RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

Development and validation of a self-updating gout register from electronic health records data

Nils Bürgisser ,[1,2,3] Denis Mongin ,[1,3] Samia Mehouachi,[1,4] Clement P. Buclin ,[2,3] Romain Guemara ,[1] Pauline Darbellay Farhoumand ,[2] Olivia Braillard ,[5] Kim Lauper ,[1,6] Delphine S. Courvoisier [1,3,4]

To cite: Bürgisser N, Mongin D, Mehouachi S, *et al*. Development and validation of a self-updating gout register from electronic health records data. *RMD Open* 2024;**10**:e004120. doi:10.1136/rmdopen-2024-004120

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/rmdopen-2024-004120).

NB and DM contributed equally.

Received 16 January 2024
Accepted 10 April 2024

Check for updates

For numbered affiliations see end of article.

Correspondence to
Nils Bürgisser;
nils.burgisser@hug.ch

ABSTRACT

Objective To develop an automatic gout register from electronic health records (EHRs) data.

Methods We analysed the EHR of all patients >18 years old from a tertiary academic hospital (2013–2022) based on six criteria: International Classification of Diseases 10 gout diagnosis, urate-lowering therapy prescription, monosodium urate crystals in joint aspiration and gout-related terms in problem lists, clinical or imaging reports. We assessed the positive and negative predictive value (PPV and NPV) of the query by chart reviews.

Results Of 2 110 902 outpatients and inpatients, 10 289 had at least one criterion for gout. The combination of joint aspiration OR diagnostic in the problem list OR≥2 other criteria created a register of 5138 patients, with a PPV of 92.4% (95% CI 88.5% to 95.0%) and an NPV of 94.3% (95% CI 91.9% to 96.0%). PPV and NPV were similar among outpatients and inpatients. Incidence was 2.9 per 1000 person-year and dropped by 30% from the COVID-19 pandemic onward. Patients with gout were on average 71.2 years old (SD 14.9), mainly male (76.5%), overweight (69.5%) and polymorbid (mean number of comorbidities of 3, IQR 1–5). More than half (57.4%) had received a urate-lowering treatment, 6.7% had a gout that led to a hospitalisation or ≥2 flares within a year and 32.9% received a rheumatology consultation.

Conclusion An automatic EHR-based gout register is feasible, valid and could be used to evaluate and improve gout management. Interestingly, the register uncovered a marked underdiagnosis or under-reporting of gout since the COVID-19 pandemic.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Gout is the most prevalent inflammatory arthritis, but it remains undertreated despite affordable and effective treatment options.

⇒ Quantifying this undertreatment and detecting its causes and risk factors to pilot quality improvement initiative requires an extensive register of gout patients.

WHAT THIS STUDY ADDS

⇒ This is the first automatic electronic health record (EHR)-based gout register, allowing frequent, inexpensive and sustainable updates.

⇒ The automated queries show high positive and negative predictive values to identify gout patients.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This register can facilitate the assessment of the adequacy of gout management and the monitoring of quality indicators following improvement projects or changes in policies.

⇒ It provides an easy platform for cohort studies or adaptive trials.

⇒ Its methodology is reproducible, facilitating the establishment of gout or other disease registers within different EHR systems.

INTRODUCTION

Gout is a chronic accumulation of monosodium urate crystals in joints and surrounding tissues. It manifests as a disease continuum, ranging from acute debilitating joint flares separated by asymptomatic intercritical period to chronic synovitis, tophi formation and progressive joint destruction.[1] Gout is the most frequent inflammatory arthritis in adult, affecting between 0.1% and 6.8% of the world's population.[2] This disease is particularly disabling, accounting for 1.3 million years lost due to disability in 2017.[3]

Despite existing guidelines on the management of the disease and widely available treatment of acute flares or chronic gouty arthritis,[4 5] gout remains alarmingly undertreated. A recent global epidemiology study reports that only 30%–50% of patients receive urate-lowering therapy (ULT) and fewer than half of them adhere to treatment.[2]

This gap highlights an urgent need to understand and address the reasons behind this undertreatment and the associated risk factors.

Despite the high prevalence of this disease, existing gout registers predominantly come from rheumatology settings[6 7] which do not fully capture the disease's spectrum.[8] One of the barriers to create larger gout registers is their labour-intensive nature, requiring manual data collection by healthcare professionals. Furthermore, administrative datasets, a common source for research studies, often lack relevant clinical information such as laboratory data and specific patient-centred outcomes.[9] Notably inflexible, these datasets cannot be tailored to answer specific research questions.[10] Built for insurance claim purposes, their ability to detect an actual patient's illness can be low depending on the condition,[11] and this is especially true for gout patient.[12 13] Building registers directly from the electronic health records (EHR) using laboratory value or International Classification of Diseases (ICD) diagnostic codes has been proven feasible for chronic kidney disease.[9] More advanced technique using natural language and machine learning techniques have shown their ability to query free text in EHR data to identify specific clinical situation such as gout flare,[14 15] recurrence of cancer[16] or aortic stenosis and its severity.[17] A combination of these methods can help solve the shortcomings of registers based on administrative datasets, streamline the process and obtain a more complete picture of gout patients.

The aims of this article are to prove the feasibility of setting up a reproducible gout register based on hospital EHR data, present the validity of its diagnostic algorithm, detail its implementation and provide an overview of the resulting register for both outpatients and inpatients.

METHODS

Study setting

The Geneva University Hospitals (HUG) is a 2000 beds French-speaking tertiary hospital, constituted of 8 hospital sites and 2 clinics. It is the only university hospital in the Canton of Geneva, which has an estimated population of 517 802 inhabitants as of the last census in December 2022. Every year, it cares for 60 000 inpatients, provides 1.2 million outpatient visits and receives close to 250 000 emergency room visits.[18] Beyond providing the standard array of care for both inpatient and outpatient, it offers specialised care to psychiatric patients, inmates and vulnerable populations.

Health and administrative data sources

A common EHR is used in every hospital, clinic and various points of care belonging to the HUG. A dedicated software contains all administrative and medical information, where any health professional can retrieve and add data transversally. These data are stored in a centralised repository and mirrored in a MongoDB database.

The HUG laboratory is accredited for joint aspiration analysis by the Swiss Accreditation Service (norm 15189), and its technicians are trained by accredited organisations, holding certificates of expertise in crystal evaluation.

Our EHR uses a problem list and problem-oriented documentation.[19] Problem lists are defined by the Royal College of Physicians Health Informatics Unit as a 'current list of a patient's problems or health issues'[20] and are contained separately from the ICD-10 diagnoses.

Inclusion criteria and time frame

All adults ≥18 years old, currently deceased or living, with any contact as an inpatient or outpatient with the Geneva University Hospital from 1 January 2013 to 31 December 2022 were included in the queries to develop the register. The year 2013 was chosen because the Swiss diagnosis-related group system was implemented in 2012. This system, used for insurance claim purposes in the inpatient setting in Switzerland, classifies patients and their diagnoses according to certain groups, which are similar in medical and economical term.[21] The German Modification of the ICD 10 (ICD-10 GM) diagnostic codes play a preponderant role in this system.

