The digital epidemiology of phenylketonuria, aka folling's disease: Retrospective analysis and geographic mapping via google trends

Ahmed Al-Imam[1, 2, 3, 4]

[1]Department of Anatomy and Cellular Biology, College of Medicine, University of Baghdad, Iraq, [2]Department of Postgraduate Medicine, School of Life and Medical Sciences, University of Hertfordshire, United Kingdom, [3]The Canadian Association for Neuroscience, [4]The Japanese Association of Anatomists

ABSTRACT

Background: Phenylketonuria, commonly known as PKU, is an inherited disorder in which there is an abnormally elevated blood level of the amino acid phenylalanine leading to several pathologies affecting multiple organs including the central nervous system and resulting in debilitating intellectual disability and other neuropsychiatric disorders. Phenylalanine is a building block of several critical proteins within the biological systems. **Aims and Objective:** To assess the digital epidemiology and geographic mapping of Phenylketonuria. **Materials and Methods:** This study is a retrospective analytic (2013-2017) of a *very large database* existing on the surface web known as Google Trends. it aims to extrapolate a statistical inference concerning the digital epidemiology and the geographic mapping of phenylketonuria. The trends database will be explored via thematic keywords specific to the condition of phenylketonuria including "Phenylketonuria [PKU]", "Phenylalanine", "Inborn errors of metabolism", "Tetrahydrobiopterin", and "Chromosome 12 (human)". **Results:** The digital epidemiology is densely clustered in countries from the developed world, eastern Europe, and Latin America. Surface web users from China appears to possess the highest interest in phenylketonuria. The contribution of the Middle Eastern and Arabic countries to the geographic mapping did not exceed 10.51% at its best. Significant changes existed for year-to-year variations of trends. Statistical outliers were also found, the strongest of which was observed during April 2016 for which there's no plausible explanation. **Conclusion:** Trends databases operating on the surface web represent potent tools of big data that can be exploited to assess the digital epidemiology and geographic mapping of countless phenomenon including rare genetic diseases and inborn errors of metabolism. There are also enormous potentials for real-time and predictive analytics of these databases when investing the application of automation in data collection and principles of machine learning.

Key words: Genetic diseases; Autosomal recessive; Inborn errors of metabolism; Phenylketonuria; Folling's disease; Phenylalanine hydroxylase; PAH deficiency; PAH gene; Phenylalanine; Tetrahydrobiopterin; Chromosome 12 (human); Google trends; Digital epidemiology; Retrospective studies; Predictive analytics; Data crunching; Knowledge discovery in databases; Arabs; Middle east; Iraq

INTRODUCTION

Phenylketonuria (PKU), also known as Folling's disease, is an inborn error of metabolism due to an autosomal recessive mutation affecting phenylalanine hydroxylase (PAH) gene located on the long arm of chromosome number-12. [1-3] PAH gene encodes a specific amino acid known as phenylalanine which is one of the few amino acids that the human body cannot synthesise on its own.[4] Hence, this type of amino acids is also known as *essential* or *indispensable amino acids*.[5, 6] Phenylalanine exists in almost all proteins and some artificial sweeteners.[1, 4, 5]

Address for correspondence:
Dr Ahmed Al-Imam, House 18/5, Al-Akhtal Street, District 318, Al-Adhamyia, 10053, Baghdad, IRAQ.
E-mail: tesla1452@gmail.com

Phenylalanine hydroxylase enzyme is responsible for the initial step in processing (hydroxylation) of phenylalanine and converting it into another amino acid known as tyrosine (4-hydroxyphenylalanine) which belongs to another category of amino acids known as *conditionally essential amino acid*.[5, 6]Tyrosine is a building block of critical monoamine neurotransmitters including dopamine and catecholamines. It is also linked to a crucial enzyme known as tyrosine kinase.[7, 9] Accordingly, any defect in phenylalanine and tyrosine metabolism, as in PKU, may lead to substantial and even fatal consequences including metabolic disturbances, intellectual disability, and other health issues.[9-11] The signs and symptoms of PKU vary from mild to severe. The most severe form of this disorder is known as classic PKU. Infants with classic PKU could exist within a spectrum from mild to severe (classic PKU).[12] Infants with classic PKU are healthy until they are a few months old, after which they will develop permanent developmental delays, intellectual disability, seizures, and behavioural-psychiatric disorders.[13-15]

