

# Progress in Characterizing the Human Exposome: a Key Step for Precision Medicine

Fernando Martin-Sanchez<sup>1,5</sup>, Riccardo Bellazzi<sup>2</sup>, Vittorio Casella<sup>2</sup>, William Dixon<sup>3</sup>, Guillermo Lopez-Campos<sup>4</sup>, Niels Peek<sup>3</sup>

<sup>1</sup> Instituto de Salud Carlos III (ISCIII), Madrid, Spain

<sup>2</sup> University of Pavia, Italy

<sup>3</sup> The University of Manchester, UK

<sup>4</sup> Queens University, Belfast, Northern Ireland, UK

<sup>5</sup> The University of Melbourne, Australia

## Summary

**Objective:** Most diseases result from the complex interplay between genetic and environmental factors. The exposome can be defined as a systematic approach to acquire large data sets corresponding to environmental exposures of an individual along her/his life. The objective of this contribution is to raise awareness within the health informatics community about the importance of dealing with data related to the contribution of environmental factors to individual health, particularly in the context of precision medicine informatics.

**Methods:** This article summarizes the main findings after a panel organized by the International Medical Informatics Association – Exposome Informatics Working Group held during the last MEDINFO, in Lyon (France) in August 2019.

**Results:** The members of our community presented four initiatives (*PULSE*, *Digital exposome*, *Cloudy with a chance of pain*, *Wearable clinics*), providing a detailed view of current challenges and accomplishments in processing environmental and social data from a health research perspective. Projects illustrate a wide range of research methods, digital data collection technologies, and analytics and visualization tools. This reinforces the idea that this area is now ready for health informaticians to step in and contribute their expertise, leading the application of informatics strategies to understand environmental health problems.

**Conclusions:** The featured projects illustrate applications that use exposome data for the investigation of the causes of diseases, health care, patient empowerment, and public health. They offer a rich overview of the expanding range of applications that informatics is finding in the field of environmental health, with potential impact in precision medicine.

## Keywords

Environmental exposure; health informatics, precision medicine

Yearb Med Inform 2020;115:20

<http://dx.doi.org/10.1055/s-0040-1701975>

## 1 Introduction

Most diseases result from the complex interplay between genetic and environmental factors. The exposome can be defined as a systematic approach to acquire large data sets corresponding to environmental exposures of an individual along her/his life [1] and associate them with specific health and disease status [2]. In its broadest sense, the exposome encompasses not only exposures to environmental stressors, but also the physical environment, socio-economic factors, the built environment<sup>1</sup>, aspects related to access to health care, and individual life habits or behaviours [3].

The health informatics community has only recently become aware of the importance of collecting environmental data in order to understand an individual's health [4]. There is a need for new digital methods and resources that select, annotate, organize, and present reliable and updated information about environmental factors affecting our health on both the population and individual/patient scales. The exposome demands a systematic research effort equivalent to what has been done to characterize the human genome (and also the human phenome) [5, 6]. Precision medicine has explicitly acknowledged the need to acquire and integrate individual-level genetic, environmental and clinical data to achieve a better understanding of multifactorial diseases and for developing

new preventive, diagnostic, and therapeutic solutions, adapted to groups of individuals with similar risk factors [7].

To address these problems, the International Medical Informatics Association (IMIA) created in August 2017 in Hangzhou (China) a Working Group on informatics aspects related to the exposome to support investigators, clinicians, and consumers navigate throughout the entire “data to knowledge” lifecycle: data collection, knowledge representation, annotation, integration with genomic and phenomic data, analytics, and visualization.

This contribution summarizes the main findings after a panel organized by the IMIA – Exposome Informatics Working Group held during the last MEDINFO, in Lyon (France) in August 2019. This panel was a follow-up, updating a very successful and well-attended panel, held at MIE 2018 in Gothenburg, Sweden.

The objective of this panel was to raise awareness within the health informatics community about opportunities for research in this area, in the context of precision medicine informatics. With that purpose, outstanding members of our community presented four on-going research projects (*PULSE*, *Digital exposome*, *Cloudy with a chance of pain*, *Wearable clinics*), providing a very detailed and complete view of current challenges and accomplishments in processing environmental and social data from a health research perspective. The four projects illustrate a wide range of research methods, digital data collection technolo-

<sup>1</sup> All the physical things constructed by humans as aids to living

gies, and analytics and visualization tools. This reinforces the idea that this area is now ready for health informaticians to step in and contribute their expertise, leading the application of informatics strategies to environmental health problems.

