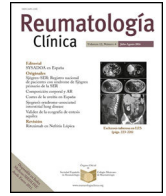




Sociedad Española
de Reumatología -
Colegio Mexicano
de Reumatología

Reumatología Clínica

www.reumatologiaclinica.org



Special Article

Biomarkers: how to consolidate them in clinical practice

María Jesús García de Yébenes Prous,* Loreto Carmona Ortells

Instituto de Salud Musculoesquelética (InMusc), Madrid, Spain



ARTICLE INFO

Article history:

Received 14 February 2024

Accepted 20 May 2024

Available online 14 July 2024

Keywords:

Biomarkers

Personalized medicine

Validation

Implementation

ABSTRACT

An inadequate biomarker validation can affect many patients' diagnosis, treatment, and follow-up. Therefore, special interest should be placed on performing these analyses correctly so that biomarkers can be applicable to patients and evidence of their clinical usefulness can be generated. A methodological work on the concept of biomarkers is presented, as well as the difficulties associated with the methodological approach to their development, validation, and implementation in clinical practice.

© 2024 Elsevier España, S.L.U. and Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Biomarcadores: cómo lograr su consolidación en práctica clínica

RESUMEN

Una validación inadecuada de un biomarcador puede tener consecuencias importantes en el diagnóstico, tratamiento y seguimiento de muchos pacientes, por lo que debe ponerse especial interés en realizar estos análisis de forma correcta para que los biomarcadores sean aplicables en los pacientes y pueda generarse evidencia sobre su utilidad clínica. Se presenta un trabajo metodológico sobre el concepto de biomarcadores, así como las dificultades asociadas al abordaje metodológico para su desarrollo, validación e implementación en práctica clínica.

© 2024 Elsevier España, S.L.U. y Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. Se reservan todos los derechos, incluidos los de minería de texto y datos, entrenamiento de IA y tecnologías similares.

Palabras clave:

Biomarcador

Medicina personalizada

Validación

Implementación

Introduction

Personalised, or precision, medicine is defined as the "emerging approach to disease prevention, diagnosis, and treatment that takes into account individual, environmental and lifestyle variability."¹ This new approach has been made possible thanks to the inroads made in several techniques, including molecular biology techniques and, in particular, the boom of -omic techniques (genomics, proteomics, metabolomics). The main feature of these techniques is that they are capable of producing a massive amount of data, which has led to greater insight into the genetic and biological determinants of disease and has significantly accelerated the discovery of biomarkers.

Biomarkers are "objectively measurable characteristics that are indicators of normal or pathological biological processes and of treatment response."² In other words, they are measurable or quantifiable parameters that generally belong to three distinct molecular categories: proteins, metabolites, and nucleic acids or genes. Similarly, there are imaging, physical, and mechanical biomarkers, and even psychological or behavioural biomarkers, among others. However, to understand the concept better, this paper will focus on biological biomarkers and all the information that can, in turn, be extrapolated to other types of biomarkers.

Because they reflect biological processes, biomarkers can be extremely useful when making decisions concerning diagnosis, treatment, disease activity, and prognosis, which is why they are sometimes categorised in terms of their clinical applications (Fig. 1):

- **Susceptibility or predisposition** biomarkers: indicate the likelihood of developing the disease,

* Corresponding author.

E-mail address: mjgarciaeyebenes@inmusc.eu (M.J. García de Yébenes Prous).

