy and limited toxicity data; a case-control genome-wide-association (GWAS) study of gout; a cohort study examining long term outcomes of gout. PPV and NPV both had five levels ranging from 60-99%.

Results The panellists in majority were male (65%) rheumatologists (93%) with an average of 19 years of practice seeing 5 to 60 patients with gout monthly. PPV was most highly weighted in decision making: the relative importance was 59% for PPV; 29% for NPV and 13% for study type. The preferred PPV was 90% or 80%, with an accompanying NPV of 70% or 80%, dependent on study type.

Conclusion Preferred positive predictive values and negative predictive values range between 70% and 90% and differ by study type. A single cut-point can be a reasonable approach for all study types if a PPV of 90% and NPV of 80% is approximated.
Significance and innovations

- In developing the ACR-EULAR classification criteria for gout, the question was what threshold to use for defining an eligible subject for inclusion in a study for gout?

- Preferred cut points for classification criteria may differ by study type. In studies (interventions) with low risk, low toxicity and low costs, misclassification is less a problem than in studies (interventions) with high risk, high toxicity and high costs.

- Conjoint analysis was successfully used to elucidate the position of cut points for classification, differentiated according to study type, which is a new approach.

- A positive predictive value of at least 90% with a negative predictive value of at least 70% was preferred for most study types; therefore, the cut point in the classification criteria for gout ideally should not produce PPV and NPV lower than these values.
Introduction

The 2015 ACR-EULAR classification criteria for gout have been established to enable a standardized identification of individuals with gout for enrollment into studies [1]. The classification criteria include 8 domains: pattern of joint/bursa involvement, characteristics of symptomatic episodes, time-course of symptomatic episodes, clinical evidence of tophus, serum urate (off treatment), synovial fluid analysis, imaging evidence of urate deposition, and imaging evidence of gout-related joint damage. Points are rewarded for presence or absence of items, the maximal possible score is 23; a threshold score of ≥8 classifies an individual as having gout. The sensitivity of the classification criteria was 0.92, and specificity was 0.89 [1]. The construction of the ACR-EULAR classification criteria for gout is similar to other recent classification systems, for rheumatoid arthritis and for systemic sclerosis [2,3].

In developing the ACR-EULAR classification criteria for gout, the question was what threshold to use for defining a subject as having gout? As long as classification criteria do not perfectly agree with the presence of MSU crystals in synovial fluid or tophi, misclassification will occur. However, in studies (interventions) with low risk, low toxicity and low costs, misclassification is less a problem than in studies (interventions) with high risk, high toxicity and high costs. Consequently, preferred cut points for classification criteria may differ by study type. Therefore, the objective of this study was to determine physician preference for the acceptable level of positive predictive value (PPV) and negative predictive value (NPV) for classification criteria for gout, given the type of study.
Methods

Design

In September and October 2014 we conducted an international web-based survey with 91 general practitioners and rheumatologists, experienced in the diagnosis and treatment of gout. Conjoint analysis was used as the framework for designing and analyzing questions. Participants were enrolled by e-mail invitation using 80 primary care contacts from the 2011 and 2012 EULAR congresses and 136 rheumatologists interested in gout from a clinical research network. They were sent an e-mail including a link to a web-based questionnaire. The sample size needed was determined as a combination of 20 choice sets answered by at least 63 panelists, based on choosing between 2 profiles with 3 attributes having 5 levels each [4].

Attributes and levels

The concepts of study type and error, the ‘attributes’ in conjoint analysis language, were defined in a consensus procedure among the members of the ACR-EULAR Gout Classification Criteria steering committee [5]. Study type was defined as having 5 levels: a phase 3 RCT of an NSAID versus prednisone for acute gout flares; a phase 3 RCT of a biologic agent for acute gout flares; a phase 2 RCT of a novel uricosuric drug of unknown efficacy with limited toxicity data; a case-control genome-wide-association (GWAS) study of gout; a cohort study examining long term outcomes of gout. It was assumed that these five study types would be representative for the kind of studies being done in gout. The error (misclassification) was defined in terms of positive predictive value (PPV) and negative predictive value (NPV), because the concepts of PPV and NPV are closest to classifying patients in daily practice. The relevant levels in PPV and NPV were defined to be 60%, 70%, 80%, 90%, 99%. A survey questionnaire was designed and pilot tested in 2 rounds among the members of the ACR-EULAR Gout Classification Criteria steering committee.
Running head: Cut point for gout classification.