## 2 Participatory Urban Living for Sustainable Environments (PULSE)

Participatory Urban Living for Sustainable Environments (PULSE) is a project funded by the European Commission within the Horizon 2020 program, with the final goal to support cities in planning and implementing innovative public health policies in the urban environment relying on an innovative “big data”-enabled information system [8].

The PULSE technological architecture is based on the integration of two main components: i) a participatory data collection system, made of an app called “PulseAir”, which allows voluntary citizens to provide information about their health, lifestyle, and exposure patterns by merging questionnaire data and wearable/home sensors signals; ii) a data analytics and decision support system made available to health care authorities and city planners, where different data sources, including data collected by PulseAir and data about health care and environment, are integrated and visualized. A peculiar aspect of the entire system is represented by the spatial enablement of the data analytics platform, i.e., the ability to add a geographic reference to data and information. This allows designing the two main components with geo-referenced components, such as maps, webGIS, and geo-analytics dashboards [9].

In terms of disease risk assessment and prevention, PULSE focuses on the link between air pollution and asthma, and on the relationships between physical inactivity and type 2 diabetes. In both cases, health risk is seen as a combination of environmental and social exposures (e.g., air pollution, poverty) and human behaviour (e.g., a sedentary lifestyle). For this reason, PulseAir collects data to profile citizens

in terms of their health risk, specifically focusing on the combination of exposome and phenome information.

Currently, the PULSE project is in its final stages of development. After the collection of data coming from more than 1500 citizens in different cities, the final version of the overall architecture is under development, and its release to the participating cities is foreseen in April 2020. In this way, each city will be able to launch Public Health Observatories (PHOs), which will be exploited to integrate, analyse, and visualize data to inform health policy decisions [10].

As exposomics is concerned, several research activities have been carried on within the PULSE project. In the following we briefly describe two of them.

### 2.1 Personal Exposure

One of the hottest topics in exposomics is the influence of air quality on health. Evaluating the individual exposure of each citizen to pollutants is a key issue. However, current technological solutions exploited by cities are unable to measure air quality with the spatial and time resolution needed for this purpose. Typically, environmental agencies install a very limited number of air quality monitors. As an example, New York City (NYC) has 18 stations for an area of 1034 km<sup>2</sup>, i.e., a density of one station per 57 km<sup>2</sup>, and the stations do not measure all pollutants. If we limit the analysis to the most important air quality pollutant, e.g., PM 2.5 (Particulate Matter), NYC has only 9 stations. PM personal exposure, expressed as the estimate of PM inhaled in one day by a given citizen, is currently only very roughly evaluated in current settings in cities, by combining PM estimates and mobility patterns (extracted for example by GPS positioning). Within the PULSE project, we are experimenting different strategies to evaluate personal exposure. On the one hand, we are combining the information coming from city stations with data collected by low cost sensors that can be installed in patients' homes. In the “city lab” of Pavia, we have installed 42 low cost Purple Air PA-II sensors, increasing the density to about 1.5 station per km<sup>2</sup>. Thanks to the PulseAir application it is possible to

estimate the inner-city trajectory of each participating citizens, and thanks to the activity tracker it is possible to derive their heart rate with high temporal resolution. The combination of this data largely increases the precision in the estimate of the exposure to all the pollutants measured by the sensors. On the second hand, we are also testing “portable” air quality sensors, i.e. sensors that can be worn by citizens when they move in the city, such as the Dunavnet DV800 sensor. Even if those types of sensors cannot yet be worn in every-day life, they can be used to assess in one day air quality in many city areas with high spatial precision. Dunavnet has been tested in New York, Barcelona, and Birmingham.

### 2.2 Transfer Learning

Deep and transfer learning to support image analytics have been exploited within PULSE to investigate the relationships between the urban landscape of city areas and well-being and healthcare indexes of citizens [11]. The goal of this activity is to design a tool for health care planners and decision-makers to find clusters of city areas that share similar urban structures and similar health care indexes, in order to plan “cluster-based” interventions, rather than relying only on the geographical locations of areas. In order to demonstrate the feasibility of this approach, images and health care data coming from New York City have been analysed.

