

Received: 18 March, 2021

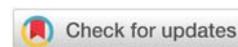
Accepted: 05 April, 2021

Published: 06 April, 2021

*Corresponding author: Iván Camilo Triana Avellaneda, Department of Medical Oncology, Fundacion Santa Fe de Bogota University Hospital, Bogotá, Colombia, Email: camilotrianaa@gmail.com

Keywords: Digital epidemiology of innovation; Big data; Data science; Machine learning

<https://www.peertechzpublications.com>



Review Article

Digital epidemiology of innovation

Iván Camilo Triana Avellaneda^{1*}, Luis Eduardo Pino¹ and Denisse Rubio Cruz²

¹Department of Medical Oncology, Fundacion Santa Fe de Bogota University Hospital, Bogotá, Colombia

²University of the Andes, Bogota, Colombia

Abstract

Bio and infotech revolution including data management are global tendencies that have a relevant impact on healthcare. Concepts such as Big Data, Data Science and Machine Learning are now topics of interest within medical literature. All of them are encompassed in what recently is named as digital epidemiology. The purpose of this article is to propose our definition of digital epidemiology with the inclusion of a further aspect: Innovation. It means Digital Epidemiology of Innovation (DEI) and show the importance of this new branch of epidemiology for the management and control of diseases. In this sense, we will describe all characteristics concerning to the topic, current uses within medical practice, application for the future and applicability of DEI as conclusion.

Abbreviations

DEI: Digital Epidemiology and of Innovation; VZV: Varicella Zoster Virus; MAIA: Medical Artificial Intelligence Assistant); HIP: Integrative cancer Healthcare Platform; IQS: Internet Query Share

Introduction

Currently Digital Epidemiology accounts for numerous definitions. One of them refers to it as a process in which one can understand the patterns of disease and the dynamics of community's healthcare, as well as their causes, to prevent and mitigate disease and promote health through digital data. (1) Indeed, that last concept of "digital data" is what makes the difference within overall epidemiology, added to the fact that these data is not collected for the traditional purpose when performing epidemiological analysis [1]. Our investigation group handles a similar concept but with some additional features, in which we add the term of "innovation". We define digital epidemiology of innovation as the use of digital information and cognitive technologies (including artificial intelligence) for the handling and collection of data that allows the analysis and comparison of patterns of disease plus the understanding of populations healthcare dynamics. Therefore, new concepts and tools can be created, based on technologies

focused on preventing and diminishing the burden of disease, and promoting health in a more efficient way. It is not only a utopic term, but it actually can also be found withing some investigations, as is depicted in our article MAIA (Medical Artificial Intelligence Assistant) as interface for a new cancer healthcare integrative platform: "A talkbot trained as a narrow artificial intelligence interface for an integrative cancer healthcare platform (HIP) is possible through the clinical and engineer integration of languages using a neural network method and other software tools. MAIA is for now a patient and physician experience improvement, but the real impact will be in the data acquisition and harmonization for advanced analytics. The final scope of MAIA HIP will be a blockchain for cancer in low and middle income countries" [2].

DEI is determined by the following components for its correct applicability and understanding:

1. Big Data- Digital Data
2. New models of medical research (adaptative trials as an example)
3. Machine Learning – Deep learning
4. Smart Disease Management Programs

The objective of this study is to explain the applicability and use of digital and innovation epidemiology for the management of many diseases, and its role in the digital transformation of the healthcare sector.

Main text

Next, and fulfilling the objective of the text, we are going to describe the 4 components in which we summarize digital and innovation epidemiology, giving the reader explanations and examples of their applicability in the real medicine.