Web-based survey

Using the web-based survey questionnaire, panelists were anonymously given 20 choice sets: pairs of two profiles each, describing a study type (see above) and a PPV and a NPV. The panelists were asked to make a choice as to which of the 2 study profiles they would prefer to enroll a patient. The question asked with each of the 20 choice sets was: ‘Given the study type and these positive and negative predictive values for classification criteria for gout used in this study, in which study do you think it is most acceptable to include patients? All other inclusion and exclusion criteria are the same for the two studies. These are your only options, please choose one by clicking one of the buttons below. You may think that neither of both studies are acceptable, but we would still like you to choose the most acceptable.’

We used the Sawtooth software (Sawtooth Software, SSI Web version 8.3.6 – Academic Lab) complete enumeration strategy to construct the choice sets, that were presented randomly with all options having an equal chance to be presented. The survey concluded with questions regarding age, gender, profession, years of experience and number of gout patients seen.

Analysis

The data were analyzed with Sawtooth software (Sawtooth Software, SSI Web version 8.3.6 – Academic Lab). A binary logit model was used to calculate the part-worth utilities (preferences) and the relative importance of study type, PPV and NPV (using the relative differences in range of the part-worth utilities of each attribute). Importance of each level of NPV given a specific study type and PPV was plotted in a line-graph to determine the point at which the function crosses the x-axis (i.e. utility = 0) above which there is preference.
Results

Sample

There were 91 surveys completed, while 4 surveys were started but not completed and these were not analysed. In 38 instances the survey was opened but not started. Most of the 91 panelists were male (65%) rheumatologists (93%), on average 49 years old with an average of 19 years of practice, seeing 5-60 gout patients monthly.

Relative importance

Among panelists, there appeared to be no preference for a certain study type. The phase 3 RCT with a biologic, the phase 2 RCT with an uricosuric, and the case-control GWAS, were each chosen 46% of the times they were available as a choice option, while the phase 3 RCT of an NSAID versus prednisone was chosen in 53% of times and the cohort study in 57% of times, which was not significantly different (p>0.05).

There was a clear preference for higher values of PPV and NPV. The PPV of 60% was chosen 19% of the times, and the PPV of 99% was chosen in 80% of the times it was presented as option. The NPV of 60% was chosen in 37% and the NPV of 99% was chosen in 65% of times. The range between the lowest proportion and the highest proportion chosen reflects the importance of the attribute. That range was 59% for PPV; 29% for NPV and 13% for study type, making study type relatively less important than PPV and NPV.

Preferences

We determined the study type and threshold for PPV and NPV most preferred by the majority, using the relation between the levels of each attribute and utility (not shown). The phase 3 RCT of a new biologic for gout was least preferred and the cohort study most preferred; reflecting into which study types panellists were more or less inclined to enrol subjects. The utilities for PPV and NPV increased gradually, with the point of indifference being at 80% and the minimal preferred value being at 90% for both, PPV and NPV. There
appeared to be an interaction effect only between PPV and study type \((p=0.012)\), pointing to statistical significant differences for PPV between study types. Accordingly, in table 1 the utilities for PPV by study type are shown; a utility of zero indicates ‘no preference’, positive values indicate preference. The utilities exceeding zero are shown in bold, indicating that for the middle three study types the preferred minimal PPV was 90%. For a phase 3 RCT on an NSAID versus prednisone and for an observational cohort study the preferred minimal PPV was 80%, or even 70%.

