Expert Review of Precision Medicine and Drug Development

Personalized medicine in drug development and clinical practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tepm20

Value of digital biomarkers in precision medicine: implications in cancer, autoimmune diseases, and COVID-19

Enrico Capobianco & Pier Luigi Meroni

To cite this article: Enrico Capobianco & Pier Luigi Meroni (2021): Value of digital biomarkers in precision medicine: implications in cancer, autoimmune diseases, and COVID-19, Expert Review of Precision Medicine and Drug Development, DOI: 10.1080/23808993.2021.1924055

To link to this article: https://doi.org/10.1080/23808993.2021.1924055

Published online: 13 May 2021.

EXPERT REVIEW OF PRECISION MEDICINE AND DRUG DEVELOPMENT
https://doi.org/10.1080/23808993.2021.1924055

Taylor & Francis
Taylor & Francis Group

EDITORIAL

Value of digital biomarkers in precision medicine: implications in cancer, autoimmune diseases, and COVID-19

Enrico Capobianco[a] and Pier Luigi Meroni[b]

[a]Institute of Data Science and Computing, University of Miami, Miami, FL, USA; [b]Immunorheumatology Research Laboratory, IRCCS Istituto Auxologico Italiano, Milan, Italy

ARTICLE HISTORY Received 1 March 2021; Accepted 27 April 2021

KEYWORDS Digital biomarkers; precision medicine; cancer; autoimmunity; Covid-19

1. Introduction: value of digital biomarkers

Based on the patient's characteristics, precision medicine (PM) aims to optimize the time of administration of the most appropriate medicine with the minimum risk of toxicity. This is a multidimensional problem due to the varied disease course and therapeutic responses of patients. General factors, such as genetics, epigenetics, environment, ethnicity, adherence, lifestyle, and diet, determine these outcomes. In clinical trials, some drugs may be not beneficial or even harmful for a given ethnic or co-morbid group. Partial response is also observed outside trials, as the most commonly used drugs show high efficacy in relatively few patients. Therefore, what we call 'imprecise medicine' is the first challenge of PM due to the assumption underlying clinical practice that disease treatment and prevention strategies developed at the population level are expected to be accurate when applied at the individual level. The complexity that drives the variations in patient profiles depends on the heterogeneity of information obtained from large volumes of genetic, serological, biochemical, and diagnostic imaging data. These represent dimensions that need harmonization and integration with lifestyle and environmental factors. The second challenge is with assessing the benefits of the data dimensions, such as diagnostic improvements, earlier interventions, increased drug efficiency, and better-targeted treatments.

To accommodate the heterogeneity of the etiologies, clinical symptoms, and treatment responses of patients in clinical practice, a revised clinical approach is recommended [1]. The first step is the development of a machine learning (ML)-assisted risk assessment model (see, for instance [2],) followed by the identification of the robust multimodal data-driven prognostic indicators (see, for instance [3],). These two efforts require new strategies for integrating heterogeneous information from different structured and unstructured data *sources* (electronic health records (EHRs), administrative databases, bioimaging archives, self-quantified measurements, etc.). Big Data has introduced a new paradigm for population-based studies that comes with challenges. For instance, the validity of such studies is based on the diagnostic accuracy used for all cases. A critical problem is the variability of the methods used to perform validations. Currently, there are challenges with validating most disease classification algorithms, and this complicates the assessment of their potential for population studies. Model validation facilitates safer interpretability of the correlations between diverse data types revealed by the models.

Data-centric perspectives of complex diseases facilitate their definition as heterogeneous processes that have multifaceted causes, courses of evolution, treatments, and patient's disease trajectories from the observed responses to treatment (see [4–6], among many other examples). These trajectories differ with each patient and, therefore, necessitate a precision approach. We emphasize the necessity of early intervention when molecular causes/patterns can still be identified. Early treatment is likely to lead to a substantial reduction in the risk of disease progression and prolonged health. Thus, it is critical to develop more inclusive digital biomarkers (DBs) [7,8] that may reflect the synergism of clinical and molecular data for identifying diseases at the early stages when interventions have optimal chances of success and future damage prevention. The DB values should be proportional to the ability to shorten the length of the trajectories during the disease course, which will reduce the temporal window of opportunity between any disease trigger and a clinical intervention before irreversible damage occurs.

