Draft GUIDANCE DOCUMENT ON GOOD *IN VITRO* METHOD PRACTICES (GIVIMP) FOR THE DEVELOPMENT AND IMPLEMENTATION OF *IN VITRO* METHODS FOR REGULATORY USE IN HUMAN SAFETY ASSESSMENT

FOREWORD

A guidance document on Good *In Vitro* Method Practices (GIVIMP) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment was identified as a high priority requirement. The aim is to reduce the uncertainties in cell and tissue-based *in vitro* method derived predictions by applying all necessary good scientific, technical and quality practices from *in vitro* method development to *in vitro* method implementation for regulatory use.

The draft guidance is coordinated by the European validation body EURL ECVAM and has been accepted on the work plan of the OECD test guideline programme since April 2015 as a joint activity between the Working Group on Good Laboratory Practice (GLP) and the Working Group of the National Coordinators of the Test Guidelines Programme (WNT).

The draft document prepared by the principal co-authors has been sent in September 2016 to all 37 members of the European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL[1]) and has been subsequently discussed at the EU-NETVAL meeting on the 10th of October 2016.

By November/December 2016 the comments of the OECD Working Group on GLP and nominated experts of the OECD WNT will be forwarded to EURL ECVAM who will incorporate these and prepare an updated version. A second round of commenting shall be concluded in the beginning of 2017. EURL ECVAM shall then prepare the final GIVIMP version, which will be submitted to OECD for proposed adoption at the OECD Joint Meeting in April 2017.

[1] https://eurl-ecvam.jrc.ec.europa.eu/eu-netval

Field Code Changed

TABLE OF CONTENTS

Field Code Changed

Field Code Changed

Field Code Changed

Glossary of important terms used in the Guidance Document[2]

Acceptance criteria	Criteria for when study results can be accepted, i.e. a set of well-defined parameters describing aspects of the *in vitro* method such as control and reference item output, acceptable range for positive and negative controls, etc. These should primarily be established based on information from existing data on the finalised *in vitro* method or described in relevant bibliographic data. However, relevant development data obtained on a version of the *in vitro* method equal to the one used for generating the study results, as well as historical data, may need to be taken into account where available.
Adverse outcome pathway (AOP)	An AOP represents the existing knowledge concerning the causal linkages between the Molecular Initiating Events (MIEs) and the cascade of intermediate or key events at subcellular, cellular, tissue, organ, whole animal, and population levels that lead to a specific adverse outcome.
Apoptosis	Process of programmed cell death generally characterised by distinct morphological characteristics and energy-dependent biochemical mechanisms. Apoptosis is considered a vital component of various processes including normal cell turnover, proper development and functioning of the immune system, hormone-dependent atrophy, embryonic development and chemical-induced cell death.
Archive	A designated area or facility (e.g. cabinet, room, building or computerised system) for the secure storage and long term, permanent retention of raw data, completed data and relevant metadata in its final form and records for the purposes of reconstruction of the activity or study. It also enables audits.
Assay	A defined laboratory procedure that produces results. Also known as testing method. An assay can be considered as a technical operation that consists of determination of one or more characteristics of a given product, process or service according to a specified procedure. Often an assay is part of an experiment. The testing result can be qualitative (yes/no), categorical, or quantitative (a measured value). It can be a personal observation or the output of a precision measuring instrument. Usually the testing result is the dependent variable, the measured response based on the particular conditions of the testing method or the level of the independent variable. Some testing methods, however, involve changing the independent variable to determine the level at which a certain response occurs: in this case, the testing result is the independent variable.
Batch	A specific quantity or lot of a test item or reference item, test system, assay reagent or other consumable, produced during a defined cycle of manufacture in such a way that it could be expected to be of a uniform character and should be designated as such.
Benchmark dose (BMD) or concentration (BMC)	A point of reference by which something can be measured. The benchmark dose/concentration approach was developed as an alternative to the use of No Observed Adverse Effect Level (NOAEL) and Lowest Observed Adverse Effect Level (LOAEL).
Best practice	A method or technique that has consistently shown results superior to those achieved with other means, and that is used as a benchmark. The term is also used to describe the process of developing and following a standard way of doing things that multiple organisations can use.
Between-laboratory assessment	Phase in which different operators from different laboratories perform (or run) the *in vitro* method independently to establishes whether or not an *in vitro* method can be successfully established in different laboratories.

[2] All terms and their descriptions should be considered as working definitions for the purpose of this Guidance Document only.

Biokinetics	Time-course of a chemical (substance and mixture) and its metabolites in a living organism, i.e., increase or decrease of substance concentration at the site of measurement due to transport or due to formation or breakdown. The term "toxicokinetics" is also often used synonymously.
Biological pathway	A number of biochemical steps, linked together in a certain order, with a start and an end. Some typical types of biological pathways are metabolic pathways and signalling pathways.
Blank control/untreated control	Separate untreated part of a test system that is kept under the original culture conditions; the untreated control provides baseline data of the test system under the conditions of the *in vitro* method. It provides the background response from the test system, obtained by treatment with only the buffer or media used for the administration of test, control and reference items. The primary purpose of a blank or untreated control is to trace sources of artificially introduced variation on the *in vitro* method results.
Bovine spongiform encephalopathy (BSE)	Bovine spongiform encephalopathy, commonly known as mad cow disease, is a fatal neurodegenerative disease (encephalopathy) in cattle that causes a spongy degeneration of the brain and spinal cord.
Carcinogenicity	The property of any agent (chemical, physical or biological agent) directly involved in causing cancer (carcinogen). Carcinogenicity results in an increased incidence of tumours, increased proportion of malignant tumours or a reduction in the time to appearance of tumours, compared with concurrent control groups. The process of carcinogenesis involves the transition of normal cells into cancer cells via a sequence of stages that entail both genetic alterations (i.e. mutations) and non-genetic events.
Coefficient of variation (CV)	A measure of spread that describes the amount of variability relative to the mean. Because the coefficient of variation is per definition unrelated to the magnitude of the mean and also unitless, it can be used instead of the standard deviation to compare the spread of data sets that have different units or different means.
Comparative genomic hybridisation analysis (aCGH)	A molecular cytogenetic method for analysing copy number variations relative to ploidy level in the DNA of a test sample compared to a reference sample, without the need for culturing cells. The aim of this technique is to quickly and efficiently compare two genomic DNA samples arising from two sources, which are most often closely related, because it is suspected that they contain differences in terms of either gains or losses of either whole chromosomes or subchromosomal regions (a portion of a whole chromosome).
Computerised systems	A group of hardware components and associated software designed and assembled to perform a specific function or group of functions.
Cytotoxicity	General cytotoxicity (or basal cytotoxicity) is the result of toxic effects on structures and functions common to all cells of the body, such as DNA, chromosomes, mitochondria, the cytoskeleton and various membranes. A large number of general (basal) cytotoxicity tests have been developed for a wide range of purposes. Tissue-specific cytotoxicity involves adverse effects on particular types of differentiated cells, in terms of either their tissue-type specific properties (e.g. hormone production, conductivity, contractility) or their modulation of general cytotoxic effects (e.g. via the metabolic activation or detoxification of xenobiotics). For this purpose target organ(system)toxicity tests have been developed.
Data	Information derived or obtained from raw data (e.g. a reported analytical result)
Data Governance	The total sum of arrangements to ensure that data, irrespective of the format in which it is generated, is recorded, processed, retained and used to ensure a complete, consistent and accurate record throughout the data lifecycle
Data Integrity	The extent to which all data are complete, consistent and accurate throughout the data lifecycle.

Data Lifecycle All phases in the life of the data (including raw data) from initial generation and recording through processing (including transformation or migration), use, data retention, archive / retrieval and destruction.

Design qualification (DQ), installation qualification (IQ), operational qualification (OQ), performance qualification (PQ) Verification of *in vitro* method equipment usually consists of design qualification installation qualification, operational qualification and performance qualification. When acquiring a new piece of equipment in an *in vitro* environment, design specifications are needed for the intended use of the equipment. Installation qualification is the first step in the process to ensure that the equipment will live up to its expectations. Operational qualification verifies that the equipment is achieving its operational requirements. For the performance qualification, the last phase, the equipment will run several times under normal operating conditions and its functions will be challenged.

ECVAM DataBase service on ALternative Methods (DB-ALM) A database aimed at implementing the communication and dissemination strategy of the Joint Research Centre on animal alternatives, as requested by the European Commission and the European Parliament.

Effective concentration 50 (EC_{50}) and Inhibition concentration 50 (IC_{50}) For *in vitro* cell and tissue culture work the terms effective concentration 50 (EC_{50}) and inhibition concentration 50 (IC_{50}) are used, in analogy to median effective dose (ED_{50}) and median lethal dose (LD_{50}) used in animal experiments. IC_{50} is used in case of an *in vitro* method where there is a decline in read-out. IC_{50} is therefore the test item concentration causing 50% inhibition of the desired activity. EC_{50} is used for read-outs that increase with concentration. EC_{50} is therefore the concentration causing 50% of maximum effect for any measured biological effect of interest.

Emulsion A stable dispersion of liquid droplets in another liquid, where the two are immiscible.

Engelbreth-Holm-Swarm (EHS) A mouse sarcoma which is a rich source of both individual basement membrane components and Matrigel often used in cell and tissue culture work.

European Chemicals Agency (ECHA) Agency of the European Union (EU) that manages technical, scientific and administrative aspects of EU chemicals legislation, notably the regulation on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).

European Directorate for the Quality of Medicines & HealthCare (EDQM) Organisation that is responsible for the European Pharmacopoeia and the European biological standardisation programme.

European Food Safety Authority (EFSA) Agency of the European Union that provides independent scientific advice in the fields of food and feed safety, animal health and welfare, plant protection and plant health and communicates on existing and emerging risks associated with the food chain.

European Medicines Agency (EMA) Agency of the European Union that is responsible for the protection of public and animal health through the scientific evaluation and supervision of medicines.

European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL) A network of highly qualified laboratories to (1) respond to some of the provisions of Directive 2010/63/EU, (2) generate *in vitro* method information that is reliable, relevant and based on current best quality and scientific practices, (3) increase the European Commission's validation capacity of *in vitro* methods and (4) provide a laboratory network knowledgeable on the routine implementation of good *in vitro* method practices for regulatory use in human safety assessment.

Foetal bovine serum (FBS) Fetal bovine serum derived from clotted blood is the most widely used growth supplement for cell and tissue culture media because of its high content of embryonic growth promoting factors. When used at appropriate concentrations it

	may supply many defined and undefined components that have been shown to satisfy specific metabolic requirements for the culture of cells and tissues.
Genetically modified micro-organisms (GMMs),	A micro-organism in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination.
Good cell culture practice (GCCP) and GCCP2.0	Guidelines developed in 2005 to define minimum standards in cell and tissue culture work. This GCCP guidance lists a set of six principles intended to support best practice in all aspects of the use of cells and tissues *in vitro*, and to complement, but not to replace, any existing guidance, guidelines or regulations. GCCP2.0 is the updated version which is currently being drafted.
Good Laboratory Practice (GLP)	A quality system applied to the conduct of non-clinical health and environmental safety testing that is intended for submission to regulatory authorities in support of the registration, licensing or regulation of chemical and related products. It concerns the organisational process and the conditions under which these studies are planned, performed, monitored, recorded, archived and reported. It ensures uniformity, consistency, reproducibility, quality and integrity of chemical non-clinical safety tests.
Hazard	An intrinsic feature of a stressor (e.g. chemical or physical in nature) to cause harm or adverse effects to human health and to the environment. It is a qualitative (for example in the case of classifications) or quantitative expression of the adverse effects elicited by a test item under defined conditions of exposure.
High performance liquid chromatography (HPLC)	High performance liquid chromatography (or high-pressure liquid chromatography) is a chromatographic technique that can separate a mixture of compounds when in solution and is used in biochemistry and analytical chemistry to identify, quantify and purify the individual components of the mixture.
High-efficiency particulate arrestance (HEPA)	High-efficiency particulate arrestance, also sometimes called high-efficiency particulate arresting or high-efficiency particulate air, is a type of air filter used to create an aseptic environment, through retention of a certain number of particles, depending of the category of the filter. Aseptic conditions are required for *in vitro* cell and tissue culture work to ensure reliability an reproducibility of results obtained.
High-throughput screening (HTS)	A high throughput method for scientific experimentation, relevant to the fields of biology and chemistry, and especially used in drug discovery. HTS involves an automated-operation platform, data processing and control software. HTS allows a researcher to quickly conduct many biochemical, genetic or pharmacological tests and obtain a large number of information from a single experiment.
In silico	The term *in silico* refers to the technique of performing experiments via computer simulations. Examples are structure-activity relationships (SAR) and quantitative structure-activity relationships (QSAR)
In vitro	The term *in vitro* (Latin for "in the glass") refers to the technique of performing a given experiment in a test tube, or, more generally, in a controlled environment outside of a living organism.
***In vitro* method endpoint**	Defined as quantitative measurable characteristics that serve as indicators of a pathologic process or related biochemical or molecular events, e.g. measured absorbance in a cytotoxicity assay or a skin irritation *in vitro* method.
***In vitro* to *in vivo* extrapolation (IVIVE)**	*In vitro* to *in vivo* extrapolation refers to the qualitative or quantitative transposition of experimental results or observations made *in vitro* to predict phenomena *in vivo*, i.e. in whole organisms.
In vivo	