Big data- Digital data

It is the first component of the DEI and from which the other three concepts branches. It is the base of the whole concept and it encompasses two terms: Utilization of digital data and the Big data. The first one, respectively, is what drove medicine to the digital epidemiology. Theoretically, data from clinical trials, systematic reviews and clinical records were the only source to perform a correct investigation in medicine; however, the digital epidemiology and of innovation changed that perspective. Nowadays, data found on the internet can be also used for population healthcare projects through algorithms and trends analysis. One example is Google Trends, when searching for influenza virus or varicella virus peak seasons in different countries [3] as some of the following studies show:

1. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization [3]: Through Google Trends and the research of people on internet, the study concluded: “ Our findings provide strong evidence that varicella zoster virus (VZV) transmission is seasonal and that seasonal peaks show remarkable latitudinal variation. We attribute the dampened seasonal cycles in chicken pox information-seeking behavior to VZV vaccine-induced reduction of seasonal transmission. In Figure 1, the trends found by the investigators of this development are evidenced. These data and the methodological approaches provide a way to track the global burden of childhood disease and illustrate population-level effects of immunization. The global latitudinal patterns in outbreak seasonality could direct future studies of environmental and physiological drivers of disease transmission [3].

2. Detecting influenza epidemics using search engine query data [4].

Frequently, countries are unable to make a proper control of epidemics due to the absence of data within their territories. The mentioned studies not only portray how useful the analysis through internet research can be, but it also is an evidence of how economical barriers can be removed. In fact, it goes beyond viral peaks detection, it is a powerful tool that permits the control and follow-up of vaccination too, as the following study shows [3,5].

Use of Internet Search Data to Monitor Impact of Rotavirus Vaccination in the United States [5]. Abstract: Google-based Internet query share (IQS) for rotavirus search terms correlated

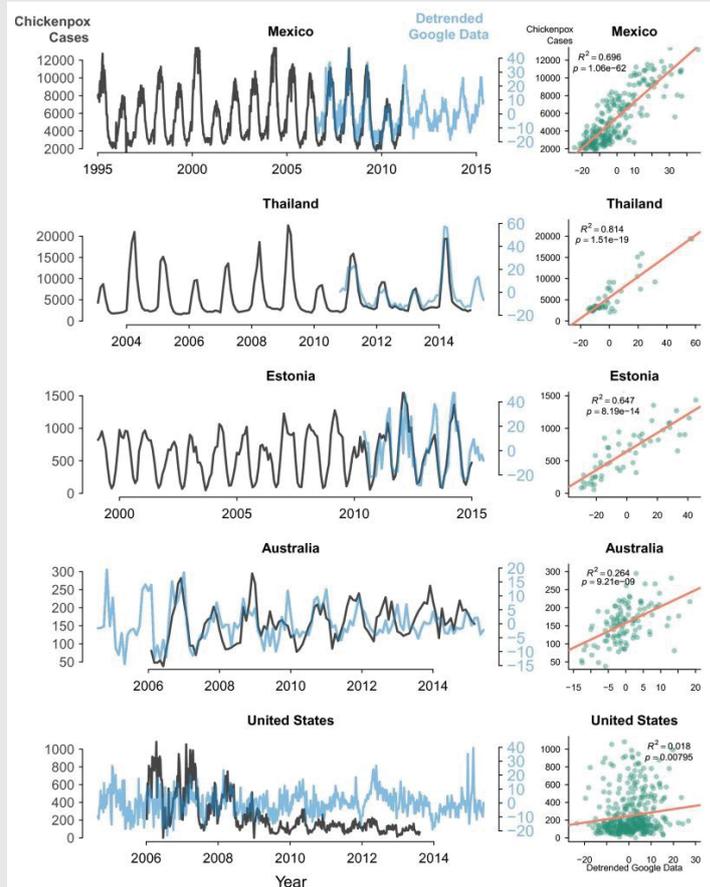


Figure 1: Relationship between chicken pox cases and information seeking. (Left) Time series of reported chicken pox cases and information-seeking behavior for chicken pox (i.e., Google Trends data) in Mexico, Thailand, Estonia, Australia, and the United States. Google data were detrended to remove long-term trends and focus on seasonal variation in information seeking. (Right) Relationship between reported cases of chicken pox and chicken pox information seeking when both were available, with applicable R² and P values [3].

well with US rotavirus laboratory detections from 2004 to 2010 ($r = 0.88$; $P < .001$), capturing the reduction observed during postvaccine years (2008–2010). IQS analysis could become an inexpensive and reliable supplement for monitoring the impact of rotavirus vaccination in the United States [5]. Figure 2a and 2b summarize the trends found in the study with the decrease previously reported when introducing the vaccine.