**Minimal preferable values for PPV and NPV**

Taking a PPV of 80% and a PPV of 90% as the smallest preferable values dependent on study type, the relation between NPV and utility for all study types is shown in figure 1 for a PPV of 80% (panel A) and for a PPV of 90% (panel B). From these data, the accompanying smallest preferable values for NPV were derived (table 2).
Discussion

In developing the 2015 ACR-EULAR classification criteria for gout, it was hypothesized that different cut-points to classify subjects as having gout may be needed, dependent on study type. According to the results of this study, a single cut-point is a reasonable approach for all study types if the cut-point can approximate a PPV of 90% and a NPV of 80%, as this exceeds the minimal preferred PPV and NPV values for all five types of study. A single cut-point is preferable, so that study samples and the underlying source population generally are identical for all studies.

The newly developed ACR-EULAR criteria for the classification of gout performed very well in the external validation sample, with a sensitivity of 0.92 and a specificity of 0.89 [1]. The PPV and NPV of these criteria will also depend on the disease prevalence, which can be different depending on (sub)populations and referral situations. Using Bayes theorem it is possible to calculate what the minimum and maximum disease prevalence needs to be for the ACR-EULAR classification criteria to lead to the preferred PPV of 90% and NPV of 80%.

Using the chosen threshold of ≥8, for a PPV of 90% the minimum pre-test probability needs to be 0.52, and for a NPV of 70% the maximum pre-test probability needs to be 0.83. This approaches the situation in clinical practice where, unlike screening, the diagnosis of gout is considered only in patients who conceivably may have gout, i.e. in patients having an indication to perform a diagnostic work-up for gout. In two clinical diagnostic studies, one in general practice and one in a referral rheumatology center, the prevalence of gout according to the presence of MSU crystals was indeed around 50% [7,8]. In some study types, such as a RCT of a biologic agent, it usually will be the case that investigators have already made a clinical diagnosis of gout when the classification criteria are applied. In such instances, the pre-test probability is plausibly within the range 0.52 and 0.83.
According to the results of this study there appeared to be no strong preference for study type, as opposed to an increasing preference for higher values of PPV and NPV. This was in line with the idea of the survey: to indicate preference for PPV and NPV, given study type. The results also do not mean that some types of study are not preferred when including subjects in studies. It rather reflected that there were less concerns about enrolling a subject into some study types, in the face of misclassification because of lower perceived risk. Illustratively, there was less hesitation to include subjects with gout in an observational cohort study than in a phase 3 RCT for a new biologic.

This study has some limitations. Although it was calculated that the inclusion of 63 panelists would be sufficient and we included 91, it is unclear in how far the results from this survey sample are generalizable to the rest of the population of physicians who may be involved in enrolling subjects with gout in clinical studies. From the sample description it seems that the panelists were quite representative of rheumatologists experienced in the diagnosis and treatment of gout. These may also be the rheumatologists who, more or less frequently, enroll subjects into gout studies. However, there were only few general practitioners participating, and we also do not know the countries in which the panelists were working. Further, being a survey with hypothetical profiles one was forced to choose from, there always may be a difference with the choices one may make in reality when considering enrollment of a potential subject into a study. However, opinion research, such as we performed here, is practically the only way to collect this kind of information. Conjoint analysis is well suited for the question we posed, although panelists may have difficulties in performing the task. We pilot tested the questionnaire in advance to check for understandability and feasibility, and it was positive that only four surveys were abandoned. Quite a number of surveys (38) had been opened but were not started, probably just out of curiosity upon receiving the e-mail. However, we do not have information that would inform about presence of selection bias from this sample nor from the larger sample of invitees.
Conjoint analysis has been successfully used before in diagnostic research, although with a different kind of objective. It has been used to elicit preferences for cardiac risk assessment of physicians and of the general public, and to rank 10 diagnostic strategies for carotid artery testing [9,10]. In fact, the conjoint analysis method is quite similar to the choice experiments that panelists used to construct the classification criteria for RA, SSc and gout [1,11-13]. As far as we are aware, conjoint analysis has not been used before to elucidate the position of a threshold, or cut point, for classification. A technically simpler alternative approach would be to just ask panelists about the absolute values of PPV and NPV (or: sensitivity and specificity, or number needed to treat) they would prefer, for certain types of study. That kind of approach was used in a study on selecting appropriate diagnostic strategies in patients suspected to have streptococcal pharyngitis, using physician perceived benefit-to-harm ratios [14]. The background clinical question for the physician was "How many additional treatment errors am I willing to make in order to treat one additional person correctly?", while the family physicians were asked to indicate their lowest and highest acceptable number-needed-to-treat with an antibiotic in order to treat one additional person who has streptococcal pharyngitis. However, comprehending risks and probabilities is difficult, also for professional physicians [15]. This may be overcome if natural frequencies are used [15]. However, we embarked on conjoint analysis as a much used and valid method to conduct choice experiments.