2. New perspectives for digital biomarkers

In the literature, DBs usually refer to data collected through digital health technologies (wearables, e-health tools, etc.) that enable artificial intelligence (AI)-assisted processing for better explanations and prediction of clinical outcomes. Analytics for large volumes of data are usually leveraged to delineate trends and infer patterns at individual and population levels. We stress that an extension of the definition of DB is needed [9–11]. While continuing with the measurement of physiological parameters, the point of leverage would be the inclusion of other types of digital information sources: examples include next-generation mobile sensors and detectors with integrated solutions from the Internet of Medical Things (IoMT) platforms. The latter represents the collection of medical devices and applications connected to healthcare IT systems through online computer networks (examples include remote patient monitoring systems, patients mHealth devices, smartphones, etc.). It is reasonable to expect that ML and AI tools will

CONTACT Enrico Capobianco ecapobianco@med.miami.edu Institute of Data Science and Computing, University of Miami, Miami, FL, USA

facilitate the processing of large amounts of data obtained from medical device-connected analytical dashboards and help doctors reach timely, reliable, and actionable decisions.

We have highlighted a few objectives that are instrumental to improving patient stratification and management in both daily practice and clinical trials:

a) Perform targeted cohort studies (focus on disease phenotypes and markers);

b) Discover and validate new biomarkers (panels/combinations) by exploiting emerging data dimensions.

c) Determine modifiable risk factors for disease onset to improve prevention in pre-disease cohorts and build accurate profiles.

d) Leverage the characterization of trajectories of disease progression while strengthening the identification of checkpoints and comorbidities.

e) Develop new analytics tools predictive of response to treatment.

We stress that there is a wealth of unused data in the databases of pharmaceutical companies and academic institutions. Coupled with new operational challenges, the benefits will come from public access to these data at some point. These include increased transparency and reproducibility of results and further development of open data initiatives and large-scale secondary analyses covering several complex diseases. This process will boost the discovery of DBs by leveraging diversity factors (age, gender, ethnicity, lifestyle, environmental exposure, and geopolitical variables) that play a role in improving the quality of life and/or reducing the risk of permanent damage.

Two critical steps should be prioritized: (a) designing trials that cover key clinical questions based on more efficient mining of multimodal outcome data and (b) enabling insights from DB validation systems to be actionable in the clinic. It is necessary to elucidate the role of DBs in assessing the significance of co-influencers of several variables obtained from heterogeneous sources that are causative or correlated with disease. An important effort in this direction is building the knowledge base with the evidence obtained from the clinical trial cohorts, responses to therapy, patient-reported outcomes (including health-related quality of life, fatigue, sleep quality, symptom scores), and other comprehensive data (serological, clinical, biochemical, and diagnostic imaging data).

In parallel, it is necessary to establish model performance with the new DBs by developing innovative scoring systems for PM approaches that are usable in various disease contexts. With the multiple signatures that need to be verified, the main challenge is with defining meaningful and clinically actionable 'cut-offs' for panels of biomarkers inclusive of DBs and dealing with variations in individuals with and without disease. The specificity of each biomarker type is expected to significantly affect the measurement of false-positive rates. In addition, tools that combine biomarkers with clinical characteristics will guide timely treatment (i.e. pre-clinical or early symptom intervention time) and assess adverse effects. This evolution naturally leads to clinical decision support systems (CDSSs) designed to identify risk profiles (e.g. with respect to disease or associated comorbidities) and assess variables predictive of treatment response. These systems are based on semi-automated rules that elaborate information cross-referenced against a knowledge base, comprising electronic medical files and collected clinical data, among others. The CDSSs can perform predictions by estimating the probabilities of relapse, designing disease trajectories, profiling risk and prognostic paths, and identifying comorbidity early warnings, among others.

