Arthritis Care & Research
Vol. 68, No. 12, December 2016, pp 1894–1898
DOI 10.1002/acr.22896

BRIEF REPORT

Survey Definitions of Gout for Epidemiologic Studies: Comparison With Crystal Identification as the Gold Standard

NICOLA DALBETH,[1] H. RALPH SCHUMACHER,[2] JAAP FRANSEN,[3] TUHINA NEOGI,[4] TIM L. JANSEN,[5] MELANIE BROWN,[6] WORAWIT LOUTHRENOO,[7] JANITZIA VAZQUEZ-MELLADO,[8] MAXIM ELISEEV,[9] GERALDINE McCARTHY,[10] LISA K. STAMP,[11] FERNANDO PEREZ-RUIZ,[12] FRANCISCA SIVERA,[13] HANG-KORNG EA,[14] MARTIJN GERRITSEN,[15] CARLO A. SCIRE,[16] LORENZO CAVAGNA,[17] CHINGTSAI LIN,[18] YIN-YI CHOU,[18] ANNE-KATHRIN TAUSCHE,[19] GERALDO DA ROCHA CASTELAR-PINHEIRO,[20] MATTHIJS JANSSEN,[21] JIUNN-HORNG CHEN,[22] MARCO A. CIMMINO,[23] TILL UHLIG,[24] AND WILLIAM J. TAYLOR[6]

Objective. **To identify the best-performing survey definition of gout from items commonly available in epidemiologic studies.**

Methods. **Survey definitions of gout were identified from 34 epidemiologic studies contributing to the Global Urate Genetics Consortium (GUGC) genome-wide association study. Data from the Study for Updated Gout Classification Criteria (SUGAR) were randomly divided into development and test data sets. A data-driven case definition was formed using logistic regression in the development data set. This definition, along with definitions used in GUGC studies and the 2015 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) gout classification criteria were applied to the test data set, using monosodium urate crystal identification as the gold standard.**

Results. **For all tested GUGC definitions, the simple definition of "self-report of gout or urate-lowering therapy use" had the best test performance characteristics (sensitivity 82%, specificity 72%). The simple definition had similar performance to a SUGAR data-driven case definition with 5 weighted items: self-report, self-report of doctor diagnosis, colchicine use, urate-lowering therapy use, and hyperuricemia (sensitivity 87%, specificity 70%). Both of these definitions performed better than the 1977 American Rheumatism Association survey criteria (sensitivity 82%, specificity 67%). Of all tested definitions, the 2015 ACR/EULAR criteria had the best performance (sensitivity 92%, specificity 89%).**

Conclusion. **A simple definition of "self-report of gout or urate-lowering therapy use" has the best test performance characteristics of existing definitions that use routinely available data. A more complex combination of features is more sensitive, but still lacks good specificity. If a more accurate case definition is required for a particular study, the 2015 ACR/EULAR gout classification criteria should be considered.**

Introduction

Information regarding the case definition of gout in epidemiologic studies is usually limited, particularly for multipurpose cohorts. A standard and accurate case definition is important for epidemiologic studies, for reasons of efficiency and validity. However, many different combinations of data available from surveys or multipurpose cohorts have been used to identify gout cases in large population studies (1), and different case definitions of disease can lead to major variation in estimates of disease

Supported by the American College of Rheumatology, European League Against Rheumatism, Arthritis New Zealand, Association Rhumatisme et Travail, and Asociación de Reumatólogos del Hospital de Cruces.

[1]Nicola Dalbeth, MD, FRACP: University of Auckland, Auckland, New Zealand; [2]H. Ralph Schumacher, MD: University of Pennsylvania, Philadelphia; [3]Jaap Fransen, PhD: Radboud University Medical Centre, Nijmegen, The Netherlands; [4]Tuhina Neogi, MD, PhD, FRCPC: Boston University School of Medicine, Boston, Massachusetts; [5]Tim L. Jansen, MD: Viecuri Medical Center, Venlo, The Netherlands; [6]Melanie Brown, MHealSc, William J. Taylor, PhD, FRACP: University of Otago, Wellington, New Zealand; [7]Worawit Louthrenoo, MD: Chiang Mai University, Chiang Mai, Thailand; [8]Janitzia Vazquez-Mellado, MD, PhD: Hospital General de Mexico, Mexico City, Mexico; [9]Maxim Eliseev, MD: Nasonova Research Institute of Rheumatology of Russia, Moscow, Russia; [10]Geraldine McCarthy, MD, FRCPI, University College Dublin