The topic of digital data correlates with Big data, since one is essential to the other. Broadly, Big data is a set of data that due to its size, volume and growth needs a different management other than Excel. (PowerData, s. f.) That is, a large amount of data, but in big data, not only the quantity matters but the organization “The concept of big data gained momentum in the early 2000s when industry analyst Doug Laney articulated the current definition of big data as the three Vs (Volume, Speed (Spanish: Velocidad), variety): Volume: Organizations collect data from various sources, for example, social media, smartphones. Speed: With the growth of the Internet of Things, data reaches businesses at unprecedented speed and must be handled in a timely manner. Variety: Data is presented in all kinds of formats: from structured numerical data in traditional databases to unstructured text documents”. (Big Data, s. f.) In medicine and demographic healthcare studies, data itself is

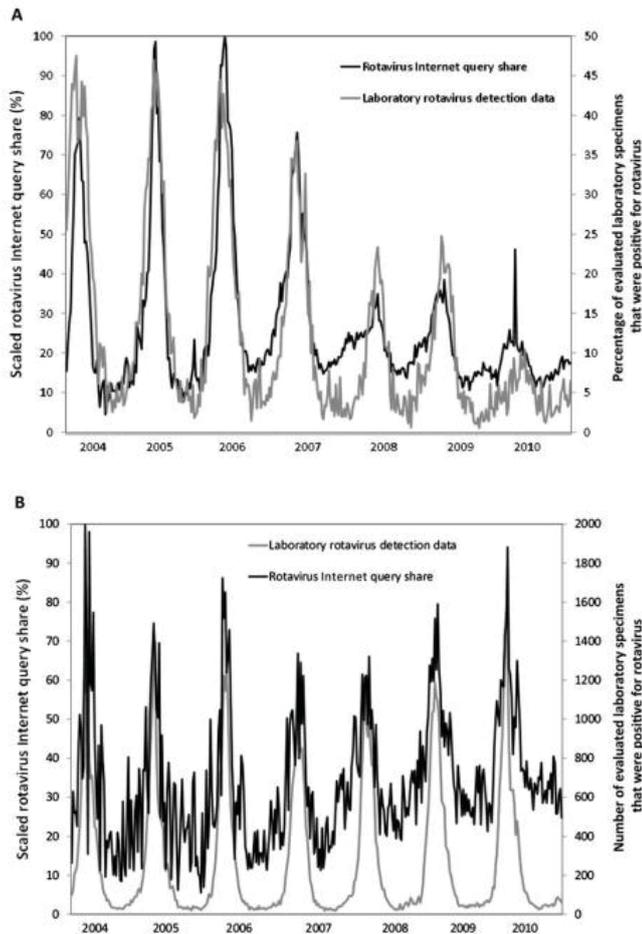


Figure 2a,b: A: Weekly US laboratory data on the percentage of evaluated specimens in which rotavirus was detected (black line) and weekly Internet query share (IQS) rotavirus data for the United States (gray line). B: Weekly UK laboratory data on the number of evaluated specimens in which rotavirus was detected (black line) and weekly IQS rotavirus data for the United Kingdom (gray line). All IQS data are scaled such that the highest point equals 100%, and the remaining data are a proportion of that value [5].

massive, and now considering that digital data also implies a component of massive information, Big data is consequently a management of medical information particularly when searching for demographic behaviors [6]. The feasibility of Big data in the medical field is clearly of worldwide use, as seen in the study SmiNet2: Description of an internet-based surveillance system for communicable diseases in Sweden [7]. Based on the results, one could say that demographic disease records of a country should be digitalized and managed as Big data, to make better and more effective decisions of public health. This trend is already happening in many countries globally, however, there are many countries that do not handle data in this way. In fact, there are many countries without population records of their diseases, in our country Colombia, that is one of the biggest challenges. Therefore, DEI initiatives and concepts can be very useful for these countries where Big Data consists of Small Data and analytics to take decisions.