In particular, an image collected by the “The National Agriculture Imagery Program” (NAIP), which acquires aerial imagery during the agricultural growing seasons in the continental United States, has been retrieved and processed to be subdivided into image square blocks of 512 meters edge. Moreover, health care data have been collected from the 500 Cities project repository. The 500 Cities project is a collaboration between the Centers for Disease Control and Prevention (CDC), the Robert Wood Johnson Foundation, and the CDC Foundation. The project provides city- and census tract-level small area estimates of many health-care indexes for the largest 500 cities in the United States. NYC is divided in 2166 census tracts. We have exploited 2017 measures about major risk behaviours that

lead to illness, pain, and early death, as well as the conditions and diseases that are the most common, costly, and preventable of all health problems.

A mapping between each image and health care data has been performed, in order to provide an estimate of health care indexes for each image square block. Then, each image has been processed with the “Painters” deep neural networks made available in the Orange software [12]. The latent variables extracted by the deep network have been used to cluster images. Finally, health care indexes of each image have been correlated with the clusters, showing that clusters (which were derived only on the basis of the urban images) can be predicted on the basis of health care data. This data analytics pipeline showed that it was thus possible to correlate urban landscape with healthcare indicators at the whole city level. In the NYC case, such correlation looks particularly relevant, probably because of social factors, which, in the US society and in large cities, make health indicators related to the urban areas where people live.

The work carried on, while demonstrating that deep neural networks designed to encode image data can be successfully reused within transfer learning approaches, also shows that it is possible to profile city areas, and that the urban structure is a component of the exposome of citizens.

### 3 Cloudy with a Chance of Pain

The weather has been thought to influence health for millennia [13]. One of the best-known beliefs is that the weather influences pain in people living with arthritis and other long-term pain conditions. Indeed, around 80% of people living with arthritis believe in such an association, and around half believe they can forecast the weather based on their symptoms.

Many researchers have tried to study this association, but results have been inconclusive [14]. Some of the limitations include small sample sizes or short durations of follow-up. Another important limitation is

the quality of the data to support the analysis, including regular information about pain or other symptoms, and high-quality information about the weather to which research participants are exposed. Many studies assume that patients stay at their postcode, or within their town or region. However, people are mobile and it would be helpful to link their moving geolocation to the local weather.

Smartphones offer new opportunities for conducting health research at scale [15]. This includes tracking symptoms on a more frequent basis, integrated into participants’ daily lives, as well as making use of sensor data from within the smartphone. Cloudy with a Chance of Pain is a national UK smartphone study designed to address the age-old question of weather and pain [16]. The study recruited over 13,000 participants in 2016 and asked them to record their daily symptoms for up to six months, while the phone’s GPS was automatically recording the local weather data. Ultimately, the study collected 5.1 million symptom scores with daily weather data accessible from the 154 UK Met Office weather stations. The analysis used a case-crossover design to compare the weather at times of increased pain to ‘control periods’ within the same month when pain was not increased. The results demonstrated that high relative humidity, stronger winds, and low pressure were associated with more painful days [16].

In addition to addressing the primary question of weather and pain, this study demonstrated how the use of consumer devices could support novel health research at scale. Thinking specifically about the exposome, the study allowed participants to move around the country and still provide accurate local weather data. The average user was linked to 9/154 possible UK weather stations (interquartile range 4–14) [16]. The most mobile person required data from 82/154 weather stations, indicating how mobile participants were during the course of the study and the importance of accounting for mobility and local weather data over the course of the study. Similar methods could be used for future studies that need to gather regular self-reported information alongside accurate exposome data, using the smartphone GPS linked to appropriate datasets to define exposure.

## 4 The Digital Component of the Exposome

The exposome has traditionally been associated with physical or socio-economical environmental factors that are specific for an individual at a given moment. Our current society lifestyle is increasingly reliant of the use of the internet and digital tools affecting almost every single aspect of our lives. This has therefore increased the amount of time spent online by citizens using different platforms such as social networks [17]. These online and digital interactions drove the definition of new elements that could be related with health aspects and the creation of new concepts such as the “Digital phenotype” [18] that was coined in 2015 to refer to the digital footprint of an individual that could eventually be related with a disease and is related with the development of the idea of defining digital biomarkers referring to objective and quantifiable data measured by digital devices [19]. However, the aforementioned concepts are related with the tools used to measure digital responses of individuals to different stimuli. On the other hand, the digital world represents a rich and new environment to which individuals are increasingly exposed and that may have health consequences.