Criteria for potential gout cases

We assessed six criteria to capture gout diagnosis (table 1; for full detail see online supplemental table 1).

Table 1 Criteria considered for gout diagnosis and their conditions

Criteria considered for gout diagnosis	Conditions
1. Problem list of the EHR	Regular expression query for gout-related terms
2. Joint aspiration result	Presence of monosodium urate crystals
3. ICD-10-GM diagnosis code	M10.00–M10.99
4. Drugs	Allopurinol, febuxostat, probenecid or lesinurad
5. Documents (any reports)	Regular expression query for gout-related terms
6. Imaging reports	Regular expression query for gout-related terms

Gout-related terms included 'gout', 'podagra', 'tophus', 'tophi', 'tophaceous' for the document, problem list and imaging report criteria. The latter included also 'double-contour'.
EHR, electronic health record; ICD-10-GM, German Modification of the International Classification of Disease 10th revision.

Refining criteria for accurate diagnosis of gout

In a first step, we selected small groups of patients to verify and refine our criteria. Sample size for these initial queries was calculated based on expected positive predictive value (PPV) and tolerating a 5% half CI. For instance, for ICD codes, assuming a 95% PPV, and accepting a CI between 90% and 100%, the computed sample size was 20 patients. The charts of these 20 patients were reviewed to refine the appropriate criteria. All M11 ICD-10-GM codes (ie, other crystal arthropathies) were excluded for the ICD10 code criteria because most results were related to calcium pyrophosphate deposition disease.

For free-text searches (problem list, medical documents and imaging reports) a list of proverbs, medication and of human body liquids was built to detect false positives. Indeed, 'gout', in French 'goutte', is a very common word, that can be used also for medication (drops), as a symptom in uro-gynaecology (as blood or urine drop), in psychiatry (a proverb 'the drop that made the vase overflow', similar to 'the straw that broke the camel back'), or even as a surname. For free-text searches (problem list, medical documents and imaging reports), presence of negation or double negation was examined in sentence related to gout, to identify situations where the text expressed an exclusion of gout diagnosis. The code and the different steps of the context analysis are described in online supplemental material and the associated code is available at https://gitlab.unige.ch/goutte/register_validation.

Allopurinol was often used in the oncology setting without a gout diagnosis. By excluding patients with an ICD-10-GM codes for leukaemia or lymphoma,[22] we were able to exclude cases of allopurinol used for an oncological indication and keep cases related to gout only.

Diagnostic algorithm

Based on the previous criteria, we used the following algorithm to identify patients with gout, as any of the following three conditions:

1. A gout diagnosis in the problem list.
2. Positive joint aspiration result for monosodium urate crystals.
3. Any combination of at least two other criteria of the following variables:
   a. ICD-10-GM codes.
   b. Drugs.
   c. Text of medical documents.
   d. Text of joint imaging reports.

Sensitivity analysis

To test how the chosen diagnostic algorithm influenced the PPV and negative predictive value (NPV) of the register, we considered two alternative algorithms, one more sensitive and one more stringent than our main algorithm:

- Algorithm 1 (more sensitive): any of the six criteria.
- Algorithm 2 (more stringent).
  - A gout diagnosis in the problem list.
  - Positive joint aspiration results for monosodium urate crystals.
  - Any combination of at least three of the remaining criteria.

Gout 'gold-standard' definition

To evaluate the accuracy of the queries to detect a real gout diagnosis and further optimise them, a randomly selected sample of charts was manually assessed. Every chart was reviewed by a physician and a research nurse. Disagreements were adjudicated by a rheumatologist. A diagnosis of gout was confirmed if documented by a physician in the patient's medical records. Any text referring to a gout or a gout-related terms (ie, tophi, podagra) was considered. If multiple differential diagnoses were mentioned, the final diagnosis established was considered. In case of a rheumatological evaluation, the diagnosis of the rheumatologist had priority.

Patients with a history of gout, without any feature of gout during any episode of medical care, were considered as having gout if established as such by a doctor in the charts.

Monoarthritis or oligoarthritis with feature of gout (rapid onset of pain, response to colchicine, non-steroidal anti-inflammatory drugs (NSAID) or corticoid and no other apparent cause) but without a specified diagnosis by the team in charge was classified as equivocal. The use of a ULT without any documented gout diagnoses was also considered equivocal.

Sample size calculation

For PPV, assuming a 95% PPV, with a precision of ±2% (93%–97% CI), the calculated sample size corresponded to at least 456 patients. We applied the same conditions for NPV, yielding the same minimal number of patients. To respect the proportions of patients for each query and to account for potential incomplete information in EHR, 518 charts were extracted for the PPV and 492 for the NPV assessment.

Selection of non-gout cases at risk of developing gout

To calculate the NPV of the criteria, patients at risk of developing gout but not detected as having gout by our algorithm were selected, based on a combination of known risk factors.[2 23 24]

1. Sex ≥65 for women and ≥40 for men.
2. Overweight or obesity (body mass index >25 kg/m^2).
3. Any of the following:
   a. Hyperuricaemia (>500 µmol/L or 8.4 mg/dL).
   b. Chronic kidney disease.
   c. Metabolic syndrome.
   d. Myocardial infarction.
   e. Deleterious use of alcohol.

The detail of the criteria used can be found in online supplemental table 2.

Inpatient and outpatient differentiation

We categorised patients as inpatients or outpatients according to the setting where they first met the condition to be detected by our algorithm. For example, a patient with a positive aspiration in the ambulatory setting was categorised as outpatient, even if they later received an ICD-10 GM code in an inpatient setting. If two criteria were required, the setting where the second criterion was met determined the patient's classification as either inpatient or outpatient.

Additional variables

In addition to the criterion necessary to determine whether a patient had gout and dates on which each of the criterion were recorded, relevant information was automatically collected. They included anthropometric and demographic data, information regarding gout episodes, clinical pathway, comorbidities, laboratory values, joint aspirations, drugs and presence of a rheumatology consultation (table 4, online supplemental table 3).

Patients were also classified as having a high gout burden if they experienced a gout episode that required hospitalisation (primary diagnosis of gout) or had at least two separate gout flares occurring at least 30 days apart but within a 1-year period, as confirmed by joint aspiration. Flare was defined as followed:

1. Discharge letter with a primary diagnosis of gout. Those patients, because they needed a hospitalisation, were considered as having had an acute flare of gout.
2. Presence of monosodium urate crystals in a joint aspiration.

Statistical analysis

We summarised data using frequencies and percentages for categorical variables and means and standard deviations (SD) for continuous variables. PPV was calculated as the proportion of gout patients according to chart review divided by the total number selected for review. Equivocal was considered as not having gout.