	The term *in vivo* ("within the living") refers to experimentation using a whole, living organism as opposed to a partial or dead organism, or an *in vitro* controlled environment. Animal testing and clinical trials are two forms of *in vivo* research.
Inhibitor or spiked up control	Mix of test item and positive control to assess any effect of inhibition of the test item on the test system endpoint measurements.
Integrated testing strategies (ITS)	Integrated testing strategies enable to significantly increase the use of non-animal testing information for regulatory decision making, and thus to minimise the need for animal testing. To this end, operational procedures are developed, tested and disseminated that guide a transparent and scientifically sound evaluation of test items in a risk-driven, context-specific and substance-tailored manner. The envisaged decision theory framework includes alternative methods such as chemical and biological read-across, *in vitro* results, *in vivo* information on analogues, qualitative and quantitative structure-activity relationships, thresholds of toxicological concern and exposure-based waiving.
Intellectual property rights (IPR)	Intellectual property refers to creations of the mind: inventions; literary and artistic works; and symbols, names and images used in commerce. Industrial Property includes patents for inventions, trademarks, industrial designs and geographical indications.
International Uniform ChemicaL Information Database (IUCLID)	A software application designed to capture, store, maintain and exchange data on intrinsic and hazard properties of chemicals (substances and mixtures). It is essential for chemical industry to comply with the new legislation which entered into force on 1 June 2007. The freely downloadable tool will assist chemical companies globally in fulfilling their obligation to submit data to the Agency under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation from 1 June 2008.
Limit of detection (LOD), Lower limit of quantification (LLOQ) and Upper limit of quantification (ULOQ)	The Limit of detection and the upper and lower limits of quantification are important parameters that need to be determined during *in vitro* method development. The LOD is the lowest amount of test item in a sample that can be detected with (stated) probability, although not quantified as an exact value. The LLOQ and ULOQ are the lowest and highest amounts of test item in a sample that can be quantitatively determined with a stated acceptable precision and accuracy, under stated experimental conditions. The usefulness and optimal throughput of an *in vitro* method may depend on the appropriate determination of the LOD and the ULOQ and LLOQ.
Lipophilicity	The ability of a chemical (substance and mixture) to dissolve in non-polar environments such as oils, lipid membranes, and non-polar solvents such as hexane or toluene.
Mass spectrometry (MS)	Analytical technique that measures the mass-to-charge ratio of charged particles. It is used for determining masses of particles, for determining the elemental composition of a sample or molecule, and for elucidating the chemical structures of molecules such as peptides and other chemical compounds.
Maximum average score (MAS)	The maximum average Draize score was utilised as the primary quantitative measurement of eye irritation potential in rabbits.
Metadata	Metadata is data that describe the attributes of other data, and provide context and meaning. Typically, these are data that describe the structure, data elements, inter-relationships and other characteristics of data. It also permits data to be attributable to an individual.
Micro-organism	Any microbiological entity, cellular or non-cellular, capable of replication or of transferring genetic material, including viruses, viroids, animal and plant cells in culture.
Minimal essential medium (MEM)	Developed by Harry Eagle, is one of the most widely used of all synthetic cell culture media for *in vitro* cell and tissue culture work.

Minimum significant ratio (MSR)	Parameter that can be used to quantify assay reproducibility and resolution (the smallest ratio between compound potency which can be detected in the *in vitro* method).
Mixture[2]	A combination of two or more chemicals (liquid or solid) that do not react with each other
Multi-component test chemicals[3]	Mixtures comprising a complex mix of individual test chemicals with different solubility and physical-chemical properties. In most cases, they can be characterised as a homologous series of test chemicals with a certain range of carbon chain length/number or degree of substitution. These materials are frequently referred to as "complex mixtures". However, in this guidance document, these are referred to as "multi-component test chemicals".
Mutual Acceptance of Data (MAD)	The OECD Mutual Acceptance of Data is a multilateral agreement which states that test data generated in any member country in accordance with OECD Test Guidelines and GLP shall be accepted in other member countries for assessment purposes and other uses relating to the protection of human health and the environment. The application of MAD avoids unnecessary and costly duplication of testing as well as non-tariff barriers to trade. In addition, it saves laboratory animals used for *in vivo* testing.
Nanomaterials	A natural, incidental or manufactured material containing particles, in an unbound state or as an aggregate or as an agglomerate and where, for 50 % or more of the particles in the number size distribution, one or more external dimensions is in the size range 1 nm - 100 nm.
Negative control	Separate part of a test system treated with an item for which it is known that the test system should not respond; the negative control provides evidence that the test system is not responsive under the actual conditions of the assay.
Omics	Omics is a general term for a broad discipline of science and engineering for analysing the interactions of biological information objects in various omes (these include genome, transcriptome, proteome, metabolome, expressome, and interactome). Some examples of 'Omics' technologies: - genomics - proteomics - metabolomics - transcriptomics
Particulates	Also known as particulate matter (PM), fine particles and soot, are tiny subdivisions of solid matter suspended in a gas or liquid. In contrast, aerosol refers to particles and/or liquid droplets and the gas together. Sources of particulate matter can be man-made or natural. Air pollution and water pollution can take the form of solid particulate matter or be dissolved. Salt is an example of a dissolved contaminant in water, while sand is generally a solid particulate.
Physiologically based pharmacokinetic, physiologically based toxicokinetic, physiologically based biokinetic (PBPK/PBTK/PBBK)	Physiologically based toxicokinetic, or alternatively referred to as physiologically based pharmacokinetic or biokinetic models, are quantitative descriptions of absorption, distribution, metabolism, and excretion (ADME) of synthetic or natural chemical substances in humans and other animal species. PBTK models are increasingly being used as an effective tool for designing toxicology experiments and for conducting extrapolations essential for risk assessments (e.g. in pharmaceutical research and drug development, and in health risk assessment for cosmetics or general chemicals).
Polymerase chain reaction (PCR)	Polymerase chain reaction is a molecular biology *in vitro* technique. Using the natural ability of DNA polymerase to synthesise a new strand of DNA

[3] Consideration is given to the consistency with the definition of "multi-component test chemicals" (or "complex test chemicals") in Chapter 1 of a draft "Guidance Document on the Use of the Harmonized System for the Classification of Chemicals which are Hazardous for the Aquatic Environment" (ENV/JM/HCL(2000)11).

	complementary to the offered template strand, millions of copies of a specific DNA sequence are made.
Positive control	Separate part of the test system treated with an item the response to which is known for the test system; the positive control provides evidence that the test system is responsive under the actual conditions of the assay.
Prediction Model	The method by which the *in vitro* endpoint value(s) is used to predict the *in vivo* equivalent activity (i.e., degree of toxicity).
Proficiency chemicals	A panel of chemicals with known and statistically well-defined responses in a particular *in vitro* method. These are items used e.g. to verify that a laboratory carries out the *in vitro* method correctly, or to validate alternative newly developed test systems intended for use with the particular *in vitro* method.
Quality assurance	A system of procedures, checks, audits, and corrective actions to ensure that all research, testing, monitoring, sampling, analysis, and other technical and reporting activities are of the highest achievable quality.
Quality assurance programme	A defined system, including personnel, which is independent of study conduct and is designed to assure test facility management of compliance with GLP.
Quality control	Documented activity which seeks to confirm that starting materials for assays (including cell lines) and key stages of individual assays and the final results, meet prescribed specifications. It should be clear that this is not the same as quality assurance. ISO 9000 defines quality control as "part of quality management focused on providing confidence that quality requirements will be fulfilled".
Quality management system (QMS)	Can be expressed as the organisational structure, procedures, processes and resources needed to implement quality management. GLP specifically refers to a quality system of management controls for test facilities and organisations to try to ensure the uniformity, consistency, reliability, reproducibility, quality, and integrity of test item non-clinical safety tests. Of all QMS regimes, the ISO 9000 family of standards is probably the most widely implemented worldwide.
Raw data	Original records and documentation, retained in the format in which they were originally generated (i.e. paper or electronic), or as a 'true copy'. Raw data must be contemporaneously and accurately recorded by permanent means. In the case of basic electronic equipment which does not store electronic data, or provides only a printed data output (e.g. balance or pH meter), the printout constitutes the raw data.
Reagent	Term used for media additives, compounds added to a system to induce a chemical reaction, anything added to get the *in vitro* method or related assays to work etc.
Reference item	Any chemical (substance and mixture) or product used to provide a basis for comparison with the test item reference items are used e.g. during method development to verify that the method classifies test items correctly, or in every experimental run of a finalised method to verify system performance according to the acceptance criteria.
Relevance	The term "Relevance" describes whether a procedure is meaningful and useful for a particular purpose.
Reliability	The term "Reliability" describes whether a procedure can be performed reproducibly within and between laboratories and over time.
Replace, Reduce, Refine (3Rs)	3Rs is the short for "Replace, Reduce, Refine". A term describing current internationally accepted strategies for minimising suffering of laboratory animals used in experimental research. The optimal solution is to replace the test method requiring animal experiments with one or several *in vitro* methods; if this is not possible at least it might be possible to modify the methods so that it is possible to Reduce the number of animals being used in each study without compromising data quality; if this is also not possible it might at least be possible to Refine the

	test method so that experiments are conducted in a way minimising stress and other impact on the animals.
Robustness	The insensitivity of test results to departures from the specified test conditions when conducted in different laboratories or over a range of conditions under which the test method might normally be used. If a test is not robust, it will be difficult to use in a reproducible manner within and between laboratories.
Safe Harbour	The international Safe Harbour Privacy Principles or Safe Harbour Privacy Principles were principles which were overturned on October 24, 2015 by the European Court of Justice, which enabled some US companies to comply with privacy laws protecting European Union and Swiss citizens. US companies storing customer data would self-certify that they adhere to 7 principles, to comply with the EU Data Protection Directive and with Swiss requirements. The US Department of Commerce developed privacy frameworks in conjunction with both the European Union and the Federal Data Protection and Information Commissioner of Switzerland. Within the context of a series of decisions on the adequacy of the protection of personal data transferred to other countries, the European Commission made a decision in 2000 that the United States' principles complied with the EU Directive - the so-called "Safe Harbour decision". However, after a customer complained that his Facebook data were insufficiently protected, the European Court of Justice declared in October 2015 that the Safe Harbour Decision was invalid, leading to further talks being held by the Commission with the US authorities towards "a renewed and sound framework for transatlantic data flows". The European Commission and the United States agreed to establish a new framework for transatlantic data flows on 2nd February 2016, known as the "EU-US Privacy Shield".
Saturation concentration	The maximum dissolved concentration of a test chemical that can be achieved under the test conditions.
Sensitivity	A measure of *in vitro* method performance that describes the proportion of all evaluated test items that are classified as positive for a particular toxicological endpoint, which are predicted as positive by the actual *in vitro* method.
Service level agreement (SLA)	A contract between a service provider (either internal or external) and the end user that defines the level of service expected from the service provider.
Short tandem repeat (STR)	Short Tandem Repeat (STR) DNA profiling, is used to uniquely identify human cell lines derived from the tissue of a single individual allowing researchers to ascertain if their cultures were misidentified or cross-contaminated.
Signal windows (SW)	A measure of the separation between the sample (positive control) and the blank (negative control/solvent) including the variability of both measurements.
Single nucleotide polymorphism analysis (aSNP)	Single nucleotide polymorphism or SNP (pronounced snip) analysis is a technique to detect a DNA sequence variation occurring when a single nucleotide - A, T, C, or G - in the genome (or other shared sequence) differs between members of a species (or between paired chromosomes in an individual). For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide.
Solid phase microextraction (SPME)	Is a solid phase extraction sampling technique that involves the use of a fiber coated with an extracting phase, that can be a liquid (polymer) or a solid (sorbent), which extracts different kinds of analytes (including both volatile and non-volatile) from different kinds of media.
Solubility limit in water	The maximum attainable concentration or concentration at thermodynamic equilibrium between aqueous pure phase and solid (or liquid or gaseous) pure phase.
Specificity	A measure of *in vitro* method performance that describes the proportion of all evaluated test items that are classified as negative for a particular toxicological endpoint, which are predicted as negative by the actual *in vitro* method.

Standard deviation (SD)	A measure that is used to quantify the amount of variation or dispersion of a set of data values.
Standard operating procedure (SOP)	A documented procedure which describes how to perform tests or activities normally not specified in detail in study plans or test guidelines.
Structure-activity relationships and quantitative structure-activity relationships (SAR/QSAR)	Structure-activity relationships and quantitative structure-activity relationships, collectively referred to as (Q)SARs, are simplified mathematical representations of complex chemical-biological interactions that can be used to predict the physicochemical and biological properties of molecules.
Study plan	A document which defines the objectives and experimental design for the conduct of the study, and includes amendments (i.e. an intended change to the study plan after the study initiation date).
Suspension	A stable dispersion of solid particles in a liquid.
Test item	A chemical (substance and mixture) or product that is the subject of a study.
Test pre-submission form (TPF) and Test submission template (TST)	For the evaluation of the readiness of an *in vitro* method to enter the EURL ECVAM validation process, the method needs to be officially submitted to EURL ECVAM by compiling as a first step the electronic version of the test pre-submission form to allow a preliminary assessment of the status of development, optimisation and/or validation of an *in vitro* method and its potential relevance with regard to the 3Rs (replacement, reduction, refinement of animal testing). If this step is satisfactory a complete submission is requested which requires the compilation of a detailed test submission template.
Test system	A test system means any biological, chemical or physical system or a combination thereof used in a study. *In vitro* test systems are mainly biological systems (e.g. cells or tissues), although some of the more recent developments in alternatives to conventional *in vivo* testing (e.g., gene arrays for toxicogenomics) may also exhibit some attributes of physical-chemical test systems, and still others, e.g., toxicometabonomics, may mainly rely on analytical methodology. Test kits, including proprietary test kits, should also be considered as test systems.
Testing method	The term testing method is used to describe both a published, well-evaluated test guideline method (e.g. OECD) and a not-fully-developed method soon ready for submission to validation bodies. A testing method is a process or procedure used to obtain information on the characteristic of a substance or agent. Toxicological testing methods generate information regarding the ability of a substance or agent to produce a specific biological effect under specified conditions.
Toxicological endpoint	A direct marker of progression to an adverse outcome - e.g. morphological or physiological changes, functional impairments, disease symptoms or death - used to describe an adverse health effect (or a probability of that adverse effect) resulting from exposure to a test item. The test system response to an exposure of a test item may be measured by a series of endpoints. The most sensitive endpoint (critical endpoint) is the one that occurs at the lowest exposure level and associated with an adverse response (committed step).
Training Set	The set of test items used to develop the prediction model for an assay. The training set items should have strong reference data (i.e., values from a recognised regulatory assay) against which the *in vitro* assay endpoint values can be compared.
Untreated control	Separate untreated part of a test system that is kept under the original culture conditions; the untreated control provides baseline data of the test system under the conditions of the assay.
Validation	Validation is the process by which the reliability and relevance of a procedure are established for a specific purpose.

