FASEB BioAdvances WILEY

Measuring and interpreting pervasive heterogeneity, poikilosis

Mauno Vihinen ⊙

Department of Experimental Medical
Science, Lund University, Lund, Sweden

Correspondence
Mauno Vihinen, Department of
Experimental Medical Science, BMC
B13, Lund University, SE-22184 Lund,
Sweden.
Email: mauno.vihinen@med.lu.se

Abstract

Measurements are widely used in science, engineering, industry, and trade. They form the basis for experimental scientific research, approach, and progress; however, their foundations are seldom thought or questioned. Recently poikilosis, pervasive heterogeneity ranging from subatomic level to biosphere, was introduced. Poikilosis makes single point measurements and estimates obsolete and irrelevant as measurands display intervals of magnitudes. Consideration of poikilosis requires new lines of thinking in experimental design, conduction of studies, data analysis and interpretation. Measurements of poikilosis must consider lagom, normal, variation extent. Measurements, measures, and measurands as well as the measuring systems and uncertainties are discussed from the perspective of poikilosis. New systematics is introduced for description of uncertainty in measurements and for types of experimental designs. Poikilosis-aware experimenting, data analysis and interpretation are discussed. Instructions are provided for how to measure lagom and non-lagom effects of poikilosis. Consideration of poikilosis can solve scientific controversies and enigmas and can allow novel insight into systems, processes, mechanisms, and reactions and their interpretation, understanding, and manipulation. Furthermore, it will increase reproducibility of measurements and studies.

KEYWORDS

experimental design, heterogeneity, lagom, measurand, measurement, measurement uncertainty, poikilosis

1 | INTRODUCTION

Measuring is a fundamental activity and technique in science, engineering, industry, and trade. Although widely utilized, not much thought is given to it. Traditionally measurements and science have aimed toward a single measurement quantity (value) with as small an uncertainty in the form of standard deviation as possible.

Measurements form the basis for experimental scientific research, approach, and progress. They should be based on deep understanding of the investigated phenomenon and its dependence on other factors for the results to be meaningful. This article is written from the perspective of life sciences, mainly biology and medicine; however, it applies also to other fields. Measurements have to be rethought to take recently introduced pervasive heterogeneity, poikilosis,[1] into account.

Well conducted experiments are expected to be valid and repeatable, but that is not always the case. Replication crisis, also called reproducibility crisis, refers to inability to replicate published studies and observations despite extensive efforts. This is far too common and seen in many fields.[2] Some examples are described here. Replication effort was conducted for 53 landmark studies in clinical cancer research published in major journals. Only in six cases (11%) the findings could be confirmed.[3] Only 39 out of 100 experimental and correlational studies in psychology could be replicated.[4] In another replication experiment, 75% of 67 articles in oncology, cardiovascular diseases and women's health failed in reproduction.[5] Two large pharmacogenomic studies showed inconsistency: out of 15 shared drugs and 471 cell lines only one compound showed "moderate" and another "fair" correlation.[6] According to a survey among 1,500 scientists from various fields, 70% had failed to replicate results of others, and >50% even their own studies.[7] There were differences between disciplines, but irreproducibility was high among all of them.

Irreproducibility originates due to many reasons. For example, antibodies, widely used reagents, are a common source for irreproducibility due to batch-to-batch variability and unspecific binding.[8] Another major reason is misuse and misunderstanding of statistical probability, which is often expressed as a p-value. In some fields, especially in medicine, observations in publications have to be supported by almost an enchanted *p*-value, typically a value <0.05. Many studies are underpowered, therefore low p-values are not reliable[9] and reproducibility is low. Single minded concentration on p-values has led to publication bias as only those studies that show "significant" *p*-value are published,[10] others are usually not even submitted to journals.

It has been presented that majority of the published statistical probabilities and interpretations based on them are wrong.[11] In addition to honest mistakes, statistical results are often misleading because of data dredging, p-hacking and p-harking (hypothesizing after results are known).[12] These practices include selection of a subset of data points supporting wanted outcome and search for a statistical test that gives low enough p-value.

Moreover, an important factor affecting reproducibility of scientific observations is poikilosis, normal, and pervasive heterogeneity in systems.[1] Every system and experiment display poikilosis at multiple levels. Since it has not been properly taken into account, replication of experiments may be challenging or impossible. As many studies are based on low numbers of investigated entities (or limited number of values even in "big data") they do not allow to chart the range of poikilosis. Therefore, replication studies may not work.

Poikilosis has much larger effect on experiments beyond reproducibility. Poikilosis makes single point measurements and estimates obsolete and irrelevant as the measured magnitudes lie within a range (interval). Inclusion of poikilosis requires new line of thinking in experimental design, in conduction of studies, data analysis, and interpretation. It is also apparent that new types of analytical and statistical approaches will be needed.

In this article, I briefly present poikilosis and its normal variation extent (lagom), discuss measurements, measures and measurands as well as the measuring systems and uncertainty from the perspective of poikilosis. Then, poikilosis-aware experimenting, data analysis and interpretation are discussed. Further, guidelines for measuring lagom and non-lagom effects of poikilosis are provided. Poikilosis has not yet been properly treated in any study, therefore its full meaning and significance have not been revealed.

2 | POIKILOSIS AND LAGOM

According to definition,[1] poikilosis is inherent pervasive variation, heterogeneity and fluctuation in living organisms, populations, ecosystems, biosphere and in their components and in processes within them. This means that most experiments aiming at and reporting a single value or score are defective as the intrinsic interval of heterogeneity in the measured entity is missed.

Heterogeneity in biological systems has often been called noise although it in fact refers to poikilosis. Noise means uncertainty, which originates from many sources and has been described in the "Guide to the Expression of Uncertainty in Measurement" (GUM) by the Joint Committee for Guides in Metrology (JCGM,[13]) for the International Organization for Standardization (ISO).

Several databases contain information about heterogeneity in different fields, some of which are listed in Table 1. There are variation data from molecular heterogeneity to genetic and epigenetic variations (substitutions, insertions, deletions, indels, chromosomal and genome wide differences, methylation status differences etc.), post transcriptional and post translational modifications of RNA transcripts and proteins, and RNA alternative splicing. Databases are also available for allometric, pharmacogenetic and -genomic and physiological differences as well as for ecosystem heterogeneities and biodiversity, and earth magnetic heterogeneity. This small sample indicates that there is already a substantial

TABLE 1 Examples of databases for information about variation

Database	URL	Reference
Ecosystems		
EarthEnv Global Habitat Heterogeneity	https://www.earthenv.org/texture	[63]
Dutch Caribbean Biodiversity Database	https://www.dcbd.nl/tags/heterogeneity	[64]
Molecules		
IDEAL	https://www.ideal-db.org	[65]
Intrinsically Disordered proteins with Extensive Annotations and Literature		
Physiology		
Organ System Heterogeneity DB	http://mips.helmholtz-muenchen.de/Organ_System_Heterogeneity	[66]
HeteroMeth: A Database of Cell-to-cell Heterogeneity in DNA Methylation	http://qianlab.genetics.ac.cn/HeteroMeth	[67]
Genetics		
dbSNP	https://www.ncbi.nlm.nih.gov/snp	[68]
LOVD	https://databases.lovd.nl/shared/genes	[69]
VariBench	http://structure.bmc.lu.se/VariBench/index.php	[70]
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar	[71]
Genomic structural variation		
European Variation Archive	https://www.ebi.ac.uk/eva	
Protein post translational modification		
PTMD	http://ptmd.biocuckoo.org	[72]
Epigenetics		
WashU Epigenome Browser	https://epigenomegateway.wustl.edu	[73]
RNA mod		
REDIportal	http://srv00.recas.ba.infn.it/atlas	[74]
Editome Disease Knowledgebase (EDK)	https://bigd.big.ac.cn/edk	[75]
Alternative splicing		
ExonSkipDB	https://ccsm.uth.edu/ExonSkipDB	[76]
ASpedia	http://combio.snu.ac.kr/aspedia	[77]
Allometry, biomass		
BAAD: a Biomass And Allometry Database for woody plants	https://github.com/dfalster/baad	[78]
Biospecimens		
EFLM Biological Variation Database	https://biologicalvariation.eu	
Biologic Variation and Desirable Specifications for QC	https://www.westgard.com/guest17.htm	
Pharmacology		
PharmVar, Pharmacogene Variation	https://www.pharmvar.org	[79]
PharmGKB, Pharmacogenomics Knowledgebase	https://www.pharmgkb.org	[80]
Magnetism		
Paleomagnetic data	https://www.ngdc.noaa.gov/geomag/paleo.shtml	

amount of poikilosis-related information, although poikilosis *per se* has not been measured.

