**ISO/DTS 9491-1:~~2022~~2023(E)**

Date: 2023-02-17

ISO/TC 276/WG 5

Secretariat: DIN

# Biotechnology — ~~Recommendations and requirements for~~ Predictive~~predictive~~ computational models in ~~personalised~~personalized medicine research — ~~Part 1: Guidelines for constructing~~Constructing, verifying and validating models

~~*Biotechnologie — Recommandations et exigences relatives aux modèles informatiques prédictifs dans la recherche sur la médecine personnalisée — Partie 1: Lignes directrices pour la construction, la vérification et la validation des modèles*~~

DTS stage

ISO Copyright Office
CP 401 • CH-1214 Vernier, Geneva
Phone: + 41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org
Published in Switzerland.

## Contents

## 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 documents 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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

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*.



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~~life sciences and ~~Health~~health research has greatly increased ~~many orders of magnitude~~ in the last decade. In combination with patient/personal-derived data, such as electronic health records, patient registries and ~~-~~databases, as well as ~~life style~~lifestyle information, this ~~Big Data~~big data holds an immense potential for clinical applications, especially for computer-based models with predictive capacities in ~~personalised~~personalized medicine. However, and despite the ever-progressing technological advances in producing data, the exploitation of ~~Big Data~~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~~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 ~~personalised~~personalized medicine.

Common standards lead to a mutual understanding and improve information exchange within and across research communities and are indispensable for collaborative work. ~~Heterogeneous~~In order to setup computer models in personalized medicine, data integration from heterogeneous and different sources ~~and recorded~~ at different times ~~shall be integrated in order to setup computer models in personalised medicine~~plays a key role. Consistent documentation of data, models and simulation results based on basic guiding principles for data management practices, such as FAIR[1] ~~or ALCOA[2]~~, (findable, accessible, interoperable, reusable)[7] 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. Such standards, together with ~~harmonised~~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 ~~visualisation~~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 ~~personalised~~personalized medicine are to ~~(1) harmonise~~:

a) harmonize the standardization efforts that refer to different data types, approaches and technologies~~, as well as to (2)~~ ;

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

An overall goal is to ~~(3)~~ 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 modelling requirements and recommendations ~~and requirements~~ for research in the field of ~~personalised~~personalized medicine, especially with focus on collaborative research, such that health-related data can be optimally used for translational research and ~~personalised~~personalized medicine ~~world-wide~~worldwide.

[1] ~~FAIR: Findable, Accessible, Interoperable, Reusable~~
[2] ~~ALCOA: Attributable, Legible, Contemporaneous, Original, Accurate~~

# Biotechnology — ~~Recommendations and requirements for predictive~~Predictive computational models in ~~personalised~~personalized medicine research — ~~Guidelines for constructing~~Constructing, verifying and validating models

## 1 Scope

This document specifies requirements and recommendations for the design, development and establishment of predictive computational models for research purposes in the field ~~personalised~~of personalized medicine. It addresses the set-up, formatting, validation, simulation, storing and sharing of computational models used for ~~personalised~~personalized medicine. Requirements and recommendations for data used to construct or required for validating such models are also addressed. This includes rules for formatting, descriptions, annotations, interoperability, integration, access and provenance of such data. ~~Computational models used in clinical, diagnostic or therapeutic purposes are excluded.~~

This document does not apply to computational models used in 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 20691:2022, *Biotechnology — Requirements for data formatting and description in the life sciences*

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

~~ISO 14155:2020, *Clinical investigation of medical devices for human subjects — Good clinical practice*~~

~~ISO 26000, *Guidance on social responsibility*~~

~~ISO 20916:2019, *In vitro diagnostic medical devices — Clinical performance studies using specimens from human subjects — Good study practice*~~

~~ISO 20186-1:2019, *Molecular in vitro diagnostic examinations — Specifications for pre-examination processes for venous whole blood — Part 1: Isolated cellular RNA*~~