NPV was calculated as the proportion of patients not having gout according to chart review divided by the total number selected for review. Equivocal was considered as having gout.

Confidence intervals (CI) were computed using the Agresti and Coull method.[25] All statistics were computed using the software R V.4.3.0.[26]

RESULTS

Process leading to the register

A total of 2110902 unique patients were seen at the hospital over 10 years, of which 10289 had at least one criterion for gout. Of these, 5151 were detected only by a single criterion other than the problem list and the joint aspiration (ie, only ULT, document, imaging report or ICD-10-GM) and were excluded from the register by our main diagnostic algorithm, yielding a final register of 5138 gout patients (figure 1). This corresponds to an incidence of 2.4 diagnoses per year per 1000 patients.

Criteria combination

In assessing the prevalence of the criteria among all patients (figure 2A), the most prevalent criteria were detection from documents alone (28.4%), use of a drug prescription alone (18.9%) or a combination of both (7.26%). It was then followed by the combination of the problem list and the presence of a diagnostic in a document (5.7%). For the gout register, we selected patients based on our diagnostic algorithm (Problem list OR Joint aspiration OR ≥2 Other criteria, figure 2B). Among the 5138 patients who were included, the combination documents/drugs (14.5%) and problem list/documents (11.5%) were the most frequent. Interestingly, 2.0% of patients had a positive aspiration for monosodium urate crystals without any gout diagnosis. Outpatients and inpatients showed relatively similar patterns of criteria presentation (online supplemental figure 1A,B), though a positive aspiration without further documentation was slightly more frequent for outpatients (3.0%) than for inpatients (1.7%).

Positive predictive value

Our diagnostic algorithm led to a PPV of 92.4% (95% CI 88.5% to 95.0%, see table 2). Results were similar in the inpatient and outpatient setting (93.3% and 92.4%, respectively, see online supplemental tables 6 and 7).

Individual criteria within our algorithm exhibited variable PPVs. The highest values were obtained for problem list and joint aspiration (97.2% (95% CI 92.9% to 98.9%) and 95.8% (95% CI 86.0% to 98.8%), respectively). In contrast, the PPVs were lower for drug (72.7% (95% CI 68.5% to 76.6)%), document (80.2%% (95% CI 75.9% to 83.8%)) and radiological reports (76.6% (95% CI 62.8% to 86.4%)). Notably, most patients without gout identified by these criteria were exclusively detected by a single criterion, for which ICD-10, drugs, documents and radiological reports yielded suboptimal VPP (online supplemental table 5).

Negative predictive value

Among the 2 110 902 patients seen at the hospital over 10 years, 15 646 had at least one risk factor for gout (online supplemental table 8). Of these, 2588 (16.5%) were found to have a gout by our diagnostic algorithm and were excluded, yielding 13 058 patients with risk factors.

Among these highly at-risk patients, NPV for all risk factors (table 3) was excellent, except for the uricaemia criterion. The overall NPV was 94.3% (95% CI 91.9% to 96.0%). Patients with a single risk factor (ie, without any other NPV criterion) yielded similar results (online supplemental table 9).

Sensitivity analysis

The more sensitive algorithm, considering any criterion as enough to diagnose patients with gout, would have included all 10 289 patients, with a PPV of 69.1% (95% CI 65.0% to 72.9%) and an NPV of 97.5% (95% CI 95.6% to 98.6%).

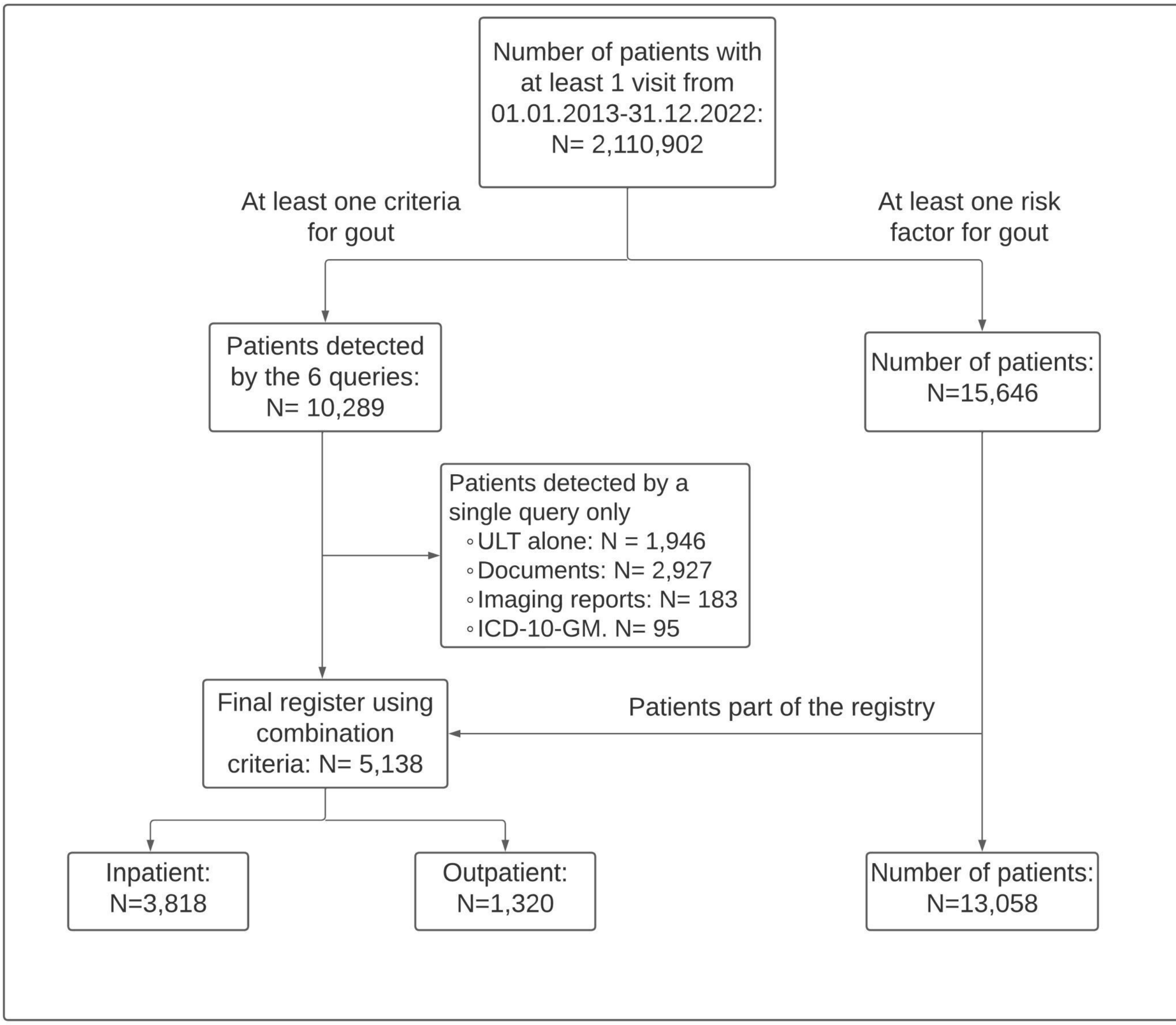

Figure 1 Flow chart of the patient selection process leading to the final gout register. ICD-10-GM, German Modification of the International Classification of Disease, 10th revision; ULT, urate-lowering therapy.