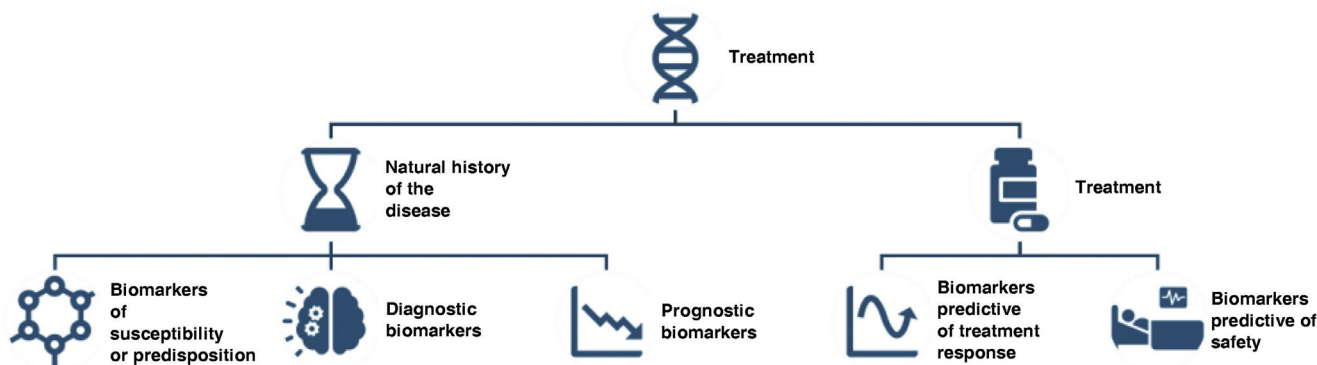


Figure 1. Types of biomarkers in health.

- **Diagnostic** biomarkers: enable the diagnosis of a specific disease to be confirmed,
- **Prognostic** biomarkers: provide information about the course of the disease (relapse or recurrence), and
- **Predictive** biomarkers of treatment response or safety.

Of course, biomarkers can also be used in research as surrogate endpoints or surrogate outcome measures.³ However, simply because they are useful in research does not necessarily translate into clinical practice.

While the same biomarker may be used for several of these applications, evidence of its value for each must be obtained.⁴ Furthermore, despite the fact that there may appear to be a certain degree of overlap regarding the definitions of the types of biomarkers based on their application, each one has its own distinctive characteristics.

Biomarker validation refers to proving the association between a biomarker and a clinical outcome using robust statistical methods. This association is independent of treatment (diagnostic or prognostic biomarkers), of the prediction of the effect of a treatment on a clinical surrogate endpoint (predictive), or of the possibility of replacing a clinical endpoint to assess treatment effects (surrogate measures).⁵ A validated biomarker can assist in targeted treatment, improve clinical diagnosis, and act as a prognostic or predictive factor for a given outcome. Hence, both analytical and clinical validation studies are essential. Inadequate validation of a biomarker can have major consequences for the diagnosis, treatment, and follow-up of many patients; consequently, special care must be exercised when performing these analyses properly so that biomarkers will be useful with respect to patients and that evidence of their clinical value can be generated.

Process of biomarker implementation in clinical practice

Adopting a biomarker in clinical practice is a complex process and entails a number of steps to guarantee that the results are both safe and reliable.^{3,4} Generally speaking, this process consists of the following stages (Fig. 2): (1) discovery, (2) development, (3) validation, and (4) proof of clinical usefulness.

Discovery

The first phase consists of identifying the biomarkers and defining them as genetic expression, level of protein, and type of tissue or fluid in which they will be measured. During this phase, the type of biomarker being looked for must be decided: diagnostic, prognostic, or predictive, and the type of biological sample in which it makes sense to measure it must be identified. It is during this step that we avail ourselves of the tools of genomics, transcriptomics,

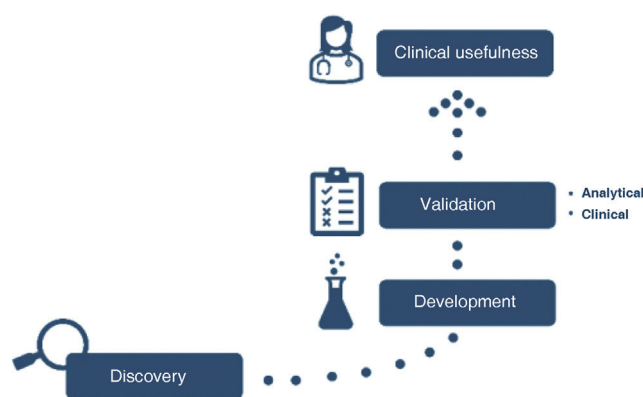


Figure 2. Process of implementation of a new biomarker.

proteomics, and metabolomics to substantiate that the biomarker and the type of sample chosen represent the physiological process we seek to quantify.

It is important for there to be a consensus in that the biomarker and type of sample chosen truly represent the physiological process we seek to quantify to ascertain whether this phase has been successfully completed.

Development

The aim of this phase is to define the experimental procedure to measure the molecule of interest and optimise the analytical method by which to obtain reliable results. These are sometimes time-consuming studies [needed] to hone [such procedures and methods]. On occasion, the importance of this phase is underestimated, given how costly the technique is, and [consequently] this stage is sometimes skipped.