Although poikilosis is pervasive, it does not mean that any extent of heterogeneity would be allowed and possible within a system. Lagom means "suitable, sufficient, allowed and tolerated extent of variation at any level in an organism, population, biological system or process".[1] Level in here does not mean any connotation of ranking or scaling. Biological processes are regulated to lagom extent by numerous active mechanisms as well as passive and intrinsic characteristics of systems.[14] Lagom indicates the range where poikilosis is at allowed and suitable extent, but it can vary at different time points and situations, thus, to be dynamic and context dependent.

Furthermore, poikilosis replaces homeostatis, which means a fixed standard set point towards which a system is returned by (negative) feedback control mechanisms, with a more relevant model of heterogeneity. As the ending -*stasis* indicates, homeostasis is based on a stable state. Even stringently regulated processes display variation. For why homeostasis is not a viable theory, see discussion on its costs, requirement for monitoring and regulation, as well as for its excessive contribution to free energy in living systems in comparison to poikilosis.[1]

3 | MEASURING PROCESS

Magnitude, quantity, is determined in a measurement process for a measurand, object under measurement (Figure 1). In metrology, the third component of the process is measure, which is a (standardized) scale for the extent of the measured entity. Measurement results are usually described with numerical values in the units of the measure and often followed by estimates of uncertainty expressed as standard deviation.

Life sciences and engineering have not generally been involved in theoretical definition of measurement, whereas in mathematics and philosophy theories of measurement have a long history (see [15]). Theories can be grouped as mathematical, realist, operationalist/conventionalist, information theoretic and model-based.[15] Representational Theory of Measurement (RTM) [16] has been the prominent theory in social sciences, psychology, mathematics, philosophy and many other fields for many decades. According to RTM, measurement is a representation of an empirical relational
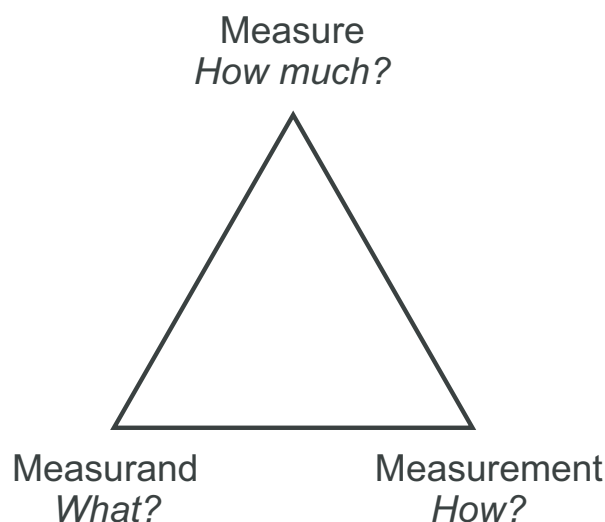

Measure
How much?

Measurand   **Measurement**
What?   *How?*

FIGURE 1 Measuring process includes measure, measurement and measurand and answers to question indicated. Poikilosis is related to measurand and measurement as the intrinsic heterogeneity of measured objects and as interval of measured quantities

system to a numerical relational system. Critique and limitations of RTM have emerged, see e.g.[17-19]

This treatise is based on the pragmatic definition of measurement in the International Vocabulary of Metrology (VIM,[20]): *process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity*.

Measurement process comprises of several steps. First, the measurand has to be defined. It has to be relevant for the investigated object or process and representative for it. Second, suitable method has to be obtained for measurements. Third, measurement process must be detailed for the entire procedure from data collection, sampling etc. all the way to data analysis and interpretation of observations. In the fourth step, the sources of error have to be charted and ways for their treatment identified.

Measurements are performed with some kind of instrument. Note that counting is not considered as a measurement. Uncertainty in measurements includes effects of noise, which according to definitions used in signal processing means modification of measurement signal during acquisition (capture), conversion, processing, transmission and storage. Many of the characteristics of the instrument contribute to uncertainty (Figure 2). The operator of the measurement can also be a source of uncertainty.

GUM defines two types of errors in measurements: random and systematic error.[13] Random error arises e.g. from stochastic or unpredictable variations, uncontrolled test and environmental conditions and varies unpredictably in replicate measurements. Systematic error remains constant in replicate measurements and can be reduced by performing multiple measurements. Neither of the error types can be completely eliminated. Poikilosis is the third component of uncertainty and affects also the quantity of the measurement being responsible for its interval. Figure 2 describes novel systematics for various forms of measurement-related uncertainties. Despite extensive literature and availability of Catalogue of Bias (https://catalogofbias.org/), the full description of extent of measurement errors and biases has been missing.

Type A evaluation of uncertainty can be obtained from probability density function derived from frequency distribution of measurements.[13] Type B uncertainty is obtained from an assumed probability density function and can include, for example, previous measurement data, data from calibration and general knowledge of the measurement system.

Poikilosis has to be taken into account in measuring process as source of variation and uncertainties, therefore results should be presented as intervals. Single measurement value with a standard deviation does not cover poikilosis and thus cannot be used to measure, for example, many biological processes. Since poikilosis is dependent on the condition of the investigated system, it is mandatory to include full details for
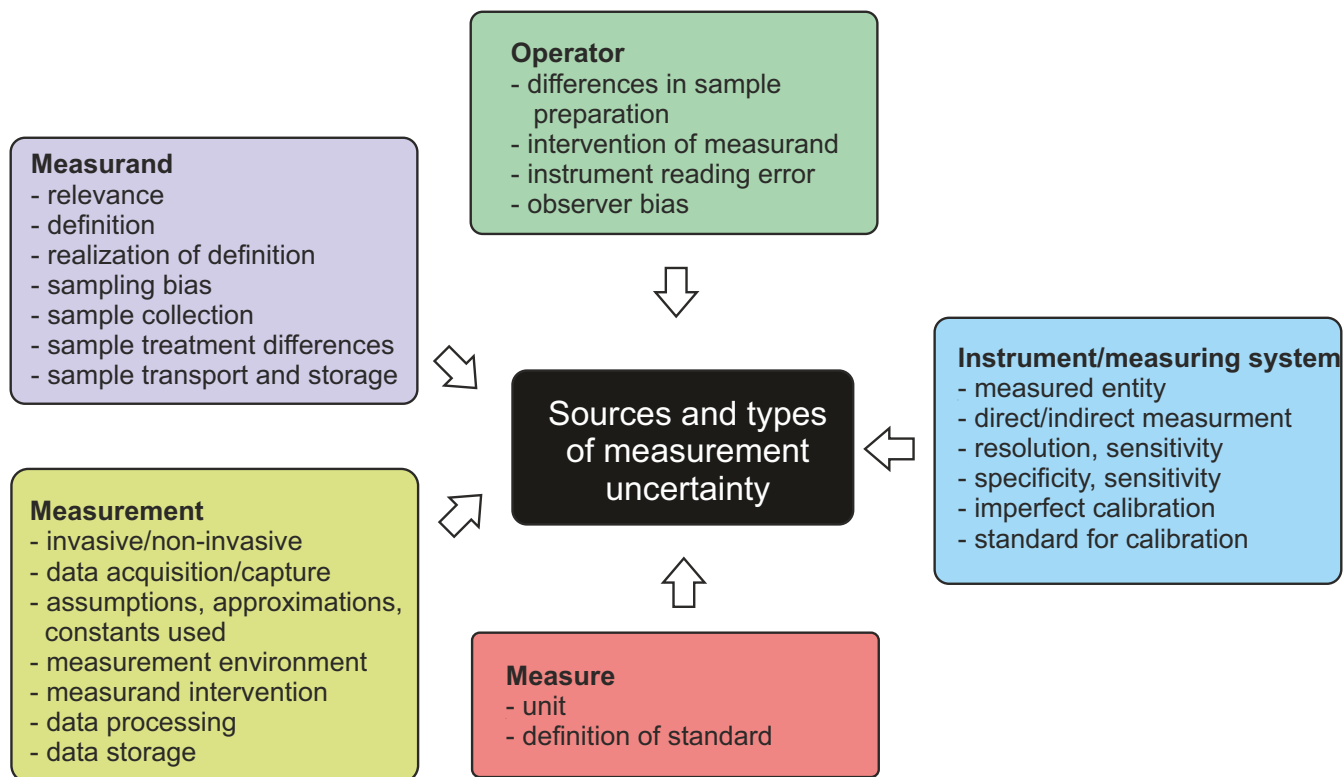
FIGURE 2 Systematics for measurement uncertainty including various types of errors, biases and poikilosis

