**eMERGE Network Proposal for Analysis**

Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | 2/14/2017 |
| **Reference Number** | NT214 |
| **Project Title** | Machine learning-based discovery of Atopic Dermatitis (AD) Sub-Populations in Adults |
| **Tentative Lead Investigator (first author)** | Al'ona Furmanchuk |
| **Tentative Senior Author (last author)** | Jonathan Silverberg |
| **All other authors** | Erin Gustafson , Will Thompson, Firas Wehbe, Jen Pacheco, Abel Kho, Kathryn Jackson |
| **Sites Involved** | Northwestern, other interested eMERGE sites |
| **Background / Significance** | Atopic dermatitis (AD) is an autosomal dominant disorder that leads to intense itch and diffuse rash. The US prevalence of AD is up to 10% in adults1 and 20% in children2. AD is associated with higher rates of allergic disease, including asthma, hay fever and food allergy, as well as multiple non-allergic comorbidities, including sleep disturbances, infections, depression, attention-deficit hyperactivity disorder, obesity, cardiovascular disease, osteoporosis and fractures, etc. AD typically starts in early childhood and disease activity persists into adulthood in >80% of patients. There is also a subset of patients who first develop clinical AD during adulthood3.  The diagnosis of AD in adults is often very challenging4-6. First, the distribution of AD lesions tends to differ from that of children. Second, it can be difficult to distinguish AD from other disorders, such as allergic contact dermatitis and cutaneous T cell lymphoma. Third, conventional histopathologic evaluation of AD lesions is often imprecise and cannot distinguish between AD and other types of eczema. Moreover, there is a subset of AD patients are higher risk for cardiovascular and cerebrovascular disease and a variety of other comorbidities.  In prior work7, the EHR-based case and control algorithms were developed in order to identify patients with and without AD. In this work, we intend to apply unsupervised clustering algorithms to the case population in order to identify clinically interesting sub-populations that may represent different forms and manifestations of AD. These sub-populations can be further analyzed for differential outcomes, responses to treatment, and genomic profiles. |
| **Outline of Project** | 1. Use cases identified using an EHR-based case selection algorithm for AD. 2. Use a comprehensive set of features such as symptoms, medications, diagnoses. These features will be largely re-used from the EHR-based case algorithm. 3. Use unsupervised algorithms to cluster AD patients into sub-populations. 4. Analyze sub-populations for statistical associations with outcomes. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | 1. \*AD signs, symptoms, and severity 2. Comorbidities. 3. \*Treatments including moisturizers, topical anti-inflammatories and then UVB, cyclosporine and methotrexate. |
| **Desired data** | 1. \*Demographics 2. \*AD severity & sub-type. 3. TBD relevant diagnoses (by ICD diagnosis codes) by date. 4. TBD relevant medications (by class or by RxNorm) by date. 5. TBD relevant lab results 6. AD signs/symptoms, medical history, & other relevant data extracted via NLP (a portable program will be developed by NU) 7. Geocoded data. |
| **Planned Statistical Analyses** | Unsupervised machine learning (model-based clustering8) for sub-phenotyping, supervised machine learning (support vector machines9, random forest10) for prediction of outcomes in AD sub-populations. Features analysis to determine similarity within cluster and key-features for outcomes of interest. |
| **Ethical considerations** | None. |
| **Target Journal** | JAMIA, or specialty dermatology clinical journal |
| **Milestones\*\*** | 1. 10/2017 : ready for secondary validation 2. 11/2017: secondary validation complete 3. 2/2018: run at all participating sites 4. 2/28/2018: 1st draft sent 5. 3/31/2018: 2nd draft sent 6. 4/7/2018: submit to journal |

**\*\*** This section should include: Timeline for completion of project, including approval, project duration, first and second draft of the paper and submission.

**References**

1. J. I. Silverberg, J. M. Hanifin. Adult eczema prevalence and associations with asthma and other health and demographic factors: A US population-based study. J. Allergy Clin. Immunol. 2013,132,1132-1138.
2. S. Nutten. Atopic Dermatitis: Global Epidemiology and Risk Factors. Ann. Nutr. Metab. 2015, 66, 8–16
3. Bannister MJ, Freeman S. Adult-onset atopic dermatitis. Australas J Dermatol. 2000, 41, 225–232.
4. A. Katsarou, M. Armenaka. Atopic dermatitis in older patients: Particular points. J. Eur. Acad. Dermatol. Venereol. 2011, 25, 12–19.
5. E. Ozkaya. Adult-onset atopic dermatitis. J Am Acad Dermatol. 2005, 52, 579–660.
6. H. C. Williams. Epidemiology of atopic dermatitis. Clin Exp Dermatol. 2000, 25, 522–530.
7. Concept sheet: “EHR Phenotyping of Atopic Dermatitis in Adults”.
8. C. Fraley, A. E. Raftery. Model-based clustering, Discriminant Analysis, and Density Estimation. J. Am. Stat. Ass. 2002, 97, 611-631.
9. V. Vapnik, S. Golowich, A. Smola, Support vector methods for function approximation, regression estimation, and signal processing. Adv. Neural Inf. Process. Syst. 1996, 9, 281-287.
10. T. K. Ho. The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998, 20, 832–844.