~~ISO/TS 20658:2017, *Medical laboratories — Requirements for collection, transport, receipt, and handling of samples*~~

~~ISO 13972:2022, *Health informatics — Clinical information models — Characteristics, structures and requirements*~~

## ~~1~~3 Terms and definitions

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

ISO and IEC maintain ~~terminological~~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/

**3.1**
**artificial intelligence**
**AI**
<system> capability to acquire, process, create and apply knowledge, held in the form of a model, to conduct one or more given tasks

[SOURCE: ISO/IEC TR 24030:2021, 3.1]

**3.2**
**molecular biomarker**
biomarker
molecular marker
detectable and/or quantifiable molecule or group of molecules used to indicate a biological condition, state, identity or characteristic or an organism~~, e.g. but not limited to, nucleic acid sequences, proteins, small molecules such as metabolites and other molecules such as lipids and polysaccharides~~

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

[SOURCE: ISO 16577:~~2016~~2022, 3.~~1144~~.28]



**3.3**
**big data in health**
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



[SOURCE: Reference [8]]

**3.4**
**community standard**
standard that reflects the results of a grass-roots standardization effort from a specific user group, and that is created by individual organizations or communities

**3.5**
**computational model**
*in silico* model
description of a system in a mathematical expression and/or graphical form highlighting objects and their interfaces ~~; object distributed processing (ODP) concept~~

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

Note 2 to entry: The computational model is similar to OMT ad UML notion of a class diagram when using the graphical form.

~~Note 2 to entry: This definition is based on ISO/IEC 16500-8:1999, 3.6.~~

[SOURCE: ISO/IEC 16500-8:1999, 3.6, modified — Admitted term added. "mathematical expression and/or" added, 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. Note 2 to entry added.]

**3.6**
**data-driven model**
model developed through the use of data derived from tests or from the output of investigated process

[SOURCE: ISO 15746-1:2015, 2.4]

**3.7**
**data harmonization**
technical process of bringing together different data types to make them processable in the same computational framework

**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 ~~individualised~~individualized computer simulation in the development or regulatory evaluation of a medicinal product, medical device~~,~~ or medical intervention



~~Note 1 to entry: definition according to: Avicenna roadmap .~~

[SOURCE: Reference [9]]

**3.11**
***in silico* ~~experiment~~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, artificial intelligence.

[SOURCE: ISO 20252:2019, 3.52~~)~~, modified — Abbreviated term “ML” added.]

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

[SOURCE ISO 16730-1:2015, 3.11]

**3.14**
**mechanism-based ~~(models)~~**
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**
**~~personalised~~personalized 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: ~~Based on the EU Health Ministers in their Council conclusions on personalised medicine for patients.~~

~~Note 2 to entry:~~ Examples for individuals' phenotypes and genotypes are molecular profiling, medical imaging~~,~~ and lifestyle data.

Note ~~3~~2 to entry: Medical decisions, prevention strategies and therapies in ~~personalised~~personalized medicine are based on this individuality.

[SOURCE: EU 2015/C 421/03[10]]

**3.18**
**raw data**
data in its originally acquired, direct form from its source before subsequent processing

[SOURCE: ISO 5127:2017, 3.1.10.04]

## ~~1~~4 Principles

### ~~1.1~~4.1 General

Research in the field of ~~personalised~~personalized medicine is highly dependent on the exchange of data from different sources~~.~~, as well as ~~harmonised~~harmonized integrative analysis of large-scale ~~personalised~~personalized medicine data (big data in health). Computational modelling approaches play a key role for understanding, simulating and predicting the molecular processes and pathways that ~~characterise~~characterize human biology. Modelling approaches in biomedical research also lead to a more profound understanding of the mechanisms and factors that drive disease, and ~~later on~~ consequently allow for adapting ~~personalised~~personalized treatment strategies that are guided by central clinical questions. Patients can greatly benefit from this development in research that equips ~~personalised~~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.



