

Understanding Disease Burden and Outcomes from the Patient's Perspective Using Disease-Focused Internet Forum Data



Thomas M¹, Akhtar A², Terkowitz J³, Powell GE⁴

¹GlaxoSmithKline, Collegeville, PA, USA; ²ZeroChaos, Orlando, FL, USA; ³Inspire, Arlington, VA, USA; ⁴GlaxoSmithKline, RTP, NC, USA

Introduction

Publicly available social media forums, including patient / disease forums such as Inspire, offer focused discussion communities and have seen exponential growth in recent years. In fact, some industry professionals claim that online forums may actually contain richer data for public health and safety-related topics compared to Facebook or Twitter, given that publicly available Facebook data is no longer freely accessible for public health purposes and Twitter data has character limitations.

For two disease areas of interest, rheumatoid arthritis and systemic sclerosis, we hypothesized that Inspire forum data may offer useful information as well as longitudinal insights into disease and patient perspectives from linked discussion threads.

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." Reaching one billion words written by its community members in 2016, Inspire has one million registered members as of Feb 2017 interacting in more than 200 online groups.¹

GSK has typically adopted a medication-focused approach to pharmacovigilance. With the onset of social media for patient safety, we initiated the current research effort to discover whether patient insights about disease exist in online disease-focused forums and what value threaded discussion data offers for drug development and patient safety-related activities.

Objective

To examine what value disease-focused internet forum data offers GSK, including characterizing data elements in posts and whether linked discussion threads provide longitudinal insights into disease and patient perspectives.

Methods

All publicly available, English language posts from the Scleroderma Foundation Support Community, and Arthritis Foundation Support Community maintained by Inspire were deidentified and provided by a third party vendor for the period of 01 Jan 2015 to 30 Nov 2015. Medications and ProtoAE language in posts were automatically coded through dictionary algorithms as part of data preparation. Discussion threads were programmatically linked by a unique identifier to facilitate longitudinal analysis. The initial retrospective analysis dataset contained 2 disease areas, RA and SS, and consisted of 5,606 discussion threads comprised of 62,806 unique posts representing 6,548 distinct authors. A random sample of 2,817 threads from the initial 5,606 threads (50%) consisting of 21,313 posts (34%), representing 3,601 unique authors (55%) was used to create the curation and analysis dataset (see Glossary below) for this research.

Expert reviewers manually curated the 21,313 randomly selected posts using pre-determined questions designed to characterize a number of post attributes associated with disease discussions. A curation guide was maintained and weekly curation meetings were held to facilitate consistent interpretation of data in posts. Draft analysis was performed when curation was 30% complete and again when curation was 100% complete.

Posts were categorized as in-scope or not by curators, based on whether an in-scope diagnosis was discussed or whether the post represented a patient with an in-scope diagnosis as follows (see Glossary): Yes, SS; Yes, RA; Yes, Other; Yes, Both (RA and SS); Probable RA; Probable SS; No (not in scope). Each in-scope post was then further curated using the pre-determined curation questions to characterize additional data elements.

Glossary of Terms Used

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 from social media data

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.

ProtoAE – terminology in posts resembling a potential adverse event

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

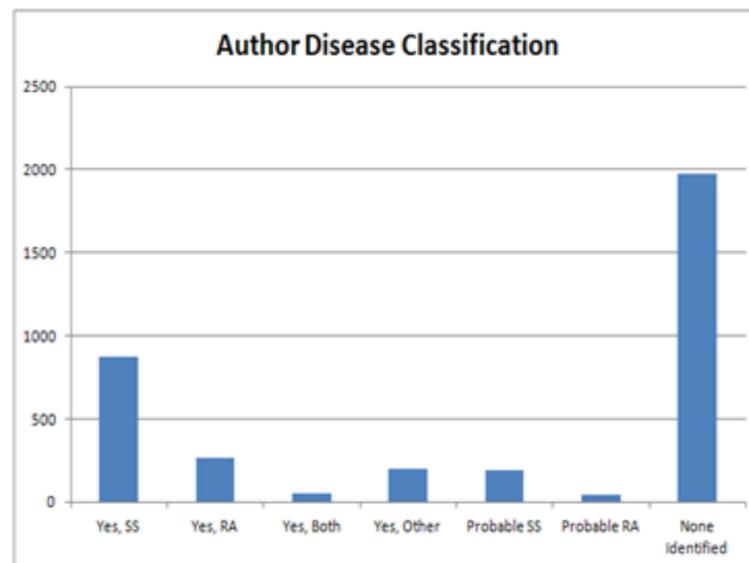
Yes, RA- rheumatoid arthritis. JRA (juvenile rheumatoid arthritis) also included.