Drugs and chemicals are paradigmatic examples of relevant factors considered to be an important element of the exposome. In a similar fashion the digital environment offers the opportunity to identify digital analogies and similarities to those well recognised components of the exposome. Drugs in the physical environment have different health effects ranging from addiction disorders to treatment of different conditions. In the digital environment it is possible to identify similar examples of how digital exposures may have different health effects. Internet addiction disorder would be an example of how an inadequate use of digital tools and contents may lead to an addiction disorder. On the other hand, digital exposures have also been applied to treating different health conditions, particularly in the mental health domain with the development of computer-based cognitive behavioural therapy (CBT). In this regard,



a clinical trial was published in 2017 comparing the use of virtual reality exposures and in vivo exposures in CBT showing that the virtual exposure was effective [20, 21]. These examples show the necessity to consider the relevance of digital exposures as another intrinsic element of the exposome. For this reason, in 2017 the concept of the digital exposome or the digital component of the exposome as “the whole set of tools and platforms (including contents) that an individual use and the activities and processes that an individual engage with as part of his digital life” [22] was introduced reflecting on the relevance of this element in the broader definition of the exposome and complementing the already existing concepts of digital phenotypes and biomarkers.

Biomedical informatics has over the years developed a vast array of methodologies and tools in areas such as natural language processing, self-monitoring, or participatory health that might be applied to characterise components of the digital exposome and may capture the digital activity of an individual. However, this very possibility of tracking the digital activities of the individuals represents a series of important challenges that range from how these data will be analysed in a longitudinal manner, how or where they will be stored, and how they will be shared. Ethics represents a major challenge in this scenario where a 24/7 surveillance of individuals is technically feasible and should be carefully considered for practical implementations of the digital exposome.

## 5 Wearable Clinics

People with long-term conditions (LTCs) typically interact with healthcare services through rigid pathways that poorly match the dynamic nature of their condition [23]. On the one hand, clinic visits may be unnecessary during times of remission, thus wasting the time and resources of patients and professionals. On the other hand, ‘one size fits all’ care is seldom timely or specific enough to arrest relapses before they lead to serious exacerbations or costly hospitalisations. “Wearable Clinics” [24, 25] are a vision for digitally transformed healthcare services that

enable new forms of collaborative care for LTC management through dynamic personal care plans that adapt to the changing state of the individuals and the world around them. The aim is to empower patients to become managers of their own care through actionable care-planning information and mobile/wearable technologies.

The development of Wearable Clinics is characterized by a number of engineering and translational research challenges: 1) The design of multimodal, adaptive sensing, and signal compression algorithms for high-resolution wearable sensing data, minimising communication demand, and maximising sensor operating lifetime; 2) The integration of passive wearable sensing modalities (e.g. accelerometer, GPS, heart rate) with active mobile sensing modalities (e.g. ecological momentary assessments through smartphones [26]); 3) The real-time prediction of disease relapse risks based on integrated data from electronic health records, passive wearable sensing, and active mobile sensing; 4) The adaptive, personalised care planning that takes into account predicted risks, individual health and care goals, and available care resources in the patient’s specific environment; and 5) Support future real-world deployment through the assessment of user acceptability, potential health and economic benefits, patient safety and data security risks, and regulatory challenges associated with clinical deployment. These challenges highlight the highly transdisciplinary nature of Wearable Clinic research and engineering: it requires the involvement of bioengineers, computer scientists, statisticians, artificial intelligence researchers, software engineers, health psychologists, health economists, and patient safety experts.

The Universities of Manchester and York are currently developing prototypes of Wearable Clinics for severe mental illnesses (in particular, schizophrenia) and ambulatory blood pressure monitoring (ABPM). Both wearable clinics consist of a patient-facing mobile app with associated sensing modalities and a clinic-facing web dashboard that summarizes patients’ status in real time. The wearable clinic app for schizophrenia integrates short smartphone questionnaires for psychosis symptom assessment [27–29], GPS sensing

for behavioural phenotyping [30–33], and real-time risk assessment using cluster Hidden Markov models [34]. The wearable clinic for ABPM is an example of multimodal wearable sensors integration and uses activity classification from accelerometer data [35, 36] to determine the optimal moments for blood pressure measurement [37].