Leveraging the described multilevel approach toward the reproducibility and generalizability of results serves the need to address the reasons for failure in clinical trials targeted at complex diseases by further consolidating the centrality of data-driven research and directing patient-focused research. The main expected outcomes are: (a) individualized patient profiles that help predict disease development and progression, for example, by influencing decisions on primary (pre-onset), secondary (post-onset impact), and tertiary (progression control) preventions; (b) identification of comorbidity trajectories with superior accuracy by exploiting patient data integration; (c) characterization of common disease features together with their variants and determination of the specificities for various factors.

3. Challenges from disease contexts

In cancer, identifying predictive biomarkers (including DBs) is an essential goal for PM approaches [10]. The focus on well-characterized processes such as mutations (KRAS mutations, Her-2 expression, etc.) reduces the complexity of cancer but may not be cost- or time-effective for potential therapeutics. The challenges are related to validation (usually in randomized controlled trials) and clinical applicability (role of measuring response and how to link it to clinical benefit), which represent two crucial factors [12]. The areas of extensive research involve radiomic assessments and evaluations of both circulating markers (tumor cells, nucleic acids, etc.) and tissue-based markers. The increased ability to detect heterogeneity depends on the states of the microenvironment (from the unperturbed state to the response to therapy), and it facilitates: (a) better patient selection for specific therapies and more precise prediction of therapeutic response through measurements of the pharmacodynamic drug effects; (b) increased chances of success of clinical trials through accurate and timely efficacy assessment, optimization of drug dose, and identification of surrogate clinical trial endpoints; (c) tailored theranostic solutions; and (d) effective cost control. The radiomic component calls for the integration of various types of markers (imaging, tissue-based, histology, pathology, molecular profiling, etc.). A critical challenge with stimulating global collaborative research is the lack of multicenter standardization approaches [13,14].

The definition of systemic autoimmune rheumatic diseases (SARDs) includes a heterogeneous group of chronic inflammatory disorders with damage to different tissues that is mediated by immune responses against self-antigens. Systemic lupus erythematosus (SLE) is a SARD prototype characterized by clinical manifestations from the skin to the kidney or central nervous system involvement. The heterogeneity of the manifestations of lupus accounts for the problem of

enrolling patients with different phenotypes in clinical trials and the frequent failure of the trials [15]. The first approach to identifying the various lupus variants took advantage of the association between clinical manifestations and some biomarkers, such as serum autoantibodies. However, this solution has several limitations related to the sensitivity and relative specificity of the biomarkers [16,17]. There is growing evidence that SLE variants can be better characterized by immunophenotyping and/or transcriptomic analysis. New biomarkers, including digitized ones, may help predict the natural disease course or the response to specific treatments [18]. Rheumatoid arthritis (RA) is another SARD with apparently more homogeneous presentations than SLE. Nevertheless, the response to therapy is variable, even for molecules belonging to the same pharmacological drug family (e.g. TNF inhibitors). Rheumatologists are still searching for reliable biomarkers or algorithms that may guide therapeutic choices [19,20]. The situation is more complicated, given the impact of additional variables (ethnic, geopolitical, and lifestyle) on the RA course and response to therapy [21,22]. The information conveyed by these variables can be leveraged at the clinical decision level. In general, the amount of data potentially useful for subtyping SARD is large and heterogeneous, and it suggests various dimensions for consideration in improving diagnosis and therapy. This implies that new analytical approaches should be employed to develop reliable algorithms.

