Case Study: Computing Complexity Scores to Identify Patients of Interest from Inspire.com Forums for Safety and Beyond

Thomas M[1], Curry A [2], Painter J[3], Akhtar A[4] , Schifano L[2], Powell GE[2]

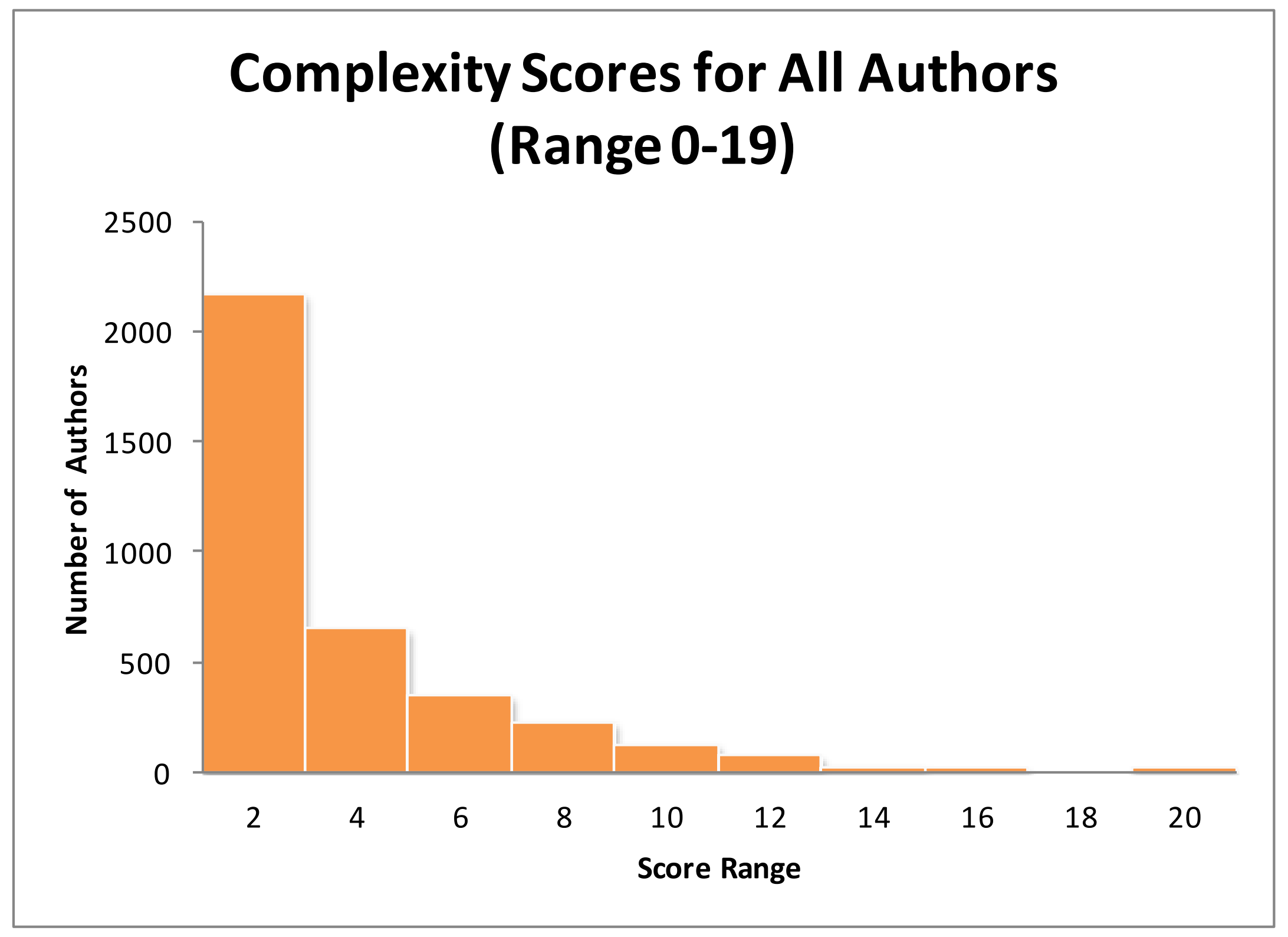
[1] GlaxoSmithKline, Collegeville, PA, USA; [2]GlaxoSmithKline, Research Triangle Park, NC, USA; [3]JiveCast, Raleigh, NC, USA; [4]ZeroChaos, Orlando, FL, USA