Validation set (test items)	The set of test items used to assess the predictive capacity of an *in vitro* method based on the performance of the endpoint values by the reference test results. Testing of the validation test items set is a principal part of *in vitro* method validation.
Vehicle or solvent control	The separate part of a test system to which the vehicle (i.e. solvent) for the test item is added without the test item; the vehicle control provides evidence for a lack of influence of the chosen vehicle on the test system under the actual conditions of the *in vitro* method.
Within-laboratory assessment	Phase in which different operators from the same laboratory perform (or run) the *in vitro* method independently and at different times to establish whether or not an *in vitro* method can be successfully established in one laboratory.
Xenobiotic	A chemical foreign to the biological system, structurally distinct from endogenous compounds present within the biological system.
Z-factor	A measure of the separation between control and sample signal which takes into account the dynamic range of the *in vitro* method and the data variation associated with the signal and control measurements. It is suitable for *in vitro* method quality assessment.

Introduction

The availability of a guidance document on Good *In vitro* Method Practices (GIVIMP) for "The development and implementation of *in vitro* methods for regulatory use in human safety assessment" was identified by the scientific and regulatory community (OECD) as a high priority requirement to reduce the uncertainties in cell and tissue-based *in vitro* method derived predictions.

An Expert Group was therefore established to develop such a guidance document. The first draft guidance document was prepared following a GIVIMP meeting on the 24th and 25th of February 2015 in Ispra, Italy (see Annex 3) with additional input from experts who could not be present at the meeting. For this first draft version expert input was received from EURL ECVAM, European receiving authorities (European Food Safety Authority EFSA, European Medicine Agency EMA, the European Chemicals Agency ECHA), from the European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL, e.g. from the Belgian, Dutch, Italian, Spanish and Swedish laboratories), from ECVAM's Stakeholder Forum (ESTAF, e.g. the European Society of *In vitro* Toxicology), from the EU and OECD Working Group on GLP (e.g. delegates from Belgium, The Netherlands, The United Kingdom, Poland, Italy, France, Singapore), from 3Rs Centres (Centre for Alternatives to Animal Testing, CAAT), from regulatory agencies (e.g. RIVM), from scientists from large industries and SMEs and from international scientists with expertise in stem cells, cell biology, Good Laboratory Practice (GLP) and *in vitro* methods.

The main authors drafting this first version are:

Maria Antonietta Antonelli[1], Maria Astridou[2], Bohumila Bichlmaier Suchanová[3], Camilla Bernasconi[4], Ann-Charlotte Bostroem[4], Gerard Bowe[4], Sandra Coecke[4], Thomas Cole[4], Hui Da Chong[5], Rita Cortvrindt[6], Laureano Cuevas[7], Agnieszka Dudra[8], Eugene Elmore[9], Chantra Eskes[10], Kristina Fant[11], Andrew Gray[12], John Harbell[13], Minne Heringa[14], Delyan Ivanov[15], Guido Jacobs[16], Rob Jaspers[17], Elisabeth Joossens[4], Ingrid Langezaal[4], Marcel Leist[18], Thomas Lucotte[19], Richard McFarland[20], Marisa Meloni[21], Anne Milcamps[4], Emma Pedersen[11], Laura Pioppo[1], Maurits-Jan Prinz[22], Annamaria Rossi[2], Laura Rossi[3], Glyn N Stacey[23], Emanuela Testai[24], Frederique van Acker[25], Sander van der Linden[4], Maurice Whelan[4], Iwona Wilk-Zasadna[26]

[1]European Medicines Agency (EMA), London, United Kingdom.
[2]European Food Safety Authority (EFSA), Parma, Italy.
[3]European Chemicals Agency (ECHA), Helsinki, Finland.
[4]European Commission, Directorate General Joint Research Centre, Directorate F-Health, Consumers and Reference Materials, F.3 Chemicals Safety and alternative Methods, The European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), Ispra, Italy.
[5]Singapore GLP Monitoring Authority, SPRING Singapore – Singapore Accreditation Council.
[6]Vito, Mol, Belgium.
[7]Centro Nacional de Microbiologia, Majadahonda, Spain.
[8]Bureau for Chemical Substances, Lodz, Poland.
[9]University of California Irvine, Trabuco Canyon, Unites States.
[10]SeCAM, Magliaso, Switzerland.

[11]SP Sveriges Tekniska Forskningsinstitut, Borås, Sweden
[12]United Kingdom GLP Monitoring Authority, The Medicines and Healthcare products Regulatory Agency (MHRA), London, United Kingdom
[13]JHarbell Consulting LCC, Dallas, United States.
[14]RIVM, The National Institute for Public Health and the Environment, Centre for Safety of Substances and Products, Bilthoven, The Netherlands.
[15]University of Nottingham, Cancer Biology Queens Medical Centre, Nottingham, United Kingdom.
[16]WIV-ISP, Brussels, Belgium.
[17]Health Care Inspectorate, Ministry of Health, Welfare and Sport , Utrecht, The Netherlands.
[18]University of Konstanz, Konstanz, Germany.
[19]ANSM, Saint-Denis, France.
[20]Policy Office of Cellular, Tissue, and Gene Therapies Center for Biologics Evaluation and Research United States Food and Drug Administration, USA
[21]Vitroscreen SRL, Milano, Italy.
[22]European Commission, Directorate General Internal Market, Industry, Entrepreneurship and SMEs, Brussels, Belgium.
[23]United Kingdom Stem Cell Bank, Advanced Therapies Division, NIBSC, Medicines and Healthcare products Agency, United Kingdom.
[24]Istituto Superiore di Sanita, Rome, Italy.
[25]Triskelion, Zeist, The Netherlands.
[26]Novartis Vaccines, Basel, Switzerland.

Acknowledgements are also made to the EURL ECVAM colleagues that reviewed this first version and gave useful additional input[4].

This first draft, currently for revision by the OECD Working Group on GLP and the nominated experts from the WNT, has been be circulated in September for review by all 37 members of the European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL). An updated version revised draft will be submitted at the beginning of January for a second OECD commenting round. EURL ECVAM will then prepare the final version for proposed adoption at the OECD meeting in April 2017.

[4] Susanne Belz, Mounir Bouhifd, Laura Gribaldo, Tomislav Horvat, Tracey Holley, Annett Janusch Roi, Roman Liska, Alfonso Lostia, Agnieszka Lidia Swiatek, Francesca Pistollata, Anna Price, Clemens Wittwehr, Andrew Worth.

Scope

There is a community desire for non-animal methods, but regulators demand validated and internationally accepted *in vitro* methods (i.e. OECD test guidelines or ISO standards). To accommodate the desires of regulatory authorities, a number of *in vitro* methods, often based on the use of human cells and tissues, were submitted to international validation bodies during the last two decennia. However, the experience gained during these validations revealed that many *in vitro* methods need serious improvements in design, robustness and reliability before they can be successfully implemented in a routine laboratory environment and generate data sets which can be used to support regulatory decisions. Therefore, OECD approached EURL ECVAM to coordinate the issuing of a guidance on Good *In Vitro* Method Practices (GIVIMP) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment. The major goal of GIVIMP consists of improving the reliability and robustness of *in vitro* methods, reducing the uncertainties of *in vitro* based predictions and therefore increasing the acceptance of the *in vitro* estimated safety measures by regulatory agencies. The scope of the GIVIMP guidance is taking into account good scientific, technical and quality practices, to ensure that the overall process, starting from *in vitro* method development up to the final *in vitro* method implementation for regulatory use becomes more efficient and effective.

This guidance document targets all players involved in the process, e.g. *in vitro* method developers, *in vitro* test system producers, validation bodies, producers of equipment, materials and reagents, *in vitro* method end-users such as EU-NETVAL test facilities, testing laboratories, large industries and small to medium enterprises as well as receiving authorities, monitoring authorities, accreditation bodies and OECD. The guidance aims to further facilitate the application of the OECD Mutual Acceptance of Data agreement for data generated by *in vitro* methods and as such contribute to avoidance of unnecessary duplicate testing. This guidance describes the areas related to *in vitro* method development, standardisation, harmonisation, and international acceptance that would benefit from more detailed scientific, technical and quality guidance.

The GIVIMP document has been written with different end users in mind, including GLP routine test facilities but also research laboratories developing new *in vitro* methods. In the latter case it is obvious that the laboratory will not be able to fully comply with this document on all points. However, it is necessary to comply with a set of good practices within the *in vitro* method life cycle so as not to jeopardise the acceptance and routine use of the *in vitro* method in a regulatory environment.

This guidance is not intended to duplicate or replace existing OECD guidance or advisory documents but rather it is complementary, addresses specific gaps and aims to collect available references and information on best scientific, technical and quality practices in one document. GIVIMP takes into account the requirements of the existing OECD guidelines and advisory documents to ensure that the guidance is complementary and fully in line with these documents.

This document is divided into 10 sections covering:

1 Responsibilities
2 Quality considerations
3 Facilities
4 Apparatus, material and reagents
5 Test systems

1 Responsibilities

1.1 *In vitro* method developers

In vitro methods are often developed without the primary aim of being used for regulatory purposes, but are rather focused on the discovery of disease pathways or investigation of mechanisms of action induced by external factors causing cell disturbance. However, these *in vitro* methods in development can form the basis for *in vitro* methods for specific toxicity endpoints, during drug and/or other chemicals safety assessment or for toxicity screening during product quality control processes.

Researchers aiming to develop *in vitro* methods suitable for regulatory testing purposes must be aware that beyond the 'short-term' repeatability, that is consonant with the good scientific work required in discovery, the quality principles for test acceptance by receiving authorities have additional requirements (OECD, 2005a). To harmonise and speed up the validation process and accelerate the acceptance of new *in vitro* methods by receiving authorities, the *in vitro* method developer should keep in mind that the quality of historical data and documentation regarding the *in vitro* method submitted will have a significant impact on the validation process.

Briefly, the *in vitro* method developer is responsible for providing a clearly written and well documented *in vitro* method description, and related standard operating procedure(s) (SOP(s)), considering all aspects described in the present guidance document.

The developer's knowledge and understanding of the *in vitro* method is the basis for establishing an approach to control the *in vitro* method and to set for instance adequate acceptance criteria for the results obtained when running an *in vitro* method.

In vitro method developers should:

- Understand the sources of variation of the *in vitro* method
- Detect the presence and degree of variation in the results
- Understand the impact of variation of the *in vitro* results on the related predictions
- Control the variation in a manner to make a sound, relevant and reliable *in vitro* method

Each developer should judge whether he or she has gained sufficient understanding of the *in vitro* method to provide a high degree of assurance to successfully propose the *in vitro* method for regulatory applications.

In vitro method developers should also take into account the Intellectual Property (IP) guidelines regarding test systems as set out on the OECD website (http://www.oecd.org/). Proposals for projects aiming at the development of new test guidelines should provide information on Intellectual Property Rights (IPR) aspects, as transparently as possible. In particular, the following information is expected to be provided: "Describe if the *in vitro* method includes components, equipment or other scientific procedures that are covered (or pending) by IPR (e.g., patents, patent applications, industrial designs and trademarks) and/or intended to remain confidential. Information should be provided on the overall availability of the IPR-protected components including whether they are commercially available or require a Material Transfer Agreement (MTA) or other licensing agreements. In addition, the possibility of providing a generic description of the IPR-covered component/test system as well as any other element intended to remain confidential should be disclosed and whether Performance Standards have been developed for the *in vitro* method."

In vitro methods proposed for regulatory use should not contain elements that are confidential to the extent that this impedes adequate scientific validation of the mechanistic relevance of the method.

When *in vitro* method developers conclude that their *in vitro* method is sufficiently developed, they can then proceed to an in-house validation. When such internal assessment is successful, they can submit the *in vitro* method to a validation body for the formal validation of the method, or, can organise the validation by themselves. In order to have the *in vitro* method considered for regulatory acceptance, *in vitro* method developers will have to contact their national coordinator to develop a project proposal for a new Test Guideline. Project proposals for new Test Guidelines need the active support of regulatory authorities in at least one member country, and have to meet a regulatory need in member countries.

