

RESEARCH

Open Access

Profiling risk factors for chronic uveitis in juvenile idiopathic arthritis: a new model for EHR-based research

Tyler S Cole^{1*}, Jennifer Frankovich^{2†}, Srinivasan Iyer¹, Paea LePendou¹, Anna Bauer-Mehren¹ and Nigam H Shah¹

Abstract

Background: Juvenile idiopathic arthritis is the most common rheumatic disease in children. Chronic uveitis is a common and serious comorbid condition of juvenile idiopathic arthritis, with insidious presentation and potential to cause blindness. Knowledge of clinical associations will improve risk stratification. Based on clinical observation, we hypothesized that allergic conditions are associated with chronic uveitis in juvenile idiopathic arthritis patients.

Methods: This study is a retrospective cohort study using Stanford's clinical data warehouse containing data from Lucile Packard Children's Hospital from 2000–2011 to analyze patient characteristics associated with chronic uveitis in a large juvenile idiopathic arthritis cohort. Clinical notes in patients under 16 years of age were processed via a validated text analytics pipeline. Bivariate-associated variables were used in a multivariate logistic regression adjusted for age, gender, and race. Previously reported associations were evaluated to validate our methods. The main outcome measure was presence of terms indicating allergy or allergy medications use overrepresented in juvenile idiopathic arthritis patients with chronic uveitis. Residual text features were then used in unsupervised hierarchical clustering to compare clinical text similarity between patients with and without uveitis.

Results: Previously reported associations with uveitis in juvenile idiopathic arthritis patients (earlier age at arthritis diagnosis, oligoarticular-onset disease, antinuclear antibody status, history of psoriasis) were reproduced in our study. Use of allergy medications and terms describing allergic conditions were independently associated with chronic uveitis. The association with allergy drugs when adjusted for known associations remained significant (OR 2.54, 95% CI 1.22–5.4).

Conclusions: This study shows the potential of using a validated text analytics pipeline on clinical data warehouses to examine practice-based evidence for evaluating hypotheses formed during patient care. Our study reproduces four known associations with uveitis development in juvenile idiopathic arthritis patients, and reports a new association between allergic conditions and chronic uveitis in juvenile idiopathic arthritis patients.

Keywords: Juvenile idiopathic arthritis, Uveitis, Allergy, Electronic health records, Text mining, Biomedical informatics

Background

Juvenile idiopathic arthritis (JIA) is the most common rheumatic disease in children, with prevalence rates similar to juvenile-onset diabetes, as high as 4.01 per 1,000 children [1]. Chronic uveitis is the most threatening co-morbid condition seen in JIA patients and affects between 2% and 38% of children with arthritis [2]. Untreated uveitis

can lead to cataracts, glaucoma, band keratopathy, retinal detachment and vision loss [3]. Most JIA patients with uveitis have asymptomatic eye disease [4] and, due to their young age, are unable to articulate and/or recognize the vision changes; because of this, clinicians must screen for uveitis routinely.

Current screening guidelines are based on the understanding of two risk factors, age and ANA status [5]. Such algorithms have been the backbone of curtailing ocular complications of uveitis [2], and the discovery of novel associations will improve risk stratification with regular screening. The knowledge embedded in clinical

* Correspondence: tylcole@stanford.edu

†Equal contributors

¹Stanford Center for Biomedical Informatics Research, Stanford University School of Medicine, 1265 Welch Road, MSOB, X-215, Stanford, CA 94305-5479, USA

Full list of author information is available at the end of the article

documents from electronic health records—used, for example, to inform therapy decisions in juvenile systemic lupus erythematosus [6]—could enable such discovery for JIA and uveitis.

With computational advances in processing unstructured clinical data, large repositories of clinical data have been used for pharmacovigilance [7], phenotypic profiling [8], and for generating practice-based evidence [9]. With structured billing and claims data complemented by the rich content of clinical text, researchers argue that much of clinical medicine can benefit from analyzing data already in clinical data warehouses [6,7,10-17]. Investigators can use this data to reveal associations and predictors for hard to detect, yet severe, disease complications and co-morbidities.

Based on clinical observations, we hypothesized that allergic conditions may be associated with uveitis in JIA patients and examined this association via an informatics approach. We tested for allergy associations by mining unstructured clinical notes and coded data. Although the methods applied have been validated in other studies [7,9,18-21], as an internal validation we reproduced previously reported associations of uveitis including age [22-26], oligoarticular-onset disease [3,22-25,27], anti-nuclear antibody (ANA) status [22-25,27], rheumatoid factor (RF) status [22,23,28], and the presence of psoriasis in the patient or in immediate relatives [29]. This study adds to a growing literature demonstrating the potential of analyzing clinical data warehouses for rapidly evaluating a clinically formed hypotheses using practice-based evidence [11,30].

Methods

Data source

Our patient population was drawn from the Stanford Translational Research Integrated Database Environment (STRIDE), containing data from 1.8 million patients from the Stanford Hospital and Clinics and the Lucile Packard Children's Hospital. Acquisition and processing of data was approved by the Stanford Institutional Review Board.