Introduction and Background

"Patients and caregivers across several thousand reported conditions are writing about their experiences [on Inspire.com], and generating relevant language that others who are facing similar experience can find."[1] Disease-focused Inspire forum data provides valuable patient insights and the ability to link authors' posts within discussion threads to create longitudinal records.[2] To further assess the value of these records, publicly available social media posts for two disease areas, rheumatoid arthritis (RA) and systemic sclerosis (SS), were retrieved from the Scleroderma Foundation Support Community and Arthritis Foundation Support Community maintained by Inspire and de-identified by a third party vendor. Using a combination of automated algorithms and human curation, data in relevant posts from linked discussion threads were characterized. Using this data, complexity scores were computed in order to identify longitudinal posts of interest for the purpose of constructing disease journeys and investigating patient-related insights.

Objective

To create and evaluate a complexity scoring methodology to systematically identify longitudinal posts of interest in order to further investigate disease-related insights.



Complexity Scores for All Authors (Range 0-19)

Glossary of Terms Used

Complexity Score – a metric to both standardize and consolidate attributes of authors in social media forums for comparison purposes

Curation- the act of manually reviewing posts that have been automatically processed by applying human judgment to further describe/ categorize certain key attributes.

Deidentified – process of removing PII (Personally Identifiable Information) from social media data.

Longitudinal Record – a series of posts from the same author over a period of time.

Probable RA or Probable SS - Patient clinically or contextually categorized as a probable patient. Scenarios include: awaiting lab test results, awaiting confirmed diagnosis from provider; initial poster of thread explicitly states their diagnosis, subsequent poster agrees "me too…" and provides symptom profile.

Yes, Other – other autoimmune diseases (list in curation guide)

Yes, RA- rheumatoid arthritis. JRA (juvenile rheumatoid arthritis) also included.

Yes, SS – systemic sclerosis. Various abbreviations also included: SD (scleroderma), LSSc (limited systemic sclerosis).

Methods

Social Media posts (Jan-01-2015 – Nov-01-2015) on SS and RA, retrieved from the Inspire forums (Scleroderma Foundation Support Community and Arthritis Foundation Support Community) were deidentified, to remove PII, and a unique identifier was assigned to individual authors in order to follow their activity in discussion threads. Expert reviewers manually curated a random sample of 2,817 threads containing 21,313 individual posts from 3,601 unique authors. Patient diagnosis was classified by the following hierarchy: "yes, both", "yes, RA", "yes, SS", "probable, RA", "probable, SS", "yes, other". Since a single post may only provide minimal insight regarding disease relevant information, we programmatically evaluated each author using a decision tree to determine their ultimate disease classification based on available posts in discussion threads. To further identify patients of interest, a weighted complexity score comprised of 28 indicators (see Table 2) was computed and applied for each author. Each indicator contributed a value of (1) to the complexity score. If an author mentioned two rare indicators, disease duration or participation in a clinical trial, they were each assigned a weight of (2). Complexity scores coupled with a specified minimum number of posts by the author, correlated to the richness of an author's cumulative posting record in the online discussion forum.

Results

Of the 21,313 curated posts, 5,559 (26%) were identified as being authored by the patient, 351 (1.6%) by family members, 15,342 (72%) were unknown, and the remaining posts were made by healthcare providers, caregivers and friends (< 1%). Of the 3,601 unique authors, 1191 (33%) indicated they or the person who was the subject of the post had been diagnosed with SS, RA or both; 203 (5.6%) were diagnosed with other autoimmune diseases; 232 (6.4%) indicated probable RA or probable SS diagnoses; 1975 (55%) did not specify a diagnosis. 15 patients of interest were subsequently identified using the following criteria: a complexity score greater than or equal to 14, and a minimum of four posts across discussion threads.
These 15 patients comprised a total of 1684 of the 21,313 curated posts (8%). Complexity score results ranged as follows: "yes, both" (1-19), "yes, RA" (1-16), "yes, SS" (1-15), "probable, RA" (1-8), "probable, SS" (1-8), "yes, other" (1-8).The highest computed complexity score was 19 and represented a patient with both RA and SS. Table 1 characterizes the top 4 patients of interest having the highest complexity scores.