## 6 Conclusions

This panel had a very special meaning. First, it was held in Lyon (France), the city where the International Agency for Research on Cancer is located. The former director (2009–2019) of this WHO institute, Dr. Christopher Wild, had coined the term exposome in 2005, when he was director of this centre. As he stated in [1]: “*There is a desperate need to develop methods with the same precision for an individual’s environmental exposure as we have for the individual’s genome. I would like to suggest that there is need for an “exposome” to match the “genome.”*” Second, according to our knowledge, this was the first panel dedicated to the subject of exposome informatics in a MEDINFO conference.

The four initiatives presented during the panel offered a very rich overview of the expanding range of applications that informatics is finding in the field of environmental health, with a potential impact on precision medicine. Table 1 summarizes the main aspects from these four projects, looking at different aspects related to their scope, data processing particularities, disease and environmental factors involved, and final users.

In terms of their objectives, the featured projects illustrate applications that use data on exposure to environmental and social risk factors for the investigation of the causes of diseases, health care, patient empowerment, and public health. Therefore, the final users of these systems can be researchers, clinical professionals, health planners and public health, and urban planning authorities, including the participation of experts in bioethics, given the challenges posed by some of these projects in terms of data privacy and security.

The studies cover aspects related to long-term, chronic, diseases and conditions, (pain, asthma, arthritis, mental health problems, hypertension), which is logical considering that it is in these pathologies that the influence of environmental factors is most important and still not well determined in many cases.

The authors have worked with a wide range of data that inform about individual's exposure to environmental and social factors that can affect health, specifically, air quality, physical activity patterns and mobility, weather, the use of digital technology and poverty. Without being exhaustive, of course, the studies included here offer very representative examples of the challenges that researchers face when processing this type of data.

In terms of data processing, the authors addressed several of the main aspects related to data collection, data processing and integration with different data sources, as well as

the most appropriate methods for their analysis, visualization, and application to health problems. Thus, in terms of data collection, the projects faced problems derived from the use of smartphones apps, wearables, fixed and mobile environmental sensors, questionnaires, and electronic medical records. These data had to be integrated in some cases with other data sources, for example, those coming from meteorological stations. Various methods were used for data management with a strong geographical component (maps, geo-reference, webGIS), as well as for data preparation (signal compression algorithms). Finally, when analysing these types of data, logically, techniques and methods of artificial intelligence played a key role (deep and transfer learning), but the authors want also to highlight the development of dashboards for patients, clinicians and health authorities, personal health plans, and tools for real-time risk prediction.

These studies illustrate the need to move forward in participatory data collection approaches (person-generated health data), new research designs, biomarker and digital therapies characterization, patient empowerment and regulatory aspects.

We hope that this collection of projects, diverse but with common aspects and objectives, will serve the reader to assess the importance of addressing studies that integrate genetic, clinical, and environmental information, all necessary for the development of precision medicine. Our working group invites the health informatics community to participate in on-going discussions and help shape the research agenda promoting the use of informatics and data science in exposome and environmental health research, particularly in the context of precision medicine.

**Table 1** Summary of the four projects according to several dimensions related to scope, goal, data processing methods, disease, and risk factors.

| Project                             | Objective  | Final users                                      | Diseases  | Social and environmental factors                       | Data collection methods  | Data processing methods                                    | Data analytics and visualization methods  | Study sample size | Other relevant info   |
|-------------------------------------|--|--|---|--|--|--|---|-------------------|---|
| <b>PULSE</b>                        | To improve urban environment.<br>- Public health observatory                     | Health care authorities and city planners        | Asthma, Type 2 Diabetes   | Poverty, air quality, physical activity and heart rate | PulseAir app with questionnaire and wearable. Environmental sensors. Portable air quality sensors. | Geo-reference, data integration, maps, WebGIS              | Deep and transfer learning, image analytics, decision support, geo-analytics dashboards   | 1500              | Participatory data collection system, patient individual risk     |
| <b>Cloudy with a chance of pain</b> | To examine the association between weather and pain                              | The public, specifically people living with pain | Arthritis, long-term pain conditions                                | Weather (humidity, low pressure, strong winds)         | Smartphone app, Smartphone GPS, 5.1 million symptom scores   | User data linked to 154 UK Met office weather stations     | Average user linked to 9 weather stations, the most mobile person linked to 82 weather stations                                     | 13,000            | Case-crossover design   |
| <b>Digital Exposome</b>             | To characterize the health impact of Internet use and other digital technologies | Researchers, bioethics experts                   | Addictive behaviour, depression and other mental health problems    | Use of Internet and digital technologies               | Smartphone apps, Questionnaires, Electronic health records   | Digital biomarkers, digital therapeutics                   | Individual digital footprints   | n/a               | Digital exposures as risk factors, but also as possible therapies |
| <b>Wearable Clinics</b>             | To develop digitally-transformed healthcare services                             | Clinicians and patients                          | Long-term conditions, chronic diseases, schizophrenia, hypertension | Mobility, other wearable sensor data                   | Electronic health records, active mobile sensing (EMA), passive wearable sensing                   | Multimodel adaptive sensing, signal compression algorithms | Dynamic and personal healthcare plans, real-time prediction of disease relapse risk, clinic dashboard, cluster hidden Markov models | 21 (first phase)  | Patient empowerment, regulatory challenges                        |