Patient population

Patients followed by the pediatric rheumatology division at Lucile Packard Children's Hospital were included in the study. The pediatric rheumatologists at our center participate in the medical management of uveitis in conjunction with a community and/or university ophthalmologist. Each clinical encounter is coded with diagnosis codes for JIA and uveitis (if present) using a standardized rheumatology clinic billing form. ICD-9 code standards have not changed during the study time period. Summaries of all clinical patient encounters are dictated by pediatric residents, rheumatology fellows, and pediatric rheumatology attending physicians using our

clinic encounter template form which includes medication reconciliation of both prescription and non-prescription medications.

Cohort selection

Cohort selection was based on both ICD-9 diagnosis codes and the contents of the processed clinical notes relating to the diagnosis as outlined in Figure 1. A dual-layered method of patient identification was used to avoid the selection of incorrectly coded patients, a common and well-documented error in the medical literature [31-33]. The ICD-9 codes and terms used to identify patients are shown in Table 1. Patients under age 16 with JIA were identified between 1/1/2000 until 12/31/2011 (N = 602) and merged with patients diagnosed with chronic uveitis to obtain a cohort of JIA patients with chronic uveitis and a cohort with JIA alone. Patients with only a diagnosis of acute uveitis (ICD-9 code 364.0) were included; those with no diagnosis of chronic uveitis (ICD-9 code 364.1) were excluded (N = 4). Since patients without uveitis tended to have fewer notes in their clinical records, patients with less than 25 clinical notes were excluded to balance this variable and prevent documentation bias. Cohort characteristics and balance are shown in Table 2.

Clinical note processing

Textual notes were processed using an optimized version of the NCBO Annotator [19-21] and 22 clinically relevant ontologies. The ontologies provide a lexicon of clinical terms, from which, using a variety of statistical and manual filters [34-38], we removed ambiguous terms including overly frequent or general terms, and misleading abbreviations. We also flagged negated terms and terms attributed to family history [39], and utilized the family history attribution in evaluation of psoriasis as a risk factor. The processing steps are summarized in Additional file 1: Figure S1. We evaluated our clinical note processing pipeline for accuracy in recognizing disease events using a gold standard corpus [40], which has been manually annotated by two annotators for 16 conditions. Overall, our event identification has 74% sensitivity and 96% specificity. Recognition of drug exposure is done in a similar manner, and an independent study [41] estimated over 84% recall and 84% precision for recognizing drugs.

We then used the terms in Table 1 and simple variants of those terms, as well as matching medication names, to determine patient-term attribution. For each patient and term, we further ordered by the time at which a term occurred in this patient-feature matrix.

Statistical analysis

The clinical notes of the two cohorts were compared for significant differences in the presence of oligoarticular