Table 1. Top 4 Patients Of Interest Using Highest Complexity Scores

Patient Age/ Gender	SS, RA, or both	Indicators mentioned	Total Curated Posts	Complexity Score
63/ Unknown	Both SS and RA	Alternative Treatments, Concomitant medications, Medical History, Treatment History, Lab Results, Disability Status, Socio-economic Status, Alcohol Use, Seeking Information, Access Concerns, Adherence Concerns, Disease Burden, Medical Device, **Disease Duration**, Products, Indication, Medically Relevant, Homeopathic Therapy	539	19
52/ Female	Both SS and RA	Alternative Treatments, Concomitant medications, Medical History, Treatment History, Lab Results, Seeking Information, Access Concerns, Adherence Concerns, Device Concerns, Disease Burden, **Disease Duration**, Products, Indication, Medically Relevant, Homeopathic Therapy	23	16
43/ Female	Both SS and RA	Alternative Treatments, Concomitant medications, Medical History, Treatment History, Lab Results, Disability Status, Seeking Information, Disease Burden, Interest in Clinical Trials, **Disease Duration**, **Clinical Trial Participation**, Products, Indication, Medically Relevant	53	16
65/ Female	RA	Alternative Treatments, Medical History, Treatment History, Disability Status, Socio-economic Status, Pregnancy, Seeking Information, Access Concerns, Disease Burden, Products, **Disease Duration**, **Clinical Trial Participation**, Indication, Medically Relevant	48	16

Table 2. 28 Unique Indicators Comprising Complexity Scores

Alternative Treatments
(ex: Physical therapy)
Disease Burden
Alcohol Use
Disability Status
Medical Device
Device Issue
Lab Results
Pregnancy
Treatment History
Medical History
Weight
Smoking
Ethnicity status
Socio-economic status
Seeking Information

Disease Duration
Clinical Trial Participation
Interest in Clinical Trials
Access Concern
Delay in Treatment Concern
Concomitant Medications
Adherence Concern
Products Identified
Side effect mentioned
Product Complaint
Homeopathic Therapy
Specified need for additional options
(e.g. need for medical alert bracelets)
Medically Relevant

Conclusions based on Complexity Scoring Applied to Inspire Forum Data

- Threaded data from Inspire.com can be leveraged to investigate author designation (e.g. patient, family member, etc.) and patient insights including diagnosis, disease duration, clinical trial participation, disease burden and disability status as they relate to two autoimmune diseases of interest.
- Creation of unique identifiers allowed us to follow a patient's voice in a deidentified fashion through the forum as he/she progressed along a disease journey.
- Examining an author's longitudinal record enabled the determination of the highest level of disease diagnosis (e.g. probable diagnosis, diagnosis etc.)
- It was possible to identify longitudinal posts of interest by computing and applying a weighted complexity score.
- It is possible to widen or narrow the range of patients of interest by adjusting the required minimums for complexity score, number of posts by an author and/or number of indicators.
- The total complexity score correlated to the richness of an author's cumulative posting record through the online discussion forum.
- There is sufficient content to create patient disease journeys which would be helpful not only to Global Clinical Safety and Pharmacovigilance but to other groups in GlaxoSmithKline.
- Additional research is necessary to more efficiently construct disease journeys for identified patients of interest, and to determine how best to leverage these insights for drug development and safety.
- Additional research is required to further assess the impact of evolving privacy regulations on extracting and using insights from social media.

[1] https://globenewswire.com/news-release/2017/02/08/915048/0/en/Inspire-grows-online-patient-community-to-one-million-strong.html (accessed 16-May-2017)

[2] Understanding Disease Burden and Outcomes from the Patient's Perspective Using Disease-Focused Internet Forum Data, Thomas M[1], Akhtar A[2], Terkowitz J[3], Powell GE[4], ISPOR, May 2017