Validation

The value of biomarkers in patient diagnosis, evaluation, and prognosis depends on the prior demonstration of the validity of its association with an illness in particular or a specific manifestation of that disease. Validation is the process that makes it possible to establish that a test performance is acceptable for its planned purpose. Internal validation examines the performance of the test on the development sample, whereas external validation probes how well it does so in an independent sample. Moreover, in the case of biomarkers, two aspects of validation must be differentiated: analytical and clinical validation.⁶

Analytical validation is the accuracy with which the biomarker identifies the result of interest (specific gene or protein) and the

Table 1
Measures of analytical validation.

Measure	Definition
Accuracy	How close the measured value is to the actual value (concentration)
Precision	Closeness between individual concentrations of repeated measurements
Sensitivity	The lowest concentration that can be accurately and precisely measured
Reproducibility	Precision of the measurement in different conditions (days, observers)
Stability	Degradation of the biomarker from the time it is extracted and until it is analysed

biological process under study. The objective is to define the biomarker's technical characteristics and not its usefulness; in other words, to establish its accuracy, precision, sensitivity, reproducibility, and stability that guarantee a measure that is consistent with the actual unknown values (Table 1).^{6,7}

This kind of validation can be impacted by various factors⁷ and must also include the determination of the range and reproducibility of detection. The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have issued methodological guidelines with respect to these processes.^{8,9}

Clinical validation seeks to demonstrate the association between the biomarker and the result of interest, in addition to its usefulness. In other words, attempts to prove if the biomarker accurately and reliably identifies a clinically or biologically defined disorder, and whether or not it is capable of discriminating between groups having different clinical or biological characteristics. It is based on the external validation and can be performed by means of various designs depending on the type of biomarker, objectives, type of sample, size, etc.^{6,10}

The main parameters of clinical validity are sensitivity (proportion of cases that test positive) and specificity (proportion of «non-cases» that test negative). Other important parameters are reproducibility, or the capacity a biomarker has to yield the same results in similar conditions, and predictive values or test performance measure in different contexts.¹⁰ The biomarker must prove to provide accurate and reliable performance in comparison with a standard reference. To do so, receiver operating characteristics (ROC) curves and their area under the curve (AUC) or discriminatory capacity between cases and non-cases are used, which must be as close to 1 as possible.⁶

Clinical usefulness

Biomarkers must be suitably validated prior to being used in clinical practice. The evaluation of clinical usefulness assesses the biomarker's capacity to provide useful information regarding diagnosis, treatment, or prevention of a given disease.¹¹ It attempts to confirm whether the results of the biomarker modify patient management; that is to say, if they are helpful to guide clinical decisions and achieve better outcomes compared with those that would be obtained if the biomarkers were not applied. To evaluate the feasibility of using the biomarker in clinical practice, the scenarios in which the biomarker improves patient outcomes must be identified. The best context in which to ascertain this is to include the biomarker to stratify the results of a diagnostic test of clinical drug trial.⁶

Despite the tremendous headway made in molecular biology and the copious amount of candidate biomarkers, the truth is that scant requirements are applied for their use. For the most part,

Table 2
Methodological approach for biomarker development and implementation.

Phase	Aspects to be considered
Clinical need	Identify an unmet need Target population, usual practice Current solutions Expected results Barriers
Analytical validation	Sensitivity and specificity Range of determination Linearity Reliability and precision Stability, interferences Consistency
Pre-analytic factors	Sample collection Processing Transportation and storage
Biological variation and clinical factors	Circadian and daily variability Age, sex, and weight Pregnancy Medications, major surgery, and other conditions
Interpretability	Distribution among healthy individuals Distribution in the target population Definition of normality-risk
Diagnostic and prognostic performance or usefulness	Patient population and selection Determination of index test and standard reference Data collection and analysis Calculation of validity parameters
Post-analytic factors	Delivery deadlines Presentation of results Quality control
Clinical and health outcomes	Morbidity and mortality PROM, quality of life Greater precision in risk evaluation Faster diagnosis Simplification of processes Costs

Source: Kolev et al.¹²

PROM: Patient Reported Outcome Measures.