Significance & Innovations

- Gout epidemiology studies are hampered by the lack of a consistently used survey definition of gout.
- This large international study, using MSU crystal identification as the gold standard, has identified a simple survey definition with good test performance characteristics.
- However, the 2015 American College of Rheumatology/European League Against Rheumatism gout classification criteria have substantially better performance characteristics than any tested survey definitions.

incidence and prevalence (2,3). The aim of this study was to construct the best-performing case definition for gout from the limited items available in survey studies and multipurpose cohorts, testing these for accuracy against monosodium urate (MSU) crystal identification as the gold standard.

Materials and methods

Data from the Study for Updated Gout Classification Criteria (SUGAR) were analyzed. The methods of this study have been described in detail (4,5). Briefly, this was a large multinational cross-sectional study of 983 consecutive rheumatology clinic patients, with at least 1 swollen joint or suspected subcutaneous tophus, designed to identify clinical features that accurately distinguish gout from nongout. At a standardized study visit, clinical features were recorded using case record forms, in addition to independent synovial fluid microscopy by a certified observer. Gout status was defined by synovial fluid or tophus aspirate microscopy result in all patients.

Items and combinations of these items used for definitions of gout in various surveys were identified from 32 studies contributing to the Global Urate Genetics Consortium (GUGC) genome-wide association study of hyperuricemia and gout (6), and were tested in the SUGAR data set. The GUGC is a large genetics epidemiology study (>140,000 participants of European ancestry). Fifteen different definitions of gout were used, including the 1977 preliminary American Rheumatism Association survey definition (7) (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.22896/abstract). Five items for survey definitions of gout were abstracted from the GUGC studies: patient self-report of gout, patient self-report of doctor diagnosis of gout, allopurinol or other urate-lowering therapy use, colchicine use, and self-report of elevated serum urate. These variables were all available in SUGAR, with the exception of self-report of elevated serum urate, so actual serum urate level was used instead. Elevated serum urate (hyperuricemia) was defined as serum urate greater than the upper limit of normal for the local laboratory.

Data from SUGAR were randomly divided into a development data subset (two-thirds of total) and test data subset (one-third of total). Items from the GUGC gout definitions were entered into a logistic regression analysis in the SUGAR development data subset to construct a data-driven case definition, using MSU crystal–defined gout/nongout status as the dependent variable and backward selection. The score for the data-driven definition was derived from the beta coefficients in this model. The data-driven case definition and definitions used in the GUGC studies (n = 10 definitions with available data in

School of Medicine and Medical Science, Dublin, Ireland; [11]Lisa K. Stamp, PhD, FRACP: University of Otago, Christchurch, New Zealand; [12]Fernando Perez-Ruiz, MD: Hospital Universitario Cruces & BioCruces Health Research Institute, Vizcaya, Spain; [13]Francisca Sivera, MD: Hospital General Universitario de Elda, Alicante, Spain; [14]Hang-Korng Ea, MD: Université Paris Diderot, Sorbonne Paris Cité, UFR de Médecine, INSERM, UMR 1132, Hôpital Lariboisière, Assistance Publique-Hôpitaux de Paris, and Hôpital Lariboisière, Paris, France; [15]Martijn Gerritsen, MD: Westfries Gasthuis, Hoorn, The Netherlands; [16]Carlo A. Scire, MD, PhD, Italian Society for Rheumatology, Milan, Italy; [17]Lorenzo Cavagna, MD: University and IRCCS Policlinico S. Matteo Foundation, Pavia, Italy; [18]Chingtsai Lin, MD, Yin-Yi Chou, MD: Taichung Veterans General Hospital, Taichung, Taiwan; [19]Anne-Kathrin Tausche, MD: University Hospital Carl Gustav Carus, Dresden, Germany; [20]Geraldo da Rocha Castelar-Pinheiro, MD: Universidade de Estado do Rio de Janeiro, Rio de Janeiro, Brazil; [21]Matthijs Janssen, MD: Rijnstate Hospital, Arnhem, The Netherlands; [22]Jiunn-Horng Chen, MD: China Medical University School of Medicine, Taichung, Taiwan; [23]Marco A. Cimmino, MD: University of Genoa, Genoa, Italy; [24]Till Uhlig, MD: Diakonhjemmet Hospital, Oslo, Norway.