New models of medical research

In the second component of DEI, the term of innovation has a more important relevance, as it is a key point in the way

we perform medical research. Currently, all scientific research must have a protocol that specifies inclusion and exclusion criteria, randomize and control the included patients, how the sample was chosen, feasibility of the study, follow-up of patients, bias, analysis of information and ethical aspects among others, depending on the type of article that will be written. However, hardly ever, the use of artificial intelligence is contemplated when performing these projects, meanwhile DEI looks for innovation and improvement of scientific production using technology and artificial intelligence. For example, the use of artificial intelligence for control of bias could be used through algorithms designed to detect bias at surveys as well as measurements when getting some values of patients. Moreover, it could avoid the loss of follow-up of patients as well as it can reduce confounding variables just by using applications from smartphones and internet search [8].

Another example that explains the relationship between artificial intelligence and DEI, is the intelligent transformation of data, to create live and dynamic information 24/7, which serves to create control panels that allow the management and optimization of disease management or pandemics like COVID-19. Figure 3 shows a control board for the “live control” of the hospital capacity for the department of Magdalena, Colombia to attend patients with SARS Cov-2. Using an artificial intelligence engine “MAIA” developed by MEDZAI0, it collected data of the hospitals in the area (in real time), transforming it onto dynamic information and created a control panel which was used as a business intelligence dashboard for supply/demand adjustment.

A main focus is the monitoring of adverse effects in clinical trial or pathologies, through digital follow-up and dynamic records at internet. This would allow to a more personalized treatment as observed on: Utilizing Digital Health to Collect Electronic Patient-Reported Outcomes in Prostate Cancer: Single-Arm Pilot Trial, which concluded: “A high compliance rate confirmed the app as a reliable tool for patients with localized and advanced prostate cancer. Nearly all participants reported that using the smartphone app is easier than or equivalent to the traditional paper-and-pen approach, providing evidence of acceptability and support for the use of remote PRO monitoring. This study expands on current research involving the value of digital health, as a social and behavioral science, augmented with technology, can begin to contribute to population health management, as it shapes psychographic segmentation by demographic, socioeconomic, health condition, or behavioral factors to group patients by their distinct personalities and motivations, which influence their choices” [9].

To give a visual explanation of the previous example, with the authorization of MEDZAI0, Figure 4 shows an intelligent control of clinical follow-up of patients after liver transplantation, an example like the one previously mentioned for prostate and clinical follow-up.

Another example of innovation in clinical trial models are the emerging designs for adaptative studies: Basket and Umbrella, for the oncology field, where the sample recruitment can result difficult for some pathologies. These



Figure 3: Control panel for optimization of hospital capacity to attend covid-19 in Magdalena, Colombia. Intelligent geolocation is observed according to occupancy alerts. Occupancy trends, summary for each hospital, occupancy alert tachometer. Image taken from: MEDZAIO.



Figure 4: Medical control panel for post-liver transplant patients, to monitor pharmacovigilance and adherence of patients to medication using artificial intelligence engines. Image taken from: MEDZAIO (Spanish version).

studies innovate the traditional designs. The basket design is based on distributing patients by common treatment without differentiating histology, while the umbrella design is the opposite, it approaches on histologic variants rather than treatments [10]. In this way, efficiency of the trial is maximized, cost are reduced, samples can be smaller as they

have more accurate calculations, calculations for necessary doses can be enhanced and thus, there are less toxicity possible effects [10]. Although these types of studies can be challenging too, especially when applying them massively and periodically, they are a clear evidence of how DEI is generating a change in the medical investigation.

Furthermore, other facilities include a randomization of patients according to internet research trends and epidemiologic vigilance through Google or other internet servers. Even totally digital trials, without on-site interventions but artificial intelligence and digital data bases, for example, to identify risk factors for suicide through internet research. All these are just some examples of the benefits of using technological tools within clinical trials, as we have experienced in our own group of Digital Epidemiology and of Innovation. Nevertheless, there are endless ideas or uses that could be given for clinical trials and that is the art of innovation.