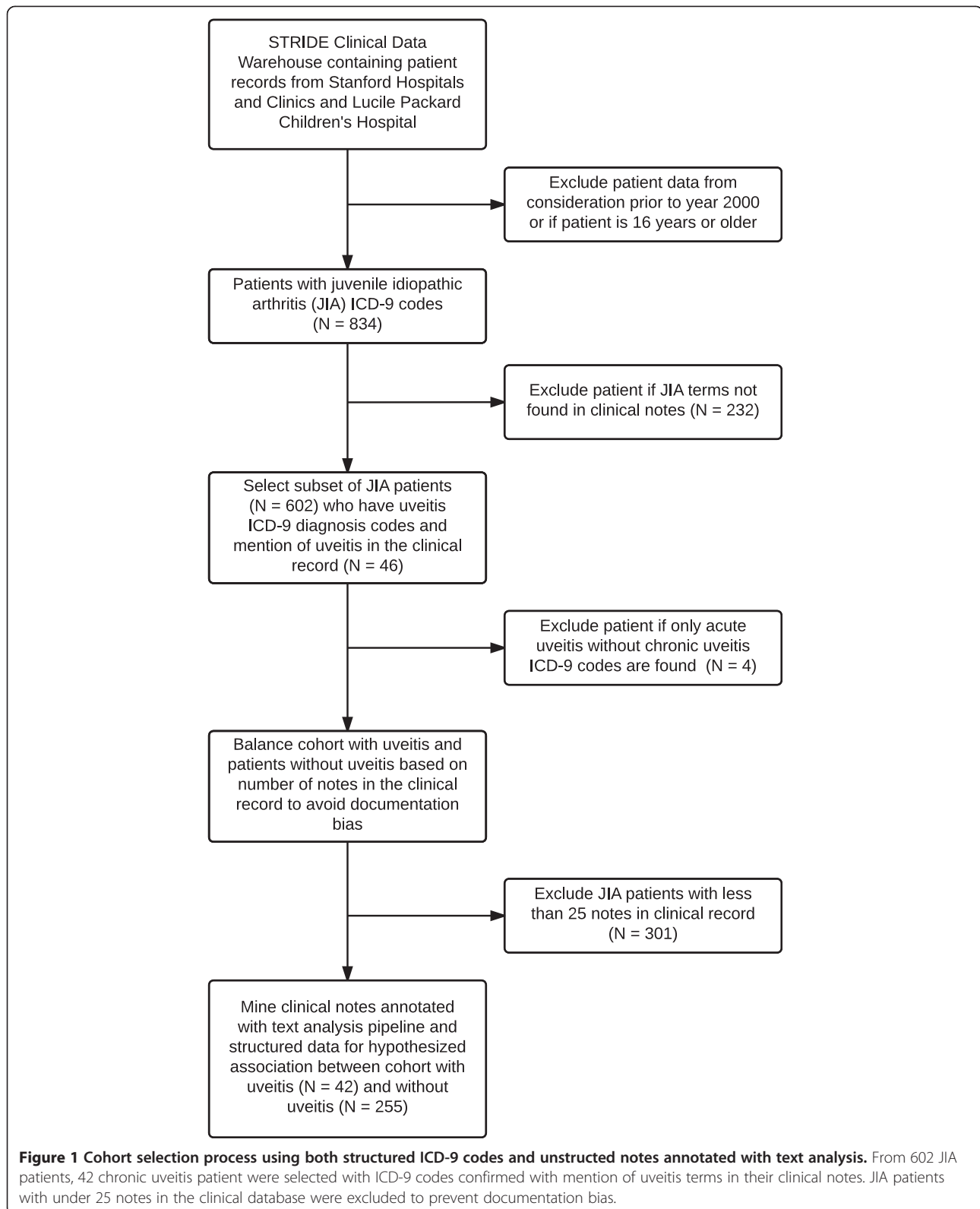


Table 1 ICD-9 codes and terms used in cohort selection and potential predictors of chronic uveitis

| Primary cohort (juvenile idiopathic arthritis) | Outcome of interest (chronic uveitis)* | Reported and hypothesized patient factors associated with uveitis |
|---|---|---|
| Juvenile Arthritis ICD 9 codes 696.0, 714.0, 714.2, 714.3, 714.9, 720.2, 720.9 | Uveitis ICD 9 codes 364.00 (acute)* 364.10 (chronic)* | Medical status terms in dictated records: Ana positive, positive ana, psoriasis, allergic, allergy, oligoarticular, oligo-onset, pauciarticular, pauci-onset, monoarthritis, monoarticular, rheumatoid factor positive, rf positive |
| Terms in dictated reports used to confirm diagnosis of juvenile arthritis: Juvenile idiopathic arthritis, jia, juvenile rheumatoid arthritis, jra, psoriatic arthritis, juvenile spondyloarthropathy, spondyloarthritis, enthesitis related arthritis, sacroiliitis, reactive arthritis and derivatives | Terms in dictated reports used to confirm the diagnosis of uveitis: Uveitis, iridocyclitis, iritis, and derivatives | Examples of allergy medications dictated in clinical records: Nasal steroids: Flonase, Nasacort Oral Antihistamines: Allegra, Zyrtec, Claritin, Clarinex, Benadryl, Xyzal Nasal antihistamines: Astelin Leukotriene inhibitors: Singulair Decongestant: Sudafed |

*This is the convention used within our institution to differentiate between acute and chronic uveitis.

disease in the first 6 months after diagnosis, ANA status, RF status, psoriasis in patient or first degree family member, allergy drugs, and allergy terms listed in Table 1.

Temporal restriction was put on the presence of diphenhydramine to be strictly prior to the first diagnosis of uveitis in the uveitis cohort to prevent confounding by its use to prevent reactions to biologics used to treat uveitis. We manually reviewed the extracted annotated terms and confirmed the presence of the terms in a non-negated context. Drugs were re-coded into a new variable representing the presence of any allergy drug. A two-tailed disproportionality test (Fisher's Exact) was used to assess association with chronic uveitis.

All variables were then examined in a multivariate logistic regression model with stepwise variable selection. Age at first arthritis diagnosis was highly associated with uveitis and was included in the logistic regression

model for adjustment. Although the cohort showed no statistical difference in race and gender (both $P > 0.7$), these variables were kept in the model for adjustment.

Cluster analysis

We then performed unsupervised hierarchical clustering based on the entire patient clinical text content prior to uveitis diagnosis. Terms from the clinical notes of the chronic uveitis patients were limited to the time period prior to the diagnosis of uveitis. The median time offset of uveitis diagnosis from arthritis diagnosis in the uveitis group (282 days) was determined, and used as the maximum time from which to extract notes in the non-uveitis cohort. Terms identified in the clinical notes were converted to medical concepts based on drug and disease ontologies (i.e. mapping similar terms to a shared concept). The presence or absence of each concept was coded

Table 2 Summary of juvenile idiopathic arthritis (JIA) cohort characteristics

| | Controls | | Cases | |
|--|--------------------------------------|-----------------------------------|-----------------------------------|---------------------|
| | JIA patients with no chronic uveitis | JIA patients with chronic uveitis | JIA patients with chronic uveitis | |
| | N = 255 | | N = 42 | |
| Race, no. (%) | | | | $P = 0.203^*$ |
| Asian | 23 (9) | | 6 (15) | |
| Black | 5 (2) | | 2 (5) | |
| Other/NA | 54 (21) | | 5 (12) | |
| White | 173 (68) | | 29 (68) | |
| Gender, no. (%) | | | | $P = 0.147^*$ |
| Female | 166 (65) | | 32 (76) | |
| Male | 89 (35) | | 10 (24) | |
| Age, years | 10.2 | | 7.2 | $P < 0.001^\dagger$ |
| Number of clinical encounter dictations | 55.4 | | 60.9 | $P = 0.54^\dagger$ |

*Difference in proportions analyzed using the Fisher's exact test.