In conclusion, at the end of the test development process, the *in vitro* method developers should be able to prove that the *in vitro* method they offer to the validation body is robust, reliable, relevant, and supported by high quality data as described in the present guidance.

1.2 Test system providers

In vitro test systems are mainly biological systems, quite often consisting of tissues or cell lines. It is difficult, if not impossible, to identify cell lines from different origins based solely on morphology and or culture characteristics. Infection or contamination of a cell line with an adventitious virus or mycoplasma may significantly change the characteristics of the cells but again such contamination may not be evident. Cell lines will also change with time in culture, and to add to all these natural hazards it is all too easy to incorrectly label or cross-contaminate different cell lines in a busy cell culture laboratory (ECACC Handbook – Fundamental Techniques for ECACC Cell Lines[5]).

The OECD GLP recommends that test system providers should adhere to a formal quality system, such as International Standard ISO 9001, and particularly Part 1 - Specification for Design/Development, Production, Installation and Servicing.

The test system provider must also provide all relevant safety information, in compliance with national and international regulations, for the transport, use and disposal, including containment in the case of an accident.

The Guidance on Good Cell Culture Practice: A Report of the Second ECVAM Task Force on Good Cell Culture Practice (GCCP) (Coecke *et al.*, 2005) provides a minimal set of information that is essential when working with cells or tissues of animal or human origin (Table 1), while the OECD GLP No 14 (The Application of the Principles of GLP to *in vitro* Studies (OECD, 2004a)) states that the characterisation of test systems can be directly fulfilled by information from the supplier.

[5] http://www.sigmaaldrich.com/life-science/cell-culture/learning-center/ecacc-handbook.html

Field Code Changed

Table 1: Examples of requirements for documentation concerning the origins of cells and tissues

	Isolated organs and tissues of animal origin (e.g. rat brain tissue)	Primary cultures of animal origin (e.g. rat hepatocytes)	All materials of human origin (e.g. cord blood)	Cell lines (e.g. Balb/c, 3T3)
Ethical and safety issues	+	+	+	Applicable, if human or involving recombinant DNA or pathogens
Species/strain	+	+	+	+
Source	+	+	+	+
Sex	+	+	+	+
Age	+	+	+	+
Number of donors	+	+	If applicable	na
Health status	+	+	+	+
Any special pre-treatment	+	+	+	+
Organ/tissue of origin	+	+	+	+
Cell type(s) isolated	+	+	+	+
Isolation technique	+	+	+	+
Date of isolation	+	+	+	+
Operator	+	+	+	+
Supplier	+	+	+	+
Informed consent	na	na	+	If human, may be applicable
Material transfer agreement	na	na	+	+
Medical history of donor	na	na	+ (if available)	If human, may be applicable (if available)
Pathogen testing	If applicable[a]	If applicable[a]	+[a]	+[a]
Shipping conditions	+	+	+	+
State of material on arrival	+	+	+	+
Cell line identification and authentication	na	na	na	+
Mycoplasma testing	na	na[b]	na[b]	+

[a]*Screening tests for animal colonies or donors of cells and tissue may be appropriate.*

[b]*May be important if material is preserved for longer term use (e.g. as feeder layers for other cultures).*

na = not applicable.

1.3 Validation bodies

In general, it is the mandate of national and international governmental validation agencies (ECVAM, ICCVAM, JACVAM, etc.) to promote and facilitate *in vitro* method validation for regulatory acceptance to replace or reduce *in vivo* animal testing.

The validation body's responsibility is to contribute to both an effective validation process and to *in vitro* method quality. The basic principle of validation is that an *in vitro* method should be produced that is fit for its intended use. The validation process consists of collection and evaluation of data, from the *in vitro* method design stage to the availability of a routine method, which establishes scientific evidence that the *in vitro* method is capable of consistently delivering quality and scientifically relevant data for the specific purpose it is designed for. Ideally the following conditions exist:

- Quality, safety, and efficacy are designed or built into the *in vitro* method.
- Quality and a sound scientific basis of the *in vitro* method are assured during the entire *in vitro* method life cycle towards regulatory acceptance.
- Each step towards a routine running of the *in vitro* method is controlled, and documented to assure that the *in vitro* method meets all scientific and quality attributes.

The validation process involves a series of activities taking place retrospectively and/or prospectively over the lifecycle of the *in vitro* method once submitted to a validation body.

Taking into account the above conditions, test developers can submit their *in vitro* methods to a validation body and ask for support in the validation process.

For example, the entire EURL ECVAM test submission process[6] follows 2 mandatory steps:

> *Step 1: Pre-submission - is based on the electronic submission of the EURL ECVAM Test Pre-submission Form (TPF). The TPF will allow EURL ECVAM to perform a preliminary assessment of the status of development of the in vitro method and the potential relevance to the 3Rs (replacement, reduction, refinement of animal testing).*
>
> *Step 2: Complete Submission - requires the compilation of a detailed Test Submission Template (TST). The TST will be made available by EURL ECVAM after successful conclusion of step 1 and the test submitter will be formally invited to complete it and providing also a SOP in the ECVAM DataBase service on ALternative Methods (DB-ALM) format.*

On the basis of a test submission, a validation body can make a final decision as to whether the submitted *in vitro* method qualifies for entering the validation process.

In "Practical aspects of designing and conducting validation studies involving multi-study trials" (Coecke *et al.* 2016), details are given that focus on practical aspects of conducting prospective *in vitro* validation studies by laboratories that are EU-NETVAL laboratories. Prospective validation studies within EU-NETVAL comprise multi-study trials involving several laboratories or "test facilities" and typically consist of two main steps:

> (1) The design of the validation study by EURL ECVAM and
>
> (2) The execution of the multi-study trial by a number of qualified laboratories within EU-NETVAL coordinated and supported by EURL ECVAM.

The approach adopted in the conduct of these validation studies adheres to the principles described in the OECD guidance document on the Validation and International Acceptance of new or updated *in vitro* methods for Hazard Assessment No. 34 (OECD, 2005a). The (Coecke *et al.*, 2016) paper, mainly focuses on the processes followed to carry out a prospective validation of *in vitro* methods involving different laboratories with the ultimate aim of generating a dataset that can support the development of an international test guideline (e.g. by the OECD) or the establishment of performance standards of *in vitro* methods.

Upon successful validation by validation bodies or other entities, an *in vitro* method can be presented to the OECD for regulatory acceptance. Once *in vitro* methods are consolidated within an OECD test guideline (TG), data produced by using those methods are mutually accepted by all receiving authorities.

[6] https://eurl-ecvam.jrc.ec.europa.eu/test-submission

Field Code Changed

1.4 Producers of equipment, materials and reagents

When performing established testing methods (validated or not), the test results can only be accepted if the equipment, materials and reagents (test and reference items, media additives, compounds added to a system to induce a chemical reaction, etc.) used, are of proven quality as established by formal testing or evaluation procedures. To be able to prove this, test end users will need to work with preferred suppliers who are selected on predefined criteria (e.g. controlled transport, technical support, assured delivery, batch selection allowed, ISO certification present, etc.). The characteristics of the supplied materials and reagents should be appropriately documented in adequate quality documents such as a certificate of analysis, batch release certificate or similar. Suppliers who cannot fulfil these criteria have to be excluded from delivering products used to run *in vitro* methods, as the quality of test data cannot be guaranteed nor be acceptable for receiving bodies. It is the responsibility of the test end-user to verify and assure the quality of the products and materials (see chapter 2.4).

1.5 *In vitro* method end-users

As the aim of this document is to provide guidance for *in vitro* methods for regulatory use in human safety assessment, the majority of end-users will be GLP compliant test facilities. If a safety study is conducted for regulatory purposes, compliance with the principles of GLP is required. GLP test facilities are covered by national and/or international GLP regulations and must adhere to the responsibilities as defined in these regulations.

The OECD Mutual Acceptance of Data (MAD) is a multilateral agreement which states that test data generated in OECD countries and full adherent countries – (Argentina[7], Brazil, India, Malaysia, South Africa and Singapore)[8] in accordance with OECD Test Guidelines and the OECD Principles of GLP shall be accepted in other member countries by regulatory bodies for assessment purposes and other uses relating to the protection of human health and the environment[9]. However, end-users should also document their competency to perform a test in compliance with a specific TG, by running the proficiency chemicals and demonstrating the overall quality assurance system of the test facility (see Section 6.5/8.6).

Non-GLP *in vitro* method users can profit from the use of the GIVIMP guidance. In these cases no regulations exist and no responsibilities are defined. It is highly recommended to apply all necessary good scientific, technical and quality practices that the guidance describes so as to reduce the uncertainties in the use of cell and tissue-based *in vitro* method: Examples of critical importance described in this document relate to guidance on test systems such as the documentation concerning the origins of cells and tissues (see Table 1), the evaluation of the performance of the *in vitro* method, adequate measures to ensure test item exposure and test item-test system compatibility, the guidance for suppliers of equipment, materials and reagents, evaluation of competence and training for executing a particular *in vitro* method, etc. Appropriate accreditation (ISO, 2015, 2005) may be requested or recommended in some other cases.

In case no specific Test Guideline harmonised template is available, all generated test data should be submitted in an easily readable format to facilitate the decision making process of

[7] Full adherence for Argentina only applies to industrial chemicals, pesticides and biocides

[8] http://www.oecd.org/env/ehs/non-memberadherentstotheoecdsystemformutualacceptanceofchemicalsafetydata.htm

[9] http://www.oecd.org/env/ehs/mutualacceptanceofdatamad.htm

Field Code Changed

Field Code Changed

the risk assessors, preferentially according to the OECD guidance document for describing non-guideline *in vitro* methods Series on Testing and Assessment No. 211 (OECD, 2014).

Data generated by non-validated *in vitro* methods, if submitted as supporting data, need to be accompanied by proof of relevance and scientific validity of the selected test system together with the description of the critical points of the test system used, including sensitivity, limits of detection etc.

1.6 Receiving authorities

Receiving authorities receive non-clinical safety data as part of regulatory submissions and they must ensure that the legal requirements are met. Receiving authorities in Europe include the European Chemicals Agency (ECHA), European Medicines Agency (EMA), European Food Safety Authority (EFSA), as well as various national agencies that are responsible for assessing safety data. Receiving authorities in the United States include the Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA).

The responsibility of the receiving authorities is to check that test data are obtained according to available OECD TG and guidance documents and they use the data accordingly in their evaluations and provisions in law. For GLP studies, they may verify whether the responsible test facility has been found in compliance by a national monitoring authority (see below) or request a study audit if the facility has not been subject to a GLP inspection by a national monitoring authority. The receiving authorities may additionally indicate to *in vitro* method developers where they see needs for new or better tests, and to validation bodies which tests deserve priority in the validation.

The majority of European regulation requiring toxicological data, allow or even encourage the use of alternative methods (e.g. Regulation (EC) No 1107/2009 for plant protection products, Regulation (EU) No 528/2012 for biocides, Regulation (EC) No 1907/2006 for industrial chemicals, known as REACH) (Heringa *et al.*, 2014). Regulation (EC) No 1223/2009 for cosmetics does not allow any *in vivo* animal testing.

As a result of these developments European, but also other international regulatory bodies, tend to accept data generated by alternative methods, including validated *in vitro* methods and, in specific cases, also non-validated *in vitro* methods, especially as supportive information or when mechanistic data are needed. They have adopted the 3Rs principles and are now proactively supporting the use and implementation of alternative methods[10].

The European Medicines Agency (EMA) has recently drafted a number of documents expressing their vision and action plans towards the implementation of the 3R principles (EMA, 2014). According to the new EMA draft guideline on regulatory acceptance of 3Rs methods, multiple and flexible approaches are considered acceptable to demonstrate scientific validity of new testing approaches and their fitness for regulatory use, either as pivotal, supportive or as exploratory mechanistic studies. Besides established formal validation processes by recognised institutions such as the Centres for the Validation of Alternative Methods (CVAMs) and The European Directorate for the Quality of Medicines & HealthCare, the EMA draft guideline is offering a new approach for submission and evaluation of a proposal for regulatory acceptance of 3R testing approaches via an EMA in-house procedure. One interesting option of this process is the collection of real-life data by using a new 3R testing approach in parallel with the data generated using existing methods.

[10] https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/in-vitro-methods

Field Code Changed

Data generated with the new method will however not be used as part of the regulatory decision making process ("safe harbour") and will be used solely for the purpose of evaluation of the new method for possible future regulatory acceptance.

1.7 GLP monitoring authorities

For studies conducted for regulatory purposes, the responsibility for evaluating the results of the study lies with the regulatory reviewer at the receiving authority. However, this evaluation can only be effective if the study data can be relied upon. The principles of Good Laboratory Practice (GLP) ensure that the quality and scientific integrity of the data can be demonstrated and the conduct of the study reconstructed.