demonstration of analytical sensitivity and accuracy in the limited number of samples are required. Nevertheless, demonstration of its diagnostic performance in an adequate clinical validation study is not currently mandatory and some authors amidst that many biomarkers are implemented too soon and lacking proper prior evaluation. The lack of rigorous validation studies can be due to a number of reasons, including the following: researchers and clinicians not knowing the suitable methodology; disparate methodological development that is not enforced by existing regulatory frameworks; regulations focused on analytical validation; lack of analysis of pre-analytic issue, and biological variation. Finally, there is a paucity of clinical validation studies, probably owing to the specific methodological requirements that entails designing and analysing this kind of study and the difficulty in finding appropriate gold standards.^{7,12}

Theoretical framework for the validation of biomarkers

Despite the fact that, for the time being, there is no common regulatory framework for the evaluation of biomarkers, some methodological proposals have been put forth to facilitate its development and implementation.^{7,12} It must be remembered that the adoption of a biomarker is a cyclical process that begins with acknowledgement of an unmet need, followed by extensive process of evaluation and approval.

Table 3
Difficulties in the validation of biomarkers and possible approaches.

Problem	Possible origin	Methodological approach
Correlated observations	Multiple observations per subject Multiple lesions per subject	Mixed linear models (structure of correlation)
Multiplicity	Multiple biomarkers or outcomes	Methods that control for false positive rate
Multiple outcomes	Interest in more than one outcome	Compound measures Prioritization of outcomes
Selection bias	Retrospective data or observational studies	Multivariate models Paired samples Propensity index

Source: Ensor.¹³

Table 2 displays the aspects to be contemplated during the various phases as a methodological approach to the development and implementation of a biomarker.¹²

Methodological considerations in the validation of biomarkers

Biomarker validation poses key methodological challenges. There are currently no regulations that have been defined regarding the evaluation and adoption of biomarkers and surrogate endpoints in the absence of robust validation data. One of the leading stumbling blocks to validation is the lack of high-quality biological samples and standardised measures of response in clinical trials.

In addition, -omic techniques make it possible to measure multiple variables (genes, single nucleotide polymorphism [SNP]) simultaneously, which means that the biomarker may comprise, not a single gene, but, for instance, a microarray of 70 genes. This multiplies the difficulty in validating biomarkers, inasmuch as the usual statistical methodology is designed for a single variable (a hypothesis for each biomarker) and not for a number of variables (one hypothesis for a set of X genes).

Therefore, validating biomarkers presents a series of differential characteristics that must be taken into account when performing statistical analyses^{13,14}; these features can be seen below (Table 3).

Intra-individual correlation

By and large, biomarker studies have multiple observations of a single parameter; for instance, tumour marker measurements of different pathology samples from the same patient or repeated measure of several pathology samples of the same marker over time (follow-up studies). In these cases, the observations (measurements of a biomarker) are not independent, but rather, correlated.

The analysis of correlated observations by means of classical statistical methods (constituted for independent observations) increases the risk of type I errors and, as a result, the likelihood of obtaining spurious associations or false positives. Consequently, methods must be used that take into account intra-individual correlations (the structure of variance-covariance), such as mixed linear models. By way of illustration, comparisons based on generalised estimation equations (GEE) adjust the dependence of the observations, thereby making more realistic p values and confidence intervals to be reached.

Multiplicity

Biomarker validation studies have several sources of multiplicity:

- Multiple markers: gene combinations, SNP.
- Multiple ways to quantify a biomarker.
- Multiple cutoffs (continuous biomarkers).

The presence of multiplicity requires that statistical procedures be used to control false positives.

In the statistical analyses, we establish a level of significance (p value) to decide if the results observed are random. The contrast of hypotheses calculates the probability of observing an association between two variables; the hypothesis of no association is rejected if the value of probability is less than a pre-set one (typically, 0.05).

The null hypothesis of validation is that the biomarker in question has no effect on the prognosis of the disease, treatment response, or the biological process. The use of more than one contrast hypothesis increases the probability of finding a statistically significant result. In other words, the more contrasts we performed, the greater the likelihood we will have of finding some kind of association (greater probability of false positives). In statistics, this is known as «multiple comparisons».

Multiple comparisons increase the probability of spurious associations (false positives); which is why the rate of type I errors must be adjusted. Two important concepts in this type of analysis are the family wise error rate (FWER) and the rate of false discovery (FDR). FWER is defined as the probability that there will be at least one false positive, while FDR is the expected proportion of type I errors or probability that a null hypothesis that has been rejected will, in fact, be true.