The epidemiology of PKU varies based on geography and ethnic grouping. In the US, PKU occurs in one for every 10,000-15,000 newly born infant.[16] However, most cases of PKU are detected after birth by newborn screening programs, and treatment is usually started promptly afterwards.[14, 15] Therefore, the exact epidemiology is not accurately estimated especially in the developing world including Arab countries and the region of the Middle East. Trends databases can be used for accurate geographic mapping and estimation of the digital epidemiology of a varied array of diseases, including rare genetic disorders and inborn errors of metabolism as in the case of PKU.[17, 18] This study will explore the digital epidemiology of PKU via retrospective analysis of *Google Trends* database. [19-22]Prospectively, it is anticipated that the digital exploration of trends database, including *Google Trends*, via automated *machine learning* and *deep thinking* algorithms will offer a substantial levitation for the quality and the level-of-evidence of epidemiologic studies, including retrospective inspection, real-time examination of trends, and potential predictive analytics.[23-25]

MATERIALS AND METHODS

Several databases, including trends databases, exist on the surface web. These can be exploited to describe or infer the digital epidemiology of a specific phenomenon including diseases due to inborn errors of metabolism. Google Trends database was used to assess the interest (attentiveness) of surface web users in PKU via thematic keywords that are synonymous for PKU. The analysis is retrospective for the past five years (2013−2017).[26]The keywords were five in

total including "Phenylketonuria [PKU]", "Phenylalanine", "Inborn errors of metabolism", "Tetrahydrobiopterin", and "Chromosome 12 (human)". Inferential statistics and hypothesis testing were applied to conclude the existence of any significant variations in web users' interest based on their location (mapped geography) and year-to-year variations (chronology). This innovative method, when combined with machine learning technologies, can be applied for real-time apprehension and predictive analytics of the epidemiological trends.

The analysis of Google Trends database reflects an extrapolation of data generated from millions of users of the visible web (surface web). Accordingly, this manuscript is considered a hybrid study of an Internet Snapshot of a trends database and a retrospective analysis of level-2a in compliance with the hierarchy of evidence-based studies imposed by the Oxford Centre for Evidence-based Medicine (CEBM).[27,28] A literature review was systematically conducted on NCBI-PubMed/Medline, EMBASE, the Cochrane Library, Scopus, EBSCO − CINAHL, Web of Science, and Google Scholar. The grey literature was also consulted. Concerning the literature review, recently published papers (past five years) were given priority. Statistical analyses were carried out using Microsoft Excel 2016 and the Statistical Package for Social Sciences (SPSS 20). The level-of-evidence of reference materials consulted for literature review has also been assessed in compliance with the Oxford CEBM via tools of critical appraisal of literature.[29, 30]