COronaVIrus Disease 19 (COVID-19) is an acute respiratory disease caused by severe acute respiratory syndrome (SARS) coronavirus (CoV) 2 and characterized by clinical symptoms ranging from minor upper airway manifestations to severe and life-threatening acute respiratory distress syndrome (ARDS), which is observed in approximately 15% of patients. Multiorgan involvement (e.g. liver, kidney, central nervous system, etc.) has been described in the most severe cases as well (see [23], among others). The major unmet need in COVID-19 management is to identify patients at risk for severe organ involvement at the early stages of the disease. It is widely accepted that comorbidities, such as diabetes mellitus, cardiovascular disorders, and obesity, are risk factors for severe COVID-19. Moreover, male sex and age of > 70 years are known risk factors. There is growing evidence that ethnicity, socioeconomic status, and lifestyle (smoking, alcohol consumption, etc.) may be additional negative risk factors [24,25]. Nevertheless, the development of severe ARDS is still unpredictable since it can occur in patients younger than the reference risk groups (> 70 years or without the above-mentioned comorbidities) [26]. Recent studies have underlined the role of genetic markers as predictors of severe COVID-19, particularly severe ARDS [27]. Interestingly, the same genetic markers and non-O blood groups have been associated with complement activation and endothelial damage, some of the main pathogenic mechanisms underlying severe COVID-19 [28]. Altogether, these findings support the effect of genetic susceptibility. Combining several biomarkers in reliable algorithms may allow (a) further stratification of the patients according to their risk of negative outcomes, and (b) timely tuning of therapeutic interventions aimed at reducing pro-inflammatory responses thought to drive multiorgan damage [29]. Of interest is the use of the same tools for emerging trends. After the acute disease, some patients report complex symptoms that persist, such as chest heaviness, breathlessness, muscle pain, palpitations, and fatigue. These are defined as 'long COVID'. While researchers and clinicians agree on its existence, the exact definition of the disease is still debated because of its heterogeneity and the different variables that can affect its presentation [30–33].

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewers Disclosure

Peer reviewers on this manuscript have no relevant financial relationships or otherwise to disclose.

Funding

This paper was supported by grant NSF 19-500 DMS1918925/1922843.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Blonde L, Khunti K, Harris SB, et al. Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther. 2018;35(11):1763–1774.
2. Banerjee A, Chen S, Fatemifar G, et al. Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: systematic review of validity and clinical utility. BMC Med. 2012;202(19):85.
3. Wang W, Kiik M, Peek N, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS ONE. 2020;15(6):e0234722.
4. Planell-Morell P, Bajekal M, Denaxas S, et al. Trajectories of disease accumulation using electronic health records. Stud Health Technol Inform. 2020;270:469–473.
5. Garibaldi BT, Fiksel J, Muschelli J, et al. Patient trajectories among persons hospitalized for COVID-19: a cohort study. Ann Intern Med. 2021;174(1):33–41. Erratum in: Ann Intern Med. 2021; 174(1):144.
   - • **The COVID-19 Inpatient Risk Calculator (CIRC) is here applied to patients based on factors present on admission to inform clinical and resource allocation decisions.**
6. Jensen AB, Moseley PL, Oprea TI, et al., Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat Comm. 5(1): 4022. 2014.
   - •• **Several pathologies were included in a real whole population study.**
7. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. Npj Digit Med. 2019;2(1):14.
8. Babrak LM, Menetski J, Rebhan M, et al. Traditional and digital biomarkers: two worlds apart? Digit Biomark. 2019;3(2):92–102.
9. Zhang L, Guo H. Biomarkers of COVID-19 and technologies to combat SARS-CoV-2. Adv Biom Sci Technol. 2020;2:1–23.
10. Guthrie NL, Carpenter J, Edwards KL, et al. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. BMJ Open. 2019;9(7):1–9.