> *GLP was developed in the 1970s in response to fraudulent scientific safety studies that were submitted to regulatory authorities in support of applications for the regulatory registration/approval of drugs to FDA. Subsequently the principles of GLP were developed by the OECD to ensure data reliability and reconstructability of safety studies. The principles apply to all non-clinical health and environmental safety studies required by regulations for the purpose of registering or licensing chemical products of various kinds. The principles have been published in 1981 as an annex to the OECD Council Decision on MAD*[11]*. The decision states that 'data generated in the testing of chemicals in an OECD Member Country in accordance with the OECD Test Guidelines (Annex I of this decision) and OECD Principles of Good Laboratory Practice (Annex II of this decision) shall be accepted in other member countries for purposes of assessment and other uses relating to the protection of man and the environment'. Since 1981 a number of additional guidance, consensus and advisory documents have been published in the OECD Series on Principles of GLP*[12]*.*

> *The Advisory Document of the OECD Working Group on GLP n° 14 specifically addresses in vitro Studies* (OECD 2004b)*:"The purpose of this document is to facilitate the proper application and interpretation of the GLP Principles for the organisation and management of in vitro studies, and to provide guidance for the appropriate application of the GLP Principles to in vitro studies, both for test facilities (management, QA, archivist, study director and personnel), and for national GLP compliance monitoring authorities."*

> *In the European Union, the principles of GLP are included in Directive 2004/10/EC, while the compliance monitoring procedures are included in Directive 2004/9/EC. GLP provisions are included in legislation for chemicals, human medicinal products, veterinary products, detergents, feed additives, food additives, genetically modified food or feed, pesticides, biocides and cosmetics (Coecke et al. 2016). The current European medical device directives do not require GLP, but the harmonised standard series ISO 10993 require that all in vitro tests "shall be conducted according to recognised current/valid best laboratory/quality practices, for example Good Laboratory Practice (GLP) or ISO/IEC 17025". However, the current proposal for a new Medical Device Regulation does require GLP where applicable.*

The role of GLP Monitoring Authorities (MAs) may vary considerably from region to region, however their responsibilities in general remains the same, i.e. to ensure the compliance of GLP studies. GLP Compliance Monitoring is intended to ascertain whether test facilities have implemented the GLP Principles and that studies are performed in compliance with the GLP principles.

The OECD expects Member countries to establish national MAs, a body or bodies responsible for monitoring the GLP compliance of test facilities within its territories and according to national legal and administrative practices. In the European Union, the principles of GLP are included in Directive 2004/10/EC, while the compliance monitoring procedures are included in Directive 2004/9/EC, where each member state is required to implement a monitoring programme, where GLP registered facilities are inspected on a

[11] http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=263&InstrumentPID=361

[12] http://www.oecd.org/chemicalsafety/testing/goodlaboratorypracticeglp.htm

Field Code Changed

Field Code Changed

regular basis, approximately every two to three years. Routine monitoring inspections will include study audits. In addition, MAs can be requested by a receiving authority to conduct specific study audits as a result of concerns raised following the review of a regulatory submission. The MA has ultimate responsibility for determining the GLP compliance status of test facilities and the acceptability of a study audit. The MA also has responsibility for taking any action based on the results of test facility inspections or study audits which are deemed necessary.

The respective national compliance MAs are also responsible for the exchange of information on the compliance of test facilities inspected, and also should provide relevant information concerning the countries' procedures for monitoring compliance. They have the responsibility to facilitate the mutual acceptance of test data (MAD) generated for submission to regulatory authorities of OECD Member countries and other countries that are full adherents to MAD.

1.8 Accreditation bodies

The International Organisation for Standardisation (ISO) is an independent, non-governmental membership organisation and the world's largest developer of International Standards with a central secretariat based in Geneva, Switzerland. The ISO story dates back to 1946 when delegates from 25 countries met at the Institute of Civil Engineers in London and decided to create a new international organization 'to facilitate the international coordination and unification of industrial standards'[13].

In this organisation, different industries define their specific technical standards and quality management requirements and issue ISO standards to guide conformity. Companies and organisations working according to ISO guidelines can ask for a conformity check and certification by independent accreditation bodies. ISO itself is not a controlling body, but has established a committee on conformity assessment (CASCO) guiding certification organisations.

While the OECD Principles of GLP and ISO/IEC 17025 (OECD, 2016a) both set out requirements for quality management systems under which testing is conducted, they have, as a result of their evolution and history, different purposes (OECD, 2016a).

The OECD Principles of GLP are used as a regulatory control mechanism to assure the quality and integrity of non-clinical health and environmental safety studies regulated under law. Such testing, for the most part, is complex and variable. The Principles are specifically designed to be applied to individual studies to accommodate the complexity and wide variety of such studies due the different scientific disciplines involved, for the different chemicals under test. The OECD Principles of GLP are therefore, out of necessity, quite general in their requirements and take the form of a set of principles.

ISO/IEC 17025 is an international standard intended to be applied by laboratory facilities conducting testing, according to established or specifically developed methodologies. The focus of the standard is on the on-going operation, monitoring and management of the laboratory itself, and on the capacity of the laboratory to produce consistent and reliable results that are scientifically valid. ISO/IEC 17025 can, in theory, be applied to any testing laboratory in any scientific discipline including those performing non-clinical testing. It is a

[13] http://www.iso.org

Field Code Changed

reliable indicator of technical competence, and many industries routinely specify laboratory accreditation for suppliers of testing services.

1.9 OECD

The OECD strives for international harmonisation of test methods for toxicity and risk assessment of new products, and issues globally accepted Test Guidelines (TGs) that are accepted by regulatory bodies of all OECD member states and MAD-adherent economies.

The availability of a guidance document on Good *In vitro* Method Practices (GIVIMP) for "The development and implementation of *in vitro* methods for regulatory use in human safety assessment" was identified by OECD as a high priority requirement by both the Working Group on GLP (WG GLP), and nominated experts from the Working Group of the National Coordinators of the Test Guidelines Programme (WNT) to reduce the uncertainties in cell and tissue-based *in vitro* method derived predictions with the OECD *in vitro* test guideline methods.

Other national and international organisations are developing policies and standards for scientific practice to assure quality in implementation of *in vitro* methods. OECD has by supporting the GIVIMP endeavour, taken the responsibility to seek consensus on the current best *in vitro* method practices and will help to ensure that *in vitro* methods used for regulatory purposes are scientifically sound, of high quality, reproducible, credible, and acceptable. Generic *in vitro* method OECD TGs incorporating performance standards are being written to allow acceptance of both non-proprietary and proprietary *in vitro* methods by regulatory agencies and to provide assurance that any *in vitro* cell and tissue culture system performs over time in a manner that is consistent with the test system as it was originally validated (Gupta *et al.* 2005; Rispin *et al.* 2006).

2 Quality Considerations

2.1 Quality assurance versus quality control

In order to adopt GIVIMP during the test development phase or the routine phase of an *in vitro* method, it is necessary to have a good understanding of the relevant QA frameworks and definitions. In particular, users should be aware that a quality assurance programme in an OECD guidance has a particular meaning (OECD, 2005a, 2004c, 1998a), that is a defined system, including personnel, which is independent of study conduct and is designed to assure test facility management of compliance with the Principles of GLP.

There are numerous definitions of quality control, but for the purposes of GIVIMP, quality control is the documented activity which seeks to confirm that starting materials for assays (including cell lines) and key stages of individual assays and the final results meet prescribed specifications. It should be clear that this is not the same as Quality Assurance, which is the overall quality system designed to assure the quality of results (see Table 2). Specific requirements may be reviewed with respect to the criticality of any change, but it must be borne in mind that a good quality system should be under ongoing review to ensure current best practice is sustained and to enable the requirement for continuous improvement.

Table 2: Examples of differences between quality assurance and quality control

Feature	Quality Assurance (QA)	Quality Control (QC)
Nature	Proactive: prevents occurrence of errors, is process oriented and is a managerial tool.	Reactive: detects an error, is product oriented and is a corrective tool.
Format	Set of SOPs to assure that the required standard is met at every stage in the process.	Review and test quality in a product or service, against a set of given requirements.
Aims	Doing things right the first and every time to prevent errors.	Detect deviations/defects that need corrective action.
Scope	Focused on continuous improvement of processes.	Typically follows the process established as part of QA function.
Responsibilities	Management but also requires commitment from all staff.	May be performed independently but overall responsibility lies with organisation delivering the product or service.
Examples of relevant activities	Creating a process. Designing templates/check lists. Identifying defects in a process (i.e. QC). Root cause analysis. Making recommendations to improve a process or service. Developing a measurement system to assess process effectiveness. Checking a process.	May include: Reviews. "Walk-throughs". Inspection. Testing (verification and validation). Monitoring control charts.

2.2 Quality Control of test system

Quality control of the *in vitro* cell and tissue test systems must be a shared responsibility of the manufacturer of for instance proprietary test systems or suppliers of cells, tissues or tissue constructs, the test facility which uses the *in vitro* test system, and the entity that submits the

in vitro method and the related test system for regulatory acceptance. Tissue constructs or cell cultures may have a short shelf life. Proprietary *in vitro* methods and the related *in vitro* systems may be relatively expensive; therefore the number of replicate systems available for quality control efforts by *in vitro* test facilities may be limited by practical considerations. In light of these considerations, the user may sometimes be dependent on the manufacturer for many of the basic elements of quality control, including cell or tissue characterisation and functional performance of the *in vitro* test system. The manufacturer should be expected to provide adequate documentation of quality control testing of representative test system from each batch manufactured.

In addition, the user must provide a quality control check in the test facility on a regular basis appropriate to the test system so that the *in vitro* method performs as expected after transport and handling of the test systems in that particular method. For cell cultures, records should be kept of cell passage, of preservation, and of quality controls.

Moreover, records recommended by GCCP or other relevant guidance documents (e.g. ISO standards, GLP) should be kept. The developer/user should confirm the key quality attributes (essential characteristics) on receipt, and periodically confirm the genetic/phenotypic stability, identity and freedom from contamination. Guidance on cell and tissue culture work is available for either general (Coecke *et al.*, 2005) or specific (Andrews *et al.*, 2015; Geraghty *et al.*, 2014; ISCBI, 2009) applications.

2.3 Quality control of consumables and reagents

Consumables (plastic ware such as flasks, cryovials, culture dishes, culture slides, tubes, cell scrapers) and *in vitro* method reagents (test and reference items, media additives, compounds added to a system to induce a chemical reaction, etc.) may have passed through a number of entities (producers, suppliers, shippers etc.) thus, their quality must be assured by good documentation, traceability, confirmation of suitable storage/shipment, identification, quality control and preparation for use in an *in vitro* method. Attention will also need to be given to the suitability of reagents and to the safety and ethical provenance of cells (Annex 2 and Annex 3).

Consumables and reagents should be reliably available, consistent in their properties and for critical reagents, alternative sources should be identified. Certain materials which are critical to the performance of a method and which may be subject to significant variation, such as conventional serum products and growth promoting reagents, should be quality controlled before use (by functional tests) and this may include in-house pre-use testing. In addition, SOPs should mention that expiry dates should be documented.

Laboratories can perform quality control checks of consumables, but the process how to do this is not always evident. Some laboratories have established procedures whereby a percentage of consumables from each lot number are evaluated prior to use in *in vitro* work. This may be especially useful for test facilities inquiring the cause of contamination. While this approach will not prevent contamination, it can provide data from any profile(s) developed during these checks, which can be used for future evaluation of potential contamination events. Ideally, sterile consumables with appropriate certificates should be used where possible. Alternatively, some consumables can be treated with ultraviolet (UV) light, gamma irradiation and/or autoclaved. These preventive measures may be useful in limiting contamination events. Other consumables, such as centrifugal filter units and filtered pipette tips, cannot be pre-treated. In the case no commercial sterile and documented centrifugal filter units and filtered pipette tips are available, establishing a method for detecting contamination from these items is very important.

Again, as far as possible, reagents that are documented as sterile should be used without further quality control. Alternatively, laboratories should also run quality control checks on reagents prior to use in *in vitro* work. These checks assist in determining if a reagent is free of contamination at that time. Negative controls and reagent blanks provide a means to detect contamination from reagents. Negative controls can then be assessed on an ongoing basis to demonstrate that they remain contaminant free. Including and assessing negative controls and reagents blanks are critical quality control steps. These controls provide a means of detecting reagent contamination and, on occasion, sporadic contamination. Because many contamination events are sporadic, negative results in these controls do not necessarily mean that samples from the same batch are contaminant free. Additionally, the detection of contamination in these controls does not mean that all batch samples have been affected.

Reagents may be selected on the basis of history of use or reference in relevant documents associated with regulatory accepted *in vitro* methods (e.g. validation reports, *in vitro* method SOPs). Established reagents should have predefined acceptance criteria, which must be met and accepted before a new batch is used. Same reagent obtained from different suppliers may each have specific and not necessarily similar acceptance criteria. Acceptance criteria must be established for different reagents considering the degree of risk they represent to the final results. This risk can be assessed based on the manufacturer/supplier's information on traceability and control on sourcing, processing, quality control (QC) and shipment. This can be addressed by:-

1) Considering the potential impact of the perceived risk to prioritise certain reagents,

2) Formally evaluate (the quality management system of) the supplier and

3) Establishing suitable service level agreements (SLA) with the provider ensuring quality, availability and shipment of the reagent. Acceptance of individual batches of reagents can be addressed by review of key elements of the certificate of analysis, compliance with specific conditions of the SLA provided by the manufacturer/ supplier or a combination of these and supplementary evaluation which may include pre-use testing to assure that individual batches are fit for purpose.

4) Consistency of lot qualification tests on critical reagents. Serum can be a critical reagent for cell-based *in vitro* methods and lot qualification testing is common practice. Standardised serum is available. However, functional tests including acceptance criteria needs to be defined.

Any reagent stocks prepared in the test laboratory for use in routine *in vitro* methods should be documented, assuring use of acceptable reagents (as above) and documenting the preparation of these stocks in batch preparation records. A batch number, storage conditions and an expiry date should be assigned (and printed on the label), which can be documented in records of performance of routine assays. It is also wise to have procedures for maintaining and controlling laboratory stocks of reagents such as minimum stock levels and identification and qualification of alternate suppliers to support continuity of testing studies etc. The stability of test items and reference items under storage and test conditions should be verified, if appropriate, and expiry dates allocated as appropriate (see GLP principles II.6.2.4). If a test item is administered in a vehicle, the homogeneity, concentration and stability of the test item in that vehicle should be determined (see GLP principles II.6.2.5).