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

Results

- Of the 3601 unique authors curated, 1191 (33%) indicated they had been diagnosed with SS, RA or both. 203 (5.6%) were diagnosed with a different autoimmune disease.
- The draft analysis performed when curation was 30% complete revealed inconsistencies in disease classification /interpretation for SS. The curation guide was updated for subsequent posts so the interpretation was made clear and the initial SS posts were recurated to remediate the inconsistency.
- Disease duration was discussed in 722 of 21,313 posts (3.4%), and of those posts 296 (41%) indicated patient had the disease for more than 10 years followed by 169 posts (23%) where patient was recently diagnosed (i.e. <1 year).

Inspire Project – Data Elements



3,601 Authors Reviewed

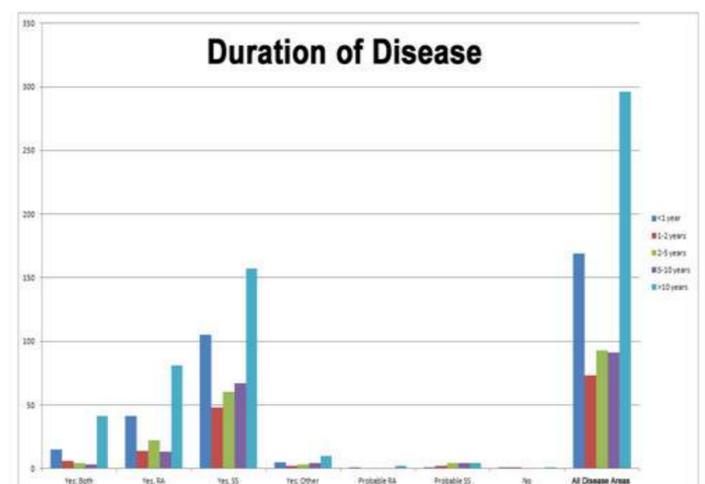
- Yes, SS: 870 (24.16%)
- Yes, RA: 267 (7.41%)
- Yes, Both: 54 (1.5%)
- Yes, Other: 203 (5.64%)
- Probable SS: 192 (5.33%)
- Probable RA: 40 (1.11%)
- None Identified: 1,975 (54.85%)

Data Elements Characterized in Posts (n=21,313 posts)



- The patient was found to be the poster in 26% (5541) of posts, family members posted in 1.6% (341) of posts. We were unable to determine the poster for 72% (15,303) of posts.
- ~13% (2762) gave medication information.
- ~3% (644) discussed potential AEs.
- ~3% (588) discussed non-medical treatments.
- ~5% (1045) discussed medical history.
- ~2% (435) discussed treatment history.
- ~2% (417) gave lab/imaging results.
- ~5% (1031) were seeking information.
- ~1% (209) discussed cost/financial concerns.
- ~8% (1778) discussed burden of disease.

Patients discussed disease duration in 722 of the 21,313 posts (3.4%), and of these 722 posts, 296 (41%) indicated they had the disease for more than 10 years.



Conclusions based on SS and RA discussion forum data

- Disease-focused Inspire forum data provides valuable patient insights and long-standing experience with the disease
- Data elements in posts from Inspire community discussions were valuable and able to be characterized
- Inspire's disease-focused internet forum data allows for linked discussion threads and creation of longitudinal posting records
- Additional research is required to further assess the value of longitudinal data for understanding the patient disease journey and how best to leverage these insights for drug development and safety.

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)