RESULTS

Concerning the geographic mapping (Figure 1 to 4) and based on "Phenylketonuria [PKU]" keyword search, the top ten contributing countries were China, Puerto Rico, Chile, Paraguay, Poland, Indonesia, United States, Iran, Sweden, and Ireland. These countries accounted for 28.10%, while Arabic and Middle Eastern countries represented 10.51% including Iran, Jordan, Turkey, Israel, United Arab Emirates (UAE), Saudi Arabia, Algeria, and Egypt. Based on "Phenylalanine" keyword search, the top ten contributing countries accounted for 26.07% including Mexico, Australia, New Zealand, United States, Paraguay, Puerto Rico, Chile, Sweden, Spain, and Italy. Countries from the Middle East and the Arab world contributed to 5.87% including Israel, Iran, Algeria, UAE, Turkey, Egypt, and Saudi Arabia. Based on "Inborn errors of metabolism" keyword search, the top ten contributing countries represented 51.97% including Taiwan, the Netherlands, Paraguay, Japan, Ecuador, Nigeria, Nicaragua, Peru, Dominican Republic, and Chile. Middle Eastern and Arabic countries represented 2.15% including Egypt, Saudi Arabia, UAE, Israel, and Turkey. Based on "Tetrahydrobiopterin"
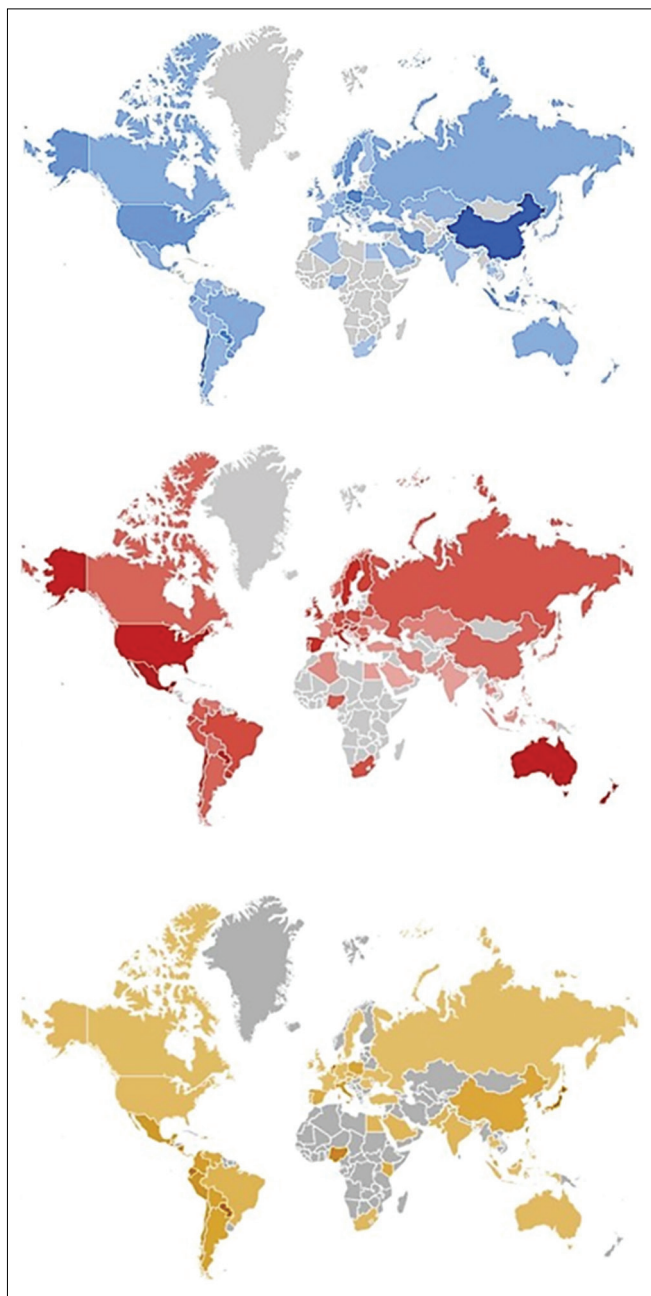
Figure 1: Geographic Mapping based on Keywords "Phenylketonuria" (upper), "Phenylalanine" (middle), and Inborn Error of Metabolism" (lower).

Figure 2: Geographic Mapping based on Keywords "Tetrahydrobiopterin" (upper) and "Chromosome 12 (human)" (lower).

Figure 3: Superimposed Mapping of the Geographic Distribution of Previous Maps (same colour coding in previous figures).