 

### ~~11.2~~4.2 Computational models in ~~personalised~~personalized medicine

#### ~~11.2.1~~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 ~~recognised~~recognized by the scientific community ~~([4], [5], [6], [7]).~~[11][12][13][14] as well as by regulatory bodies such as the European Medicines Agency (e.g. EMA guideline on PBPK reporting[3] ~~),~~[15]), or the US Food and Drug Administration (FDA~~) [8], [9].~~) [16][17] Computational models are integrated in different fields in medicine and drug development expanding from disease modelling, molecular biomarker research to assessment of drug efficacy and safety. *In silico* approaches are also expanding in neighbouring fields, such as pharmacoeconomics ~~([10], [11]).~~[18][19] analytical chemistry ~~([12], [13])~~[20][21] and biology that are out of scope of this document ~~[14], [15].~~[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 ~~personalised~~personalized medicine.



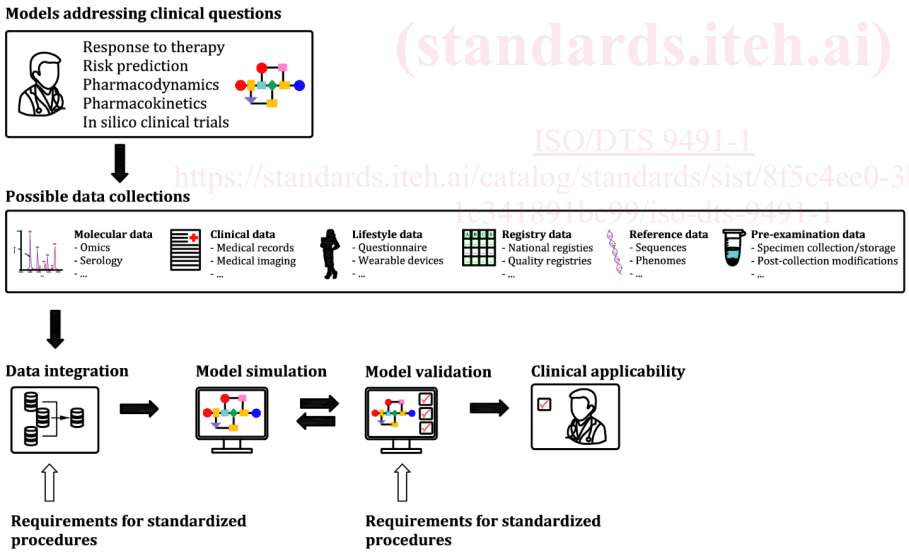


**Figure 1 — Modelling approach for ~~personalised~~personalized medicine**

[3] ~~https://www.ema.europa.eu/en/reporting-physiologically-based-pharmacokinetic-pbpk-modelling-simulation~~

### ~~11.2.2~~4.2.2 Cellular systems biology models

#### ~~11.2.2.1~~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 ~~[16][24]~~ or, alternatively, on static interacting networks ~~[17], [18].~~[25][26] Data-driven approaches ~~[4], [19][11][27]~~ require sufficiently rich and quantitative time-course data to train and to validate the model. Due to its often black-box nature, the model validation process in data-driven approaches relies on performance tests against known results.

#### ~~11.2.2.2~~4.2.2.2 Challenges

The challenges are as follows:

— Creation of models that balance the level of abstraction with comprehensiveness to make modelling efforts reproducible and reusable (abstraction ~~vs~~versus size).

— Development of prediction models that can be adopted easily to individual patient profiles.

— Efficient parameter estimation tools to cope with population and disease heterogeneity.

— Overfitting of the model to the experimental/patient data and optimization methods for model predictions in a realistic parametric uncertainty.

— Flexibility in models to cope with missing data (e.g~~,~~. diverse patient profiles).

— Scaling from cellular to organ and to organism levels (e.g~~,~~. high clinical relevance, high hurdles for regulatory acceptancy).



