statistically significant association between proximity to highways and disease phenotypes, however there was a trend for higher incidence of discoid rash (OR 1.5) neurological disorders (OR 1.4) and antiphospholipid antibodies (1.4). Analysis of genome-wide methylation data revealed 3 methylation sites that were significantly hypomethylated in patients who resided in a high risk zone (P value < 1.7 E-07, Table 1). These three sites belonged to a single gene, *UBE2U,* which encodes one of the E2 enzymes involved in the ubiquitination of proteins and histones, as well as DNA repair. To replicate these findings, CpG Sites for all the *UBE* gene family were analysed in a control group. Our 3 top hits were not replicated, however one CpG site (cg22352634) belonging to the first intron of the gene *UBE2U* was found hypomethylated in controls who resided in a high risk zone, with a p value of 0.0044.

Conclusions Hypomethylation of *UBE2U* was associated with residing close to a highway in our SLE patients, however this was not seen in a control cohort, and suggesting increased susceptibility for exposures in patients with SLE. Additional work is warranted to confirm these findings, examine other potentially relevant exposures, and determine whether these epigenetic changes are associated with increased UBE2U expression.

CE-49 RHEUMATIC AND NON-RHEUMATIC AUTOIMMUNE DISEASES IN SLE OFFSPRING

Julie Couture, Sasha Bernatsky, Susan Scott, Christian Pineau, **Evelyne Vinet***. *Department of Medicine, MGill University, Canada*

10.1136/lupus-2016-000179.127

Background Autoimmune diseases (AID) have familial aggregation and frequently share a common genetic predisposition. Only few small uncontrolled studies have evaluated the risk of AID in SLE offspring, with inconsistent results. In a large population-based study, we aimed to determine if children born to mothers with SLE have an increased risk of rheumatic and non-rheumatic AID compared to children born to mothers without SLE.

Materials and methods The "Offspring of SLE mothers Registry (OSLER)" includes all women who had ≥1 hospitalisation for delivery after SLE diagnosis, identified through Quebec's universal healthcare databases (1989–2009). OSLER also includes a randomly selected control group of women, matched at least 4:1 for age and year of delivery, who did not have a diagnosis of SLE prior to or at the time of delivery. We identified children born live to SLE mothers and their matched controls, and ascertained rheumatic (i.e. juvenile idiopathic arthritis, SLE, systemic sclerosis, Sjögren's disease, inflammatory myositis, systemic vasculitis) and non-rheumatic (i.e., type 1 diabetes, inflammatory bowel disease, psoriasis, celiac disease, autoimmune thyroid disease, myasthenia gravis, multiple sclerosis) AID based on ≥1 hospitalisation or ≥2 physician visits with a relevant diagnostic code, at least 2 months apart but within 24 months. The study interval spanned from birth to the first of the following: event of interest, age 18, death, or end of study. We performed multivariate analyses to adjust for maternal age, education, and ethnicity, as well as calendar year of birth and sex of the child.

Results 509 women with SLE had 719 children, while 5824 matched controls had 8493 children. Mean follow-up was 9.1 (SD 5.8) years. Compared to controls, children born to mothers with SLE had similar records of rheumatic diagnoses [0.14% (95% CI: 0.01, 0.90) vs 0.19% (95% CI: 0.11, 0.32)]. However, there was a trend towards more non-rheumatic AID in offspring of mothers with SLE versus controls [1.11% (95% CI: 0.52, 2.27) vs 0.48% (95% CI: 0.35, 0.66)]. The most frequently observed non-rheumatic AID were Crohn's disease (0.56% in SLE offspring, versus 0.19% in control children) and type 1 diabetes (0.42% in SLE offspring, versus 0.22% in control children).

In multivariate analyses, children born to mother with SLE had a substantially increased risk of non-rheumatic AID compared to controls (OR 2.62, 95% CI: 1.10, 6.24), while results were inconclusive for the risk of rheumatic AID (OR 0.78, 95% CI: 0.10, 5.92).

Conclusions Our novel data suggest that, compared to children from the general population, children born to women with SLE have an increased risk of non-rheumatic AID. Our effect estimate for the risk of rheumatic AID is inconclusive. Further study of these children, throughout late childhood, adolescence, and adulthood, would be additionally enlightening.

CE-50 NOVEL INFORMATICS APPROACHES TO AUTOMATE CASE-IDENTIFICATION OF LUPUS IN AN ELECTRONIC HEALTH RECORD

Jinoos Yazdany*, Sara G Murray. *Department of Medicine, University of California, San Francisco, United State of America*

10.1136/lupus-2016-000179.128

Background Electronic health records (EHRs) can play an important role in generating data on the natural history, treatment, and outcomes of systemic lupus erythematosus (SLE). A key issue in using EHRs for SLE research is accurately identifying populations of patients with the disease. This is especially important because traditional definitions that rely on coding systems such as ICD9 have had poor specificity in previous studies. We aimed to develop and test disease classification algorithms to define a population with SLE in the EHR. We analysed both traditional definitions that used structured data (ICD-9 codes, medications, laboratories) and machine learning algorithms that used the entirety of information in the EHR, including unstructured data from clinical notes.