keyword search, the top ten contributing countries accounted for 67.38% including Grenada, Philippines, Lebanon, United Kingdom, Singapore, India, Morocco, Puerto Rico, United States, and Czechia. The Middle East and the Arab world contributed to 5.35% including Lebanon, UAE, Iran, Saudi Arabia, Egypt, Turkey, and Israel. Based on "Chromosome 12 (human)" keyword search, the top ten contributing countries accounted for 79.91% including Colombia, Belarus, United States, Costa Rica, Canada, Philippines, Australia, Italy, France, Spain, and Germany. There was no country from the Middle
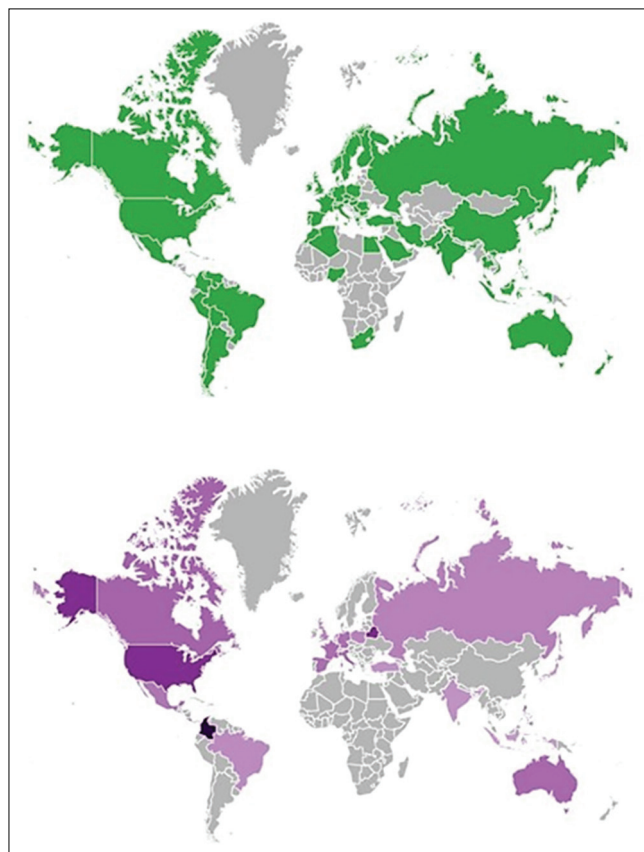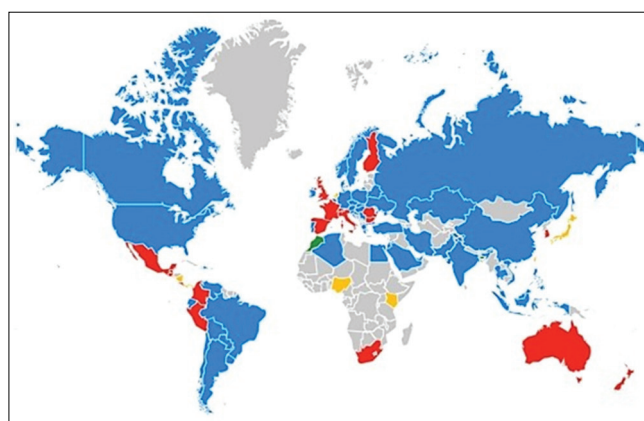
East except for Turkey (1.28%) while no data signals were existent from the Arab countries.

The relative interest of surface web users in PKU was variable in time (2013-2017) and for each of the five utilised keywords (Figure 5). The keywords averaged 39.85 (Phenylketonuria [PKU]), 33.24 (Phenylalanine), 8.21 (Inborn errors of metabolism), 2.79 (Tetrahydrobiopterin), 0.93 (Chromosome 12 (human)). statistical outliers were
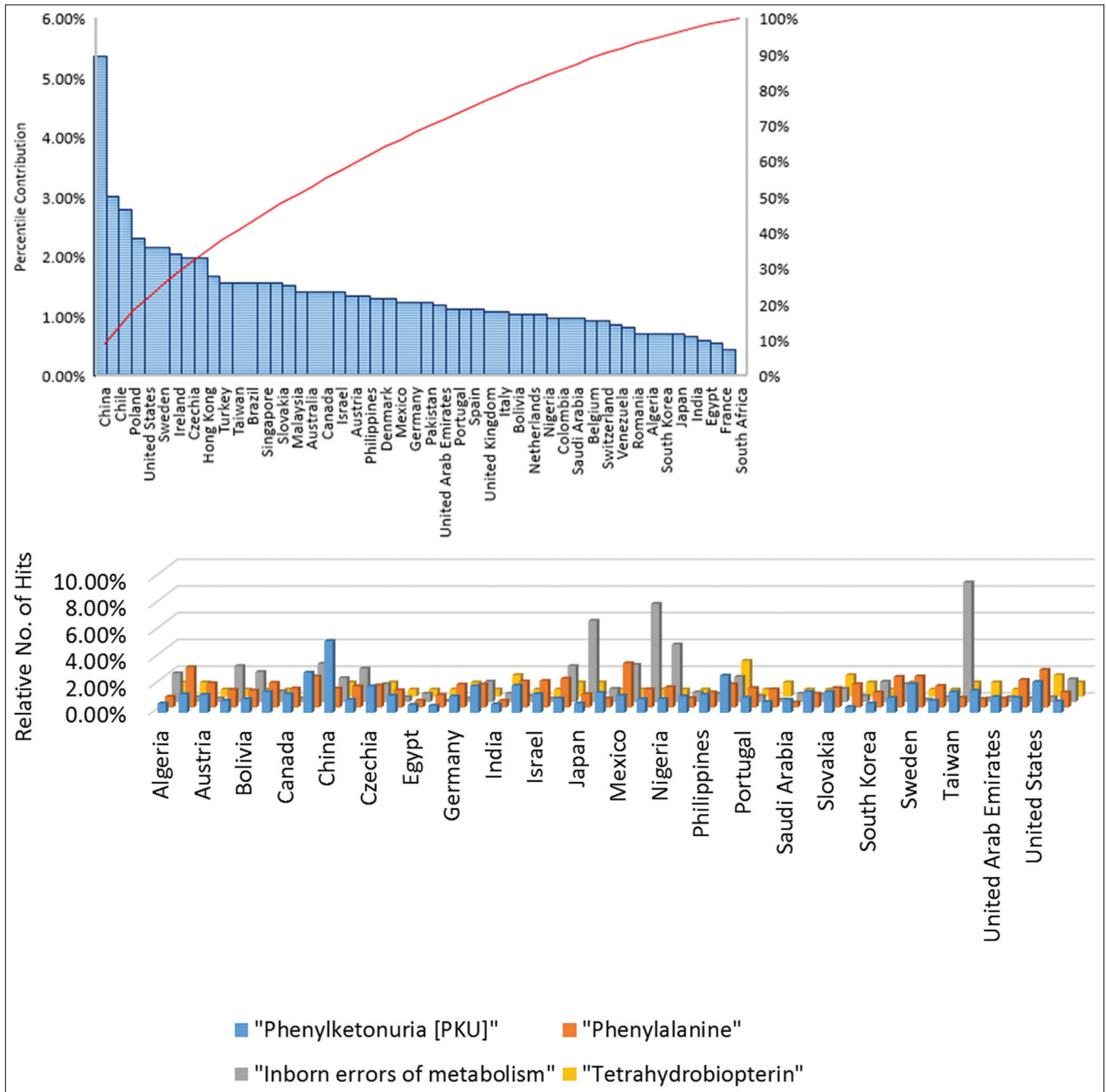
Figure 4: Geographic Mapping based on Keywords Search: Pareto Chart (above) and Column Chart (below).