2.4 Data management

Before beginning to collect data from *in vitro* test procedures, it is important to assess the format of collection, the complexity involved and requirements for traceability, storage,

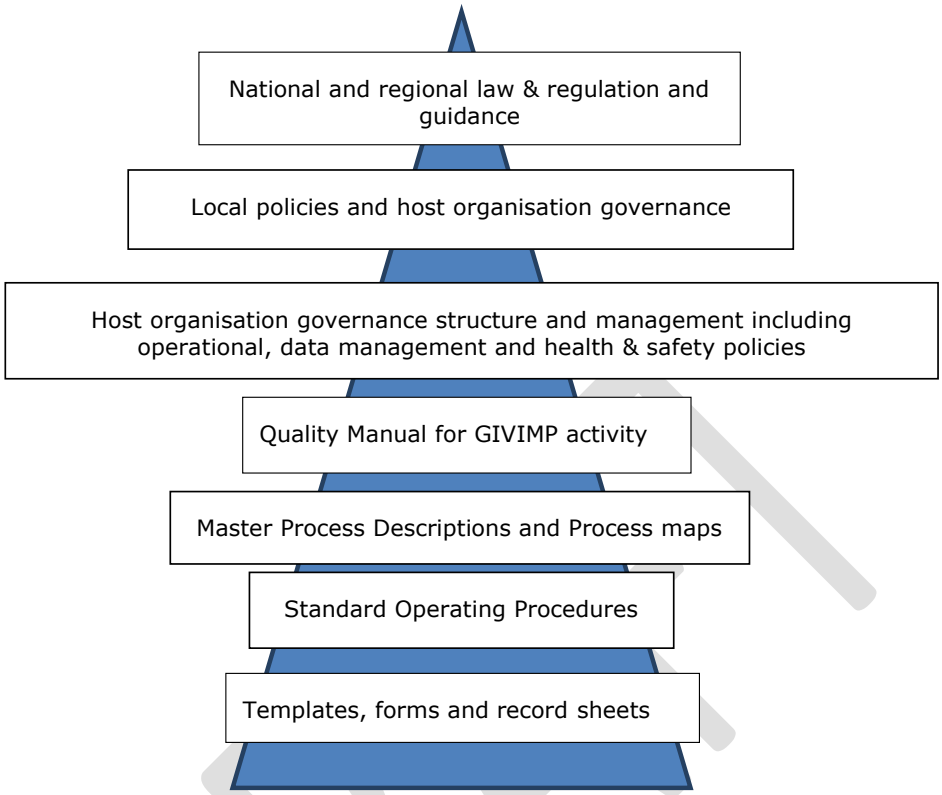
verification and transmission of data. Data should be recorded concurrently with the performance of the procedures. Specific standards may apply for data from regulatory testing and manufacturing (Coecke *et al.*, 2005; FDA, 2003; OECD, 1999). Data from material provided by tissue donors may also be subject to the requirements of data management and control under local, regional, national or international rules and regulations such as the EU Directive on Data Protection (NB. national and regional rules should be consulted as these may vary). It should be ensured that data reported accurately reflect the results obtained during experimental work, by performing adequate quality control of the data.

2.5 Types of documentation

Documentation in a quality system typically involves documents and records at several levels (Figure 1). The overarching document is a high-level, accurate description of the types of work performed by the organisation or group, key policies and standards adopted for delivering the work and the structure of the quality system used to deliver all aspects of test performance and provision of results. In some systems, this may be called a "quality manual". Another level may include overviews of procedures referring to the various specific testing methods involved at the next level. Finally, supporting the SOPs, there will be formal record sheets for test and control data and templates for reporting results. Another example for descriptive and prescriptive documents is provided in the WHO handbook on quality practices in Biomedical Research (WHO, 2013).

In many formal quality systems, including GLP, there will need to be a document control convention which assures that all documents are developed and approved in a formal process, that active versions are accurately dated, authored and approved with specific version numbers that will avoid inadvertent use of obsolete testing methods.

Figure 1: Hierarchy of governance and management documents



NB: this schematic overview is intended to reflect some general aspects of documentation hierarchy and is not a prescriptive requirement of GIVIMP or GLP. Another example for descriptive and prescriptive documents is provided in the WHO handbook on quality practices in Biomedical Research (WHO, 2013). QA documentation may be structured in different ways as required in different organisations and regulatory jurisdictions.

2.6 Staff training and development

Training files are helpful to capture all procedures in which staff should be trained before they are considered competent. These may include regulatory requirements of the work, specific controlled procedures (e.g., SOPs, testing methods), use and storage of documentation, as well as general training in best practice such as indicated in GCCP (2005). When new staff is recruited to work in the laboratory, it is important to guide the staff and review and document any training requirements before assignment to carry out any tasks. A written training plan might serve as a basis for a formal training record, which should be reviewed periodically by their line manager. It may be helpful to demonstrate competence in each procedure by recording individual elements of training in a particular SOP, until the trainee is considered competent and the training records are signed to authorise that they can perform the procedure independently. Training may include:

1) self-study of testing methods and SOPs and recording that the document has been read and understood,

2) correct performance of the procedure, witnessed by a qualified supervisor a minimum number of times and notification from the trainer or supervisor that the staff member is competent to perform the task independently.

A list of core training for *in vitro* cell culture laboratory staff is given in GCCP (Coecke *et al.*, 2005) and special aspects of training are also referred to in other sections of this document where relevant.

It is good practice to record all training in staff training records or periodic competency reviews, which should be regularly updated. In some circumstances, the host organisation or relevant professional bodies may run or support attendance at supplementary training and education which should also be documented to demonstrate maintenance of ongoing professional development and support assurance that current best practice is maintained in testing procedures. Annual review of staff performance is also a useful tool for considering ongoing training needs.

2.7 Assurance of data integrity

Data integrity is fundamental for any quality system and starts with good quality documentation (see Table 3).

Table 3: Key criteria to be addressed for data generated under a GLP environment (ISPE, 2012)

Criteria		GLP
Attributable	Who acquired the data or performed the action	Initials/signature/UserID
Legible	Data must be recorded permanently in a durable medium and be readable	
Contemporaneous	Documented at the time of activity	Date and time (stamp)
Original	Original or true copy	
Accurate	No errors, no undocumented changes	QC, QA audits
Complete	All data available	QC, QA audits
Consistent	Traceability, dataflow, date timestamps	
Enduring	Recorded on paper or electronically	
Available	Accessible for the lifetime of the record	Archiving

For validation of certain documents, such as spread sheets, there may be specific procedures to adopt good practices e.g. (Esch *et al.*, 2010; OMCL, 2009). It may be necessary to have an audit trail of modifications.

If data is translated between different recording methods, systems and/or data bases and in particular critical phases like manual or semi-automatic transfer (e.g. Excel™ files to database for kinetic assessment, combination of information obtained from 2 or 3 databases to one database), correct resolution of pre- and post-translation data should be reviewed and confirmed by a qualified person. For handwritten data, translated into an excel sheet, it is also advisable for the changes to be verified by the same person who has made the observations. These issues are of special concern where data are exchanged between countries (e.g. for pharmacokinetic assessment). When data translation occurs between different software or data base systems their compatibility and inability to be altered in translation should be tested and will need to involve appropriate validation procedures (see i) below)

Data Integrity (DI) closely correlates with good scientific practices. The same principles should be applied when using either paper based and electronic systems, or both. It should be assured that the obtained information is unchanged from the source, and has not been intentionally modified, altered or destroyed. To ensure DI for both systems, the following components of this process should be taken into consideration.

a) Sample integrity: Information and decisions based on data from samples used must be robust, so sample control should be ensured from point of entry into the laboratory through to data review and data release.
b) Integrity of materials/solutions: The right standard of materials/solutions must exist to ensure a valid reference, so the process for standardisation and requirements for re-standardisation should be established e.g. documentation of receipt, calculation and preparation, traceability, assignment of expiry or retest dates etc. Reference data needs to be complete (e.g. lot number, reference standard data). For further details see Section 4.2.
c) Instrument and equipment: The data generated by instruments and equipment must be robust. The equipment unit(s) must be suitable for their "intended use", therefore the equipment must be calibrated or qualified by appropriately trained personnel. Calibration should, where appropriate, be traceable to national or international standards of measurement. In order to operate within specification, scheduled maintenance of the equipment and instruments once again must be performed by qualified personnel. For examples and more information see Section 4.1.
d) Documentation and result reporting: Records must be clear and accurate. All activities should be recorded at the time they are performed (see Annex 2). Records should also be chronological, traceable, and readily retrievable. Original documents must be clearly identifiable (e.g. time stamps, watermarks) and standardised, predefined, authorised forms and templates should be used wherever possible and applicable. Records should be signed and dated allowing for clear identification, no pencil, no recording on loose 'post-it' notes, "white-out" paste or scrap sheets of paper should occur. Any corrections written on documents should be signed and dated by a trained staff member. Transcriptions if performed need to be attached to the original results (full traceability) and reviewed. Chronology of recorded data and correct page numbering must be ensured.
e) Proficiency of the analyst/operator: The person producing the data using samples, materials/solutions, instruments and equipment, documenting and reporting results must be relied upon to produce robust data. Therefore, the ability of the person to perform a task accurately and proficiently should be ensured, e.g. through appropriate training (see section 2.6). The person must also understand that destroying, deleting, hiding or using selected data without scientific justification (or even falsifying data) are not acceptable. Facility management policies should assure a working

environment in which that when a mistake has been made, the individual concerned should feel comfortable to report and explain it and not conceal it.

f) Method validation and verification: The *in vitro* method has to be validated or verified "suitable for its intended use".

g) Effective review and verification: A clear definition and understanding of raw data should be ensured. There needs to be traceability to the testing method used, source data and verification of raw data. SOPs need to be in place for data handling, record retention and good documentation practices and deviation handling etc.

h) Additional considerations for electronic data: In addition to comments made above, if a system is required to maintain electronic data, it should be managed by unique user identity and password combination. If the system does not permit this, a paper-based log must be in place to record who uses the generic user and password combination, or who uses the unprotected equipment. Paper records can be reviewed for any amendments or crossings out/deletions plus the signature/date of and the reason for doing so. This is to be replicated in an electronic system in the same way by use of an electronic log (audit trail). Where there are multiple users, an approach to periodically review who has system access must be in place. There must be a periodic user account review procedure. There should be procedures in place for assigning access rights to each user. The level of access should be in line with the tasks that have to be performed.

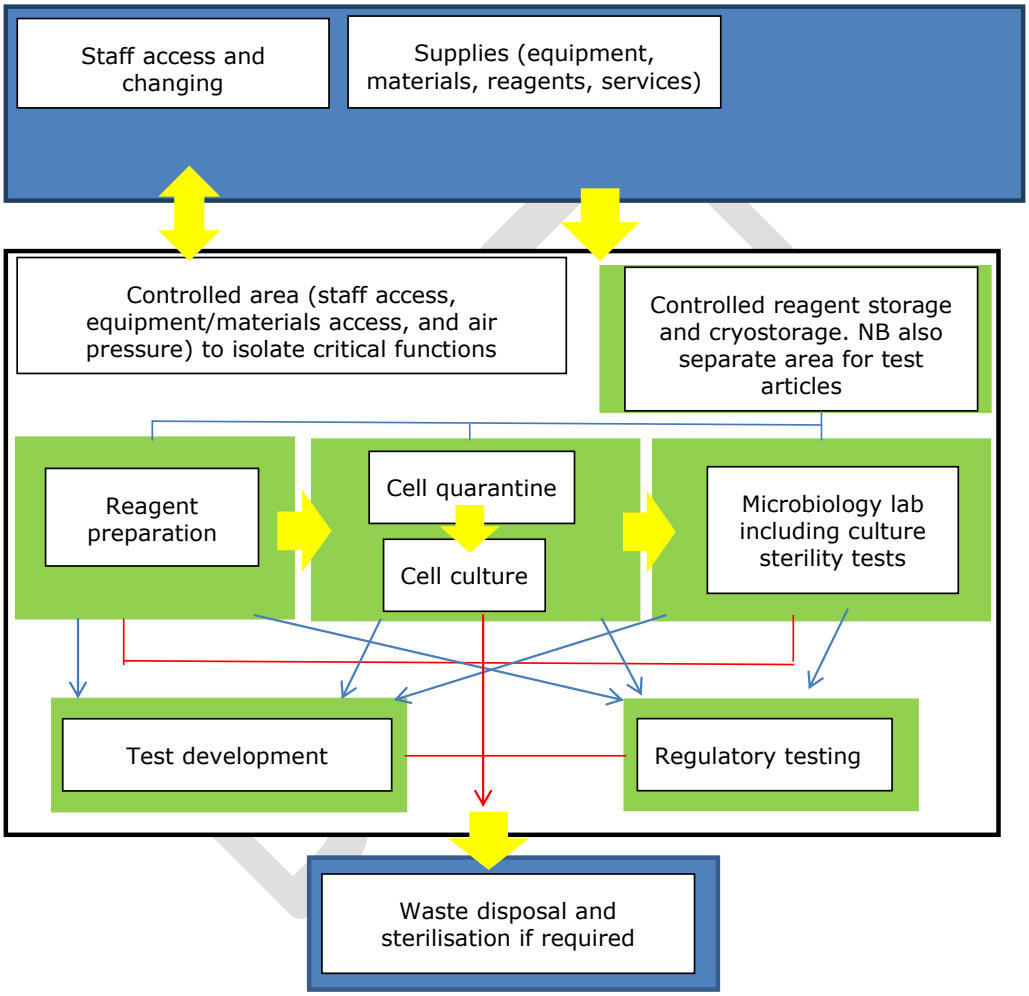
i) Data storage, archive & retrieval: Data must be stored in a safe and secure place for paper-based systems and in protected folders for electronic systems. An approach must be in place to ensure that data are protected against loss, damage or overwriting. Manipulation of stored paper or electronic records must not be possible. Electronic records must be held in a format that is not readily corruptible and protected from deliberate or accidental alteration (e.g. CFR 21 part 11, GLP: see OECD GLP Guidance Document 17).