The more stringent algorithm would have included only 3746 patients, with a PPV of 96.9% (95% CI 93.5% to 98.6%) and an NPV of 91.6% (95% CI 88.8% to 93.7%).

Gout register

The 5138 patients detected by our algorithm were mostly old men, frequently overweight (table 4). The vast majority had at least one comorbidity (83%), hypertension (69.4%), cardiovascular and ischaemic diseases (stroke, heart failure, ischaemic heart and peripheral vessel disease, 53%) being the most prevalent. As of 31 December 2022, 34.5% of the patients were recorded as deceased. Most patients were Swiss citizen (68.8 %) or came from other European countries (22.6 %). At time of detection by the algorithm, 74.3% were categorised as inpatients. Patients were mostly detected in the department of medicine (27.8%), followed by geriatrics (19.5%). 92% of the patients had a document referring the gout diagnosis, and 18.6% had a joint aspiration positive for monosodium urate crystals. A gout diagnosis was documented in the problem list in half of the case overall (53.3%) but reached 78.2% in the outpatient setting. Around half the patients (49.2%) had an ICD10 code corresponding to gout. 6.7% of the patients had a gout that led to a hospitalisation or at least two flares within a year. Concerning drugs, 57.0% of the patients had received a ULT at any one time, most frequently allopurinol, and 48.3% received colchicine. Uricosurics (probenecid and lesinurad) were almost never prescribed. Only one-third of the patients (33.0%) had a rheumatology consultation.

Our algorithm revealed a 30% decrease in yearly gout diagnoses, falling from 2.9 to less than 2 per 1000 patient-years before and since the COVID-19 pandemic in 2020 (figure 3). This decline remained consistent over time and was primarily observed in the inpatient setting (online supplemental figure 2).

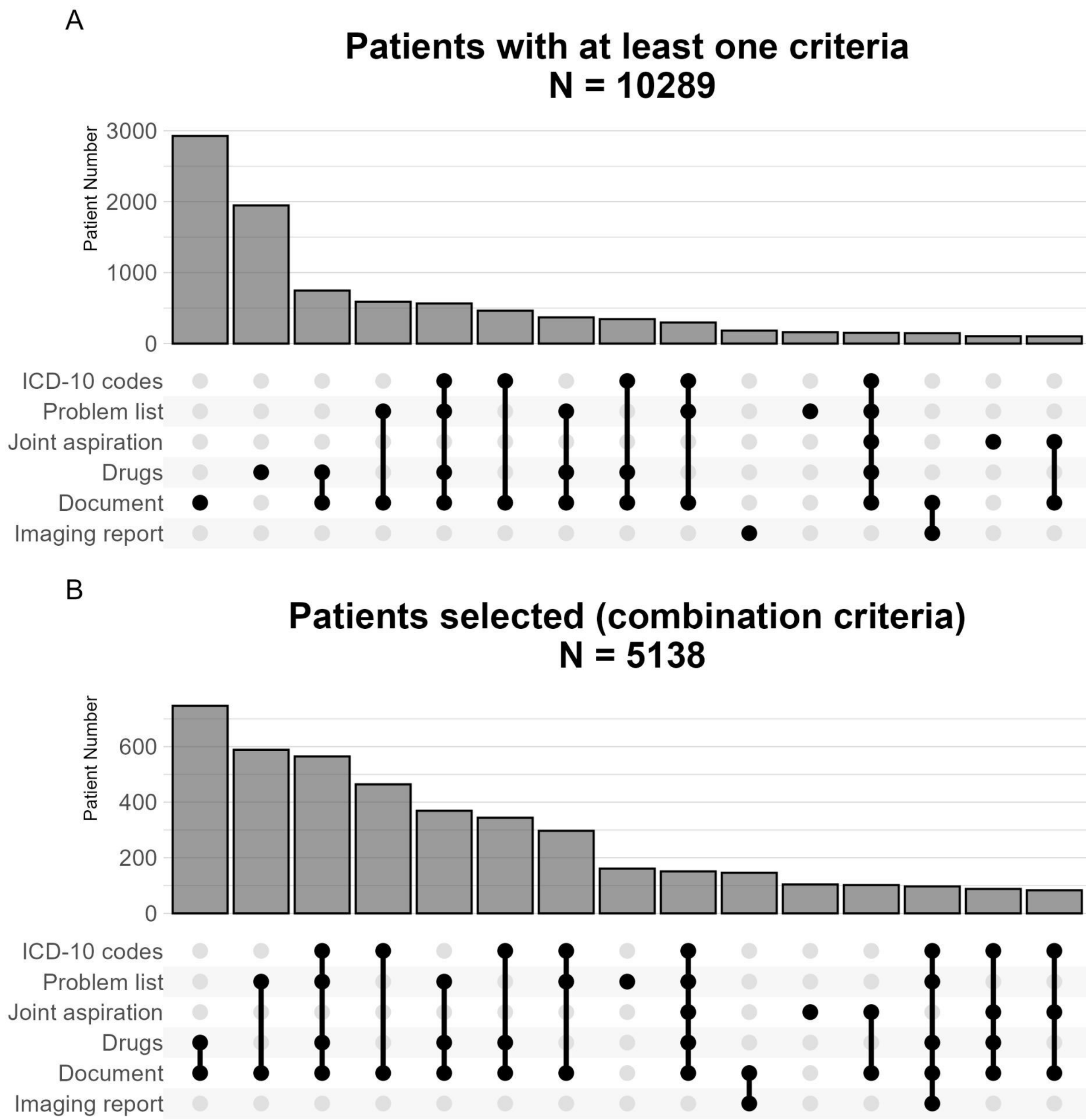

Figure 2 (A, B) Upset-plot of the six criteria identifying gout patients in the electronic health record of the Geneva University Hospital. (A) The combinations of criteria present among patients selected by at least one criteria (n=10 289). (B) The combinations of criteria present among patients selected by the final algorithm used for the register (problem list OR aspiration OR≥2 other criteria). Rare combinations of criteria are not displayed. Stratification by setting (inpatient/outpatient) can be found in online supplemental figure 1. ICD 10, International Classification of Diseases 10th revision.

When studying how our algorithm's criteria have evolved over time to initially diagnose patients with gout (second detection in the case of combination of criteria, online supplemental figure 2), we observed a rise in the problem list over time, introduced gradually in our hospital since 2011. Notably, there was a decreasing trend in joint aspiration as the initial detection method, particularly in inpatient settings since the onset of the COVID-19 pandemic in 2020.