also detected (Figure 6) during January 2013, March 2013, October and November 2013, August 2014, November 2014, January-March 2015, May 2015, February-April 2016, November 2016, February and March 2017, June and July 2017, and November-December 2017. The highest peak of interest was recorded during the middle of April 2016 and in relation to the keyword "Phenylketonuria [PKU]" (Figure 5). Modelling via linear regression confirmed the existence of a positive correlation between all of the mapped keywords, the strongest was for "Phenylketonuria [PKU]" versus "Phenylalanine" (R^2 score=0.2806, *p-value*<0.001) (Figure 7).

The analysis of variance and covariance (ANOVA) confirmed the existence of a significant difference among the keywords (p<0.001).Statistical inference was extrapolated for year-to-year variation of trends of each keyword (Table 1). Most of these were clustered for the two keywords "Phenylketonuria [PKU]" and "Phenylalanine". However, significant changes in trends were also observed for the keyword "Inborn errors of metabolism" for the year 2013 versus 2016 (p=0.061). Similarly, significant changes in trends for the keywords "Tetrahydrobiopterin" for the years 2014 versus 2017 (p=0.002), 2015 versus 2017 (p<0.001),

and 2016 versus 2017 (p<0.001). Concerning the keyword "Chromosome 12 (human)", significant change in trends were existent for 2014 versus 2015 (p<0.001), 2014 versus 2016 (p=0.002), and 2014 versus 2017 (p=0.002).

DISCUSSION

Phenylketonuria is a rare inborn disease that may lead to severe neurobehavioral dysfunction resulting in a substantial economic burden to the parents and caregivers. In China, the median value for the financial hardship of

Figure 5: Line Graph of Trends (2013-2017).

Figure 6: Boxplot Presentation for Change in Trends (2013-2017).

