

# Rosacea Patients Are at Higher Risk for Obstructive Sleep Apnea: Automated Retrospective Research

Peter L. ELKIN<sup>1,2</sup>, Sarah MULLIN<sup>1,2</sup>, Sylvester SAKILAY<sup>1,2</sup>

<sup>1</sup>*Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, SUNY, USA*

<sup>2</sup>*Western NY VA, Department of Veterans Affairs*

**Abstract.** Using big data science we employ NLP and a novel interface the BMI Investigator to answer clinically meaningful questions. The use case presented is the association between Rosacea and Obstructive Sleep Apnea.

**Keywords.** Big data science, information Retrieval, NLP, Rosacea, Obstructive Sleep Apnea

## 1. Introduction

Automated retrospective research is a goal of an EHR driven research program. Here with this application we can go from clinical observation to science in ten minutes using a large cohort driven observational database linked to codified clinical notes using SNOMED CT. In this study we show that patients with Rosacea are at significantly increased risk for Obstructive Sleep Apnea (OSA).

Rosacea is a chronic skin condition that typically affects the face.[2][3] It results in redness, pimples, swelling, and small and superficial dilated blood vessels.[2] Obstructive sleep apnea (OSA) is the most common type of sleep apnea and is characterized by repeated episodes of complete or partial obstructions of the upper airway during sleep, OSA is usually associated with a reduction in blood oxygen saturation.

The BMI Investigator is a computer human interface built in .Net which allows simultaneous query of structured data such as demographics, administrative codes, medications (coded in RxNorm), laboratory test results (coded in LOINC) and formerly unstructured data in clinical notes (coded in SNOMED CT). The SNOMED CT codes are stored using Berkley DB and NoSQL Elastic Search, and the structured data is stored in SQL using the OMOP / OHDSI format. We have integrated genomic data into the software platform and are working to integrate image features into the query engine.

## 2. Methods

We use the high throughput phenotyping – Natural Language Processor (HTP-NLP) developed at the University at Buffalo.[4]. The BMI Investigator application was written in .Net and was created using the user-centered design development method with a team of programmers, physicians, graduate students and faculty [5,6]. We tested the system on

a population of 212,343 patients in our outpatient practices at the UBMD practice plans age 18 and older. The data for this trial was from 2010 to 2015. The data used in the system was judged by the IRB to be IRB Exempt #587570. The system allows users to use Boolean logic and parentheses to construct their queries.

Genomic data is presented as gene abnormalities that are used in clinical medicine and polymorphisms that have been identified are stored in a separate set of tables and they are also used to match to our patients who are included in the precision oncology project [7]. In an exploratory analysis, which took five to ten minute, we entered four queries into the BMI Investigator (Rosacea, Rosacea and OSA, OSA, total eligible cases in the database) and used a Pearson Chi-Square test to compare the results.

### 3. Results

The dataset has 212,343 patients. 211,764 patients did not have Rosacea. 5443 patients had OSA without Rosacea. 580 patients had OSA with Rosacea and 51 patients had OSA with Rosacea. The chance of OSA without Rosacea was 5443/211,764 (2.6%). The chance of OSA with Rosacea was 51/580 (8.8%). The Pearson Chi-Square test showed a p-value of <0.001. The relative risk of OSA given Rosacea was 3.4 times those patients without Rosacea. The number needed to test was only 12 to get one positive case.

### 4. Conclusions

The BMI investigator is an evolving tool linked to observational EHR data that has the ability to facilitate rapid retrospective research which can be clinically relevant. In this case we have shown evidence that patients with Rosacea are at increased risk for Obstructive Sleep Apnea and with a low number needed to test (NNT) they should all be asked the screening questions for OSA and if positive they should be considered for polysomnography (a sleep study). Many more studies and quality and safety projects are possible using the EHR based data and the BMI Investigators.

### References

- [1] Sand, M; Sand, D; Thrandorf, C; Paech, V; Altmeyer, P; Bechara, FG (4 June 2010). "Cutaneous lesions of the nose". *Head & Face Medicine*. **6**: 7. doi:10.1186/1746-160X-6-7. PMC 2903548. PMID 20525327.
- [2] Tüzün Y, Wolf R, Kutlubay Z, Karakuş O, Engin B (February 2014). "Rosacea and rhinophyma". *Clinics in Dermatology*. **32** (1): 35–46. doi:10.1016/j.clindermatol.2013.05.024. PMID 24314376.
- [3] van Zuuren, EJ; Fedorowicz, Z (September 2015). "Interventions for rosacea: abridged updated Cochrane systematic review including GRADE assessments". *The British Journal of Dermatology*. **173** (3): 651–62. doi:10.1111/bjd.13956. PMID 26099423.
- [4] Schlegel DR, Crowner C, Lehoullier F, Elkin PL. HTP-NLP: A New NLP System for High Throughput Phenotyping. *Stud Health Technol Inform*. 2017;235:276-80.
- [5] Beuscart-Zephir MC, Elkin P, Pelayo S. Human factors engineering for clinical applications. *Stud Health Technol Inform*. 2006;124:685-90.
- [6] Husser CS, Buchhalter JR, Raffo OS, Shabo A, Brown SH, Lee KE, PL Elkin. Standardization of microarray and pharmacogenomics data. *Methods Mol Biol*. 2006;316:111-57.
- [7] Elkin PL, Mullin S, Sakilay S. *Biomedical Informatics Investigator*. *Stud Health Technol Inform*. 2018;255:195-199