Some available examples in literature include:

- Addressing Bias in Artificial Intelligence in Health Care [8].
- Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales [11].

Machine learning - Deep learning

Machine learning and Deep learning approaches allow the introduction of a concept barely used among epidemiologists, yet of utmost importance. They refer to neural networks and predictive analytical assessments, which we believe is the future of the investigation field. A neuronal network is a model that analyzes different pathways to reach a particular outcome. In other words, it analyzes distinct sceneries and variables and allocates weights to the criteria to find the best options. Imagine having the ability to tell an oncologic patient the probability of death according to his/her age, weight, sex, type of mutation and type of drug that will be used. But even better, imagine having the possibility to observe a change in that probability of death whether another drug is more suitable for that individual, considering that the same drug for other people with different variables could show totally distinct survival probabilities. This is not science fiction, in fact, it just requires the use of neural networks to treat patients depending on their very own probabilities of death and survival (real world predictive analysis), as seen on Predictive Analytic of Survival Using a Neural Network in a Real World Cohort of Advanced Non-Small Cell Lung Cancer Patients. [12]. Table 1 shows the results of the previously mentioned neural network. Figure 5 shows a diagram of how a neural network works so that the reader has a visual aid to understand the concept.

Recent algorithms plus the use of artificial intelligence to detect diseases are based on these conceptual margins of Machine Learning y Deep learning. Some of the available examples in the literature are:

- Dermatologist-level classification of skin cancer with deep neural networks [13].
- Inception v3 de Google: "Scientists at New York University demonstrated that an algorithm of Google can distinguish between the two more prevalent subtypes of lung cancer with 97-99% accuracy. Recently, for their diagnosis of adenocarcinoma and squamous lung cell carcinoma requires an experienced pathologist who could identify the tumor by visual inspection." [14,15].

Table 1: A Neural Network in a Real World Cohort of Advanced Non-Small Cell Lung Cancer Patients. Take from [2].

Gender	Age(Years)	Exposure to cigarette	Mutation/PDL-1	Probability of overall survival	Probability of dying
M	21	No	PDL1: 1% AL 49%	1.20%	98.60%
F	28	No	EGFR mut	19.10%	80.90%
F	40	No	EGFR mut	13.80%	86.20%
F	41	No	PDL1: 1% AL 49%	8.30%	91.70%
F	44	No	EGFR mut	7.80%	92.20%
F	48	No	EGFR mut	5.10%	94.90%
M	49	Si	PDL1>=50%	14.20%	85.60%
F	50	Si	EGFR mut	47.80%	52.20%
M	52	Si	EGFR mut	5.60%	94.40%
F	53	no	EGFR mut	3.50%	96.50%
F	53	Si	EGFR mut	41.50%	58.50%
F	54	no	No Oncogenic Addiction-No PDL 1 Expression	4.50%	95.50%

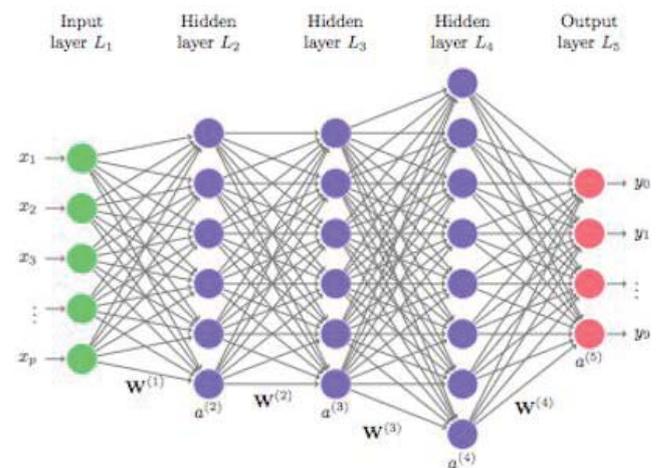


Figure 5: Diagram of a neural network (deep learning). Image take from: <https://towardsdatascience.com/its-deep-learning-times-a-new-frontier-of-data-a1e9ef9fe9a8>

As mentioned, these algorithms could be also used to improve the internal and external validity of the clinical studies, to perform predictive epidemiological vigilance, to control diseases and to help taking actions based on predictions in public health field. Likewise, neural network could be applied to health economics to predict how much to adjust costs for drugs or a new technology so that they are cost-effective according to personal and demographic variables. Finally, those algorithms can be used to train medical decision support systems. At this moment, some investigation groups are working on these topics, for more information: <https://www.epidemiologyserviceit.com>.