classical PKU was estimated to be 75% of the total annual family income.[31]Phenylketonuria represents one of the most prevalent autosomal recessive hereditary disorders in Turkey.[32] On the other hand, the PAH gene in Iranian Azeri Turkish population was studied, it confirmed that *Variable Number of Tandem-Repeat* (VNTR) alleles with thirteen repeats showed 3% frequency in the studied population.[33]In Lebanon and based on a study of 2921 patients, Phenylketonuria and methylmalonic acidemia were found to be the most frequent among aminoacidopathies and organic acidemias.[34]In 2016, Abiri and colleagues reported two patients from Iran with two co-existing inborn errors of metabolism including a girl affected by classic PKU and maple syrup urine disease (MSUD), and a male patient affected with PKU and Sandhoff disease.[35] In the southwest of Iran and the Khuzestan Province, molecular analysis confirmed detection of thirteen different mutations affecting the PAH gene which was anticipated due to the ethnic heterogeneity of the specified region.[36]In western Iran, Alibakhshi and co-workers confirmed a distinct difference existing concerning the characteristics of PAH mutations between the Kermanshah province and other areas of Iran. It has been suggested that Kermanshah, also known as the gateway to Asia, may have a unique population distribution of PAH gene mutations as it lies
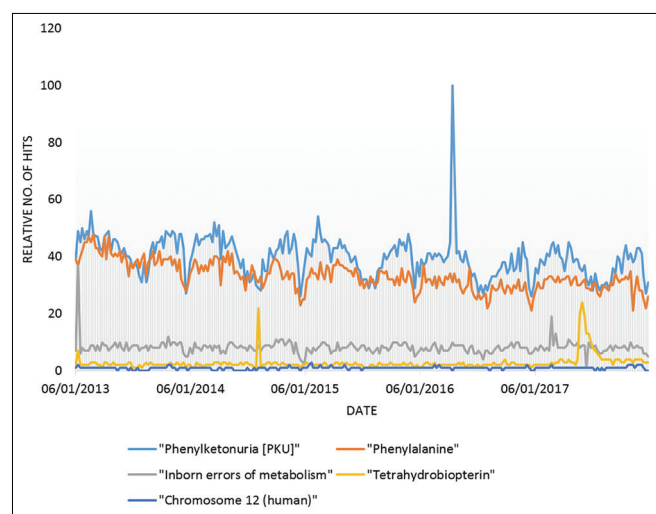

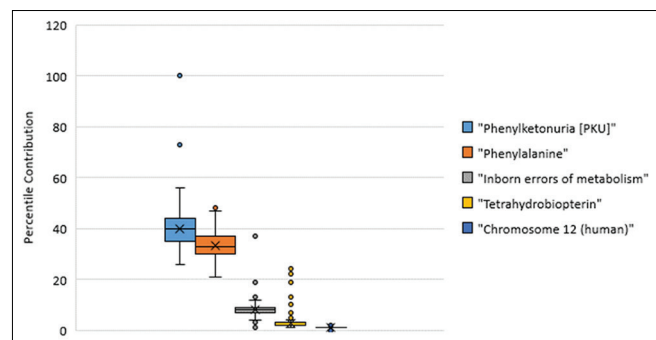**Figure 7:** Linear Regression of Keywords: "Phenylketonuria [PKU]" versus "Phenylalanine".

Table 1: Statistical inference (based on p-value calculations) for year-to-year variations in trends				
	"Phenylketonuria [PKU]"	"Phenylalanine"	"Inborn errors of metabolism"	"Tetrahydrobiopterin"	"Chromosome 12"
2013 vs. 2014	0.096	0.000	0.363	0.457	0.172
2013 vs. 2015	0.012	0.000	0.168	0.364	0.014
2013 vs. 2016	0.009	0.000	0.061	0.313	0.122
2013 vs. 2017	0.000	0.000	0.337	0.000	0.080
2014 vs. 2015	0.484	0.036	0.354	0.273	0.000
2014 vs. 2016	0.146	0.000	0.047	0.255	0.002
2014 vs. 2017	0.000	0.000	0.873	0.002	0.002
2015 vs. 2016	0.302	0.000	0.269	0.870	0.228
2015 vs. 2017	0.001	0.000	0.543	0.000	0.527
2016 vs. 2018	0.305	0.159	0.142	0.000	0.658

on the route of major ancient movements of the Caucasian people toward the Mediterranean basin.[37] In Europe, the amounts of total protein prescribed for patients with PKU was varied between European countries.[38] Pena and colleagues also reported similar findings. There were widely different numbers of protein substitutes available for PKU patients from various European countries.[39] Our study is the first of its kind daring to assess the digital epidemiology and geographic mapping of PKU via Trends databases. Digital epidemiological analysis has proven successful via Googles Trends. Future studies, both observational and experimental, can incorporate collateral data of digital epidemiology retrieved from trends databases. Case reports and case-control studies can be adequately correlated with change in trends of digital epidemiology. Trends databases of the surface web can be integrated into a predictive early warning system to anticipate any change in the interest of surface web users in a specific disease including those of low prevalence and incidence. Real-time and predictive analytics can be achieved by exploiting the use of machine learning algorithms and data crunching technologies.

