International Standard

ISO 9491-1

Biotechnology — Predictive computational models in personalized medicine research —

Part 1: Constructing, verifying and validating models

Biotechnologie — Modèles informatiques prédictifs dans la recherche sur la médecine personnalisée —

Partie 1: Construction, vérification et validation des modèles

Second edition
2026-06

Reference number
ISO 9491-1:2026(en)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 276, *Biotechnology*.

This second edition cancels and replaces the first edition (ISO/TS 9491-1:2023), which has been technically revised.

The main changes are as follows:

— normative references in Clause 2 have been consolidated, updated and revised;

— update and clarification of terminology including the alignment with the terminology of ISO/TS 9491-2;

— updated to match the latest developments in the domain;

— bibliography has been revised and updated;

— editorial revision and clarification of wording.

A list of all parts in the ISO 9491 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

The capacity to generate data in life sciences and health research has greatly increased in the last decade. In combination with patient/personal-derived data, such as electronic health records, patient registries and databases, as well as lifestyle information, this big data holds an immense potential for clinical applications, especially for computer-based models with predictive capacities in personalized medicine. However, and despite the ever-progressing technological advances in producing data, the exploitation of big data to generate new knowledge for medical benefits, while guaranteeing data privacy and security, is lacking behind its full potential. A reason for this obstacle is the inherent heterogeneity of big data and the lack of broadly accepted standards allowing interoperable integration of heterogeneous health data to perform analysis and interpretation for predictive modelling approaches in health research, such as personalized medicine.

Common standards lead to a mutual understanding and improve information exchange within and across research communities and are indispensable for collaborative work. In order to setup computer models in personalized medicine, data integration from heterogeneous and different sources at different times plays a key role. Consistent documentation of data, models and simulation results based on basic guiding principles for data management practices, such as FAIR (findable, accessible, interoperable, reusable)[6] or ALCOA (attributable, legible, contemporaneous, original, accurate), and standards can ensure that the data and the corresponding metadata (data describing the data and its context), as well as the models, methods and visualizations, are of reliable high quality.

Hence, standards for biomedical and clinical data, simulation models and data exchange are a prerequisite for reliable integration of health-related data[7]. Such standards, together with harmonized ways to describe their metadata, ensure the interoperability of tools used for data integration and modelling, as well as the reproducibility of the simulation results. In this sense, modelling standards are agreed ways of consistently structuring, describing, and associating models and data, their respective parts and their graphical visualization, as well as the information about applied methods and the outcome of model simulations. Such standards also assist in describing how constituent parts interact, or are linked together, and how they are embedded in their physiological context.

Major challenges in the field of personalized medicine are to:

a) harmonize the standardization efforts that refer to different data types, approaches and technologies;

b) make the standards interoperable, so that the data can be compared and integrated into models.

An overall goal is to FAIRify data and processes in order to improve data integration and reuse. An additional challenge is to ensure a legal and ethical framework enabling interoperability.

This document presents computational modelling requirements and recommendations for research in the field of personalized medicine, especially with focus on collaborative research, such that health-related data can be optimally used for translational research and personalized medicine worldwide. The recommendations are primarily oriented towards the application of computational modelling in the biotechnology domain (e.g. biomolecular and cellular research, as well as in clinical trials and drug development), but also can be applied in other fields of personalized medicine research.

Biotechnology — Predictive computational models in personalized medicine research —

Part 1: Constructing, verifying and validating models

1 Scope

This document specifies requirements and recommendations for the design, development and implementation of predictive computational models for research purposes in the field of personalized medicine and health product development.

This document addresses the set-up, formatting, validation, simulation, storing and sharing of computational models used for personalized medicine. Requirements and recommendations for data used to construct or required for validating such models are also specified. This includes rules for formatting, descriptions, annotations, interoperability, integration, access and provenance of such data.

This document does not apply to computational models used for standard routine clinical, diagnostic or therapeutic purposes.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 20691, *Biotechnology — Requirements for data formatting and description in the life sciences*[1)]

ISO 20387:2026, *Biotechnology — Biobanking — General requirements for biobanks*[2)]

ISO 23494-1, *Biotechnology — Provenance information model for biological material and data — Part 1: Design concepts and general requirements*

ISO 23494-2, *Biotechnology — Provenance information model for biological material and data — Part 2: Common Provenance Model*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

1) https://fairsharing.org/3533

2) Under preparation. Stage at the time of publication: ISO/FDIS 20387:2026.