Dr. Dalbeth has received consulting fees, speaking fees, and/or honoraria from Crealta, Cymabay, Menarini, Pfizer, Takeda, and Teijin (less than $10,000 each) and from Ardea Biosciences/AstraZeneca (more than $10,000). Dr. Schumacher has received consulting fees, speaking fees, and/or honoraria from AstraZeneca, Metabolex, Novartis, and Regeneron (less than $10,000 each). Dr. Jansen has received consulting fees, speaking fees, and/or honoraria from AbbVie, AstraZeneca, Bristol-Myers Squibb, Menarini, and Roche (less than $10,000 each). Dr. Stamp has received consulting fees from AstraZeneca (less than $10,000). Dr. Perez-Ruiz has received consulting fees, speaking fees, and/or honoraria from AstraZeneca, Menarini, Novartis, Pfizer, and Sobi (less than $10,000 each). Dr. Gerritsen has received consulting and/or speaking fees from Menarini and Sobi (less than $10,000 each). Dr. Tausche has received consulting and/or speaking fees from Ardea Biosciences/AstraZeneca, Berlin-Chemie, Menarini, and Novartis (less than $10,000 each). Dr. Cimmino has received consulting fees from Menarini (less than $10,000). Dr. Uhlig has received consulting fees from AstraZeneca and Sobi (less than $10,000 each). Dr. Taylor has received consulting fees, speaking fees, and/or honoraria from Pfizer (less than $10,000) and educational grants from AbbVie, Pfizer, and Roche (less than $10,000 each).

Address correspondence to Nicola Dalbeth, MD, FRACP, Department of Medicine, University of Auckland, Private Bag 92019, 85 Park Road, Grafton, Auckland, New Zealand. E-mail: n.dalbeth@auckland.ac.nz.

Submitted for publication December 9, 2015; accepted in revised form March 22, 2016.

SUGAR, including 7 composite definitions) were applied to the SUGAR test data subset, and the sensitivity and specificity of each definition was calculated (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.22896/abstract). The 2015 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) gout classification criteria were also applied to the test data subset (8,9). Data were analyzed using SPSS, version 22.

Results

Development data subset. In the development data subset, all 5 items (patient self-report, patient self-report of doctor diagnosis, allopurinol or other urate-lowering therapy use, colchicine use, and elevated serum urate) independently contributed to the regression model (Table 1). Using these data, a score for the case definition was derived from the 5 items: self-report of gout (3 points), self-report of doctor diagnosis of gout (2 points), colchicine use (1 point), urate-lowering therapy use (2 points), and hyperuricemia (3 points) (Table 1). The points were derived from rounding the beta coefficient from the multivariate model to the nearest 0.5 and multiplying by 2. A cut point of >5 for the data-driven SUGAR survey definition provided maximal sensitivity and specificity according to the receiver operating characteristic curve (Figure 1).

Test data subset. The sensitivity and specificity for the data-driven SUGAR survey definition, along with individual items, other definitions from GUGC studies, and the 2015 ACR/EULAR gout classification criteria were calculated in the SUGAR test data subset (Table 2). Self-report of gout had the best overall performance as a single item (sensitivity 80%, specificity 72%). Use of urate-lowering therapy as a single item had high specificity (91%) but very low sensitivity (36%). For all tested GUGC definitions, the simple definition of "self-report of gout or urate-lowering therapy use" had the best test performance characteristics of existing definitions, with a sensitivity of 82% and a specificity of 72%.

Table 1. Regression model of individual survey items using the development data subset and data-driven SUGAR survey definition of gout*

Item	Odds ratio (95% CI)	β coefficient (SE)	Points
Self-report of gout	4.1 (2.4–6.8)	1.40 (0.26)	3
Self-report of doctor diagnosis of gout	3.1 (1.8–5.1)	1.12 (0.26)	2
Hyperuricemia	5.3 (3.3–8.4)	1.67 (0.24)	3
Colchicine use	1.6 (1.0–2.6)	0.49 (0.24)	1
Urate-lowering therapy use	2.2 (1.2–3.9)	0.77 (0.31)	2

* For the data-driven SUGAR survey definition of gout, a cut point >5 points provided optimal sensitivity and specificity in the development data subset. Regression-model chi-square test 305, df 5, $P < 0.001$, R^2 0.54. SUGAR = Study for Updated Gout Classification Criteria; 95% CI = 95% confidence interval.