Disease control with digital tools

Although the control of diseases is an essential topic within epidemiology, in the past, people did not relate it with technological tools. Epidemiology relies on analysis of data to understand the dynamics of diseases and generate measures



of public health for their control [1]. Indeed, DEI has a new perspective to perform all these activities using digital data, technology and artificial intelligence as seen along this article. Some examples include:

- Technology permits education propagation and support people in the distance without an on-site control, something crucial in the management of chronic diseases with which patients need the continuum of medical interventions and medications [16]. For example, medical supervision via internet for a patient diagnosed with a depressive disorder, in order to watch out for suicidal tendencies.
- Artificial intelligence and smartphone applications do also allow for a continuous follow-up of the outpatient, for an intelligent epidemiologic vigilance, for a continuous tracking of adverse effects and thus, for a reduction of morbimortality. These is not surreal, in fact, in Spain there is an overview of publications about the latest evidence of technology applications for the control of diseases: “Revisión de intervenciones con nuevas tecnologías en el control de las enfermedades crónicas.” It concludes that “the use of technologies of information and communication in the monitoring of physiologic variables for the detection and follow-up of the cardiovascular pathology had better clinical outcomes, a reduction in mortality and a reduction in the use of medical services.” [16].
- Epidemic and pandemic control through digital data is another way of controlling diseases with DEI [3].
- Algorithmic prediction will provide prevention of diseases and the mitigation of the burden of the disease, which is the final objective of the intelligent/predictive epidemiology [2].
- Intelligent follow-up of the mortality of chronic pathologies and survivors and survival analysis [17].

Finally, we want to mention that although DEI is not designed for digital medical records, it turns to be an essential tool to develop digital studies based on those sources. With the digital medical records, artificial intelligence can collect all the information with neither the patient's intervention nor the need for someone to extract data, can compare patients records with the verbal answers of a survey to find mistakes and can make a control of population's diseases like for example, for the report of diseases of obligatory notification and then making a digital follow-up of them [18–21].

Conclusion

The Digital Epidemiology and of Innovation (DEI), is a new branch of epidemiology that will transform the medical investigation by using digital data and artificial intelligent and by improving the analysis and understanding of diseases. Indeed, it will improve health outcomes globally. In this article we described the concept and used examples based on the experience of our group of digital epidemiology and of

innovation, however, we must highlight that this is just the beginning of a new term with multiple pathways and ideas to explore.

Authors' contributions

All authors worked in the conception and design of this study, in the writing, edition and publishing of this article.