There are several methods by which to adjust for the rate of type I errors. The choice of one or another of several factors, such as the exploratory or confirmatory nature of the analysis, or the expected number of contrasts. Decision trees¹⁵ and even on-line calculators for multiple comparisons have been elaborated to choose the appropriate method.¹⁶

Conventional adjustment methods control the rate of false positives or FWER and are based on rejecting the null hypothesis with lower p values. One of the most widely used is Bonferroni's method that consists of dividing the usual level of significance (α) by the number of comparisons made (n), rejecting the null hypotheses whose p values are less than the α/n coefficient instead of α . By lowering the level of significance, the probability of false positives is also lower, however, the statistical power to detect real differences is also reduced, which limits its applicability.^{15,17}

FDR adjustment methods have become a usual alternative to Bonferroni's correction, particularly in genomic studies. The aim of controlling for FDR is, from among the tests declared as significant, to manage to achieve a proportion of true null hypotheses to be less than a specified value. The Benjamin-Hochberg procedure controls FDR by using the q-value, the FDR analogue to the p value. The q-value is defined as the expected proportion of false positives among all the tests that are equal to or more extreme than the one observed. A p value of 5% means that 5% of all the tests will be false positives; a q value of 5% means that 5% of the significant results will be false positives.^{15,17}

These methods are based on different suppositions and have different levels of complexity and power, but all point toward balancing the compensation between type I and type II errors.

Multiple outcome measures

The use of different outcome measures is also a source of multiplicity.

In the literature, it is not uncommon to find studies in which various outcome measures are used simultaneously without correcting for multiplicity. For example, outcomes measures such as overall survival, progression-free survival, remission rates, and the rate of stable disease may all be used as outcomes measures in the

same study. This issue of multiplicity becomes even greater when the definition of «time to event» intervals varies depending on the date chosen as the onset of exposure (date of diagnosis, date of treatment initiation, date of surgery, etc.).

In these cases, it is difficult to reach a consensus as to the validity of a biomarker or regarding what decision to make when the findings of different outcome measures are contradictory. Several approaches have been proposed, including the following:

- Use an outcome measure as «fundamental», adjust the analysis for multiple comparisons, and consider the remaining outcomes as biologically related.
- Combine different outcome measures in a single outcome to evaluate.
- Prioritize outcomes and carry out the analysis of the first one. If the results are significant, regard the analysis as finalized; if the results are not significant, move on to the second outcome, and so forth.

It is essential that no more contrasts of hypothesis be performed than necessary. One outcome must be chosen, and the results obtained be accepted in order to achieve reproducibility and data coherence.

Selection biases

Retrospective designs are often used in molecular epidemiology for reasons of feasibility, given that retrospective data are generally more easily accessible and make it easier to conduct «time to event» analyses. Be that as it may, retrospective designs entail limitations that must be taken into account.

On the other hand, the different levels of biomarkers (e.g., normal versus overexpression) may not be homogenous compared to other possible study outcome predictors. The fact that clinical efficacy predictors are not balanced across the various levels of a biomarker is not a selection bias in the strictest sense of the term, although it can confound associations. The consequence is a problem of inference and limitation on the part of the researchers to establish the prognostic and predictive value of a given biomarker.

In consequence, validation of these biomarkers must address these issues by means of the use of statistical methods that make it possible to control confusion, such as:

- Multivariate models to control for the different confounding factors.
- Paired samples for those factors that can affect results.
- Propensity index to simulate the conditions of a controlled clinical trial, the gold standard to estimate an effect.

Types of designs

In addition to the afore commented methodological challenges, the fact that there may be differences in validation study designs depending on the type of biomarker that is the object of analysis must also be kept in mind.

In the case of prognostic biomarkers, the aim is to demonstrate an association between the biomarker, or its change over time, and the appearance of a clinical outcome irrespective of treatment. Initially, retrospective studies can yield enough data for this analysis. Nevertheless, the validation must be carried out in a number of centres or using re-sampling techniques if only a single centre is available. Once the multicentre validation is confirmed, the ultimate proof of a biomarker's clinical usefulness must be attained through prospective, randomised clinical trials. Prognostic biomarkers are fairly easy to identify, but only rarely are multicentre validations conducted.⁵

Predictive biomarkers are those in which the baseline value or the changes over time predict the efficacy or toxicity of a given therapy. The statistical analysis of these biomarkers demand randomised, controlled clinical trials, ideally having an interaction study design to determine whether the status of the biomarker modifies the association between treatment and the observed effect.

