Background and aims Systemic autoimmune diseases (SAD) are characterised by a wide spectrum of demographic patterns with respect to the ethnic differences, age at diagnosis and especially gender distribution. Studying the distribution of these diseases across geographic regions using a big data-driven approach may help obtain a more “high-definition resolution” of these complex diseases.

Methods We explored the potential of the Google search engine to collect and merge cohorts (>100 patients) of patients with systemic lupus erythematosus (SLE) reported in the Pubmed library. We made a text-word search in Google between 8th and 15th May 2015 using SLE and “100...100000000 patients” and “site:http://www.ncbi.nlm.nih.gov/pubmed”. We collected the available data about study design, country, ethnicities, age and gender, clinical features and immunological markers.

Results We merged the data of 153 SLE cohorts including 71 000 patients; gender was detailed in 130 cohorts;88% women(female:male ratio, 8:4). mean age at onset (29.89 ±3.48), at diagnosis (32.33 ±2.99).The countries contributing the most cohorts were the USA (31), Japan (8) and Spain (5). The main clinical features included arthritis in 72%, haematological abnormalities in 62%, malar rash in 50%, photosensitivity in 48%, renal involvement in 38%, oral ulcers in 34%, serositis in 30% and neurological involvement in 14%. Haematological abnormalities included lymphopenia in 43%, leukopenia in 38%, thrombocytopenia in 13% and hemolytic anaemia in 4%.Positive autoantibodies included ANA in 91%, dSDNA in 62%, anti-Ro/SSA in 35%, anti-RNP in 25%, antiSm in 21% and anti-La/SSB in 15%.

Conclusions There is a clear trend of association between the percentage of women diagnosed with SLE and some indicators of development of each country. The gap between women and men diagnosed with SLE is wider in countries with the highest frequencies of women working and women with high study degrees, and those countries with more taxes and a higher percentage of protected geographical areas.