the entire measurement and sample collection and treatment process. Poikilosis-aware measuring requires a mental shift to cover the full range and extent of variations and their significance, especially for lagom extent.

4 | MEASURAND

Measurand is the object or event of measurement, see measurement process triangle in Figure 1. Several aspects of measurand are pertinent for measurements, namely relevance of measurand, choice (sampling), collection and treatment of samples. Only some measurements are direct, i.e. measuring the object of interest. In biology and medicine, many measurands are indirect and measure some entity related to the investigated phenomenon, because direct measurement of the actual measurand, for example, enzyme activity or protein or metabolite concentration in cells or tissues of interest cannot be directly studied. These biomarkers are typically measured from blood, saliva, urine, sweat, or tears. For example, it is not possible to take biopsy from the brain of a living person. The relevance of the biomarker is essential for the biological process, component, or reaction that is of interest. When measuring biomarkers instead of the actual object-related data, conversions, assumptions and computations are made and can be sources of error (Figure 3).

Unless properties of a single individual entity are measured, scientific measurements are related to an entire population or on samples of it. It is often impossible and many times unnecessary to measure every individual, entity, element, unit, group, or data item in a population. Results can be generalized to the entire population from a representative unbiased sample. There are two main categories of sampling methods, namely probability and nonprobability sampling, both of which can be used with or without replacement. In sampling without replacement, an element can be selected only once, whereas when sampling with replacement, an element can appear more than once. An example of the latter sampling type is a fruit fly experiment where an insect is released after measurement and could be caught again.

Sampling methods are well developed but have to be properly applied. In probability sampling, every unit or individual in a population has a chance to be selected. In nonprobability sampling some elements of the population have no chance to be selected at all. Probability sampling methods include systematic, simple random, stratified sampling, cluster, multistage, and probability proportional to size sampling. There are many nonprobability sampling methods including convenience (accidental, grab, or opportunity), purposive (judgemental), consecutive (total enumerative), quota, snowball, voluntary, minimax, line intercept, panel, and theoretical sampling.
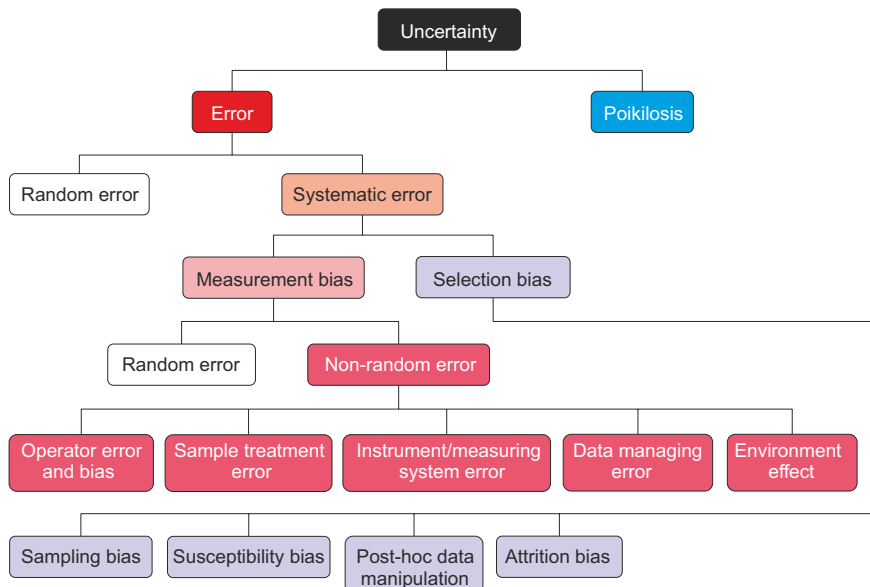
Bias in sampling always leads to errors and uncertainty in measurements. In medicine, sampling bias has been called ascertainment bias. Multiple types of biases are related to survey studies originating, for example, from selection and exclusion criteria such as pre-screening, coverage, questionnaire formulation and order or questions, self-selection, survivorship, and others. STRANGE framework was presented to prevent sampling bias in animal studies by considering social background, trappability and self-selection, rearing history, acclimation and habituation, natural changes in responsiveness, genetic make-up, and experience.[21] Existing sampling methods are valid tools from the perspective of poikilosis provided that the heterogeneity of the system is appropriately sampled.

Another important area of sample related bias originates from how the samples are collected, treated, transported, and stored. Cell and tissue samples are examples in which extensive treatments are performed before the samples are measured, therefore they can be sources for various forms of measurement uncertainty (Figure 3).

Several calculators and approaches are available for sample size estimation to define how many replicates are needed to obtain statistically significant results. These methods require information about population and sample size, estimate of error rate and wanted confidence level in different experimental setups. These estimates of numbers of parallel experiments do not consider poikilosis. Coverage of interval of values characteristic for poikilosis may require more samples than traditionally expected.

Many measurements have an assumption that the investigated system is at a steady state. Such measurements are the easiest to perform. Transition states are often more interesting as they allow to follow how the system changes. Since many transitions are fast and difficult to capture in heterogeneous samples, large datasets are needed.

5 | MEASURES AND MEASUREMENTS

Standard measures are usually not contributing towards uncertainty of measurements as they are well defined. The International Bureau of Weights and Measures (French: Bureau international des poids et mesures, BIPM) is an intergovernmental organization that has defined the most essential and basic measurement standards (etalons). International System of Units (SI, abbreviated from the French Système international (d'unités)) provides official metric base units for measures and standards for defining the measures. Many additional measures are used and obtained from instruments and they may contribute to uncertainty or measurement results.

No measurement is exact and thus all measurements contain uncertainty by being composites of actual measured entities and uncertainties. In addition to measurement errors and biases, poikilosis adds to uncertainty (Figure 3).

In science, a typical measurement setup is to investigate the effects of a perturbation or alteration. Measurement quantity M is the sum of effective variation (E) and uncertainty (U) as follows:

$$M = E \pm U.$$

Effective variation is the introduced variation V reduced by R

$$E = V - R,$$

where R is the sum of reversing, attenuating, buffering and correcting factors and processes.[14] The mechanisms responsible for the reduced effects are called TARAR countermeasures after tolerance, avoidance, repair, attenuation, and resistance.[14]

Effects of most alterations and perturbations in normal biological systems are depleted. Measurement of a perturbation effect on measurand reveals effective variation E and its uncertainty U. E is the measurable quantity.

Uncertainty has traditionally been calculated as root sum of squares for individual components of type A and type B uncertainty, the former describing systematic error and the latter random error. There are other formulations for inclusion of more complicated cases of uncertainty (see e.g. [22]). How poikilosis should be considered in the calculation of combined uncertainty may have to be defined case by case.