†Difference in means analyzed with a Student's t-test.

into a binary matrix, from which a patient distance (Euclidian) matrix was calculated. Unsupervised hierarchical clustering based on inter-patient distance was then performed. Terms previously used in cohort selection, univariate analysis, or multivariate analysis were excluded from the cluster analysis to understand effects of residual textual features beyond those in Figure 2.

Software

All statistical analysis was performed using R version 2.15.1 [42].

Results

Confirming known risk factors of uveitis

In our data, age at diagnosis in this cohort reflects the age when patients are first seen at our hospital with the diagnosis of JIA. Since some patients transfer their care on referral, the mean age at diagnosis in our cohort is

biased towards an older age. The onset of arthritis was earlier in patients who develop uveitis (onset at 7.23 versus 10.17 years, $P < 0.001$). There was no significant difference in sex between the two groups ($P = 0.147$), with more females overall.

ANA positivity mentions were increased in chronic uveitis patients (adjusted OR 5.38, 95% CI 2.48–12.45, $P < 0.001$). Oligoarticular onset pattern was identified as a risk factor (adjusted OR 3.66, 95% CI 1.71–8.25, $P < 0.001$) based on data from the first 6 months after JIA diagnosis. There was also a positive association between uveitis and psoriasis in the patient or family member (adjusted OR 2.74, 95% CI 1.21–6.1, $P = 0.03$). An inverse relationship was noted between RF positivity and uveitis (adjusted OR 0.18, 95% CI 0.01–0.96, $P = 0.03$).

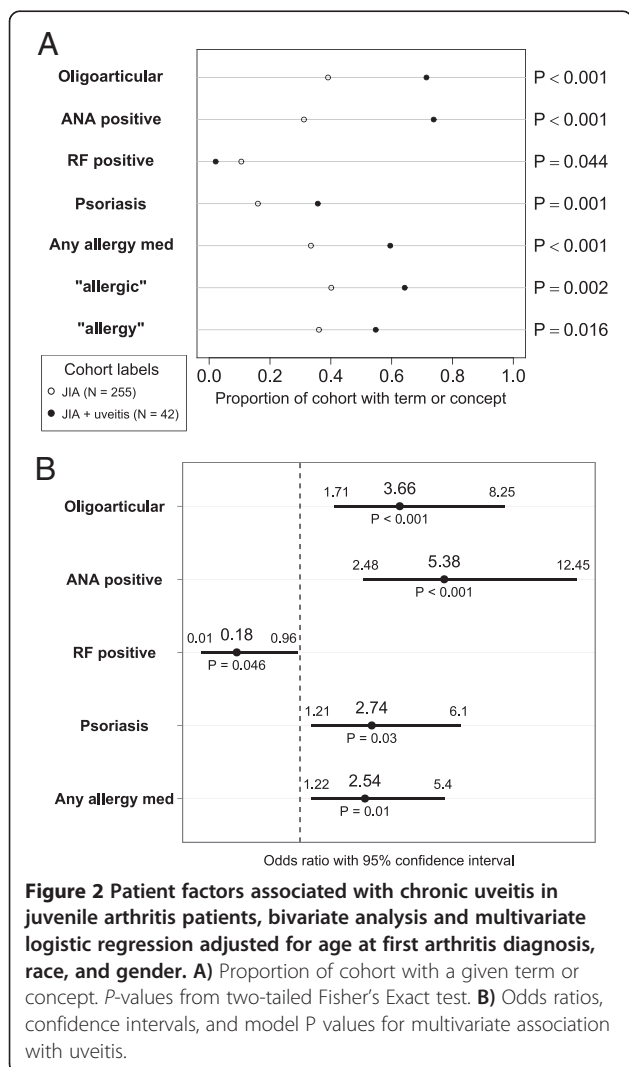
Testing the hypothesized association between allergies and uveitis

The concept of “allergy” is diffuse, so we examined the hypothesis regarding association of allergy with chronic uveitis in two ways. First, we checked for non-negated mentions of “allergy” and “allergic” within the clinical notes. Second, we created a list and checked for mentions of allergy medications commonly used to treat nasal allergy symptoms in children. We included both over-the-counter and prescription medications, outlined in Table 1.

Allergy drugs were grouped to increase statistical power given the large variety of allergy drugs used and the small patient cohort size. Allergy medication documentation was used as a proxy for nasal allergies; medication reconciliation is a routine part of the medical interview in our pediatric rheumatology clinic and is documented in the clinic visit dictations.

With bivariate analysis, the presence of an allergy drug showed the strongest association with chronic uveitis (OR 2.92, 95% CI 1.47–5.91, $P < 0.001$). Bivariate analysis also showed the presence of terms “allergic” (OR 2.68, 95% CI 1.34–5.55, $P = 0.002$) and “allergy” (OR 2.14, 95% CI 1.08–4.27, $P = 0.016$) to be associated. After finding a positive association with the allergy terms, we subsequently evaluated sinusitis due to its association with allergies. However, terms relating to sinusitis were not statistically associated (OR 1.95, 95% CI 0.94–3.8, $P = 0.051$).