Classification criteria inform about the presence or absence of disease. Misclassification error can be made in classification criteria, and usually is being made. For research purposes, when developing a threshold for classification criteria the consequences of the error should be taken into account. This may mean that in classification criteria, like in clinical practice, there may be a need to use different cut points to define presence or absence of disease, for different types of study types (interventions). Having said that, it would be preferable for classification criteria to use a single uniform threshold for defining the condition of interest, instead of a range of different relevant cut points. We have shown that
the range of preferred PPV and NPV for a variety of study types can be achieved by a single chosen cutpoint for the 2015 ACR-EULAR Gout Classification Criteria when the pre-test probability is at least equivocal (50%) for the presence of gout. This means that the proposed threshold can be used in all types of clinical studies for gout.
References


Running head: Cut point for gout classification.

**Table 1** Utilities for PPV by study type.

<table>
<thead>
<tr>
<th>Study type</th>
<th>PPV 60</th>
<th>PPV 70</th>
<th>PPV 80</th>
<th>PPV 90</th>
<th>PPV 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 3 RCT NSAID</td>
<td>-1.57</td>
<td>-0.73</td>
<td>0.77</td>
<td>0.92</td>
<td>1.78</td>
</tr>
<tr>
<td>Phase 3 biological drug</td>
<td>-3.09</td>
<td>-1.13</td>
<td>0.0073</td>
<td>0.89</td>
<td>1.76</td>
</tr>
<tr>
<td>Phase 2 RCT uricosuric drug</td>
<td>-1.53</td>
<td>-1.29</td>
<td>-0.62</td>
<td>0.64</td>
<td>1.63</td>
</tr>
<tr>
<td>Case-control GWAS</td>
<td>-1.43</td>
<td>-0.55</td>
<td>-0.29</td>
<td>0.52</td>
<td>1.14</td>
</tr>
<tr>
<td>Cohort study</td>
<td>-1.58</td>
<td>0.16</td>
<td>0.58</td>
<td>1.22</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Utilities were calculated using a binary logit model, with study type, positive predictive value (PPV), negative predictive value (NPV), and an interaction between study type and PPV. Utilities above 0 (above the point of indifference) are printed in bold.

**Table 2** Smallest preferable PPV and NPV by study type.

<table>
<thead>
<tr>
<th>Study type</th>
<th>PPV at least</th>
<th>NPV at least</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 3 RCT NSAID</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Phase 3 RCT biological drug</td>
<td>90%</td>
<td>70%</td>
</tr>
<tr>
<td>Phase 2 RCT uricosuric drug</td>
<td>90%</td>
<td>70%</td>
</tr>
<tr>
<td>Case-control GWAS</td>
<td>90%</td>
<td>70%</td>
</tr>
<tr>
<td>Cohort study</td>
<td>80%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Smallest preferable PPV was derived from Table 1 and smallest preferable NPV was derived from figure 1.
Taking a PPV of 80% and a PPV of 90% as the smallest preferable values dependent on study type, the relation between NPV and utility for all study types is shown for a PPV of 80% (panel A) and for a PPV of 90% (panel B).