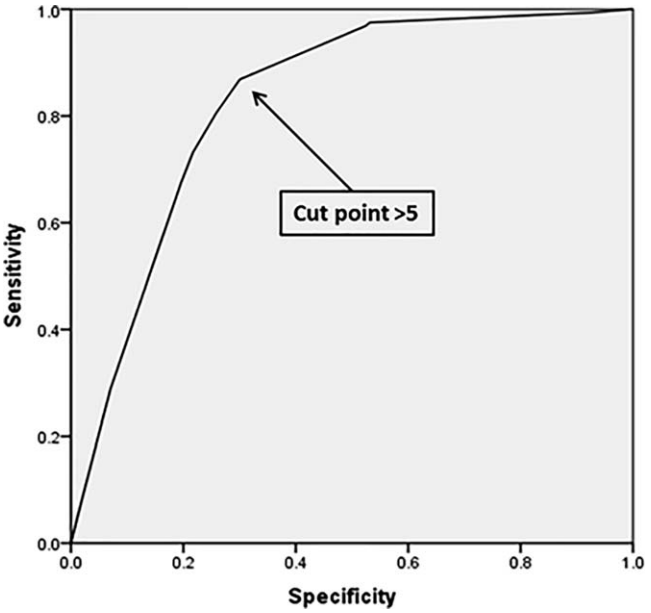

Figure 1. Receiver operating characteristic curve for the data-driven Study for Updated Gout Classification Criteria survey definition in the development data subset. Area under the curve (95% confidence interval) for curve 0.83 (0.78–0.88).

The data-driven SUGAR survey definition had a sensitivity of 87% and a specificity of 70% in the test data subset. Overall, this performance was similar to the simple definition of "self-report of gout or urate-lowering therapy use." The simple definition of "self-report of gout or urate-lowering therapy use" and the data-driven SUGAR survey definition both performed better than the 1977 American Rheumatism Association survey criteria (sensitivity 82%, specificity 67%). Of all tested definitions, the 2015 ACR/EULAR gout classification criteria had the best performance (sensitivity 92%, specificity 89%).

Discussion

This analysis has identified that a simple definition, "self-report of gout or urate-lowering therapy use," has the best (although not without limitations) test performance characteristics of existing survey definitions, with a sensitivity of 82% and a specificity of 72%. Given the design features of SUGAR, the specificity is likely to be an underestimate of test performance for population studies, and these values are therefore helpful in estimating worst-case misclassification rates from population studies. A more complex combination of features available from routinely collected data is more sensitive, but still lacks very high specificity. Importantly, none of these survey definitions perform as well as the 2015 ACR/EULAR gout classification criteria. However, the 2015 ACR/EULAR gout classification criteria require a patient interview for typical clinical characteristics of gout, physical examination, and laboratory testing, with or without imaging assessment (8,9). For large multipurpose epidemiologic studies, particularly those with general cohorts not focused on gout or

Table 2. Performance of individual items and composite survey definitions in the SUGAR test data subset*

	Sensitivity, %	Specificity, %	Youden index†
Single items used in GUGC study			
Self-report of gout	80	72	0.52
Self-report of doctor diagnosis of gout	80	69	0.49
Hyperuricemia	85	60	0.45
Colchicine use	48	76	0.24
ULT use	36	91	0.27
Composite definitions reported in GUGC study			
Self-report of gout or ULT use	82	72	0.54
Hyperuricemia and ULT use	31	94	0.25
Gout-specific medications (colchicine or ULT)	61	72	0.32
Self-report of gout or gout-specific medications	87	61	0.48
Self-report of gout and gout-specific medications	53	83	0.36
Self-report of gout or hyperuricemia	96	50	0.46
1977 preliminary ARA survey criteria	82	67	0.49
New composite definitions			
Data-driven SUGAR survey definition	87	70	0.57
2015 ACR/EULAR gout classification criteria	92	89	0.81

* SUGAR = Study for Updated Gout Classification Criteria; GUGC = Global Urate Genetics Consortium; ULT = urate-lowering therapy; ARA = American Rheumatism Association; ACR = American College of Rheumatology; EULAR = European League Against Rheumatism.
† Youden index = sensitivity + specificity − 1 (perfect test is 1; test no better than chance is 0).

established before the 2015 gout classification criteria were published, such detailed information may not be feasible or available, and for this reason survey definitions may be required.