Measurement errors have several sources starting from operator bias and errors and intervention of the measurand by operator and observer bias (experimenter bias, in some fields called allegiance bias) (Figure 3). Instrument/measuring system can intervene the measurand, for example, via sensors. Intervention of the measurand depends also whether the measurement is invasive or noninvasive. Invasive instruments likely affect the measurand more than noninvasive methods. The instrument or measuring system is dependent on the type of measurement. Only some instruments measure directly the measurand such as a tape measure the height of a study subject. Instead, most biological and medical instruments measure indirectly. For example, most barometers measure the air pressure alterations based on compaction or expansion of an aneroid capsule of thin metal, movements of which are converted indirectly to indicator movement via mechanical levers. Similarly, most scientific instruments make indirect measurements that are based on various assumptions, approximations and constants to obtain the final measurement quantity. Many instruments are dependent on computers and are prone to data processing and storage errors and, for example, errors in curve fitting. Measurement environment may have a substantial effect on the outcome.

Instrument or measuring system realizes the definition of a measure into practice. The accuracy and resolution of measurements depend on how well this realization is made. Calibration is an important process to prevent systematic error and is dependent on reliable standards and their implementation.

The discussion above in this section relates to errors and biases within measurement process and measuring device. Poikilosis adds uncertainty to the measurement as the measured object has a spectrum of values within an interval instead of a single value. In addition to population variation, measured entities show within subject heterogeneity due to poikilosis.

Scientific measurements are typically standardized or harmonized to have as low an uncertainty as possible. Depending on the experiment and field, there are different ways to achieve this. Standardization and harmonization, in the form of standard operating procedures (SOPs) and other systematic protocols, reduce sample to sample variation. In biology, samples can be harmonized, for example, using homogeneous cell lines or inbred strains, or by synchronizing cells, reactions or processes within them. However, none of these approaches reduces poikilosis within the samples. Single cell studies do not suffer from sampling bias (however, even they are somehow selected) but require a large number of cells because of poikilosis.

Standardization of samples, experiments, and conditions can be taken even too far and thereby undermine the relevance of observations. An example comes from animal research. Mouse studies are widely used model systems which are performed in standardized conditions using genetically uniform inbred strains of certain age, standardized husbandry and environmental conditions etc. These actions are good to homogenize experiments, however, they miss biological variation and can lead to results that cannot be generalized and verified, for example, in other strains or conditions.[23] Reproducibility of animal models is generally rather poor.[24,25] Simulations indicated that standardization of animal studies contributed to poor reproducibility.[26] Laboratory mice models in immunology may have different immune phenotypes in comparison to wild animals. Various naturalization strategies have been presented to improve translational potential of studies with laboratory animals.[27]

6 | EXPERIMENTAL DESIGN

The choice of experimental design is instrumental for every investigation. The design depends on many factors. A researcher has to consider what is the scientific question, what kind of data are available, and what is possible to obtain, the extent of resources, and what is legal and allowed to do. Figure 4 presents a new and systematic taxonomy for experimental designs ranging from observational, descriptive and analytical, methods to various types of experimental study designs. Analytical studies contain a control which is missing from descriptive study designs.

Many studies are observational because of necessity, analytical, or experimental investigation may not be possible for example due to small number of available cases. Studies of rare diseases are an example, there may be just a few known cases in the entire world, therefore large experimental studies are not possible in such cases.

In the beginning of experimental design, investigated variables (independent, dependent and control variables) are chosen along with estimated amount of data. A good design should facilitate obtaining valid and reliable results that can be replicated.

Random controlled trials (RCTs) are considered as the most reliable design type for guiding medical decision making on treatments, drugs and others. Interestingly, analysis of effect estimates did not find significant differences
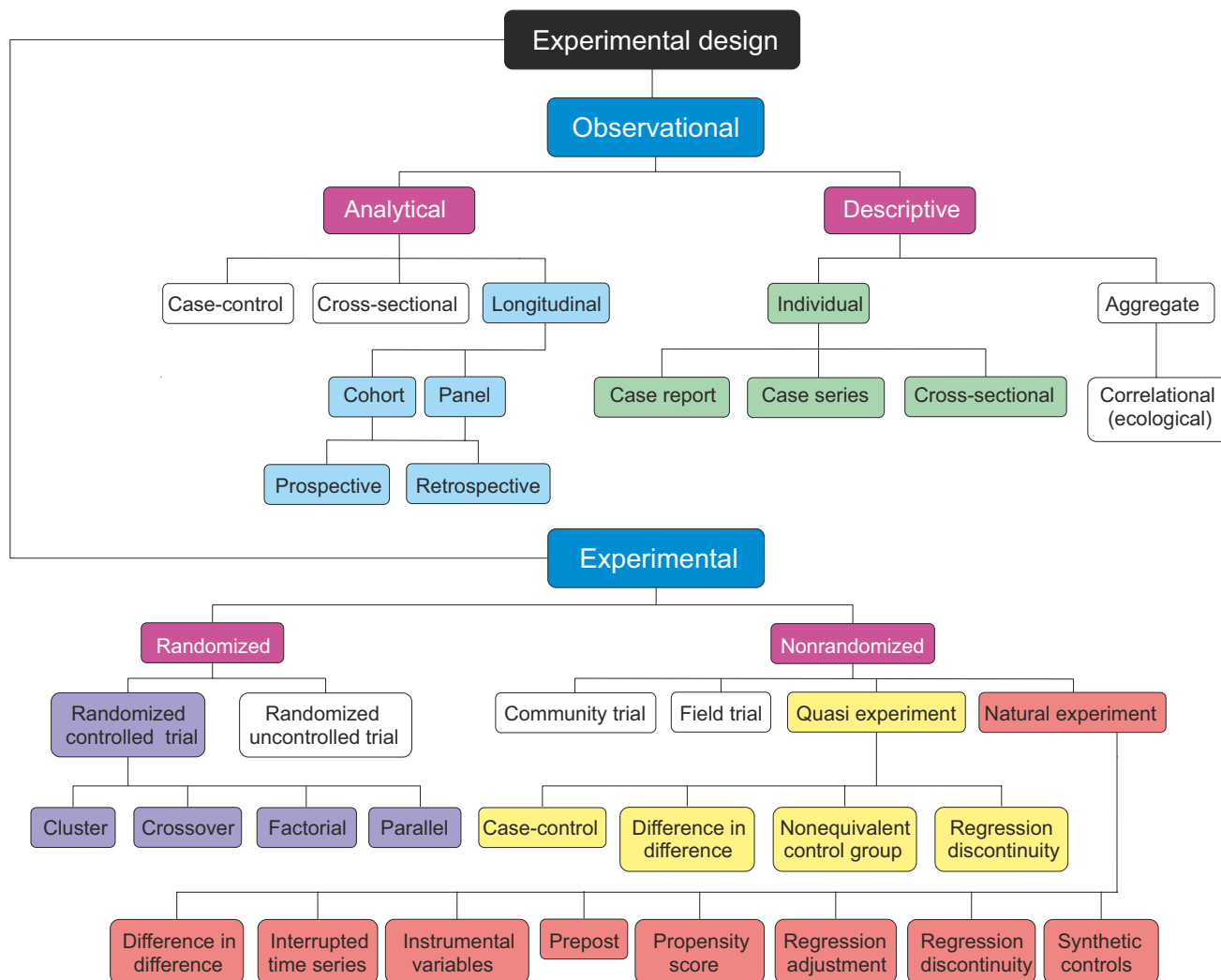
FIGURE 4 Systematics for types of experimental designs. The types at most detailed levels are not comprehensive, instead present a sample of available designs. Although descriptive studies report heterogeneity cases, they cannot provide full account of poikilosis. Analytical experimental designs can consider some details of poikilosis, however, only certain experimental study designs allow systematic study of poikilosis with controlled perturbations. These studies allow to define whether the observations are of lagom or non-lagom extent and thus indicating the status of the system. Experimental studies can also be designed to address interactions of levels

between RCTs and observational studies.[28] Experimental designs have been ranked in medicine into a hierarchy (evidence pyramid) according to increasing evidence they are supposed to provide.[29] On the top of the pyramid are systematic reviews, which are often based on meta-analyses of large RCTs. There are about 200 ranking schemes (http://cjblunt.com/hierarchies-evidence/), the Grading of Recommendations Assessments, Development, and Evaluation (GRADE) [30] being probably the most widely used. Although these schemes can be valuable, the outcome has to be estimated case by case. It has been argued that the hierarchies rank used methods instead of data produced by the methods and that interpretations based on evidence hierarchies are poor.[31]

Systematic reviews and meta-analyses are not without flaws, therefore a classification of spin (overstatement of effects) has been released.[32] Analysis of the 10 most cited RCT articles revealed biased results,[33] thus it is important to know about limitations and biases in all types of experimental designs instead of considering certain types or experiments to be without bias or to be better and more reliable by definition.