DISCUSSION

This study demonstrates the feasibility and the relevance of building a clinical gout register through automated queries on EHR data, encompassing outpatients and inpatients across diverse care settings. Out of over 2 million patients, 5138 were definitively diagnosed with gout, reflecting an incidence rate of 2.4 per 1000 patients annually.

The fine-tuning of our criteria on a small subset of patients together with the careful estimation of the validity of various algorithms allowed us to propose an efficient algorithm with excellent PPV and NPV, ensuring accurate identification of gout patients. The incidence of newly diagnosed patients up to 2019 (pre-COVID-19) is comparable to previously developed medical records databases across the world.[2] It is twice that reported in

Table 2 Presence of a written gout diagnosis (defined gold standard) for each criterion and associated positive predicted value (PPV) (a patient can appear in multiple criteria) among 518 manually reviewed charts

Criterion	Total	Gout	No gout	Equivocal	PPV (95% CI)
Problem list	141	137	4	0	97.2% (92.9% to 98.9%)
Joint aspiration	48	46	2	0	95.8% (86.0% to 98.8%)
ICD-10-GM codes	132	121	10	1	91.7% (85.7% to 95.3%)
Drugs	462	336	29	97	72.7% (68.5% to 76.6%)
Documents	383	307	68	8	80.2% (75.9% to 83.8%)
Radiology reports	47	36	8	3	76.6% (62.8% to 86.4%)
Combination of criteria					
Problem or aspiration or ≥2 other criteria	262	242	16	4	92.4% (88.5% to 95.0%)

ICD-10-GM, German Modification of the International Classification of Disease 10th revision.

studies exclusively relying on ICD-10 codes.[27] This aligns with our finding that over half of our diagnosed patients did not have ICD-10 codes for gout. Our approach based on a combination of various criteria offers thus a less biased mean to study this disease and underscores the value of using a multifaceted approach. Our reported incidence could still be underestimated for two reasons. First, our VPN is 94%, meaning that among our 13058 patients at risk not detected by our algorithm, an additional 744 may have gout. Second, gout tends to be underdocumented, as proven by the high number of undocumented suspected gout (ie, use of allopurinol without any gout diagnosed).

The decline in gout diagnoses since the onset of the COVID-19 pandemic, the limited use of ULT or the low rate of rheumatology consultation show that our approach offers insights into diverse aspects of patient care.

The yearly incidence of patients diagnosed by our algorithm decreased from 2020 onwards, corresponding to the beginning of the COVID-19 pandemic, in line with findings from a recent study in England.[27] This decrease could be the consequence of three different factors: the decrease of diagnosis by healthcare professionals, the change of population attending the hospital or an actual decrease of the gout incidence. First, since the start of the pandemics, the disruption of medical education of physicians and medical students could have affected their ability to detect and diagnose gout.[28 29] Second, the lack of access to healthcare for gout patients due to the COVID-19 pandemics, with unrecognition of the disease and inability to refill prescription drug could explain the inflexion of diagnoses,[30] with potential lasting effect.[27] Third, although it seems unlikely that SARS-CoV-2 affected directly the occurrence of gout, it imposed a great toll on patient with cardiovascular risk factors.[31 32] This population of patient, particularly prone to gout, could have been reduced.

In our register, the best criterion (highest PPV) to detect a gout diagnosis was the problem list, which was gradually introduced in 2011 and is now used in every department. Problem lists keep track of all current and past diagnoses, they centralise the usually scattered relevant medical information's and are used to familiarise oneself with a new patient.[33] In 2022, it was a prevalent inclusion criterion in the register for both outpatient and inpatient. The outpatient setting saw higher prevalence, as it was introduced earlier in our EHR. It is, however,

Table 3 Presence of a written gout diagnosis (defined gold standard) among patients with risk factors for gout but not detected by the algorithm and associated negative predictive value (NPV) (a patient can appear in multiple risk factors) from 492 manually reviewed charts

Criterion	Total	Gout	No gout	Equivocal	NPV (CI 95%)
Age ≥65 for women OR≥40 for men AND Overweight (body mass index >25 kg/m^2):					
Uricaemia >500 µmol/L	41	8	33	0	80.5% (66.0% to 89.8%)
Chronic kidney disease	154	10	142	2	92.2% (86.9% to 95.5%)
Metabolic syndrome	79	6	73	0	92.4% (84.4% to 96.5%)
Myocardial infarction	196	2	193	1	98.5% (95.6% to 99.5%)
Deleterious use of alcohol	136	3	132	1	97.1% (92.7% to 98.9%)
Total (any risk factor)	492	24	464	4	94.3% (91.9% to 96.0%)

NPV was calculated by dividing the number of patients not having gout by the total number of patients, meaning that equivocal cases were classified as gout in NPV analysis.

Table 4 Characteristics and stratification per setting of care of the patients forming the final register

Setting Variables	Overall	Outpatient	Inpatient
Patient in the register	5138	1320	3818
Total number of patients	2 110 902	1 806 981	860 049
Incidence (per 1000-person year)	2.4	0.7	4.4
Age (mean (SD))	71.22 (14.89)	66.25 (14.88)	72.93 (14.50)
Men (%)	3931 (76.5)	1062 (80.5)	2869 (75.1)
BMI (mean (SD))	28.33 (6.54)	28.47 (6.06)	28.28 (6.71)
BMI 25–30 kg/m^2	1381 (36.9)	374 (37.0)	1007 (36.8)
BMI≥30 kg/m^2	1220 (32.6)	351 (35.1%)	869 (31.8%)
Number of death (%) at 31 December 2022	1771 (34.5)	286 (21.7)	1485 (38.9)
Criterion ever present			
Problem list	2741 (53.3)	1032 (78.2)	1709 (44.8)
Joint aspiration	957 (18.6)	236 (17.9)	721 (18.9)
ICD-10-GM Codes	2528 (49.2)	485 (36.7)	2043 (53.5)
Drugs	2807 (54.6)	663 (50.2)	2144 (56.2)
Documents	4728 (92.0)	1157 (87.7)	3571 (93.5)
Radiology reports	698 (13.6)	182 (13.8)	516 (13.5)
Urate-lowering therapy (ever)			
Allopurinol	2813 (54.7)	695 (52.7)	2118 (55.5)
Febuxostat	209 (4.1)	64 (4.8)	145 (3.8)
Probenecid	9 (0.2)	6 (0.5)	3 (0.1)
Lesinurad	1 (0.0)	1 (0.1)	0 (0.0)
None	2189 (42.6)	592 (44.8)	1597 (41.8)
Colchicine (ever)	2484 (48.3)	602 (45.6)	1882 (49.3)
Rheumatology consultation (ever)	1690 (32.9)	376 (28.5)	1314 (34.4)
High gout burden*	344 (6.7)	79 (6.0)	265 (6.9)
≥2 flare/year	76 (1.5)	26 (2.0)	50 (1.3)
Hospitalisation due to gout	294 (5.7)	63 (4.8)	231 (6.1)
Comorbidities (ICD-10-GM codes)†			
Number (median (IQR))	3 (1–5)	3 (1–5)	3 (2–5)
Any comorbidity	4266 (83.0)	997 (75.5)	3269 (85.6)
Hypertension	3564 (69.4)	814 (61.7)	2750 (72.0)
Dyslipidaemia	1822 (35.5)	453 (34.3)	1369 (35.9)
Diabetes	1524 (29.7)	355 (26.9)	1169 (30.6)
Cardiovascular diseases	2722 (53.0)	566 (42.9)	2156 (56.5)
Liver disease	603 (11.7)	149 (11.3)	454 (11.9)
Kidney disease (stage≥3)	1818 (35.4)	403 (30.5)	1415 (37.1)
Psychiatric disorder	1983 (38.6)	448 (33.9)	1535 (40.2)
Alcohol use disorder	733 (14.3)	172 (13.0)	561 (14.7)
Organ transplant	84 (1.6)	33 (2.5)	51 (1.3)
Malignancies	1145 (22.3)	306 (23.2)	839 (22.0)
Disorder of purine and pyrimidine metabolism	107 (2.1)	22 (1.7)	85 (2.2)
Department at first algorithm detection			
Medicine	1428 (27.8)	544 (41.2)	884 (23.2)
Geriatrics	1002 (19.5)	46 (3.5)	956 (25.0)
Surgery	718 (14.0)	264 (20.0)	454 (11.9)