Limitations of this study include recruitment for SUGAR from rheumatology clinics. Patients presenting to secondary care may not be representative of people with gout in a community or general-population setting due to disease severity or comorbid conditions. It is also likely that the predictive properties of all definitions will differ in a general population cohort in which the majority of participants do not have gout. In addition, although the specificity of all of these case definitions is likely to be even higher among general-population nongout controls, it is likely that the same order of specificity values we observed in SUGAR would hold true in a general-population sample. SUGAR did not collect information about self-report of elevated serum urate, and this variable may have different properties than hyperuricemia defined by a laboratory test. It is also possible that different serum urate cut points may alter the sensitivity and specificity of a survey definition. This study has a number of strengths. SUGAR is a large multinational study designed specifically to identify features that classify gout. The case definition of gout using the pathologic gold standard of crystal identification is a major strength. The findings of this study are likely to be widely applicable, noting that the items self-report of gout or urate-lowering therapy use are available in many surveys and multipurpose cohorts.

In summary, a simple definition of "self-report of gout or urate-lowering therapy use" has the best test performance of existing survey definitions for epidemiologic gout studies. If a more accurate case definition is required for a particular study, the 2015 ACR/EULAR gout classification criteria should be considered.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of Ole Slot (Copenhagen University Hospital, Glostrup, Denmark); Joung-Liang Lan, Chien-Chung Huang, Po-Hao Huang, Hui-Ju Lin, and Su-Ting Chang (China Medical University Hospital, Taichung City, Taiwan); Anne Madigan (Dublin, Ireland); Yi-hsing Chen (Taichung City, Taiwan); Alain Sanchez-Rodríguez and Eduardo Aranda-Arreola (Mexico City, Mexico); Viktoria Fana (Copenhagen, Denmark); Panomkorn Lhakum and Kanon Jatuworapruk (Chiang Mai, Thailand); Dianne Berendsen (Nijmegen, Netherlands); Femke Lamers-Karnebeek (Amsterdam, Netherlands); Olivier Peyr (Paris, France); Ana Beatriz Vargas dos Santos (Rio de Janeiro, Brazil); Fatima Kudaeva (Moscow, Russia); Angelo Gaffo (Birmingham, Alabama); Douglas White (Hamilton, New Zealand); Giovanni Cagnotto (Pavia, Italy); and Juris Lazovskis (Sydney, Nova Scotia, Canada) with data collection, crystal examination, or patient referral. The authors are grateful to Eliseo Pascual (Alicante, Spain) for help with MSU observer certification.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Dalbeth had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Dalbeth, Schumacher, Fransen, Neogi, Jansen, Taylor.
Acquisition of data. Neogi, Brown, Louthrenoo, Vazquez-Mellado, Eliseev, McCarthy, Stamp, Perez-Ruiz, Sivera, Ea, Gerritsen, Scire, Cavagna, Lin, Chou, Tausche, Castelar-Pinheiro, Janssen, Chen, Cimmino, Uhlig.
Analysis and interpretation of data. Dalbeth, Schumacher, Fransen, Neogi, Jansen, Taylor.

REFERENCES

1. Kuo CF, Grainge MJ, Zhang W, Doherty M. Global epidemiology of gout: prevalence, incidence and risk factors. Nat Rev Rheumatol 2015;11:649–62.
2. Bardin T, Bouee S, Clerson P, Chales G, Flipo RM, Liote F, et al. Prevalence of gout in the adult population of France. Arthritis Care Res (Hoboken) 2016;68:261–6.
3. Richette P, Clerson P, Bouee S, Chales G, Doherty M, Flipo RM, et al. Identification of patients with gout: elaboration of a questionnaire for epidemiological studies. Ann Rheum Dis 2015;74:1684–90.
4. Taylor WJ, Fransen J, Dalbeth N, Neogi T, Schumacher HR, Brown M, et al. Performance of classification criteria for gout in early and established disease. Ann Rheum Dis 2016;75:178–82.
5. Taylor WJ, Fransen J, Jansen TL, Dalbeth N, Schumacher HR, Brown M, et al. Study for Updated Gout Classification Criteria: identification of features to classify gout. Arthritis Care Res (Hoboken) 2015;67:1304–15.
6. Kottgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet 2013;45:145–54.
7. Wallace SL, Robinson H, Masi AT, Decker JL, McCarty DJ, Yu TF. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum 1977;20:895–900.
8. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2015;74:1789–98.
9. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheumatol 2015;67:2557–68.