## References

- Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev.*: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2005;14(8):1847-50.
- Patel CJ, Ioannidis JP. Studying the elusive environment in large scale. *JAMA* 2014 Jun 4;311(21):2173-4.
- Wild CP. The exposome: from concept to utility. *Int J Epidemiol* 2012 Feb;41(1):24-32.
- Martin Sanchez F, Gray K, Bellazzi R, Lopez-Campos G. Exposome informatics: considerations for the design of future biomedical research information systems. *J Am Med Inform Assoc* 2014;21(3):386-90.
- Thomas DC, Lewinger JP, Murcray CE, Gauderman WJ. Invited commentary: GE-Whiz! Ratcheting gene-environment studies up to the whole genome and the whole exposome. *Am J Epidemiol* 2012; 175: 203-7; discussion 208-9.
- Manrai AK, Cui Y, Bushel PR, Hall M, Karakitsios S, Mattingly CJ, et al. Informatics and Data Analytics to Support Exposome-Based Discovery for Public Health. *Annu Rev Public Health* 2017 Mar 20;38:279-94.
- National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington (DC): National Academies Press (US); 2011.
- Pala D, Pagán J, Parimbelli E, Rocca MT, Bellazzi R, Casella V. Spatial Enablement to Support Environmental, Demographic, Socioeconomics, and Health Data Integration and Analysis for Big Cities: A Case Study with Asthma Hospitalizations in New York City. *Front Med (Lausanne)* 2019 Apr 24;6:84.
- Pala D, Rocca MT, Casella V. Advantages and difficulties of using spatial enablement to support public health in cities: The pulse case study, GISTAM 2019 - Proceedings of the 5th International Conference on Geographical Information Systems Theory, Applications and Management 2019;322:29.
- Pala D, Holmes J, Pagán J, Parimbelli E, Rocca MT, Casella V, et al. Agent-based models and spatial enablement: a simulation tool to improve health and wellbeing in big cities. In: Riaño D, Wilk S, ten Teije A, editors. *Artificial Intelligence in Medicine. AIME 2019. Lecture Notes in Computer Science*, Vol. 11526. Cham: Springer; 2019. p.79:83.
- Bellazzi R, Caldarone AA, Pala D, Franzini M, Malovini A, Larizza C, et al. Transfer learning for urban landscape clustering and correlation with health indexes. *Proceedings of the 17th International Conference, ICOST 2019, New York City, NY, USA*; 2019.
- Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, et al. Orange: Data Mining Toolbox in Python. *J Mach Learn Res* 2013;14: 2349-53.
- Hippocrates. On Airs, Waters and Places [Internet]. <http://classics.mit.edu/Hippocrates/airwatpl.html>
- Smedslund G, Hagen KB. Does rain really cause pain? A systematic review of the associations between weather factors and severity of pain in people with rheumatoid arthritis. *Eur J Pain* 2011 Jan;15(1):5-10.
- Check Hayden E. Mobile-phone health apps deliver data bounty. *Nature* [Internet]. 2016 Mar 21 [cited 2018 May 16];531(7595):422-3.
- Dixon WG, Beukenhorst AL, Yimer BB, Cook L, Gasparrini A, El-Hay T, et al. How the weather affects the pain of citizen scientists using a smartphone app. *npj Digit Med* [Internet]. 2019 Dec 24 [cited 2019 Nov 20];2(1):105.
- <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>. [cited 2019 Nov 7].
- Jain SH, Powers BW, Hawkins JB, Brownstein JS. *Nat Biotechnol* 2015 May;33(5):462-3.
- Dorsey ER, editor. *Digital biomarkers*. Basel: Karger. Available from: <https://www.karger.com/Journal/Home/271954>
- Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger É, Forget H, Loranger C, Roucay FX. Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. *Br J Psychiatry* 2017 Apr;210(4):276-83.