The variables shown in Figure 2A were incorporated into a multivariate logistic regression model to assess adjusted associations. Age at first arthritis diagnosis was included in the model to account for residual confounding given its strong association. The results of this analysis are shown in Figure 2B. The presence of an allergy drug (OR 2.54, 95% CI 1.22–5.4, $P = 0.01$) remains associated in the multivariate analysis. There was an overall 14% prevalence of uveitis in our JIA cohort. With estimation of 50% sensitivity for allergy detection and 25% sensitivity of early uveitis detection after referral for continued



ophthalmologic evaluation prior to other indications for referral, this would represent a number needed to screen of 95.

The terms “allergy” and “allergic” are independently associated with uveitis (OR 2.14 and 2.68), but less strongly as compared to allergy medications (OR 2.92). Patients who have “allergy” or “allergic” in their clinical notes also tend to have mentions of an allergy drug, therefore the allergy term variable is eliminated in the multivariate model building with stepwise variable selection. When allergy terms are included and allergy medications are excluded in model building, the association with allergy terms remains significant (OR 2.3, 95% CI 1.11–4.96). Manual verification in the 25 chronic uveitis patients with identified allergy drug mentions confirmed direct mentions of allergy conditions in 20 patients. In 3 patients, allergic symptoms and/or consultation with an allergist were directly mentioned. In 2 patients’ clinical text, only direct mention of the allergy drug was identified. Given mandatory medication reconciliation during clinic visits at our institution, and potentially unreliable documentation of allergic conditions such as seasonal allergies due to their high prevalence, mentions of allergy drugs may be a more reliable indicator of allergic conditions. Rheumatologists, ophthalmologists, and pediatricians at our institution do not prescribe the allergy medications identified in our study for general anti-inflammatory purposes.

Cluster analysis indicating presence of additional associations with uveitis

As a qualitative look into the potential of such analytics on clinical data warehouses to discover additional factors associated with chronic uveitis, we clustered our cohort of patients based on the terms in their clinical record blinded to a future diagnosis of uveitis (Figure 3). In order to investigate if residual associations within the clinical record were strong enough to cluster patients prior to uveitis diagnosis, terms relating to RF status, ANA status, oligoarticular subtype, psoriasis, and allergy were excluded, as were demographic variables and terms already used in cohort selection.

We find that the patients who ultimately develop uveitis do tend to cluster together—even after the exclusion of the strongly associated variables that were the focus of this study—indicating that there are additional predictors within clinical documentation to be investigated. 20 uveitis patients form the main cluster, with only 3 non-uveitis patients included. Influential features of clinical interest include mentions of medications such as beclomethasone, commonly used in pediatric asthma and nasal allergies, and mentions of conditions such as contact dermatitis, folliculitis, and hypersensitivity ($P < 0.02$).

Discussion

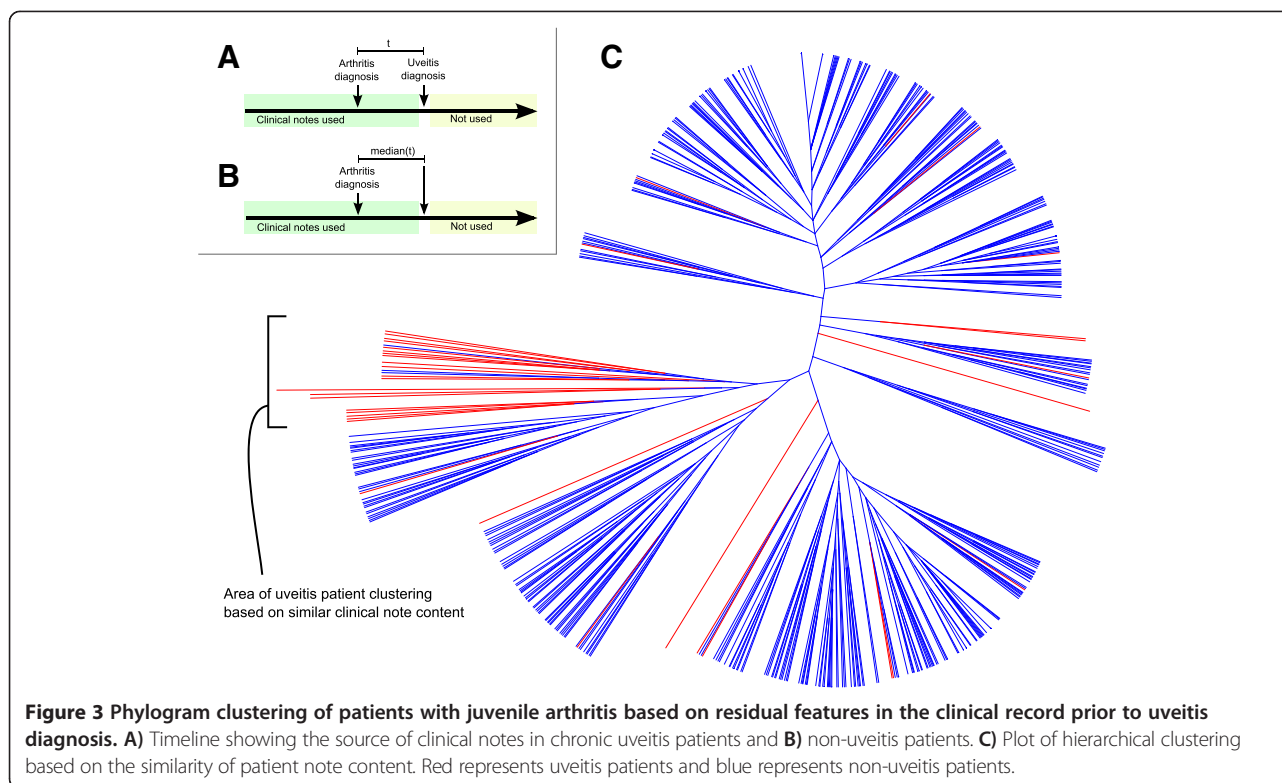
We examined a hypothesis of a possible association between allergic conditions and uveitis in JIA patients by querying a clinical data warehouse. To validate our methods in the current use case, we demonstrated known positive and negative associations with uveitis in JIA. The onset of arthritis has been demonstrated to be earlier in patients who develop uveitis [26], as shown in the current study. The established associations between uveitis and positive ANA status [22-25], oligoarticular onset pattern of JIA [3,22-25,27], and psoriasis (in patient or first degree family member) [29] is supported by our data. An inverse relationship has been established between RF positivity and uveitis [22,23,28], also confirmed in our study. The ability to also identify negative associations helps to provide an internal validation of the study approach.