Continued

Table 4 Continued

Setting Variables	Overall	Outpatient	Inpatient
Acute medicine	681 (13.3)	32 (2.4)	649 (17.0)
Primary care	223 (4.3)	111 (8.4)	112 (2.9)
Psychiatry	53 (1.0)	20 (1.5)	33 (0.9)
Other	1033 (20.1)	303 (23.0)	730 (19.1)

*High gout burden is defined as patients who had a gout leading to a hospitalisation or ≥2 flares proven by joint aspiration occurring at least 30 days apart but within a 1-year period.
†Details of the ICD-10 codes used to assess comorbidities can be found in online supplemental table 3. There is no missing value for sociodemographic variables, except for the BMI (27.1%). The disciplines included in the various departments can be found in online supplemental table 10.
BMI, body mass index; ICD-10-GM, German Modification of the International Classification of Disease 10th revision.

a manually created and edited tool, prone to inaccuracies, accumulation of duplicate, lack of update and incompleteness.[34] Efforts to maintain their quality are warranted.

Despite being the gold standard, it is noteworthy that the presence of monosodium urate crystals in the synovial fluid did not yield a 100% PPV. Indeed there can be a lack of consensus between operators in analysing synovial fluid, even in an accredited laboratory resulting in false positive monosodium urate crystals results.[35 36]

Despite its rather good PPV, in agreement with what has been reported in the literature,[13 37] the use of ICD codes alone was not sufficient to build a register of gout patients. Indeed, half of the patients identified did not have an ICD gout diagnosis, either because they were never hospitalised, wrongly coded or not coded at all. Studies have shown mixed result for the use of ICD codes as predictor of a gout diagnosis.[11 13 37] In the Swiss healthcare system, ICD codes are documented by specialised coders, based on a written diagnosis in the EHR, either as a problem list or in the final report. This could lead to under-reporting of the disease and lack of proper billing.

The estimated PPV of the drug criteria was not optimal partly due to the lack of gout diagnosis by a clinician in the EHR (our defined gold standard). Many patients had already been prescribed ULT outside this hospital, probably due to a history of gout. However, ULT might have been prescribed for other reasons such as kidney stone without gout, oncological indication or inappropriately for asymptomatic hyperuricaemia. In our study, most charts did not provide other diagnosis explaining the need for a ULT. Indication alert prompting the documentation of gout in the problem list, triggered by the prescription of a ULT, could help solve this shortcoming.[38]

The query in documents was complicated by the fact that gout, a prevalent word in French, is used commonly by patients and healthcare professionals. We used a combination of regular expression and natural language processing to exclude situations where the word gout was used in a negative context (eg, psychiatry) or referring to drugs (eg, drops of vitamins) and analyses (eg, drops of blood). Although some researchers have proposed artificial intelligence-based models to extract data from EHR with great success,[39 40] queries-based algorithms like ours

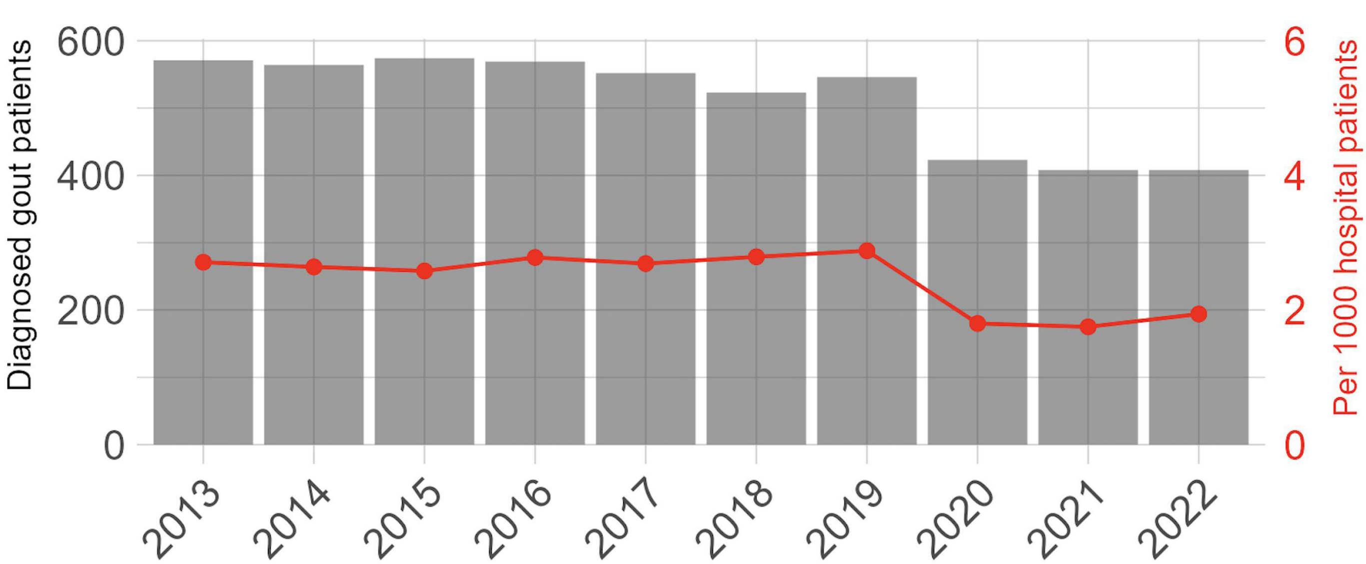

Figure 3 Evolution in time of number of gout patient per year (left y axis) and corresponding incidence (red line, right y axis).

have also succeeded with the advantage of simplicity and easy reproducibility.[41] Use of advanced natural language processing or large language model could help find the correct diagnosis when multiple differential diagnoses are mentioned, though preliminary efforts in our hospital yielded lower predictive values. Indeed, despite best optimisation of the regular expression filters, during an episode of acute arthritis, gout was often considered and mentioned in clinical documentation, especially admission records, before being discarded as the final diagnosis.