References

1. Salathé M (2018) Digital epidemiology: What is it, and where is it going? *Life Sciences. Society and Policy* 14: 1. [Link: https://bit.ly/3fIQ9xc](https://bit.ly/3fIQ9xc)
2. Pino L, Triana I, Mejia J, Ospina A, Camelo M, et al. (2019) P2.16-22 Predictive Analytic of Survival Using a Neural Network in a Real World Cohort of Advanced Non Small Cell Lung Cancer Patients. *J Thorac Oncol* 1. [Link: https://bit.ly/3mjlcj6](https://bit.ly/3mjlcj6)
3. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ (2016) Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proc Natl Acad Sci U S A* 113: 6689-6694. [Link: https://bit.ly/3wujGAs](https://bit.ly/3wujGAs)
4. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012-1014. [Link: https://bit.ly/39IfSSv](https://bit.ly/39IfSSv)
5. Desai R, Lopman BA, Shimshoni Y, Harris JP, Patel MM, et al. (2012) Use of Internet Search Data to Monitor Impact of Rotavirus Vaccination in the United States. *Clin Infect Dis* 54: e115-e118. [Link: https://bit.ly/3ueMhbc](https://bit.ly/3ueMhbc)
6. Shilo S, Rossman H, Segal E (2020) Axes of a revolution: Challenges and promises of big data in healthcare. *Nat Med* 26: 29-38. [Link: https://bit.ly/31J8yBE](https://bit.ly/31J8yBE)
7. Rolfhamre P, Janson A, Arneborn M, Ekdahl K (2006) SmiNet-2: Description of an internet-based surveillance system for communicable diseases in Sweden. *Eurosurveillance* 11: 15-16. [Link: https://bit.ly/39Ebeow](https://bit.ly/39Ebeow)
8. Parikh RB, Teeple S, Navathe AS (2019) Addressing Bias in Artificial Intelligence in Health Care. *JAMA* 322: 2377. [Link: https://bit.ly/3rPF3Zu](https://bit.ly/3rPF3Zu)
9. Tran C, Dicker A, Leiby B, Gressen E, Williams N, Jim H (2020) Utilizing Digital Health to Collect Electronic Patient-Reported Outcomes in Prostate Cancer: Single-Arm Pilot Trial. *J Med Internet Res* 22: e12689. [Link: https://bit.ly/31MSoYc](https://bit.ly/31MSoYc)
10. Ordoñez JM, Comas JM, Erustes M, Solergasto R (2018) Estudios adaptativos: Diseños Basket Y Umbrella en on- cología 6.
11. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L (2013) Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales. *PLOS Computational Biology* 9: 11.
12. Pino L, Triana I, Mejia J (2019) MAIA (Medical Artificial Intelligence Assistant) as interface for a new cancer healthcare integrative platform. *JCO Global Oncology*. [Link: https://bit.ly/3dw1j5K](https://bit.ly/3dw1j5K)
13. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542: 115-118. [Link: https://bit.ly/3dwENJT](https://bit.ly/3dwENJT)
14. Algoritmo de Google detecta el cáncer de pulmón en un 99%. (s. f.). [Link: https://bit.ly/3mkYXKH](https://bit.ly/3mkYXKH)
15. Velasquez P (2020) Google crea algoritmo para detectar el cáncer de pulmón. *CONSULTORSALUD*. [Link: https://bit.ly/3miwvNH](https://bit.ly/3miwvNH)
16. Agencia de Evaluación de Tecnologías Sanitarias (2005) Revisión De Intervenciones Con Nuevas Tecnologías En El Control De Las Enfermedades Crónicas. 58.



17. Samerski S (2018) Individuals on alert: Digital epidemiology and the individualization of surveillance. Life Sciences Society and Policy 14: 13. [Link: https://bit.ly/3mvxd6x](https://bit.ly/3mvxd6x)
18. Molina-Rueda MJ, Cabrera-Castro N, Onieva-García M, Lopez B, Abreu-González P (2014) The electronic health record (Diraya): A resource in epidemiological surveillance. Gac Sanit 28: 341-342. [Link: https://bit.ly/3cMWoy6](https://bit.ly/3cMWoy6)
19. Onieva-García MÁ, López-Hernández B, Molina-Rueda MJ, Cabrera-Castro N,

- Mochón-Ochoa M, et al. (2015) Aportación de la historia clínica digital a la vigilancia de enfermedades de declaración obligatoria. Rev Esp Salud Publica 89: 515-522. [Link: https://bit.ly/31JNc7l](https://bit.ly/31JNc7l)
20. Big Data (2021) Qué es y por qué importa. (s. f.). [Link: https://bit.ly/39Daksv](https://bit.ly/39Daksv)
21. Power Data G (2020) Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad. [Link: https://bit.ly/39Jxz41](https://bit.ly/39Jxz41)

Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROMEO, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services
(<https://www.peertechz.com/submission>).

Peertechz journals wishes everlasting success in your every endeavours.

Copyright: © 2021 Triana Avellaneda IC, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.