After confirming these associations, we present evidence supporting a clinically formed hypothesis about an association between allergic conditions and uveitis. This study highlights the potential of text analytics on unstructured data in clinical data warehouses to examine hypotheses formed during clinical care using practice-based evidence. Without such a computational approach, such hypotheses might be impractical to answer using traditional chart review studies.

We argue that data-mining of electronic medical records—which researchers currently use to inform therapy decisions [6] and enable phase IV surveillance [9,18,19]—should be extended to learn associations and predictors of hard to detect, yet severe, disease complications. Taking such an approach also allows a spectrum of variables to be assessed. Understudied subgroups such as children, the elderly, underrepresented ethnic groups, and pregnant women can be investigated with this approach.

Despite the efficacy of such text-based analyses demonstrated in pharmacovigilance, off-label drug use, and in studying chronic conditions [7,9,18-21], this study has several potential limitations that warrant discussion. Although text analysis techniques achieve 97% accuracy in detecting negated terms, 93% accuracy in detecting drug mentions, and 86% accuracy in recognizing disease conditions in validation studies [7,9,19,20], events occurring outside of the hospital can lead to false negatives. Additionally, it is possible that there is increased reporting of allergic conditions and allergy medications among the chronic uveitis sub-group due to a higher level of concern given the eye disease complication. For these reasons, we feel that a prospective study must validate our findings before allergic conditions can be used as a clinically useful predictor.

In order to account for under-reporting bias arising from fewer visits in patients with less co-morbidity (i.e., the non-uveitis cohort), cases and controls were matched on the number of clinical dictations. The unbalanced JIA



cohort without uveitis had approximately half of the number of notes as the cohort with uveitis, and initial analysis including patients with fewer notes revealed stronger positive associations in the uveitis cohort. This was interpreted as falsely strong associations since we could not ensure that patients with less clinical record content were truly negative for a given factor or negative due to less thorough documentation. However, it is possible that patients with fewer notes have less severe disease, biasing the study population to those patients with more severe JIA. This tradeoff to ensure similar medical record information content must be recognized.

Finally, this study does not address causation. Indeed, the association between allergy and uveitis is unanticipated since autoimmune disorders are thought to have a Th₁/Th₁₇ bias while allergic disorders tend to be associated with a Th₂ cytokine profile. However, this immune classification may be an oversimplification in patients with complex and overlapping diseases, such as those with both uveitis and JIA. Recent investigation suggests that Th₂ cytokines in the anterior chamber of the eye distinguish patients with idiopathic uveitis from controls using cluster analysis methods [43]. Furthermore, we argue that allergies may be a surrogate risk factor possibly reflecting a heightened immune response to antigens in certain tissues, or a predisposition to both symptomatic and asymptomatic sinusitis where associated bacterial antigens may be the driver of an immune reaction.

If this type of research is performed widely and reliably, then it will become a key aspect of meaningfully using electronic medical records, summarizing practice-based evidence, and can help prioritize prospective clinical trials [6,11,30].

Conclusion

We report an association between allergic conditions and chronic uveitis in JIA patients. We uncovered this association via analyzing the clinical notes using a validated text analytics data-mining pipeline. If the association is confirmed in prospective studies, it may inform pathogenic mechanisms and help guide clinicians in uveitis screening in patients with allergic conditions. This study shows the potential of analyzing clinical data warehouses to gain new clinical insights and to rapidly prioritize hypotheses formed during clinical care using practice-based evidence.