A system must be in place for the archive of electronic and paper data and records (see 10.3). It should be considered and tested how the data and records can be retrieved and used to re-create a complete accurate picture on the rare occasions that this is required.

3 Facilities

A very important aspect in the process of *in vitro* method development to *in vitro* method routine use is the need for a well-designed facility to ensure that good quality results are produced in a safe and efficient manner. The types of laboratory areas which might need to be separated are indicated in Figure 2.

Figure 2: A typical "onion ring" structure used to separate different operational areas.



It may not be possible or acceptable to separate such laboratory functions but they are among the considerations for separation (physical or process/training) that someone establishing or running the facility should be aware of and there may be other issues which must be considered for different kinds of work. It is wise to avoid physical contact between materials transfers (blue arrows) and waste removal (red arrows) so that there is very low risk of contamination from waste affecting reagents, cultures and test materials.

As stated in OECD No 14 (OECD, 2004c), facilities should be suitable to maintain the assay, resulting data and all archived items.

Facilities must be fit and suitable for the purpose of the work; that is, size, construction, and location should be appropriate, and the building should allow for the separation of activities.

Both test developer facilities and routine use facilities should ensure to be designed or adapted to have separate areas for similar but unrelated work and sample preparation. In addition, buildings should be validated for the required functionality (e.g. air handling) and properly monitored (e.g. air pressure differences, flow patterns, etc.), with easily accessible results. Validation of facilities should be formally documented. Environmental control systems should be regularly maintained and serviced, with full records of maintenance and any modifications to demonstrate appropriate upkeep and function. All the necessary permits should be in place before any activities are initiated. Finally, there should be dedicated areas for data storage and archiving.

3.1 Containment

Cell lines and primary tissues may carry a variety of different microorganisms or pathogens, which can potentially cause human disease, pose hazard to employees and distort the *in vitro* method results. Cell lines and primary tissues should be handled at biosafety (hazard) level 2, unless the cells are known to be specific pathogen-free. This level of containment is also appropriate for monoclonal antibody-containing supernatants and cell homogenates. Access to level 2 facilities should be restricted to authorised personnel only, and specific risk assessment and training activities should be followed according to the national legislation on Level 2 containment (Coecke *et al.*, 2005; Geraghty *et al.*, 2014).

As a minimum, all cell and tissue work should be performed in a Class II biological safety cabinet as even screened tissues or cell cultures may carry infectious agents not covered by virological screening. Depending on the kind of test items in the *in vitro* method, several subtypes of Class II biological safety cabinets must be considered. Horizontal (Class I) flow cabinets, where the airflow is directed at the operator, are intended for sterile media preparation and not for use with tissues and cell cultures due to the inability to exclude all infectious agents in screening and to avoid the widespread laboratory contamination with mycoplasma or bacterial and fungal contaminants. For this reason they should only be used if the cells are known to be pathogen-free. Class I cabinets are also not appropriate when potentially toxic chemicals might be used (which is often the case for *in vitro* methods for regulatory use). In the case of using toxic chemicals it is recommended to let the outflowing air pass through a volatile organic compound-filter (active charcoal) after a High-Efficiency Particulate Arrestance (HEPA) filter.

If microscopes or other equipment are to be installed in biological safety cabinets, the cell containment equipment should be checked for flow disruption as well as operator and cell culture protection.

Splashes and aerosols carry contamination and infection risks, which can not only endanger the operator, but also compromise the integrity of the *in vitro* method (i.e., cross contamination of cell lines or introduction of adventitious agents). Therefore, all procedures should aim at minimising aerosol production. Any procedures likely to produce aerosols should be contained (biosafety cabinet) or the material should be rendered harmless.

As contaminated working surfaces can lead to microbial contamination or cross-contamination between cell lines and pose a risk to the *in vitro* method quality, working surfaces should be easy to clean, resistant to acids, alkalis, solvents and disinfectants. There should be appropriate documented procedures for disinfection of work surfaces, safety cabinets and equipment.

Higher containment levels may be required depending on the biosafety risk level of the biological agents handled. If *in vitro* work is to be performed with group 3 or 4 human hazards, which can cause severe human disease and may be a serious hazard to employees or spread to the community, then separated facilities, appropriate levels of biosafety, such as air filtration and negative pressure differences will need to be maintained. Groups 3 and 4 are more complex in complying with specific facility requirements and personnel skills. Therefore, *in vitro* methods for regulatory use in human safety assessment should be developed to require mainly level 2 biosafety.

3.2 Level of separation to avoid cross-contamination

Measures should be taken to ensure adequate separation of different biological agents and *in vitro* studies taking place in the same physical environment (OECD, 2004a). The integrity of each test system and study should be protected by spatial or temporal separation from other studies and test systems to avoid cross-contamination and mix-up. Air flow requirements, for example, are opposite for spaces where human cell lines are used (flow out of the space to avoid their contamination) and for spaces where pathogenic micro-organisms are used for tests (flow into the space to avoid spread of the pathogens).

Tissues and cells from different studies can be kept in the same incubator provided that they are labelled appropriately, placed in a different position and none of the test items are volatile enough to cause contamination. Tissues and cells from different species or *in vitro* methods where yeast and bacteria are used would require a higher level of separation. The most important issue here is to separate the areas used for cell culture/tissue and microbiological culture (see Figure 2). Other degrees of separation may be achieved using the specific requirements described elsewhere for quarantine of untested material.

Temporal separation of test systems should be used in Class II biological safety cabinets. This can be achieved by handling only one cell line at a time, followed by cleaning and decontamination of the cabinet, working surfaces and related equipment.

Rooms and areas used for preparation and mixing of test and reference items with vehicles should allow aseptic working conditions in order to minimise the risk of microbial contamination of the test system.

Appropriate training should be given to the authorised personnel regarding the necessary precautions to prevent contamination and cross-contamination.

When performing molecular biology techniques and especially PCR-based assays, which are high sensitivity methods, extreme care should be taken in facility design and operation in order to avoid false-positive results. False-positive results can originate, for example, from sample-to-sample contamination, from carry-over of nucleic acid from previous amplification of the same or similar target. Cloned DNA or virus infected cell cultures may represent other source of contamination[14].

The greatest threat of contamination lies in laboratories that manipulate amplified or cloned DNA; laboratories exclusively performing real-time PCR and properly discarding all amplified products without opening the reaction tubes or sealed plates are less liable to contamination.

[14] https://www.gov.uk/government/publications/smi-q-4-good-laboratory-practice-when-performing-molecular-amplification-assays

Field Code Changed

It is vital that the correct workflow is followed in a molecular laboratory in order to minimise contamination and ensure good laboratory practises are followed. It is the responsibility of all laboratory staff to ensure that the workflow is followed. The equipment, consumables and laboratory coats should be dedicated to each area. If possible it is helpful to colour code racks, pipettes and laboratory coats in the different areas to be able to easily monitor movement between the different areas. Powder-free gloves should be used throughout the process in all the different areas as the powder on powdered gloves might result in assay inhibition. It is particularly important to always use powder-free gloves in the pre-PCR area, as this is prone to contamination by RNases.

Perform physical separation of pre- and post-PCR assay stages. Between these two areas the work flow should be uni-directional and the relative air pressure and direction should differ. It is recommended for personnel working with post-PCR assay stages to not work with pre-PCR parts later the same day.

PCR reactions should be set up in a separate room from that used for post-PCR manipulation. Bench areas should be wiped daily with hypoclorite solution following use and contaminated areas should be additionally decontaminated with ultra-violet radiation if fitted. Hypochlorite solution (20% or greater) should not be applies to stainless steel (types 304/347, 316 and 400 series) as it leads to corrosion with repeated use. Reagents should be taken from clean storage into the pre-PCR area and never taken or shared with post-PCR areas.

Where possible, the PCR facility should be organised in four discrete rooms/areas: (Requirements may vary with assay format e.g. real time PCR does not require post-PCR analysis).

- **Reagent preparation clean room**: it is free from any biological material such as DNA/RNA, cloned material, etc.). Primers and reagents aliquoting is recommended to minimise contamination consequences. The air pressure should be positive and blow out of the room. The clean areas must be kept free of amplicon at all times, to ensure this occurs there should be no movement back from the dirty area to the clean area. If under extreme circumstances a consumable or reagent needs to be moved backwards it must be thoroughly decontaminated with bleach and ethanol. Returning racks should be soaked in 1% bleach overnight before soaking in distilled water and placing in the clean area. To ensure minimal movement between areas during the running of molecular assays, it is optimal to have dedicated storage (freezer, fridge and room temperature) for each area.
- **Nucleic acid extraction room**: in this area samples are processed, reverse transcriptase step of RT-PCR is performed and DNA or cDNA and positive controls are added to the PCR reaction mix (prepared in the Reagent preparation clean room). The air pressure should be positive and blow out of the room. If chemicals are stored in this area appropriate facilities and storage requirements should be in place.
- **Amplification room**: PCR machines are housed in this room. It may contain an area (cabinet with air pressure slightly positive) for the nested PCR.
- **Product analysis room**: post-PCR manipulations such as agarose gel electrophoresis are performed in this area. It is thus a contaminated area and therefore no reagents, equipment, coats, etc. used in this room should be used in any other PCR areas. The air pressure should be negative and blow into the room.

3.3 Air handling, water supply, environmental control, heating and cooling

Air handling systems should be operated to ensure that the correct environment is maintained for the type of work conducted in the laboratory. These systems should be subject to regular

maintenance by qualified personnel. In case *in vitro* work involves serious human pathogens, the laboratory should operate with specific trained personnel, using biosafety level 3 and 4 and the room should be kept at negative pressure to guard against infection spread. In contrast, work with cell and tissue cultures, for example, may require positive pressure relative to other laboratory areas to minimise the risk of test system contamination from the outside. When HEPA filters are used in differential pressure isolation rooms, the filters and their fittings and seals need to be thoroughly examined and tested at regular intervals (e.g. annually). Decontamination should be carried out before servicing is carried out. The air handling system should also be designed to account for exhaust air from the Class II biological safety cabinets that are vented to the outside of the building.

Cell culture work requires high-quality ultra-pure water, which is usually deionised via reverse osmosis, followed by passage through a series of carbon and micropore filters eliminating organic materials and pyrogens. Tissue culture grade water should be controlled for pH, conductance and total organic carbon, as well as absence of endotoxins. Note that pyrogens can be deleterious to cell cultures at concentrations below the level of detection for organic carbon. Where small qualities of purified water are required for cell culture this may be obtained by obtaining water for irrigation (WFI) or other medically approved pure-water preparations (Stacey and Davis, 2007).

Heating, cooling and humidity should be adequate for the comfort of laboratory occupants and for operation of laboratory equipment, and should not adversely affect test system survival/behaviour and test item stability. For example, in some cases (e.g. preparation of microscopic slides) specific humidity might be required.

Many tissue culture media components are sensitive to white light (especially sun light). The blue wavelengths are of particular concern. Filters can be used in the room and laminar flow cabinet light to reduce this exposure where necessary.

Mid to long term storage of media is usually best at temperatures below ambient lab temperatures. Accordingly, an optimal solution may be to store all cell culture media at 4˚C (refrigerator) or frozen (electric freezer) as recommended by the manufacturer. There may be exceptions to this general opinion but manufacturers' instructions should always be consulted.

3.4 Cell and tissue culture transportation and cryostorage

Because cells and tissue in culture are often transported across the world, it is very important to keep these test systems as healthy as possible during the long transport times. A quality cell culture incubator makes it possible to transport valuable and delicate cell cultures in a temperature-controlled environment, so that they are less likely to become damaged during the transport process. However, one should note that the incubator for transport has limited space (2-3 plates or flasks) and adequate sealing of plates is necessary to avoid leakage during transport. Because of their fragile nature, live cells and tissues cannot be shipped like a regular package in a cardboard box or shipping envelope. They must be transported in the special temperature-controlled environment of a mini-cell culture incubator, if they are expected to reach their destination in good condition. Technologies have provided such solutions during the last decennia. Alternatively, good low temperature transport practices can allow also for safe transport of cells and tissue systems. Cells are usually shipped on dry ice. Ideally, temperature should also be monitored (e.g. by using data-logger) during transportation, especially for long distance transport.

Cryostorage systems should ensure the long term preservation of the stored test system. For cryopreserved cell cultures, the viability of mammalian cells is progressively lost within

months at -80˚C, thus, long term storage below the glass transition point of water (-136˚C) is recommended. While true for mammalian cells, this is not the case for bacteria or yeast.