Materials and methods We created a repository of patients with possible SLE (based on relevant ICD-9 codes, positive auto-antibodies, and/or mention of "SLE" or "lupus" in the text of a clinical note). We combined 300 patients from that repository with 1000 randomly selected adult patients in our EHR as our training set. These patients were reviewed by domain experts for a diagnosis of SLE and confirmed cases were used as a gold standard for training our machine learning algorithms. We calculated the test characteristics for various definitions of SLE using only structured data. Finally, we compared this to a series of supervised machine learning algorithms based on support vector machines (SVMs) that used text features extracted from clinical notes in addition to structured fields. All SVM algorithms were trained and validated using 10-fold cross-validation.

Results One hundred thirty-seven patients met criteria for SLE. The test characteristics of both the structured and supervised ML algorithms are shown in the Table 1. A single ICD-9 code for 710.0 had a precision/positive predictive value of 79%. In contrast, machine learning algorithms greatly outperformed structured definitions in terms of precision, with precision/positive

Abstract CE-50 Table 1 Performance of traditional structured definitions and supervised machine learning algorithms in case identification of SLE in the electronic health record

Structured Definitions				Supervised Machine Learning			
	Recall		Precision		Recall		Precision
	(Sensitivity)	Specificity	(PPV)		(Sensitivity)	Specificity	(PPV)
Single ICD9 710.0	0.99	0.97	0.79	All ICD-9 codes and counts[1]	0.89	0.99	0.89
Single ICD9 710.0	0.96	0.98	0.86	All ICD-9 codes and counts	0.90	0.99	0.89
+ any lupus medication				+ NLP of clinical notes[2]			
Single ICD9 710.0	0.93	0.98	0.87	All ICD-9 codes and counts	0.91	0.99	0.92
+ any lupus medication				+ NLP of clinical notes			
+ any positive lupus-serology				+ all serologic data[3]			
				+ all medication data			
				All ICD-9 codes and counts	0.85	0.99	0.96
				+ NLP of clinical notes			
				+ all serologic data			
				+ all medication data			
				+ demographics[4]			

[1] Supervised Machine Learning algorithms included all available ICD-9 codes for patients as well as counts and locations in the medical records in which they were found (i.e. clinical encounters, problems lists, medications orders, etc.)
[2] All text data from clinical notes associated with a patient's medical record were included in the ML algorithm
[3] Serologic data included ANA, double-stranded DNA, anti-Smith antibody, anti-RNP, SSA, and SSB
[4] Demographic information included age, gender, race/ethnicity, insurance status, and employment status

predictive value approaching 96% in the most comprehensive algorithm.

Conclusions In an EHR-based data repository, a single ICD-9 was highly sensitive for SLE. Machine learning algorithms processed a multitude of structured and unstructured EHR data, allowing improved precision/positive predictive value. Further validation across different health systems will be necessary prior to implementing these algorithms on a national basis.

Acknowledgements The Rheumatology Research Foundation provided funding for this work.

CE-51 UTILITY OF LINKAGE OF A SURVEILLANCE REGISTRY TO STATE BIRTH RECORDS FOR CATEGORIZATION OF RACE AND ETHNICITY: MICHIGAN LUPUS EPIDEMIOLOGY & SURVEILLANCE (MILES) PROGRAM

[1,2,3]**Emily C Somers***, [1,3]Wendy Marder, [1]Martha Ganser, [4]Peter DeGuire, [5]Charles G Helmick, [6]Caroline Gordon, [7]Lu Wang, [1]Emily E Lewis, [1]W Joseph McCune. *[1]Internal Medicine, University of Michigan, USA; [2]Environmental Health Sciences, University of Michigan, USA; [3]Obstetrics and Gynaecology, University of Michigan, USA; [4]Michigan Department of Health and Human Services, USA; [5]Centers for Disease Control and Prevention; [6]Rheumatology, University of Birmingham; [7]Biostatistics, University of Michigan, USA*

10.1136/lupus-2016-000179.129

Abstract CE-51 Table 1 Comparison of race/ethnicity classification among SLE cases as part of a lupus surveillance registry: medical records versus Michigan birth files. Data from birth files represent the gold standard, as race/ethnicity was self-identified

		BIRTH CERTIFICATE (BC) – "gold standard"									
		Race								Ethnicity	
Medical	Records (MR)	Black	White	Asian	NHOPI	AIAN	Other	Unknown	Total	Hispanic	Not Hispanic
Race											
	Black	**962**	17	1	0	1	0	2	983	–	–
	White	9	**563**	2	3	2	0	2	581	–	–
	Asian	0	4	**5**	2	0	0	0	11	–	–
	NHOPI	0	3	0	**0**	0	0	0	3	–	–
	AIAN	0	0	0	0	**1**	0	0	1	–	–
	Other	2	28	1	0	0	**1**	0	32	–	–
	Unknown	3	15	4	0	0	0	**0**	22	–	–
Ethnicity											
	Hispanic	–	–	–	–	–	–	–	–	**20**	12
	Not Hispanic	–	–	–	–	–	–	–	–	12	**1589**
Total		976	630	13	5	4	1	4		32	
	Difference (BC-MR)	−7	49	2	2	3	−31	−17		0	
	Classification (BC/MR)	0.99	1.08	1.18	1.67	4.00	0.03	0.19		1.00	
	Sensitivity (of MR)	98.6	89.4	38.5	0.0	25.0	100.0	0.0		62.5	
	95% CI for sensitivity	(97.6, 99.2)	(86.7, 91.7)	(13.9, 68.4)	(0.0, 52.2)	(0.6, 80.6)	(2.5, 100.0)	(0.0, 60.2)		(43.7, 78.9)	

Abbreviations: NHOP = Native Hawaiian and Other Pacific Islander; AIAN = American Indian/Alaska Native