With randomization of study objects, it is possible to avoid many biases. Blinded, especially double blinded, studies are highly valued in medicine as many sources of bias and error are excluded, for example, when even the health care personnel treating the individual do not know whether the treatment is a drug or placebo. RCT is considered as gold standard in clinical medicine and sociology, but not equally relevant in many other fields.

Study designs have different capabilities to quantify and measure poikilosis, explain its significance and to be used in

—WILEY—

FIGURE 5 Capabilities of experimental designs to address poikilosis-related features
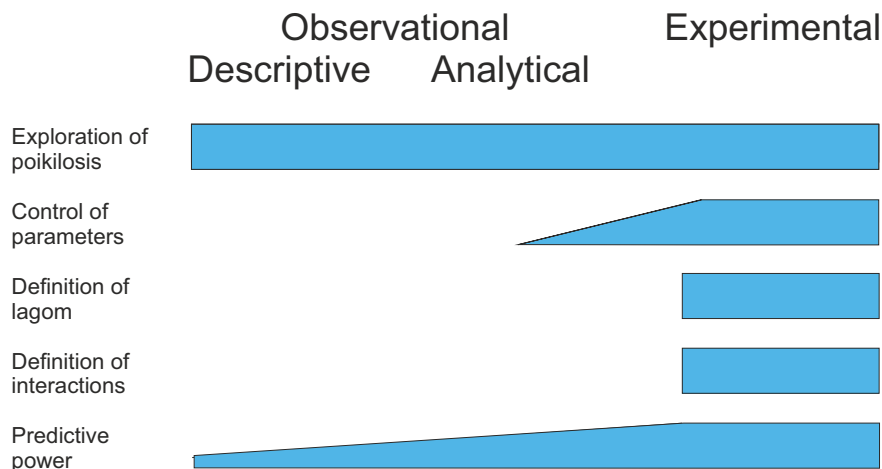
predictions (Figure 5). Only experimental studies can control and test poikilosis; however, observational studies can also chart and analyze poikilosis. Descriptive studies can explain and identify heterogeneity in individuals and cohorts and range from anecdotal examples up to detailed aggregates. Analytical approaches can define poikilosis intervals in controlled setups, for example to obtain reference values for clinical laboratories.

Analytical experiment designs have a potential to describe some aspects of poikilosis. In experimental study designs poikilosis can be defined based on perturbations and control of heterogeneity. Randomized controlled trials are in theory the best suited study design, however randomized uncontrolled trials and quasi experiments could also be poikilosis-aware. In natural experiments, community and field trials the contribution of poikilosis may be difficult to discern from other factors, errors and biases. Experimental studies can be used for investigation of connected levels and poikilosis in them.

Data from descriptive studies can be used for developing prediction methods, mainly for classification purposes. Data obtained from experimental study designs lend to predictive power also for mechanism-oriented predictors and for regression.

In addition to experimental design, size and representativeness of studied sample, and evaluation whether the measurements are at lagom level are crucial for reliable measurements. Understanding effects and mechanisms of relevant TARAR countermeasures is a key for accurate interpretation of observations.

7 | DATA ANALYSIS AND INTERPRETATION OF MEASUREMENTS

Data processing and analysis are integral parts of many measurements and many instruments are connected to or contain a computer. As mentioned above, various assumptions, constants and approximations in the instruments and measurements can contribute to uncertainty of results. Further data analysis is needed when analyzing and combining observations for several measured entities. Suitable statistical test has to be used for these purposes, but without data dredging or p hacking. In data analysis, the most common assumption is to have normally distributed data, although that is not always the case.

7.1 | Distribution

Experimental data sets show a very large number of different types of data distributions, such as continuous (including normal), discrete, joint, mixed discrete/continuous, and other distributions. Although experimental data are generally assumed to have normal, also called Gaussian, distribution, this form of distribution is quite rare, even compared in frequency to that for unicorns[34] based on analysis of 440 studies. In a more recent paper for a study of 693 distributions in many fields, only 5.5% were close to normality.[35]

To find out the type of distribution, it is recommended to visualize the dispersion of data and perform computational analyses. Normal distribution has a symmetric bell-shape. From the distribution can be detected modality of the set, whether it is unimodal with a single maximum or multimodal when there is more than one peak. Quantile-quantile (Q-Q) plot and similar visualizations are useful and can be produced with many analysis packages. Q-Q plot is a probability plot, which shows a graphical comparison of two probability distributions by plotting their quantiles against each other. Normal probability plot is a special case of Q-Q plot for normal distribution.

There are additional relevant features to study from distributions. Skewness is a measure for the asymmetry of a distribution. Instead of being symmetrical, a distribution can have negative or

positive skew. Kurtosis reveals how tailed the distribution is since it describes the shape of a probability distribution. Leptokurtic distributions have fat tails with many cases towards the end of the scale. Platykurtic distribution has thin tails compared to normal distribution, which is called mesokurtic. There are mathematical tests for skewness and kurtosis, the third and fourth central moments are widely used for this purpose.

Visualization can indicate an additional property, whether the data are homo- or heteroscedastic. The variance in heteroscedastic distribution is different across elements while in homoscedastic case the variance is at the same range for all elements.

The most advanced statistical tests are for normally distributed data. In the case of non-normally distributed data there are some options. First, although normal distribution is rather rare event, methods designed for this distribution do not demand to have a perfect fit.

Second, certain statistical tests for normally distributed data are robust and can thus be applied to non-normal distributions. These tests include t-tests (1-sample, 2-sample, paired), ANOVA and some others.

Use of non-parametric tests is a third option. However, the number of these tests is limited and their power and specificity are generally considered to be lower than for parametric tests.

Fourth, it may be possible to convert the data to normal distribution by performing a transformation. A mathematical function is then applied to each point in the data set. If done, it is important to keep in mind that interpretations apply only to the converted data.

Statistical methods are needed also for the analysis of poikilosis. As poikilosis means heterogeneity, an interval of values has to be considered instead of a single number per measured entity. Available methods are not well suited for this purpose. Intervals at lagom extent could possibly in some studies be reduced to single values allowing use of traditional statistical methods. New experimental and statistical approaches will be needed to fully investigate poikilosis and its extent.

7.2 | Outliers

Many types of biases and errors lead to occurrence of spurious data points that are often called for outliers. It is essential to find out if there is a measurement error, typically a random error, preferably by rerunning the analysis, although this is not always possible. Repeated experiments can reveal whether there is an outlier, measurement error or a wrong assumption of the data. The numbers of outliers are reduced by performing multiple replications. The presence of many putative outliers may indicate heavy-tailed distribution and that the measurements considered as deviant are not outliers.

Treatment of outliers can have a substantial effect on descriptive statistics. Thus, it is important to have consistent definition for them. Their numbers should be compared to the expected number per sample size and distribution, for example, as an estimate with binomial distribution for normally distributed data set. Several model-based methods have been developed for oulier detection. All these methods are subjective and depend on some assumptions. They also require normally distributed data and none of them considers effects of poikilosis.

There are two ways to deal with the outliers. Either they are deleted or trimmed. They can be transformed with winsorization to the nearest non-outlier value or be replaced by another estimate, such as median or mean or utilizing regression model by imputation. Transformation of data can be a solution in some cases.

From poikilosis point of view, definition of outliers is equally important. There is a possibility that putative outliers are true observations and an extreme value is for another lagom state or it is entirely non-lagom but real observation. Only controlled experiments with perturbations can explain such cases.