There are several strengths of our study. First, we provided detailed performance metrics of our diagnosis algorithm, using conservative choices to ensure accuracy. Indeed, we used an at-risk population for the NPV, and equivocal cases were considered as not having gout in the PPV or as having gout in the NPV calculation. Second, we proposed sensitivity analysis regarding two alternative diagnosis algorithms. Third, by providing a detailed procedure and choosing commonly documented variables, we facilitate the implementation of similar registers in other hospitals. Last, our approach uniquely sets apart this register from other registers, such as the CORRONA and EHR-based RISE register which are confined to rheumatology practices.[6 7] We included a very diverse population of outpatients and inpatients, from all specialties of an academic tertiary hospital, including vulnerable population (uninsured, migrants, inmate, etc), thereby providing a more comprehensive and varied dataset than typically seen in specialty-specific registers.

The main limitation of this register is selection bias, which could limit generalisability of the results. As in all hospital EHR-based study (eg, chronic kidney disease register[9]), this register only contains patients who used medical resources in the ecosystem of the Geneva University Hospitals. It does not assess patients consulting only in private clinics or practices, nor those who did not seek medical attention. This risk is mitigated by the fact that the Geneva University Hospitals are the only public hospital in the region, providing free inpatients and outpatients care to vulnerable population and inpatient care for the majority of the regional population. This is further confirmed by the high incidence of new cases reported by our method. Another limitation is the use of a single hospital for the register, due to the use of different EHR systems in Switzerland. Furthermore, the gout diagnosis by a physician in the hospital can be seen more as a silver standard since patients could have been diagnosed outside of the hospital, which may certainly bias prevalence studies. Finally, because of the nature of the gold standard used (diagnosis based on elements in the EHR), there is also a possibility that certain patients diagnosed with gout (according to their physician's assessment during their stay or outpatient visit) may not have been suffering from gout or that some patients with gout were missed because of missing information in the EHR. It is mitigated by the fact that more than half of the patients were under a ULT, 30% were seen by a rheumatologist and 20% had a gout that was proven by joint aspiration of monosodium urate crystals.

To allow the reproducibility of the register in other EHR, each of the criteria will need to be adapted to suit the needs of the specific EHR. For instance, not all EHR contain a problem list section. Nevertheless, we described a free text query method highlighting the potential wording issues that could arise in any language, allowing any setting to apply it to its own section containing the current list of a patient's health issues. In any case, a review of charts, according to our method to properly assess the accuracy of various combination of criteria to detect gout, could be reproduced in any hospital with an EHR.

This study proves the feasibility of implementing an EHR-based register with excellent PPV and NPV for detecting gout patient. The use of criteria based on several variables allowed to detect gout diagnosis otherwise missed by ICD codes or explicit diagnosis alone. The automatic nature of the query makes this register inexpensive and sustainable, facilitating the assessment of the adequacy of gout management, the monitoring of indicators following quality improvement projects, and the detection of gout patients to be included in new studies or trials. Also, the decline of gout diagnoses since 2020, especially evident in inpatient settings, prompts questions about how the pandemic may have affected healthcare access, patient behaviours and diagnostic approaches.

Author affiliations
[1]Division of Rheumatology, Geneva University Hospitals, Geneva, Switzerland
[2]Division of Internal Medicine, Geneva University Hospitals, Geneva, Switzerland
[3]University of Geneva Faculty of Medicine, Geneva, Switzerland
[4]Quality of Care Division, Geneva University Hospitals, Geneva, Switzerland
[5]Division of Primary Care Medicine, Geneva University Hospitals, Geneva, Switzerland
[6]Geneva Center for Inflammation research, University of Geneva Faculty of Medicine, Geneva, Switzerland

Acknowledgements We thank Dr Anne-Marie Rassinoux and Mr Emmanuel Durand from the direction of the information system of the HUG for providing access to the data lake containing the EHR datas. A feasibility study of the creation of the register was presented at the EULAR 2023 conference.

Contributors Designed research: KL and DSC. Performed research: NB, DM and SM. Analysed data: DM and NB. Wrote the paper: NB, DM, KL and DSC. Critical revision of the paper: all authors. KL and DSC are responsible for the overall content as guarantors, accepting full responsibility for the finished work and/or the conduct of the study, having access to the data, and controlling the decision to publish.

Funding This project was funded by the Private Foundation of the Geneva University Hospitals, a not-for-profit foundation.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and the creation and use of the register for quality improvement programs has been approved by the Geneva ethics commission (CCER 2023-00129). The need for consent was waived by the Geneva Ethics Committee because this study qualifies as a quality improvement initiative.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been

ORCID iDs
Nils Bürgisser http://orcid.org/0009-0007-3375-1985
Denis Mongin http://orcid.org/0000-0002-4801-8395
Clement P. Buclin http://orcid.org/0009-0000-3763-7863
Romain Guemara http://orcid.org/0000-0001-7776-4533
Pauline Darbellay Farhoumand http://orcid.org/0000-0003-4108-4266
Olivia Braillard http://orcid.org/0000-0003-2720-3900
Kim Lauper http://orcid.org/0000-0002-4315-9009
Delphine S. Courvoisier http://orcid.org/0000-0002-1956-2607