3.1
artificial intelligence
AI
<discipline> research and development of mechanisms and applications of *AI systems* (3.2)

Note 1 to entry: Research and development can take place across any number of fields such as computer science, data science, humanities, mathematics and natural sciences.

[SOURCE: ISO/IEC 22989:2022, 3.1.3]

3.2
artificial intelligence system
AI system
engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives

Note 1 to entry: The engineered system can use various techniques and approaches related to artificial intelligence to develop a model to represent data, knowledge, processes, etc. which can be used to conduct tasks.

Note 2 to entry: AI systems are designed to operate with varying levels of automation.

[SOURCE: ISO/IEC 22989:2022, 3.1.4]

3.3
big data
extensive datasets — primarily in the data characteristics of volume, variety, velocity, and/or variability — that require a scalable technology for efficient storage, manipulation, management, and analysis

Note 1 to entry: Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big data extensive datasets.

EXAMPLE High volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points (see reference [8] for additional information).

[SOURCE: ISO/TR 24291:2021, 3.2, modified — EXAMPLE added.]

3.4
community consensus standard
standard that reflects the results of a consensus standardization effort from a specific domain-specific expert group outside of recognized standard defining organizations and their technical committees

Note 1 to entry: Created by domain-specific professional societies, scientific standardization initiatives, individual organizations or research communities (often in collaboration with industry partners)

Note 2 to entry: Often publicly available, open and not proprietary

3.5
computational model
in silico model
description of a biological system in either a mathematical expression or graphical form, or both, that is implemented and studied with a computer highlighting objects and their interactions

Note 1 to entry: An object distributed processing (ODP) concept.

[SOURCE: ISO/IEC 16500-8:1999, 3.6, modified — Admitted term added. “biological”, “mathematical expression or”, “, or both, that is implemented and studied with a computer” added, “interfaces” changed to “interactions” and “as such it is similar to the OMT and UML notion of a class diagram” deleted from the definition. “An object distributed processing (ODP) concept” moved to Note 1 to entry.]

3.6
data-driven model
model developed through the use of data derived from tests or from the output of investigated process or from real world data or routinely acquired primary care data

[SOURCE: ISO 15746-1:2015, 2.4, modified — "or from real world data or routinely acquired primary care data" added]

3.7
harmonization of data concepts
data harmonization
process of reconciling differences in semantics, structure and syntax of similar data concepts

Note 1 to entry: Harmonization can include the establishment of a single pervasive definition for each data concept (i.e. standardization), but can also encompass flexible approaches in which definitions can be understood to grow closer without becoming identical.

[SOURCE: ISO/TR 25100:2012, 2.1.4, modified — "harmonisation" replaced by "harmonization", "may" in Note 1 to entry replaced by "can".]

3.8
data integration
systematic combining of data from different independent and potentially heterogeneous sources, to create a more compatible, unified view of these data for research purpose

[SOURCE: ISO 5127:2017, 3.1.11.24]

3.9
genome-wide association studies
GWAS
testing of genetic variants across the genomes of many individuals to identify genotype–phenotype associations

3.10
in silico clinical trial
use of computer modelling and simulation(s) to mimic human experimentation in the development or regulatory evaluation process of a medicinal product (e.g. medical device) or medical intervention, under defined conditions using verified and validated models

Note 1 to entry: It is a subdomain of 'in silico medicine', the discipline that encompasses the use of individualised computer simulations in all aspects of the prevention, diagnosis, prognostic assessment, and treatment of disease.

[SOURCE: Reference [9], modified — Note 1 to entry added.]

3.11
in silico approach
computer-executable analyses of *mathematical model(s)* (3.13) to study and simulate a biological system

3.12
machine learning
ML
computer technology with the ability to automatically learn and improve from experience without being explicitly programmed

EXAMPLE Speech recognition, predictive text, spam detection, or optimizing model parameters through computational techniques, such that the model's behaviour reflects the data or experience.

[SOURCE: ISO 20252:2019, 3.52, modified — Abbreviated term "ML" added and EXAMPLES changed to "Speech recognition, predictive text, spam detection, or optimizing model parameters through computational techniques, such that the model's behaviour reflects the data or experience.".]