Additional file

Additional file 1: Figure S1. Generation of the patient-feature matrix. This process (1) starts by downloading ~5.6 M strings for every term in ontologies from both UMLS and BioPortal as well as all trigger terms from NegEx and ConText, (2) uses term frequency and syntactic type information (e.g., predominant noun phrases) from MedLine to prune the set of strings into a clean lexicon, (3) applies the lexicon directly against the textual notes using exact string matching, (4) applies NegEx and ConText rules to identify negated terms and family history contexts respectively, (5) applies UMLS and BioPortal mappings and semantic type

information to normalize terms into concepts that are grouped by drug, disease, device, or procedure, (6) and results finally in the patient-feature matrix. Each row of the matrix represents a single patient's note and the timestamps of the notes induces a temporal ordering over the entire patient-feature matrix.

Abbreviations

ANA: Anti-nuclear antibody; EHR: Electronic health record; ICD-9: International Classification of Disease version 9; IL: Interleukin; JIA: Juvenile idiopathic arthritis; OTC: Over-the-counter; RF: Rheumatoid factor; STRIDE: Stanford Translational Research Integrated Database Environment; Th: T helper cell.

Competing interest

The authors declare that they have no competing interest or financial disclosures.

Authors' contributions

TSC: Coordination of data collection, study design, data analysis, manuscript drafting and revision. JF: Hypothesis development, clinical correlation, study conception, manuscript drafting and revision. SI: Data collection, clinical text analysis workflow, statistical and computational consulting. PLP: Data collection, clinical text analysis workflow, computational advice, and critical revision of manuscript. AB-M: Statistical consulting, critical revision of manuscript. NHS: Clinical text analysis workflow, methodology for interpreting unstructured EHR data, manuscript drafting and revision, principal investigator. All authors read and approved the final manuscript.

Acknowledgements

The National Center for Biomedical Ontology U54-HG004028, NIH Clinical Translational Science Award 5UL1RR025744-03, NIH Training Grant Number T32 AR050942 from NIAMS/NIH, Stanford Medical Scholars Program, Northern California Chapter Arthritis Foundation, and Lucile Packard Foundation for Children's Health.

Author details

¹Stanford Center for Biomedical Informatics Research, Stanford University School of Medicine, 1265 Welch Road, MSOB, X-215, Stanford, CA 94305-5479, USA. ²Division of Rheumatology, Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, USA.