- Godino JG, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowle JH, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. *Lancet Diabetes Endocrinol* 2016;4(9):747-55.
- Lopez-Campos G, Merolli M, Martin-Sanchez F. Biomedical Informatics and the Digital Component of the Exposome. *Stud Health Technol Inform* 2017;245:496-500.
- de Silva D. Evidence: Helping people help themselves. A review of the evidence considering whether it is worthwhile to support self-management. London; 2011.
- The University of Manchester. *The Wearable Clinic: Connecting Health, Self and Care*, (2017). [https://www.herc.ac.uk/research\\_project/wearable-clinic-connecting-health-self-care/](https://www.herc.ac.uk/research_project/wearable-clinic-connecting-health-self-care/) [cited 2019 Nov 17].
- The University of Manchester. What is The Wearable Clinic? (video); 2019. <https://www.youtube.com/watch?v=LPQwjYH4pBw> [cited 2019 Nov 17].
- Bell IH, Lim MH, Rossell SL, Thomas N. Ecological Momentary Assessment and Intervention in the Treatment of Psychotic Disorders: A Systematic Review. *Psychiatr Serv* 2017;68:1172-81.
- Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, et al. A Comparison of Two Delivery Modalities of a Mobile Phone-Based Assessment for Serious Mental Illness: Native Smartphone Application vs Text-Messaging Only Implementations. *J Med Internet Res* 2013;15:e60.
- Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. *BMC Psychiatry* 2012;12:172.
- Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Lavery L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. *BMC Psychiatry* 2013;13:34.
- Difrancesco S, Fraccaro P, Van der Veer SN, Alshoum B, Ainsworth J, Bellazzi R, et al. Out-of-home activity recognition from GPS data in schizophrenic patients. *Proc IEEE Symp Comput Med Syst* 2016. p. 324-8.
- Fraccaro P, Lavery-Blackie S, Van der Veer SN, Peek N. Behavioural Phenotyping of Daily Activities Relevant to Social Functioning Based on Smartphone-Collected Geolocation Data. *Stud Health Technol Inform* 2019;264:945-9.
- Zheng Y, Fraccaro P, Peek N. The minimum sampling rate and sampling duration when applying geolocation data technology to human activity monitoring. In: Riaño D, Wilk S, ten Teije A, editors. *Artificial Intelligence in Medicine. AIME 2019. Lecture Notes in Computer Science*, Vol. 11526. Cham: Springer; 2019. p. 233-38.
- Fraccaro P, Beukenhorst A, Sperrin M, Harper S, Palmier-Claus J, Lewis S, et al. Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: a systematic review. *J Am Med Inform Assoc* 2019;26:1412-20.
- Hulme W, Stockton-Powdrell C, Lewis S, Martin GP, Bucci S, Parsia B, et al. Cluster Hidden Markov Models: An Application to Ecological Momentary Assessment of Schizophrenia. *IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*; 2019. p. 99-103.
- Zebin T, Sperrin M, Peek N, Casson AJ. Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. *Conf Proc IEEE Eng Med Biol Soc* 2018 Jul;2018:1-4.
- Zebin T, Scully PJ, Peek N, Casson AJ, Ozanyan KB. Design and Implementation of a Convolutional Neural Network on an Edge Computing Smartphone for Human Activity Recognition. *IEEE Access* 2019;7(10.2019):133509-20.
- The University of Manchester. *The Wearable Clinic: 24 hour blood pressure monitoring video*, (2019). <https://www.youtube.com/watch?v=OGp-JCPEo9Zc> [cited 2019 Nov 17].

### Correspondence to:

Fernando Martin-Sanchez  
 Instituto de Salud Carlos III, Pab 7  
 Avda. Monforte de Lemos 5  
 28029 Madrid  
 Spain  
 Tel: +34 8222479  
 E-mail: [fmartin@isciii.es](mailto:fmartin@isciii.es)