8 | MEASURING AND DEFINING LAGOM

When measuring poikilosis it is essential to identify what is lagom extent of heterogeneity in the investigated condition. When measurement escapes from the lagom extent, the system may be in transition state from one lagom level to another (Figure 6). Lagom is context dependent, thus measuring conditions have to be harmonized and then systematically perturbed. By perturbing controlled measurement conditions, it will be possible to define lagom extent of poikilosis. The system is at lagom state when measurements after perturbation stay within certain limits. After excessive perturbation, the system escapes the lagom level and further increased perturbations have ever increasing effects.

The capability of a system to stay at lagom level is due to systemic and active mechanisms that restrict variation extent. In biology these are called TARAR mechanisms[14] and they limit extent of variation in normal conditions. Despite perturbations the measured entity remains within certain range that indicates the extent of lagom. The interval of lagom poikilosis is determined with repeated measurements along with different extents of perturbation. Measurements within lagom extent are biologically equal. Once perturbation is too large for TARAR mechanisms to restrict or small enough, the system either enters to another lagom level with higher or lower extent of heterogeneity or goes to non-lagom level (Figure 6).

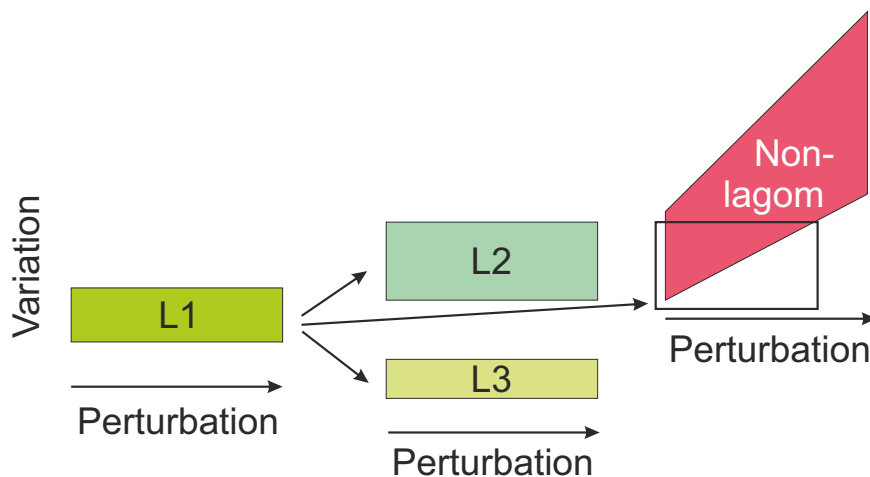Studies of intra-and inter-individual variations of biomarkers in healthy subjects are examples of collected

FIGURE 6 Measuring lagom. System stays at lagom state L1 with the help of active and passive TARAR mechanisms despite perturbations. Once the perturbation is large enough, the system leaves state L1 and moves either to another lagom state (L2 or L3) or becomes non-lagom when the extent increases along increased perturbation. TARAR countermeasures limit the extent of variation at lagom levels L1 to L3. The box in the non-lagom indicates region where the measurements are of lagom L2 extent

heterogeneity data, see also Table 1. Reference values used e.g. in clinical diagnosis are defined in this way by measuring a sample of population. It is possible that such results refer to one or several lagom states, this cannot be estimated from the data items as the data collection is not systematic enough and does not include perturbation.

Many of the reported biomarker reference values are based on quite small populations and some of them are outdated.[36] Inter-individual variation is wide for many biomarkers and many times close to population variation, such as in the analysis of 21 hematological parameters.[37] In addition to cohort/population variation there are day-to-day and within-day variations, which can be large, exemplified by circulating cell-free DNA levels[38] and inflammation biomarkers in type 2 diabetes patients and healthy controls,[39] respectively.

Transition states, even in well controlled systems, are difficult to measure because the transition can be very quick from one lagom level to another. Measurements in such cases are mainly for one or the other lagom state. It is likely that the lagom extents for different lagoms of a system can overlap, but the extreme ends of the distributions are either higher or lower than for the previous lagom state, see L2 and L3 in Figure 6. This can be seen when the interval of measurements is different from the measurements in the previous lagom state.

A system is at non-lagom situation when increased perturbations lead to ever increasing values for the measurand. This can be because the system may not be able to return to lagom extent of poikilosis without external actions, as in medicine, or changes in the environment, etc. The consequences of non-lagom heterogeneity depend on the level, effects to the system, how linked the level is to others, and so on. Non-lagom is widely variable and increases along increased perturbation, that is, TARAR countermeasures cannot control the extent of variation (Figure 6).

Only for cases with very large deviations from lagom, a single reliable measurement can indicate non-lagom extent. Therefore, it is important to follow the progress in measured parameter over time, for example, for patients, and to perform perturbations and remeasurements, when possible. Excessive non-lagom heterogeneity leads to disease in organisms, possibly to impaired functionality and eventually to death. Once non-lagom extent is large enough, the system cannot return to normal state without some kind of intervention. In medicine, this can be established with a treatment, for example, with a drug that reduces the extent of poikilosis and reconstitutes the system to normal extent of heterogeneity.[40] Curative treatment is not available for all non-lagom variations and diseases caused by them.

Non-lagom variation affects other connected levels, such as molecules, processes, pathways, cells or tissues and when excessive causes some of the connected levels to enter to non-lagom level. Then, in the most severe instances, a domino-like effect follows and leads to systemic disease and even to death.

Relevant tools and approaches at systems level to follow, control, perturb, and measure poikilosis simultaneously at many levels are largely missing. Once available, systems biological studies would provide highly interesting multidimensional information and possibilities to understand and simulate systems at all levels of poikilosis. Method development would be necessary to make full use of such data.

9 | DOCUMENTATION OF EXPERIMENTS

Poikilosis and its effects cannot be properly considered and understood unless experiments are comprehensively

described. Without full details even those studies that are solid cannot be retested. At the moment, most published articles include insufficient details to allow reproduction of the analyses and comprehension of poikilosis and its origins. Among 268 biomedical publications with experimental data there was only one (0.37%) that reported the full protocol.[41]

Guidelines and recommendations have been published for description of many types of experiments and computational approaches. These guidelines, checklists and instructions need to consider how to cover and explain poikilosis. The documentation of experiments and measurements has to include all the relevant details for samples, their collection, selection and treatment, for measurements and instruments used for this purpose, followed by the full account of data analysis and computational approaches.

Several minimum reporting requirements and other guidelines are available from FAIRsharing at https://fairsharing.org/. Some guidelines will be discussed in here; however, there are many more available – and to follow. The Biological Variation Data Critical Appraisal Checklist (BIVAC) is systematics for variation description in clinical medicine samples.[42] There are altogether 14 items to be included. The European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) has released a checklist for the assessment of publications of biological variation data.[43] The Standards for Reporting Diagnostic Accuracy (STARD) provides instructions how to systematically describe relevant details in publications.[44] The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) Statement and its extensions are instructions for how to conduct and disseminate observational studies [45] and the CONsolidated Standards of Reporting Trials (CONSORT) 2010 statement explains systematics for reporting randomized trials.[46] Enhancing the QUAlity and Transparency of health Research (EQUATOR) Network distributes more than 400 reporting guidelines for various types of studies, evaluations, reviews, protocols etc.[47]

All the guidelines mentioned above are for studies in human or with human samples. Many of these guidelines are equally relevant for studies on other organisms. There are some specific instructions for researchers working on other organisms, those for veterinary clinical pathology[48] and Animals in Research Reporting *In Vivo* Experiments (ARRIVE) guidelines for animal research[49] being examples.

Thus, instructions and guidelines for proper reporting are available for many types of studies and can be applied also to related investigations. The problem is that these guidelines are not followed or just some part of them are followed, animal studies being an example where progress in systematic description has been slow despite availability of recommendations for a decade.[24] As replication of poikilosis studies demand for even more detailed description of the methods and study protocols than currently customary, it is essential to increase the quality of publications in this respect. Most journals allow supplements where the details can be distributed. Systematic reporting with proper metadata will facilitate also computerized data analyses.