REFERENCES

1 Dalbeth N, Choi HK, Joosten LAB, *et al*. Gout. *Nat Rev Dis Primers* 2019;5:69.
2 Dehlin M, Jacobsson L, Roddy E. Global epidemiology of gout: prevalence, incidence, treatment patterns and risk factors. *Nat Rev Rheumatol* 2020;16:380–90.
3 Safiri S, Kolahi A-A, Cross M, *et al*. Prevalence, incidence, and years lived with disability due to gout and its attributable risk factors for 195 countries and territories 1990–2017: a systematic analysis of the global burden of disease study 2017. *Arthritis Rheumatol* 2020;72:1916–27.
4 Choate EA, Kaeley GS, Brook J, *et al*. Ultrasound detects synovitis in replaced and other surgically operated joints in rheumatoid arthritis patients. *BMC Rheumatol* 2020;4:8.
5 Richette P, Doherty M, Pascual E, *et al*. 2016 updated EULAR evidence-based recommendations for the management of gout. *Ann Rheum Dis* 2017;76:29–42.
6 Schlesinger N, Etzel CJ, Greenberg J, *et al*. Gout prophylaxis evaluated according to the 2012 American college of rheumatology guidelines: analysis from the CORRONA gout registry. *J Rheumatol* 2016;43:924–30.
7 Yazdany J, Bansback N, Clowse M, *et al*. The rheumatology informatics system for effectiveness (RISE): a national informatics-enabled registry for quality improvement. *Arthritis Care Res (Hoboken)* 2016;68:1866–73.
8 Jatuworapruk K, De Vera R, Estrella AM, *et al*. The APLAR gout registry: a multinational collaboration to better understand people with gout in the Asia-Pacific. *Int J Rheum Dis* 2023;26:1432–4.
9 Navaneethan SD, Jolly SE, Schold JD, *et al*. Development and validation of an electronic health record–based chronic kidney disease registry. *Clin J Am Soc Nephrol* 2011;6:40–9.
10 Cowper DC, Hynes DM, Kubal JD, *et al*. Using administrative databases for outcomes research: select examples from VA health services research and development. *J Med Syst* 1999;23:249–59.
11 Quan H, Li B, Saunders LD, *et al*. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique Dually coded database. *Health Serv Res* 2008;43:1424–41.
12 Singh JA. Veterans affairs databases are accurate for gout-related health care utilization: a validation study. *Arthritis Res Ther* 2013;15:R224.
13 Malik A, Dinnella JE, Kwoh CK, *et al*. Poor validation of medical record ICD-9 diagnoses of gout in a veterans affairs database. *J Rheumatol* 2009;36:1283–6.
14 Osborne JD, Booth JS, O'Leary T, *et al*. Identification of gout flares in chief complaint text using natural language processing. *AMIA Annu Symp Proc* 2020;2020:973–82.
15 Zheng C, Rashid N, Wu Y, *et al*. Using natural language processing and machine learning to identify gout flares from electronic clinical notes. *Arthritis Care & Research* 2014;66:1740–8.
16 Ling AY, Kurian AW, Caswell-Jin JL, *et al*. Using natural language processing to construct a metastatic breast cancer cohort from linked cancer registry and electronic medical records data. *JAMIA Open* 2019;2:528–37.
17 Solomon MD, Tabada G, Allen A, *et al*. Large-scale identification of aortic stenosis and its severity using natural language processing on electronic health records. *Cardiovasc Digit Health J* 2021;2:156–63.
18 Geneva University Hospital (HUG). Annual report of 2021. 2021. Available: https://panorama.hug.ch/sites/default/files/2022-04/HUG_rapport_gestion_2021.pdf
19 Shah AD, Quinn NJ, Chaudhry A, *et al*. Recording problems and diagnoses in clinical care: developing guidance for healthcare professionals and system designers. *BMJ Health Care Inform* 2019;26:e100106.
20 Shah A, Quinn N, Chaudhry A, *et al*. Guidance for recording problems and diagnoses in electronic health records. Available: https://theprsb.org/wp-content/uploads/2019/10/Final_Diagnosis_Recording_Guidance.pdf [Accessed 18 Mar 2024].
21 Holzer B. Swissdrg – Das Wichtigste in Kürze. *Schweiz Ärzteztg* 2012;93:1079–81.
22 Winnard D, Wright C, Taylor WJ, *et al*. National prevalence of gout derived from administrative health data in Aotearoa New Zealand. *Rheumatology (Oxford)* 2012;51:901–9.
23 Evans PL, Prior JA, Belcher J, *et al*. Obesity, hypertension and diuretic use as risk factors for incident gout: a systematic review and meta-analysis of cohort studies. *Arthritis Res Ther* 2018;20:136.
24 Dalbeth N, Phipps-Green A, Frampton C, *et al*. Relationship between serum Urate concentration and clinically evident incident gout: an individual participant data analysis. *Ann Rheum Dis* 2018;77:1048–52.
25 Agresti A, Coull BA. Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician* 1998;52:119.
26 R Core Team. R: A language and environment for statistical ## computing. R foundation for statistical computing. Vienna, Austria, 2021. Available: https://www.R-project.org/
27 Russell MD, Massey J, Roddy E, *et al*. Gout incidence and management during the COVID-19 pandemic in England, UK: a nationwide observational study using opensafely. *Lancet Rheumatol* 2023;5:e622–32.
28 Young K, Yeoh S-A, Putman M, *et al*. The impact of COVID-19 on rheumatology training—results from the COVID-19 global rheumatology alliance trainee survey. *Rheumatol Adv Pract* 2022;6:rkac001.
29 Connolly N, Abdalla ME. Impact of COVID-19 on medical education in different income countries: a scoping review of the literature. *Med Educ Online* 2022;27:2040192.
30 Tai V, Robinson PC, Dalbeth N. Gout and the COVID-19 pandemic. *Curr Opin Rheumatol* 2022;34:111–7.
31 Adair T. Premature cardiovascular disease mortality with overweight and obesity as a risk factor: estimating excess mortality in the United States during the COVID-19 pandemic. *Int J Obes* 2023;47:273–9.
32 Maung KK, Marques-Vidal P. Impact of COVID-19 pandemic on cardiovascular diseases hospitalisation, management and mortality in Switzerland. *Open Heart* 2023;10:e002259.
33 Wang ECH, Wright A. Characterizing outpatient problem list completeness and duplications in the electronic health record. *J Am Med Inform Assoc* 2020;27:1190–7.
34 Brown AR, McCoy AB, Wright A, *et al*. Decluttering the problem list in electronic health records. *Am J Health Syst Pharm* 2022;79:S8–12.
35 Lumbreras B, Pascual E, Frasquet J, *et al*. Analysis for crystals in synovial fluid: training of the analysts results in high consistency. *Ann Rheum Dis* 2005;64:612–5.
36 von Essen R, Hölttä AM, Pikkarainen R. Quality control of synovial fluid crystal identification. *Ann Rheum Dis* 1998;57:107–9.
37 Dehlin M, Stasinopoulou K, Jacobsson L. Validity of gout diagnosis in Swedish primary and secondary care - a validation study. *BMC Musculoskelet Disord* 2015;16:149.
38 Grauer A, Kneifati-Hayek J, Reuland B, *et al*. Indication alerts to improve problem list documentation. *J Am Med Inform Assoc* 2022;29:909–17.
39 Riskin D, Cady R, Shroff A, *et al*. Using artificial intelligence to identify patients with migraine and associated symptoms and conditions within electronic health records. *BMC Med Inform Decis Mak* 2023;23:121.
40 Macri CZ, Teoh SC, Bacchi S, *et al*. A case study in applying artificial intelligence-based named entity recognition to develop an automated ophthalmic disease registry. *Graefes Arch Clin Exp Ophthalmol* 2023;261:3335–44.
41 Huang SD, Bamba V, Bothwell S, *et al*. Development and validation of a computable phenotype for Turner syndrome utilizing electronic health records from a national pediatric network. *medRxiv* 2023.