Received: 6 August 2013 Accepted: 26 November 2013

Published: 3 December 2013

References

- Gowdie PJ, Tse SM: Juvenile idiopathic arthritis. *Pediatr Clin North Am* 2012, **59**(2):301–327.
- Cassidy J, Kivlin J, Lindsley C, Nocton J, Section on R, Section on O: Ophthalmologic examinations in children with juvenile rheumatoid arthritis. *Pediatrics* 2006, **117**(5):1843–1845.
- BenEzra D, Cohen E, Behar-Cohen F: Uveitis and juvenile idiopathic arthritis: a cohort study. *Clin Ophthalmol* 2007, **1**(4):513–518.
- Anesi SD, Foster CS: The importance of recognizing juvenile idiopathic arthritis-associated uveitis and preventing blindness from it. *Arthritis Care Res (Hoboken)* 2012, **64**(5):653–657. doi:10.1002/acr.21599.
- National Guideline C: Best Evidence Statement (BEST). Screening for Uveitis in Children with Juvenile Idiopathic Arthritis (JIA). <http://www.guideline.gov/content.aspx?id=38247>. Accessed 11/21/2012.
- Frankovich J, Longhurst CA, Sutherland SM: Evidence-based medicine in the EMR era. *N Engl J Med* 2011, **365**(19):1758–1759.
- Lependu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchyska T, et al: Pharmacovigilance using clinical notes. *Clin Pharmacol Ther* 2013, **93**(6):547–555.
- Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al: Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc* 2013, **20**(e1):e147–154.
- Leeper NJ, Bauer-Mehren A, Iyer SV, Lependu P, Olson C, Shah NH: Practice-based evidence: profiling the safety of cilostazol by text-mining of clinical notes. *PLoS One* 2013, **8**(5):e63499.
- Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al: 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff* 2011, **30**(4):581–589.
- Kohane IS, Drazen JM, Campion EW: A glimpse of the next 100 years in medicine. *N Engl J Med* 2012, **367**(26):2538–2539.
- Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr: Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009, **301**(8):831–841.
- Fineberg HV: Shattuck Lecture. A successful and sustainable health system—how to get there from here. *N Engl J Med* 2012, **366**(11):1020–1027.
- Evans RS, Lloyd JF, Pierce LA: Clinical use of an enterprise data warehouse. *AMIA Annu Symp Proc* 2012, **2012**:189–198.
- Roelofs E, Persoon L, Nijsten S, Wiessler W, Dekker A, Lambin P: Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. *Radiother Oncol* 2013, **108**(1):174–179. doi:10.1016/j.radonc.2012.09.019.
- Jensen PB, Jensen LJ, Brunak S: Mining electronic health records: towards better research applications and clinical care. *Nat Rev Gen* 2012, **13**(6):395–405.
- Cohen KB, Hunter LE: Chapter 16: text mining for translational bioinformatics. *PLoS Comput Biol* 2013, **9**(4):e1003044.
- Liu Y, Lependu P, Iyer S, Shah NH: Using temporal patterns in medical records to discern adverse drug events from indications. *AMIA Summ Translat Sci Proce* 2012, **2012**:47–56.
- Lependu P, Iyer SV, Fairon C, Shah NH: Annotation analysis for testing drug safety signals using unstructured clinical notes. *J Biomed Seman* 2012, **3**(Suppl 1):S5.
- Jonquet C, Shah NH, Musen MA: The open biomedical annotator. *Summ Translat Bioinform* 2009, **2009**:56–60.
- Shah NH, Bhatia N, Jonquet C, Rubin D, Chiang AP, Musen MA: Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinformatics* 2009, **10**(Suppl 9):S14.
- Cassidy JT, Sullivan DB, Petty RE: Clinical patterns of chronic iridocyclitis in children with juvenile rheumatoid arthritis. *Arthritis Rheum* 1977, **20**(2 Suppl):224–227.
- Kurrie R, Kanzy EJ, Racenberg J, Lang W, Seiler FR: BMA 031—a TCR-specific monoclonal antibody for clinical application. *Transplant Proc* 1989, **21**(1 Pt 1):1017–1019.
- Saurenmann RK, Levin AV, Feldman BM, Rose JB, Laxer RM, Schneider R, et al: Prevalence, risk factors, and outcome of uveitis in juvenile idiopathic arthritis: a long-term followup study. *Arthritis Rheum* 2007, **56**(2):647–657.
- Heinz C, Mingels A, Goebel C, Fuchsluger T, Heiligenhaus A: Chronic uveitis in children with and without juvenile idiopathic arthritis: differences in patient characteristics and clinical course. *J Rheumatol* 2008, **35**(7):1403–1407.
- Nordal EB, Songstad NT, Berntson L, Moen T, Straume B, Rygg M: Biomarkers of chronic uveitis in juvenile idiopathic arthritis: predictive value of antihistone antibodies and antinuclear antibodies. *J Rheumatol* 2009, **36**(8):1737–1743.
- Prieur AM, Chedeville G: Prognostic factors in juvenile idiopathic arthritis. *Curr Rheumatol Rep* 2001, **3**(5):371–378.
- Saurenmann RK, Rose JB, Tyrrell P, Feldman BM, Laxer RM, Schneider R, et al: Epidemiology of juvenile idiopathic arthritis in a multiethnic cohort: ethnicity as a risk factor. *Arthritis Rheum* 2007, **56**(6):1974–1984.
- Guillaume S, Prieur AM, Coste J, Job-Deslandre C: Long-term outcome and prognosis in oligoarticular-onset juvenile idiopathic arthritis. *Arthritis Rheum* 2000, **43**(8):1858–1865.
- Pincus T, Sokka T: Evidence-based practice and practice-based evidence. *Nat Clin Pract Rheum* 2006, **2**(3):114–115.
- Maclntyre CR, Ackland MJ, Chandraraj EJ, Pilla JE: Accuracy of ICD-9-CM codes in hospital morbidity data, Victoria: implications for public health research. *Aust N Z J Public Health* 1997, **21**(5):477–482.
- Zhan C, Miller MR: Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003, **12**(Suppl 2):i58–63.
- O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM: Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005, **40**(5 Pt 2):1620–1639.
- McCray AT, Burgun A, Bodenreider O: Aggregating UMLS semantic types for reducing conceptual complexity. *Stud Health Technol Inform* 2001, **84**(Pt 1):216–220.
- Bodenreider O, McCray AT: Exploring semantic groups through visual approaches. *J Biomed Inform* 2003, **36**(6):414–432.
- Parai GK, Jonquet C, Xu R, Musen MA, Shah NH: The lexicon builder Web service: building custom lexicons from two hundred biomedical ontologies. *AMIA Annu Symp Proc* 2010, **2010**:587–591.

37. Xu R, Musen MA, Shah NH: A comprehensive analysis of five million UMLS metathesaurus terms using eighteen million MEDLINE citations. *AMIA Annu Symp Proc* 2010, **2010**:907–911.
38. Wu ST, Liu H, Li D, Tao C, Musen MA, Chute CG, et al: Unified medical language system term occurrences in clinical notes: a large-scale corpus analysis. *J Am Med Inform Assoc* 2012, **19**(e1):e149–e156.
39. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG: A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* 2001, **34**(5):301–310.
40. Uzuner O: Second i2b2 workshop on natural language processing challenges for clinical records. *AMIA Annu Symp Proc* 2008, **6**:1252–1253.
41. Marshall MS, Boyce R, Deus HF, Zhao J, Willighagen EL, Samwald M, et al: Emerging practices for mapping and linking life sciences data using RDF — a case series. *Web Semant Sci Serv Agents World Wide Web* 2012, **14**:2–13.
42. R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012.
43. Curnow SJ, Falciani F, Durrani OM, Cheung CM, Ross EJ, Wloka K, et al: Multiplex bead immunoassay analysis of aqueous humor reveals distinct cytokine profiles in uveitis. *Invest Ophthalmol Vis Sci* 2005, **46**(11):4251–4259.

doi:10.1186/1546-0096-11-45

Cite this article as: Cole et al.: Profiling risk factors for chronic uveitis in juvenile idiopathic arthritis: a new model for EHR-based research. *Pediatric Rheumatology* 2013 **11**:45.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