3.13
mathematical model
set of equations that describes the behaviour of a physical system

[SOURCE: ISO 16730-1:2015, 3.11]

3.14
mechanism-based
approach in computational modelling that aims for a structural representation

3.15
model validation
comparison between the output of the calibrated model and the measured data, independent of the data set used for calibration

[SOURCE: ISO 14837-1:2005, 3.7]

3.16
model verification
confirmation that the mathematical elements of the model behave as intended

[SOURCE: ISO 14837-1:2005, 3.8]

3.17
molecular biomarker
biomarker
molecular marker
detectable and/or quantifiable molecule or group of molecules used to indicate a biological condition, state, identity or characteristic of an organism (e.g. an individual)

EXAMPLE Nucleic acid sequences, proteins, small molecules such as metabolites, other molecules such as lipids and polysaccharides.

[SOURCE: ISO 16577:2022, 3.4.28, modified — "or an" changed to "of an" and "(e.g. an individual)" added to definition.]

3.18
personalized medicine
precision medicine
medical model using characterization of individuals' phenotypes and genotypes for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention

Note 1 to entry: Examples for individuals' phenotypes and genotypes are molecular profiling, medical imaging and lifestyle data.

Note 2 to entry: Medical decisions, prevention strategies and therapies in personalized medicine are based on this individuality.

[SOURCE: EU 2015/C 421/03,[10] modified — Notes 1 and 2 to entry added and "(e.g. molecular profiling, medical imaging, lifestyle data)" deleted from definition.]

3.19
phenotype
set of observable characteristics of an organism resulting from the interaction of its genotype with the environment

[SOURCE: ISO 4454:2022, 3.14, modified — Note 1 to entry deleted.]

3.20
raw data
data in its originally acquired, direct form from its source before subsequent processing

[SOURCE: ISO 5127:2017, 3.1.10.04]

4 Principles

4.1 General

Research in the field of personalized medicine is highly dependent on the exchange of data from different sources, as well as harmonized integrative analysis of large-scale personalized medicine data (big data in health research). Computational modelling approaches play a key role for understanding, simulating and predicting the molecular processes and pathways that characterize human biology. Modelling approaches in biomedical research also lead to a more profound understanding of the mechanisms and factors that drive diseases, and consequently allow for adapting personalized treatment strategies that are guided by central clinical questions. Patients can greatly benefit from this development in research that equips personalized medicine with predictive capabilities to simulate in silico clinically relevant questions, such as the effect of therapies, the response to drug treatments or the progression of disease.

4.2 Computational models in personalized medicine

4.2.1 General

Computational models have the potential to translate in vitro, non-clinical and clinical results (and their related uncertainty) into descriptive or predictive expressions. The added value of such models in medicine and pharmacology has increasingly been recognized by the scientific community,[11][12][13][14] as well as by regulatory bodies such as the European Medicines Agency (e.g. EMA guideline on PBPK reporting[15]), or the US Food and Drug Administration (FDA).[16][17] Computational models are integrated in different fields in medicine as well as in the development of drugs and other health products, expanding from disease modelling, molecular and physiological biomarker research to assessment of drug and medical device efficacy and safety. In silico approaches are also expanding in neighbouring fields, such as pharmacoeconomics,[18][19] analytical chemistry[20][21] and biology that are out of scope of this document[22][23].

Model creation starts with a clinical question and the collection of data (see Figure 1). The data employed need harmonized approaches for data integration to start the model construction. The initial model usually undergoes several refinement and improvement iterations to enhance predictive capabilities. Common standards (see 4.3.3) should be used for the model building and curation process. Accuracy measurements and validation processes are key, and should be transparent, while model output and function should ideally be interpretable or explainable.

A number of computational modelling approaches in pre-clinical and clinical research already address these questions in detail (see 4.2.2 to 4.2.6) and, therefore, play a leading role for the future development of personalized medicine.