### ~~11.2.3~~4.2.3 Risk prediction for common diseases

#### ~~11.2.3.1~~4.2.3.1 General

Predictive models stratify patients into distinct subgroups at different levels of risk for clinical outcomes (risk prediction for disease). By training the algorithm on clinical data, phenotypic or genotypic, ~~one can identify~~ subgroups can be identified which have identifiably different patterns of clinical markers ~~and by.~~ By then identifying which patterns a patient fits best, the model can place a particular patient within the most similar trajectory, thereby also stratifying the patient to a particular level of risk. Clinical markers used in such models can be any health feature, ~~tokenised~~tokenized as to be analysable by the model, from phenotypic data such as disease history symptoms, treatment and other exposure data, family history, ~~lab~~laboratory data ~~and so on,~~ etc., to genetic data.

#### ~~11.2.3.2~~4.2.3.2 Challenges

The challenges are as follows:

— Understanding the possible implication to patients at an individual ~~-~~level~~, what~~. What can be inferred? How to test the inference made?

— Limited replication of genetic associations and poor application of diverse populations (e.g~~,~~. too poorly represented to be of interest for specific analyses), specifically of mixed or non-European ancestry.

— Varying transparency of methodological choices and reproducibility.

— Limited cellular/tissue context and harmonized functional data availability across populations/studies.

— Missing environmental information coupled to genetic data.

### ~~11.2.4~~4.2.4 Disease course and therapy response prediction

#### ~~11.2.4.1~~4.2.4.1 General

Prediction of the disease behaviour (mild ~~vs.~~versus severe, stable ~~vs.~~versus progressive) early in the disease course based on specific molecular biomarkers can allow an improved timing of therapy introduction, as well as the choice of therapy scheme (targeted therapy~~) [24].~~.[28] Ideally, these models can provide a prediction of ~~'~~multi-factorial~~'~~ diseases at unprecedented resolution, in a way that clinicians can use the information in their daily decision-making.

#### ~~11.2.4.2~~4.2.4.2 Challenges

The challenges are as follows:

— Harmonization and standardization of clinical information for measuring the disease of interest.

— Developing transparent and quality-controlled workflows for molecular data generation and interpretation in clinical settings.

— Harmonization and application of existing and upcoming pre-examination workflow standards (including specimen collection, storage and nucleic acid isolation), as well as developing feasible ring trial formats and ~~External Quality Assurance~~external quality assurance (EQA) schemes for given molecular analysis types.

— Transparent reduction of contents and definition of appropriate marker sets and dynamic models to foster clinical translation.

— Developing intuitive visualization results and insights into molecular analyses, as well as critical appraisal of limitations of models by physicians.

### ~~11.2.5~~4.2.5 Pharmacokinetic/-dynamic modelling and *in silico* trial simulations

#### ~~11.2.5.1~~4.2.5.1 General

Pharmacokinetic/pharmacodynamic (PK/PD) models ~~[25, 26~~[29][30] can usefully translate *in vitro*, non-clinical and clinical PK/PD data into meaningful information to support decision-making during drug development. At the individual level, drug PKs can either be described by non-compartmental analysis~~,~~ and compartmental PK modelling or by physiologically-based PK (PBPK) modelling. At the population level, population PK have become the most commonly used top-down models that derive a pharmaco-statistical model from observed systemic concentrations. PK/PD modelling involves on the one hand a quantification of drug absorption and disposition (PK) and on the other hand a description of the drug-induced effect (PD). PK/PD models and quantitative systems pharmacology (QSP) both aim for mechanistic and quantitative analyses of the interactions between a drug and a specific biological system ~~[27].~~[31]

~~Today,~~ PK and PBPK modelling are currently used for simulations for virtual patient populations in *in silico* clinical trials ~~(ISCTs)~~. The concept is that computer simulations are proposed as an alternative source of evidence to support drug development to reduce, refine, complement~~,~~ or replace the established data sources including *in vitro* experiments, *in vivo* animal studies~~,~~ and clinical trials in healthy volunteers and patients.