Even when full descriptions of methods and approaches are available, repeatability may be difficult to achieve in different laboratories. Some 100,000 *Caenorhabditis elegans* worms were needed to reproduce studies on effects of drug-like molecule treatments in three laboratories [50] despite very close collaboration and harmonization of approaches, joint purchase of reagents etc. For example, differences in temperature fluctuations of incubators and lab benches had significant effects. Internal replication of studies at US Defense Advanced Research Projects Agency (DARPA) indicated numerous factors ranging from reagents to flow rates etc. as instrumental for reproducibility.[51] The authors present a list of items to follow including documentation of reagents, following how experiment is done by other researcher, stating ranges instead of single numbers (i.e. poikilosis), testing cells before shipping, double checking protocols, having a designated person for communication and keeping data analysis software and pipelines up to date. Replication studies demand for close and intense collaboration, communication and attention to all details.

Special attention has to be paid to the choice of research objects. Misidentification of cell lines is a common problem in literature. ICLAC Register of Misidentified Cell Lines version 10 (iclac.org/databases/cross-contaminations/) lists 537 misidentified cell lines where no authentic stock is known and 71 cell lines where authentic stock is known to exist. Publications based on or using these cell lines are of suspect. Over 32,000 articles report studies on these cell lines and they were estimated to be cited half a million times.[52] Literature is thus littered with irrelevant and irreproducible studies and citations to them.

10 | DISCUSSION

As poikilosis affects all systems, it should be regarded when designing and conducting measurements. Current systems, practices and methods have not been designed and optimized for this purpose. Although some of them may be appropriate for the task in some cases, there is need for new solutions throughout the measurement process and data analysis.

Interval arithmetic could be used for poikilosis-related calculations of heterogeneity in some mathematical operations. Probabilistic and fuzzy logic are potential approaches as they provide mathematical means to represent vagueness, uncertainty and imprecise information, that is, poikilosis data. Fuzzy logic is already used for clinical diagnosis of some diseases[53,54] as well as in several other applications in medicine and life sciences.[55-57]

When heterogeneity in a system is not properly treated, researchers make overoptimistic estimates of the significance of results due to statistical significance filter.[58] In such cases, the magnitude of effect can be exaggerated in small studies. Consideration of poikilosis throughout experimental studies and measurements demands for new types of analysis methods and statistical analyses. Some of the existing methods can be used, especially when there is very low level of heterogeneity; however, inclusion of more extensive poikilosis will only be possible with new analytical methods.

Consideration of poikilosis in experimental design affects the choice of the study type as well as how to measure and estimate the heterogeneity. Therefore, existing methods for evaluating required sample sizes have to be updated. To cover intrinsic poikilosis of systems, it is likely that larger samples are needed than currently estimated by the tools. Poikilosis need to be included also to meta-analyses where the hierarchy of evidence should be based on evidence not just on the types of experimental designs with which the measurements are obtained.

Measurements of poikilosis provide a continuum for the investigated phenomena. Still, many systems and predictions are considered to have binary dichotomy. Pathogenicity model (PM)[59] is an example of how to deal with and interpret continuous data. PM displays the continuum in healthy-diseased spectrum.[59] For example, heterogeneity among individuals with similar medical condition provides means for making diagnosis and to stratify individuals to drug responders and non-responders, and possibly to identify and stratify those who could benefit from a drug and those at risk of adverse drug reactions.

Poikilosis-aware predictors have to accommodate for heterogeneity and continuum in the data based on which the tool is developed, as well as in the interpretation of predictions. This can be achieved in several ways. Still, most existing predictors are binary and thus too simplistic. There are some prediction methods that account for the heterogeneity, for example, by having more than two classes in variant effect prediction. PON-P2 variant tolerance/pathogenicity predictor achieves this by having three categories for benign, disease-related and cases of unclassified, unsure or heterogenous outcome.[60] Another example of a predictor geared towards inclusion of poikilosis is PON-PS method for phenotypic severity that classifies variants into benign, mild/moderate and severe disease-causing groups.[61]

It will be necessary to rethink also quality control approaches to include poikilosis. By considering poikilosis at different levels and stages of quality assessment process, consequences of poikilosis could be understood. External Quality Assurance Systems (EQASs) are widely used for laboratory accreditation. Although highly standardized, even these schemes should consider and test for effects of poikilosis.

Different experimental designs can be used for different purposes. Hypothesis generation and hypothesis testing and verification require different experiments, amounts of data etc. It is essential to consider whether the tested system is at lagom level, otherwise the results are difficult or impossible to compare and interpret.

Measurements are involved in various types of experiments. Exploratory observations can be validated by larger studies to confirm the findings. Many measurements are done for analytical purposes, or to aid in predictions or for optimization of processes. Many initial studies are hypothesis-generating by nature. Their replication could be done in different ways as suggested by.[62] They propose a multi-tiered approach including exact, close or direct replication experiments followed by extended replications for partial, systematic or differential, or conceptual replication. Quasi replications e.g. in different species further validate the original study. Inclusion of poikilosis in experiments and experimental design significantly contributes to reproducibility provided that large enough studies are performed. As experiences from animal studies show,[23] the experiment/measurement setup should be tested by introducing various types of heterogeneity.

Poikilosis accounts for uncertainty in measurements and occurrence of intervals in measurements. It is a normal phenomenon but largely neglected. Measurements of poikilosis demand for novel approaches in experimental design, conduction of experiments, data analysis and statistics. Without including poikilosis in measurements, obtained results cannot be fully understood nor are they reliable. What needs to be done depends on the case. Consideration of poikilosis can open new ways to solve controversies and enigmas and can allow novel insight into systems, processes, mechanisms and reactions and their interpretation, understanding and manipulation as well as increase reproducibility of measurements and experiments.

CONFLICT OF INTEREST
The authors have stated explicitly that there are no conflicts of interest in connection with this article.

ORCID
Mauno Vihinen https://orcid.org/0000-0002-9614-7976

REFERENCES
1. Vihinen M. Poikilosis - pervasive biological variation. *F1000Res*. 2020;9:602.
2. Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res*. 2015;116:116-126.
3. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature*. 2012;483:531-533.

4. Psychology. Estimating the reproducibility of psychological science. *Science* (New York, N.Y.). 2012;349, aac4716.

5. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712.

6. Haibe-Kains B, El-Hachem N, Birkbak NJ, et al. Inconsistency in large pharmacogenomic studies. *Nature*. 2013;504:389-393.

7. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533:452-454.

8. Baker M. Reproducibility crisis: blame it on the antibodies. *Nature*. 2015;521:274-276.

9. Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. The fickle P value generates irreproducible results. *Nat Methods*. 2015;12:179-185.

10. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess*. 2010;14:iii, ix–xi, 1–193.

11. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.

12. Motulsky HJ. Common misconceptions about data analysis and statistics. *J Pharmacol Exp Ther*. 2014;351:200-205.

13. JCGM. (2008) JCGM 100:2008 GUM 1995 with minor corrections. Evaluation of measurement data — Guide to the expression of uncertainty in measurement.

14. Vihinen M. Functional effects of protein variants. *Biochimie*. 2020;180:104-120.

15. Tal E. Measurement in science. Zalta EN. *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). 2020. https://plato.stanford.edu/archives/fall2020/entries/measurement-science/

16. Scott D, Suppes P. Foundational aspects of theories of measurement. *J Symbolic Logic*. 1958;23:113-128.

17. Michell J. *An Introduction To the Logic of Psychological Measurement*, L. Erlbaum Associates; 1990.

18. Savage CW, Ehrlich P. *Philosophical and Foundational Issues in Measurement Theory*. L. Erlbaum Associates. 1992.

19. Mitchell DJ, Tal E, Chang H. The making of measurement: editors' introduction. *Stud Hist Philos Sci*. 2017;65–66:1-7.

20. JCGM. JCGM 200:2008 International vocabulary of metrology - Basic and general concepts and associated terms (VIM). 2008.