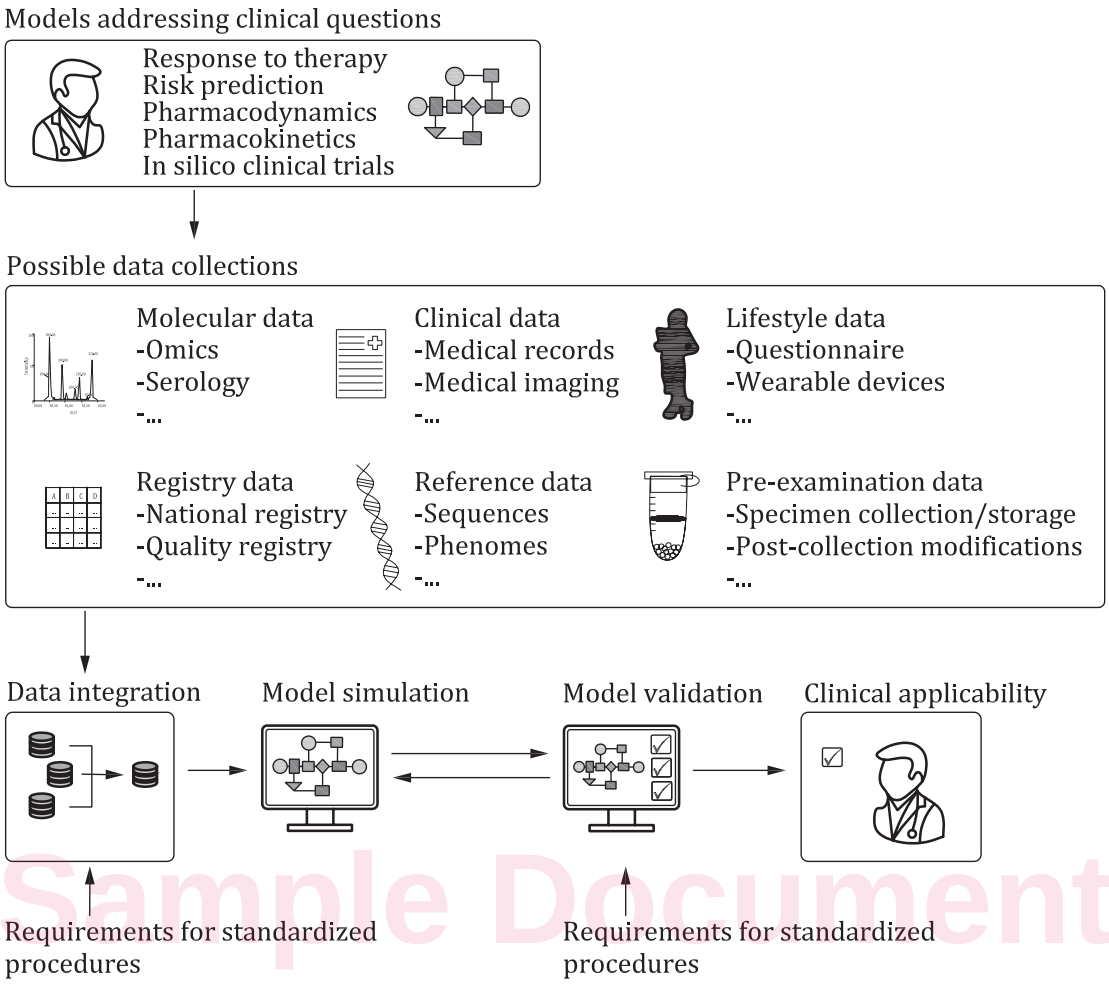

Figure 1 — Modelling approach for personalized medicine

4.2.2 Cellular systems biology models

4.2.2.1 General

For the simulation of complex dynamic biological processes and networks, models can be either data-driven ("bottom-up") or mechanism-based ("top-down").

Mechanism-based concepts aim for a structural representation of the governing physiological processes based on model equations with limited amount of data, which are required for the base model establishment[24] or, alternatively, on static interacting networks.[25][26] Data-driven approaches[11][27] require sufficiently rich and quantitative (e.g. time-course) data to train and to validate the model. Due to the occasional black-box nature of data-driven approaches, the model validation process relies on performance tests against known results.

4.2.2.2 Challenges

The challenges are as follows:

a) the creation of models that balance the level of abstraction with comprehensiveness to make modelling efforts reproducible and reusable (abstraction versus size);

b) the development of prediction models that can be adapted easily to individual patient profiles;

c) efficient parameter estimation tools to cope with population and disease heterogeneity;