One alternative is to use the so-called «selection» designs that only include patients who are positive for the biomarker and can therefore confirm the biomarker's usefulness to identify those populations in which the therapy can be efficacious. However, these designs cannot actually establish whether the biomarker is predictive or not, inasmuch as there is no information available regarding the lack of efficacy in individuals who are negative for the biomarker in question.

In light of the difficulty inherent in performing randomised studies to validate predictive biomarkers, oftentimes retrospective validation is conducted by evaluating previous clinical trials to gauge whether the treatment has a differential effect based on the presence of a biomarker. As a result, validation of a predictive biomarker is costly and calls for randomised trials and meta-analyses.⁵

Conclusion

The tremendous inroads made in the field of biomarkers facilitates making decisions concerning the diagnosis, treatment, and prognosis of the disease and levels the path toward personalised, or precision, medicine. Nevertheless, the development process and implementation of biomarkers is complex and requires a number of steps to guarantee the safety and reliability of the results. Moreover, diagnostic performance studies must be an essential requirement for the implementation of a biomarker. These studies display distinct characteristics and methodological requirements that must be considered when properly undertaking the validation of biomarkers and being able to guarantee their usefulness in clinical practice.

Funding

This work has not received any funding whatsoever.

Conflict of interests

The authors have no conflict of interests to declare.

References

1. Mestre-Ferrandiz J., Nuño-Solinis R. Del Llano Nuñez-Cortes A. Del Llano Señaris F J.E. Los biomarcadores como motor de la Medicina de Precision en Oncología 2023. 1 ed. Madrid, Fundacion Gaspar Casal. [accessed 11 Dic 2023]. Available from: <https://fundaciongasparcasal.org/wp-content/uploads/2023/03/Los.biomarcadores.2023.web.2as.pdf>.
2. Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV AIDS*. 2010;5(6):463–6.
3. Gimenez P, Peel C. Biomarcadores en la práctica clínica. Validación y verificación. *Dianas*. 2014;3:e20140914.
4. Aronson JK, Ferner RE. Biomarkers—A general review. *Curr Protoc Pharmacol*. 2017;76, 9.23.1–17.
5. Buysse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points—the challenge of statistical validation. *Nat Rev Clin Oncol*. 2010;7(6):309–17.
6. Ou FS, Michiels S, Shyr Y, Adjei AA, Oberg AL. Biomarker discovery and validation: statistical considerations. *J Thorac Oncol*. 2021;16(4):537–45.
7. Kraus VB. Biomarkers as drug development tools: discovery, validation, qualification and use. *Nat Rev Rheumatol*. 2018;14(6):354–62.
8. European Medicines Agency. Qualification of novel methodologies for drug development: guidance to applicants 2014. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf. Fecha de acceso: Febrero 2024.
9. Food and Drug Administration. Bionalytical method validation. Guidance for industry 2018. Available from: <https://www.fda.gov/files/drugs/>

- [published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf](#). Fecha de acceso: Febrero 2024.
10. Hayes DF. Biomarker validation and testing. *Mol Oncol*. 2015;9(5):960–6.
 11. Navanandan N, Searns J, Ambroggio L. Methodology of phases of biomarker discovery. *Hosp Pediatr*. 2023;13(7):e181–5.
 12. Kolev M, Horn MP, Semmo N, Nagler M. Rational development and application of biomarkers in the field of autoimmunity: a conceptual framework guiding clinicians and researchers. *J Transl Autoimmun*. 2022;5:100151.
 13. Ensor JE. Biomarker validation: common data analysis concerns. *Oncologist*. 2014;19(8):886–91.
 14. Sargent DJ, Mandrekar SJ. Statistical issues in the validation of prognostic, predictive, and surrogate biomarkers. *Clin Trials*. 2013;10(5):647–52.
 15. Menyhart O, Weltz B, Györfy B. MultipleTesting.com: a tool for life science researchers for multiple hypothesis testing correction. *PLoS One*. 2021;16(6):e0245824.
 16. Multiple testing correction. A tool for life science researchers for multiple hypothesis testing correction. Available from: www.multipletesting.com. [accessed Abr 2024].
 17. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol*. 2014;67(8):850–7.