CONCLUSION

The digital epidemiological analysis is feasible to be applied via trends databases. There have been some significant changes in the trends versus time. Geographic mapping of the top contributing countries was restricted mainly to the developed world, as well as some Latin nations, and eastern Europe. However, data signals also existed in the Middle East and Arab countries including the UAE, Egypt, Saudi Arabia, Turkey, Israel, and Iran. This study may have some limitations due to the sole reliance on Google Trends as a representative of the trends databases of the surface web. Further, the analysis was retrospectively specific for a restricted period (2013-2017). Data collected from Google Trends might be occasionally faulty or misleading as some web users might be deploying the use of virtual private networks and internet protocol masking. Subsequent studies should incorporate more than one database. Additionally, ventures into the deep web and the darknet should be attempted to recover any relevant data including those on geographic mapping whenever feasible.

ACKNOWLEDGEMENTS

None

CONFLICTS OF INTEREST

The authorshave no competing interests.

REFERENCES

1. Blau N. Genetics of phenylketonuria: then and now. Human mutation. 2016 Jun 1;37(6):508-15.

2. Al-ImamA. A Prototype for Digital Epidemiological Studies of Inborn Errors of Metabolism: Demonstration via Retrospective Analysis of Phenylketonuria. Available at https://www.researchgate.net/publication/324953838_A_Prototype_for_Digital_Epidemiological_Studies_of_Inborn_Errors_of_Metabolism_Demonstration_via_Retrospective_Analysis_of_Phenylketonuria (accessed 5 May 2018).Hendriksz CJ, Walter JH. Update on phenylketonuria. Current Paediatrics. 2004 Oct 1;14(5):400-6.

3. Hendriksz CJ and Walter JH. Update on phenylketonuria. Current Paediatrics 2004;14(5):400-406.

4. Kaufman S. Studies on the mechanism of the enzymatic conversion of phenylalanine to tyrosine. Journal of Biological Chemistry 1959;234(10):2677-2682.

5. Aristoy MC and Toldrá F. Essential amino acids. Handbook of Analysis of Edible Animal By-Products. Nollet, LML and Toldrá, F (eds) CRC Press, Boca Raton, FL, USA. 2011:123-135.

6. Volpi E, Kobayashi H, Sheffield-Moore M, Mittendorfer B and Wolfe RR. Essential amino acids are primarily responsible for the amino acid stimulation of muscle protein anabolism in healthy elderly adults. The American journal of clinical nutrition 2003; 78(2):250-258.

7. Wurtman RJ and Fernstrom JD. Control of brain monoamine synthesis by diet and plasma amino acids. The American journal of clinical nutrition 1975; 28(6):638-647.

8. Patschinsky T, Hunter T, Esch FS, Cooper JA and Sefton BM. Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation. Proceedings of the National Academy of Sciences 1982;79(4):973-977.

9. Hanley WB. Adult phenylketonuria. The American journal of medicine 2004;117(8):590-595.

10. Villasana D, Butler IJ, Williams JC and Roongta SM. Neurological deterioration in adult phenylketonuria. Journal of inherited metabolic disease 1989;12(4):451-457.

11. Mitchell JJ, Trakadis YJ and Scriver CR. Phenylalanine hydroxylase deficiency. Genetics in medicine 2011;13(8):697.

12. Guttler F. Hyperphenylalaninemia: Diagnosis and classification of the various types of phenylalanine hydroxylase deficiency in childhood. Hyperphenylalaninemia: Diagnosis and classification of the varioustypes of phenylalanine hydroxylase deficiency in childhood.1980(Suppl. 280).

13. Scriver CR. The PAH gene, phenylketonuria, and a paradigm shift. Human mutation 2007;28(9):831-845.

14. deBaulny HO, Abadie V, Feillet F and de Parscau L. Management of phenylketonuria and hyperphenylalaninemia. The Journal of nutrition 2007;137(6):1561S-1563S.

15. Abadie V, Berthelot J, Feillet F, Maurin N, Mercier A and de Baulny Ogier H. Management of phenylketonuria and hyperphenylalaninemia: The French guidelines. Archives de pediatrie: Organeofficiel de la Societefrancaise de pediatrie 2005;12(5):594-601.