11. Dominietto MD, Capobianco E. Expected impacts of connected multimodal imaging in precision oncology. Front Pharmacol. 2016;7:451.
12. Amin S, Bathe OF. Response biomarkers: re-envisioning the approach to tailoring drug therapy for cancer. BMC Cancer. 2016;16(1):850.
13. Mankoff DA, Pryma DA, Clark AS. Molecular imaging biomarkers for oncology clinical trials. J Nucl Med. 2014;55(4):525–528.
14. Chennubhotla C, Clarke LP, Fedorov A, et al. An assessment of imaging informatics for precision medicine in cancer. Yearb Med Inform. 2017;26(1):110–119.
15. Allen ME, Rus V, Szeto GL. Leveraging heterogeneity in systemic lupus erythematosus for new therapies. Trends Mol Med. 2021;27 (2):152–171.
16. To CH, Petri M. Is antibody clustering predictive of clinical subsets and damage in systemic lupus erythematosus? Arthritis Rheum. 2005;52(12):4003–4010.
17. Bossuyt X, De Langhe E, Borghi MO, et al. Understanding and interpreting antinuclear antibody tests in systemic rheumatic diseases. Nat Rev Rheumatol. 2020;16(12):715–726.
    - **A review paper that stresses the need of improving the predictivity/specificity of the most common diagnostic tests for systemic autoimmune disorders.**
18. Nagafuchi Y, Shoda H, Fujio K. Immune profiling and precision medicine in systemic lupus erythematosus. Cells. 2019;8(2):140.
19. Smolen JS, Aletaha D. Rheumatoid arthritis therapy reappraisal: strategies, opportunities and challenges. Nat Rev Rheumatol. 2015;11(5):276–289.
20. Bartels CM, Rosenthal A, Wang X, et al. Investigating lupus retention in care to inform interventions for disparities reduction: an observational cohort study. Arthritis Res Ther. 2020;22(1):35.
21. Barton JL, Trupin L, Schillinger D, et al. Racial and ethnic disparities in disease activity and function among persons with rheumatoid arthritis from university-affiliated clinics. Arthritis Care Res (Hoboken). 2011;63(9):1238–1246.
22. Zou Y, Cheung PP, Teoh LK, et al., Sociodemographic factors as determinants of disease, disability and quality of life trajectories in early rheumatoid arthritis: a multi-ethnic inception cohort study. Int J Rheum Dis. 23(1): 55–64. 2020.
    - **Significance of sociodemographic variables that may affect both course and presentation of RA.**
23. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA. 2020;323 (20):2052–2059. Erratum in: JAMA. 2020;323(20):2098.
24. Lowe KE, Zein J, Hatipoglu U, et al. Association of smoking and cumulative pack-year exposure with COVID-19 outcomes in the cleveland clinic COVID-19 registry. JAMA Intern Med. 2021; e208360. DOI:10.1001/jamainternmed.2020.8360.
25. Price-Haywood EG, Burton J, Fort D, et al. Hospitalization and mortality among black patients and white patients with Covid-19. N Engl J Med. 2020;382(26):2534–2543.
    - **The role of ethnicity in COVID-19 clinical course.**
26. Cunningham JW, Vaduganathan M, Claggett BL, et al. Clinical outcomes in young US adults hospitalized with COVID-19. JAMA Intern Med. 2020:e205313. DOI:10.1001/jamainternmed.2020.5313.
27. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, et al. Genome-wide association study of severe Covid-19 with respiratory failure. N Engl J Med. 2020; 383 (16):1522–1534.
    - **The clinical course of COVID-19 is influenced by genetics.**
28. Valenti L, Griffini S, Lamorte G, et al. Chromosome 3 cluster rs11385942 variant links complement activation with severe COVID-19. J Autoimmun. 2021;117:102595.
29. Alunno A, Najm A, Machado PM, et al. EULAR points to consider on pathophysiology and use of immunomodulatory therapies in COVID-19. Ann Rheum Dis. 2021;0:1-9.
30. The Lancet editorials. Facing up to long COVID. Lancet. 2020;396 (10266):1861.
31. Nature editorials. Let patients help define long-lasting COVID symptoms. Nature. 2020;586(7828):170.
32. Cortinovis M, Perico N, Remuzzi G. Long-term follow-up of recovered patients with COVID-19. Lancet. 2021;397(10270):173–175.
33. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet. 2021 Jan 16;397(10270):220–232.
    - **One of the first studies addressing the COVID-19 consequences issue.**