21. Webster MM, Rutz C. How STRANGE are your study animals? *Nature*. 2020;582:337-340.

22. Bell S. Measurement good practice guide No. 11 (issue 2). A beginner's guide to uncertainty of measurement. 2001.

23. Voelkl B, Altman NS, Forsman A, et al. Reproducibility of animal research in light of biological variation. *Nat Rev Neurosci*. 2020;21:384-393.

24. Enserink M. Sloppy reporting on animal studies proves hard to change. *Science* (New York, N.Y.). 2017;357:1337-1338.

25. Reckelhoff JF, Alexander BT. Reproducibility in animal models of hypertension: a difficult problem. *Biol Sex Differ*. 2018;9:53.

26. Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol*. 2018;16:e2003693.

27. Graham AL. Naturalizing mouse models for immunology. *Nat Immunol*. 2021;22:111-117.

28. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014, Mr000034.

29. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71-72.

30. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924-926.

31. Blunt CJ. *Hierarchies of evidence in evidence-based medicine*. PhD degree, The London School of Economics and Political Science; 2015.

32. Yavchitz A, Ravaud P, Altman DG, et al. A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *J Clin Epidemiol*. 2016;75:56-65.

33. Krauss A. Why all randomised controlled trials produce biased results. *Ann Med*. 2018;50:312-322.

34. Micceri T. The unicorn, the normal curve, and other improbable creatures. *Physchol Bull*. 1989;105:156-166.

35. Blanca MJ, Arnau J, Lopez-Montiel D, Bono R, Bendayan R. Skewness and kurtosis in real data samples. *Methodology*. 2013;9:78-84.

36. Carobene A. Reliability of biological variation data available in an online database: need for improvement. *Clin Chem Lab Med*. 2015;53:871-877.

37. Coşkun A, Carobene A, Kilercik M, et al. Within-subject and between-subject biological variation estimates of 21 hematological parameters in 30 healthy subjects. *Clin Chem Lab Med*. 2018;56:1309-1318.

38. Madsen AT, Hojbjerg JA, Sorensen BS, Winther-Larsen A. Day-to-day and within-day biological variation of cell-free DNA. *EBioMedicine*. 2019;49:284-290.

39. Mallard AR, Hollekim-Strand SM, Ingul CB, Coombes JS. High day-to-day and diurnal variability of oxidative stress and inflammation biomarkers in people with type 2 diabetes mellitus and healthy individuals. *Redox Rep*. 2020;25:64-69.

40. Vihinen M. Strategy for disease diagnosis, progression prediction, risk group stratification and teatment - Case of COVID-19. *Front Med (Lausanne)*. 2020;7:294.

41. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible research practices and transparency across the biomedical literature. *PLoS Biol*. 2016;14:e1002333.

42. Aarsand AK, Røraas T, Fernandez-Calle P, et al. The biological variation data critical appraisal checklist: A standard for evaluating studies on biological variation. *Clin Chem*. 2018;64:501-514.

43. Bartlett WA, Braga F, Carobene A, et al. A checklist for critical appraisal of studies of biological variation. *Clin Chem Lab Med*. 2015;53:879-885.

44. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.

45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61:344-349.

46. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18.

47. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Open Med*. 2008;2:e49-50.

48. Freeman KP, Baral RM, Dhand NK, Nielsen SS, Jensen AL. Recommendations for designing and conducting veterinary clinical pathology biologic variation studies. *Vet Clin Pathol*. 2017;46:211-220.

49. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. 2010;8:e1000412.

50. Lithgow GJ, Driscoll M, Phillips P. A long journey to reproducible results. *Nature*. 2017;548:387-388.

51. Raphael MP, Sheehan PE, Vora GJ. A controlled trial for reproducibility. *Nature*. 2020;579:190-192.

52. Horbach S, Halffman W. The ghosts of HeLa: how cell line misidentification contaminates the scientific literature. *PLoS One*. 2017;12:e0186281.

53. Thukral S, Rana V. Versatility of fuzzy logic in chronic diseases: a review. *Med Hypotheses*. 2019;122:150-156.

54. Arji G, Ahmadi H, Nilashi M, et al. Fuzzy logic approach for infectious disease diagnosis: a methodical evaluation, literature and classification. *Biocybern Biomed Eng*. 2019;39:937-955.

55. Nobile MS, Votta G, Palorini R, et al. Fuzzy modeling and global optimization to predict novel therapeutic targets in cancer cells. *Bioinformatics (Oxford, England)*. 2020;36:2181-2188.

56. Sheehan T, Bachelet D. A fuzzy logic decision support model for climate-driven biomass loss risk in western Oregon and Washington. *PLoS One*. 2019;14:e0222051.

57. Liu F, Heiner M, Gilbert D. Fuzzy Petri nets for modelling of uncertain biological systems. *Brief Bioinform*. 2018:bby118.

58. Loken E, Gelman A. Measurement error and the replication crisis. *Science (New York, N.Y.)*. 2017;355:584-585.

59. Vihinen M. How to define pathogenicity, health, and disease? *Hum Mutat*. 2017;38:129-136.

60. Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS One*. 2015;10(2):e0117380.

61. Niroula A, Vihinen M. Predicting severity of disease-causing variants. *Hum Mutat*. 2017;38:357-364.

62. van der Staay FJ, Arndt SS, Nordquist RE. The standardization-generalization dilemma: a way out. *Genes Brain Behav*. 2010;9:849-855.

63. Tuanmu M-N, Jetz W. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecol Biogeograph*. 2015;24:1329-1339.

64. Hunting ER, van Soest RWM, van der Geest HG, Vos A, Debrot AO. Diversity and spatial heterogeneity of mangrove associated sponges of Curaçao and Aruba. *Contr Zool*. 2008;77:205-215.

65. Fukuchi S, Sakamoto S, Nobe Y, et al. IDEAL: intrinsically disordered proteins with extensive annotations and literature. *Nucleic Acids Res*. 2012;40:D507-511.

66. Mannil D, Vogt I, Prinz J, Campillos M. Organ system heterogeneity DB: a database for the visualization of phenotypes at the organ system level. *Nucleic Acids Res*. 2015;43:D900-906.

67. Huan Q, Zhang Y, Wu S, Qian W. HeteroMeth: a database of cell-to-cell heterogeneity in DNA methylation. *Genomics Proteomics Bioinformatics*. 2018;16:234-243.

68. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29:308-311.

69. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD vol 2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32:557-563.

70. Sarkar A, Yang Y, Vihinen M. Variation benchmark datasets: update, criteria, quality and applications. *Database*. 2020:baz117.

71. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46:D1062-d1067.

72. Xu H, Wang Y, Lin S, et al. PTMD: A database of human disease-associated post-translational modifications. *Genomics Proteomics Bioinformatics*. 2018;16:244-251.

73. Li D, Hsu S, Purushotham D, Sears RL, Wang T. WashU epigenome browser update 2019. *Nucleic Acids Res*. 2019;47:W158-w165.

74. Picardi E, D'Erchia AM, Lo Giudice C, Pesole G. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res*. 2017;45:D750-d757.

75. Niu G, Zou D, Li M, et al. Editome Disease Knowledgebase (EDK): a curated knowledgebase of editome-disease associations in human. *Nucleic Acids Res*. 2019;47:D78-D83.

76. Kim P, Yang M, Yiya K, Zhao W, Zhou X. ExonSkipDB: functional annotation of exon skipping event in human. *Nucleic Acids Res*. 2019.

77. Hyung D, Kim J, Cho SY, Park C. ASpedia: a comprehensive encyclopedia of human alternative splicing. *Nucleic Acids Res*. 2018;46:D58-d63.

78. Falster DS, Duursma RA, Ishihara MI, et al. BAAD: a biomass and allometry database for woody plants. *Ecol*. 2015;96:1445.

79. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Twist GP, Klein TE, Miller NA. The evolution of PharmVar. *Clin Pharmacol Ther*. 2019;105:29-32.

80. Whirl-Carrillo M, McDonagh EM, Hebert JM, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther*. 2012;92:414-417.