16. United States National Library of Medicine. Phenylketonuria. Available at https://ghr.nlm.nih.gov/condition/phenylketonuria#synonyms(accessed 4 May 2018).

17. Salathe M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS computational

biology 2012;8(7):e1002616.

18. Bakker KM, Martinez-Bakker ME, Helm B and Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. Proceedings of the National Academy of Sciences 2016;113(24):6689-6694.

19. Choi H and Varian H. Predicting the present with Google Trends. Economic Record 2012;88(s1):2-9.

20. Carneiro HA and Mylonakis E. Google trends: A web-based tool for real-time surveillance of disease outbreaks. Clinical infectious diseases 2009;49(10):1557-1564.

21. Pelat C, Turbelin C, Bar-Hen A, Flahault A and Valleron AJ. More diseases tracked by using Google Trends. Emerging infectious diseases 2009;15(8):1327.

22. Seifter A, Schwarzwalder A, Geis K and Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. Geospatial health 2010;4(2):135-137.

23. Nasrabadi NM. Pattern recognition and machine learning. Journal of electronic imaging 2007;16(4):049901.

24. Amarasingham R, Patzer RE, Huesch M, Nguyen NQ and Xie B. Implementing electronic health care predictive analytics: Considerations and challenges. Health Affairs 2014;33(7):1148-1154.

25. Gandomi A and Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015;35(2):137-144.

26. Microsoft. Google Trends. Available at https://trends.google.com/trends/explore?date=2013-01-01%202018-01-01&q=%2Fm%2F05sgs,%2Fm%2F09gwd,%2Fm%2F048fwq,%2Fm%2F041wq4,%2Fm%2F099x5f(accessed 4 May 2018).

27. Sackett DL and Rosenberg WM. On the need for evidence-based medicine. Journal of Public Health 1995;17(3):330-334.

28. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) levels of evidence (background document). Oxford Center for Evidence-Based Medicine 2011 Sep 16.

29. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS and Grimmer KA. A systematic review of the content of critical appraisal tools. BMC medical research methodology 2004;4(1):22.

30. Crowe M and Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. Journal of clinical epidemiology 2011;64(1):79-89.

31. Wang L, Zou H, Ye F, Wang K, Li X, Chen Z, et al. Household financial burden of phenylketonuria and its impact on treatment in China: A cross-sectional study. Journal of inherited metabolic disease 2017;40(3):369-376.

32. Mazlum B, Anlar B, Kalkanoglu-Sivri HS, Karli-Oguz K, Özusta S, et al. A late-diagnosed phenylketonuria case presenting with autism spectrum disorder in early childhood. The Turkish journal of pediatrics 2016;58(3):318.

33. Bagheri M, Rad IA, Jazani NH, Zarrin R and Ghazavi A. Molecular genetic analysis of the variable number of tandem-repeat alleles at the phenylalanine hydroxylase gene in Iranianazeri Turkish population. Iranian biomedical journal 2015;19(3):183.

34. Karam PE, Habbal MZ, Mikati MA, Zaatari GE, Cortas NK and Daher RT. Diagnostic challenges of aminoacidopathies and organic acidemias in a developing country: A twelve-year experience. Clinical biochemistry 2013;46(18):1787-1792.

35. Abiri M, Talebi S, Uitto J, Youssefian L, Vahidnezhad H, Shirzad T, et al. Co-existence of phenylketonuria either with maple syrup urine disease or Sandhoff disease in two patients from Iran: Emphasizing the role of consanguinity. Journal of Pediatric Endocrinology and Metabolism 2016;29(10):1215-1219.

36. Alibakhshi R, Moradi K, Mohebbi Z and Ghadiri K. Mutation analysis of PAH gene in patients with PKU in western Iran and its association with polymorphisms: Identification of four novel mutations. Metabolic brain disease 2014;29(1):131-138.

37. Aguiar A, Ahring K, Almeida MF, Assoun M, Quintana AB, Bigot S, et al. Practices in prescribing protein substitutes for PKU in Europe: No uniformity of approach. Molecular genetics and metabolism 2015;115(1):17-22.

38. Pena MJ, de Almeida MF, van Dam E, Ahring K, Belanger-Quintana A, Dokoupil K, et al. Protein substitutes for phenylketonuria in Europe: Access and nutritional composition. European journal of clinical nutrition 2016;70(7):785.