Storage in the vapour phase of liquid nitrogen is generally advised for all cells and necessary for potentially infectious cells and tissues. This eliminates the chances of transfer of pathogenic material between vials which can occur in the liquid phase of nitrogen (Coecke *et al.*, 2005), Appendix 1). It is also considered safer as liquid nitrogen can enter storage vials if they are stored in the liquid phase and cause them to explode upon thawing. If vials need to be stored in the liquid phase, protection wrapping may be considered.

Cryostorage requires temperature and liquid nitrogen level monitoring to ensure that the test system stocks are at optimal storage temperature. Cryostorage vessels can be fitted with alarms and data loggers and liquid nitrogen levels recorded at regular intervals (e.g. weekly).

In cases when ultra-low electrical -150˚C freezers are used, CO_2, liquid N_2 or electrical backup systems need to be in place to guard against loss of power supply.

Storing valuable test system stocks in more than one cryostorage location is recommended for security/back up purposes, and off-site storage may also need to be considered in disaster recovery plans for the facility.

3.5 Quarantine for new test systems

All new cells and tissues should be quarantined in the laboratory and in storage until determined free of contaminating microorganisms (see Section 5). Early checks of cell authentication are also recommended to avoid wasted time and resources on unauthentic cell lines. If a separate quarantine laboratory is not available, a quarantine biological safety cabinet, a dedicated incubator and liquid nitrogen tank can be used. Alternatively, other steps can be taken to minimise contamination risks, such as handling the quarantine cells last on each day, rigorous post-manipulation disinfection of the work areas and placing cultures for incubation in a filter-sealed container into the general incubator (Geraghty *et al.*, 2014). Cells procured from a cell bank may carry a certificate of analysis for contamination tests performed. The certificate will list the tests performed and may give details of testing methods. As a minimum, a mycoplasma test should be performed upon receipt.

4 Apparatus, material and reagents

4.1 Apparatus

Apparatus, including validated computerised systems, used for the generation, storage and retrieval of data, and for controlling environmental factors relevant to the study should be suitably located and of appropriate design and adequate capacity. In general, all apparatus used should be operated and maintained correctly by trained staff to ensure correct use and knowledge on the procedures for equipment maintenance and calibration.

Computerised systems should be formally validated prior to use in a GLP environment (see 10.1). Apparatus should be periodically inspected, cleaned, maintained, and calibrated according to SOPs and records of these activities should be maintained (OECD, 1998b).

As stated in OECD document *Number 14* (OECD, 2004b), the commonly observed, routine requirements for apparatus used in a GLP environment apply equally to apparatus used for *in vitro* development work, like laminar flow cabinets and incubators, there are some specific points and issues of particular importance for the latter. As an example, equipment such as microbalances, plate readers, centrifuges, micropipettes, laminar air flow cabinets, fridges and freezers, water baths, and incubators should be regularly maintained and calibrated (traceable to international standards) where possible, to ensure the integrity and reliability of the results. For each type of equipment, critical parameters (e.g. supply of gases for mass spectrometry; liquid nitrogen levels in storage containers, CO_2 levels in incubators; or functioning of filters to ensure sterility of the air in addition to the airflow in the laminar flow cabinets) should be identified as requiring continuous monitoring or the setting of limit values together with installation of alarms.

Centrifuges which are routinely used in cell and tissue culture work (sub-culture work, cryopreservation etc.) produce aerosols and therefore it is important to consider models that have sealed buckets. Ideally, one should consider working with models where the condition of the load can be observed without opening the lid. Besides the containment issues for centrifuges, when documenting their operation, it is necessary to specify centrifugation speeds as x g rather than rpm (unless the rotor radius is stated), incubation conditions, time and volumes of centrifugation with tolerances when relevant, and any other information that enables the accurate reproducibility of procedures. In addition, procedures should be established on how to respond in case of an emergency (e.g. broken tube).

Working with cell and tissue culture *in vitro* requires a strictly controlled environment for cell growth. This is achieved using specialised incubators which provide the correct and controlled growth conditions (temperature, humidity, CO_2 levels), which should be checked (and logged) on a regular basis. To provide the required humidity, incubators are frequently equipped with a water basin at the bottom. This combination of high humidity and temperature increases the risk of bacterial or fungal contaminations. To reduce the risk of microbial contamination, copper-coated incubators are now available. Antifungal or bactericidal agents can be added to incubator water trays to reduce the risk of bacterial and fungal growth (but any possible impact on the *in vitro* method to be carried out should be checked and documented). Incubators with self-sterilising cycles may also be used, although this does not replace regular cleaning and maintenance.

Similarly, waterbaths used to thaw and/or to warm up stored solutions like medium and frozen stocks, or to defrost vials of cryopreserved cells and tissues, carry a high risk of introducing contamination. The use of bactericidal and fungicidal agents in waterbaths can

aid in the control of contamination, but the impact on the test system should be checked and documented. Alternatively, cleaning procedures and regular change of water may be a better choice.

A laminar air flow cabinet can be considered as a critical piece of equipment for cell and tissue culture work, since, when it is used correctly (see 3.1), it ensures a clean working environment protection for both the operator and for cells/tissues and other materials and reagents.

For equipment such as refrigerators and freezers, temperatures should be checked regularly and preferably logged: simple data loggers are available to log the temperature at set intervals. In addition to the regular recording of temperatures, an alarm system to alert staff when acceptable operating limits are exceeded is desirable, and a backup system should be in place, such that materials may be transferred from one fridge/freezer to another, in case of malfunction.

For all equipment used during an experiment, acceptable operating limits should be set, monitored and recorded. Equipment should be fit for purpose with respect to sensitivity and selectivity. Prior to use, it should be established that the equipment is functioning according to the (suppliers') specifications and it should be qualified and validated for its intended use, e.g. via a formal DQ/IQ/OQ/PQ process. Since DQ/IQ/OQ/PQ are not specifically defined in GLP, and its application is not always harmonised, several test facilities follow the OMCL guidelines for qualification of equipment (OMCL, 2011). The equipment needs to be maintained and calibrated at regular intervals depending on the type of equipment and the frequency of use. As an example, pipettes or micropipettors may need to be calibrated more frequently than centrifuges.

It may be necessary to have separate procedures for regular checks (e.g. daily checks of pipettes used) and complete calibration (monthly or quarterly, depending on the frequency of use). The standard operating procedures need to describe how to deal with data produced with a potentially uncalibrated pipette. When equipment such as a pipette is out of specifications during a calibration, it is important to determine how to interpret data that have been generated since the most recent successful calibration and determine the impact of the potential deviation to the outcome of the study. Therefore, it is crucial to record every piece of equipment, uniquely identified, that has been used during the performance evaluation of an *in vitro* method. In general, facility practices should ensure that equipment is within specifications before the start of study and throughout the experimental phase to avoid the rejection of the *in vitro* study. Nevertheless, some test facilities use periodical calibrations for pipettes (e.g. every 3 months), given their low failure rate.

To enable broader use of a new method, successful transfer to a range of equipment and different laboratories should be demonstrated. This increases the robustness of the method. To increase transferability, preference should be given to the use of generally widely available equipment. In addition, the impact of the use of a certain type or brand of equipment on the outcome of the individual assays or the overall *in vitro* method needs to be determined. An *in vitro* method should specify the acceptable limits that the equipment should meet to be used for a specific *in vitro* method[15].

[15] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=48908

Field Code Changed

4.2 Materials and reagents

Critical materials and reagents can influence the outcome of *in vitro* studies and it is therefore important to identify and control them in each experimental phase. In using the term critical materials and reagents we refer to highly complex preparations used as a fundamental component in the *in vitro* method, which are not currently definable in terms of composition, stability or biological activity. Examples include processed tissue and cell culture preparations, which may contain viable cells. As with all reagents, the standards of preparation and testing of kits should be clear and traceable and they should be used before their expiry date. Further guidance on complex cell systems such as 3D culture, is given in Annex 2. For GLP studies, preparation of substrates and use of reagents and kits should be documented in sufficient detail to allow for complete reconstruction of these activities.

Other examples of critical reagents are serum and growth factors. Each of the critical reagents should either be sourced from a reputable supplier, who accompanies the shipment with a certificate of analysis, or one should ensure that there are appropriate quality controls (see Annex 2). These controls may include in-house growth or functional characterisation and the controls should be done by trained personnel and according to procedures described at *in vitro* method development laboratories or in the standard operating procedures (SOPs) of routine or GLP facilities. It is recommended to perform a batch validation to reduce the introduction of unknown variables into a culture system that can interfere with assay or overall *in vitro* method performance. For this purpose a batch is tested first and when approved, a large quantity of the batch can be acquired to reduce variability during the performance of a certain number of assays.

The use of good quality assurance practices applies to all laboratory reagents and consumables. For instance, preparation of reagents should follow a standardised and fully documented methodology that lists supplier information, lot numbers of component chemicals and reagents, dates of preparation, and the names of the staff involved in the preparation. Substances should be tested against reagents and standards of known reactivity before being released for use, and the results of these tests must be recorded (on, for example, special reagent preparation forms). All prepared solutions should have unique lot numbers (a laboratory-based system is acceptable as long as the provided numbers are actually unique). This information log prevents duplication and should comprise lot numbers, dates, descriptions, expiry dates, and signatures. For both reagents and reagents mixtures, the container should be inert to the stability of the substance or mixture and clearly labelled with the following details:

- contents or identity
- potency (titre, concentration, or activity, for example)
- storage temperature
- preparation date
- unique lot number
- container number (in case there are multiple containers of the same lot)
- date first opened (as appropriate)
- expiration date (determined by experimentation or reference to manufacturer's recommendation)
- signature of the person who prepared the contents.

Much of this information can be recorded separately with a unique identification number log.

Even when reagents are sourced from a reputable supplier, it remains important to assure the stability of the reagents during shipment conditions, in addition to the storage. For example,

reagents shipped frozen should arrive frozen and this should be documented on the receiving document. The presence of a data logger is the best practice in these cases.

Storage should be done according to the manufacturer's specifications in the supplied certificate. Most solutions which come in a large quantity should be aliquoted, in order to minimize the number of times a bottle is opened and thus minimize risk of contamination (and spread of contamination). This is particularly important for solutions which require storage below 0°C, in order to avoid repeated freeze-thaw cycles. When reagents need to be thawed and possibly frozen again, it is recommended to determine the number of freeze-thaw cycles that the reagents can withstand (EMEA, 2011; FDA, 2001; Viswanathan *et al.*, 2007). Stability of aliquots should be verified in the lab performing the *in vitro* method and not rely solely on literature data.

In these cases, quality controls need to be performed according to pre-defined procedures described in SOPs. Normally, stability of the analyte in the studied matrix is evaluated using at least triplicate samples of the low and high concentrations, which are analysed immediately after preparation and after the applied storage conditions that are to be evaluated. The thawed samples are analysed against a calibration curve, obtained from freshly prepared calibration standards, and the obtained concentrations are compared to the nominal concentrations. The deviation should be within previously established acceptance criteria (usually ±20% for large molecules). It is absolutely necessary that the number of cycles in the freeze/thaw stability evaluation should equal or exceed that of the freeze/thaw cycles of study samples.

When pipetting problematic volatile/viscous liquids or suspensions, it is strongly recommended to use positive displacement pipettes. Certain chemicals may exhibit non-specific adsorption to the plastic tips of pipettes and the use of low-binding materials (including glass) or acoustic droplet ejection (Ekins *et al.*, 2013; Grant *et al.*, 2009) can be utilised to alleviate these issues.

4.3 Basal medium

Depending on the circumstances, the basal culture medium can be serum-supplemented (as in traditional cell culture methods) or serum-free, but supplemented with additives necessary for obtaining satisfactory cell proliferation and production, or for maintaining a desired differentiation status. Many slightly different formulations exist under the same general medium names, such as Minimum Essential Medium (MEM), and even subtle changes in the medium formulation can substantially alter the characteristics of certain cells and tissues. In many cases, these variations are deliberate for specific applications. Therefore, the medium to be used should be precisely specified, and it is important to check that new supplies of medium meet the required specifications (Coecke *et al.*, 2005).

4.3.1 The use of serum in cell culture

The use of serum has been discouraged in recent years due to the necessity of animals to produce it, while *in vitro* methods are mostly developed to replace animal use. Furthermore, serum contains many undefined factors and these could change every time a new batch of serum is ordered, even if it is from the same source. In this respect, serum starvation, use of serum-free media and serum replacements has become standard practice in many laboratories.

Despite the availability of serum-free media and serum replacements, serum is still used in a lot of *in vitro* development work. Animal serum can be derived from adult, new born or

foetal sources. Bovine sera are most commonly used, and during the last few decades, foetal bovine serum (FBS) has become the standard supplement for cell culture media. However, each new batch of serum may contain different concentration of growth factors and hormones, Therefore new batches should be tested on a relevant range of cell lines for cell attachment, spreading, cloning efficiency, growth rates and activity in functional assays (Geraghty *et al.*, 2014).

Cell lines which have been derived or cultured in serum-containing media long-term may become dependent on the multitude of growth factors present in serum and may experience a phenotypic drift upon abrupt serum withdrawal. This may manifest as growth arrest or activation/inactivation of various signalling pathways. Serum can also be used to mimic the protein binding occurring in the blood *in vivo*, but this is a process for which specific attention and calculation should be taken into account. Specific test item aspects need consideration in choosing to work with serum: 1) If the test item is known to bind to protein, its effect might not be seen unless a very high concentration of test item is used (see section 6.3 on biokinetic parameters); 2) If the test item antagonises an endogenous circulating hormone or factor, the serum might contain such hormone or factor and may thus affect the assay results.

Test developers determine serum specifications that meet their particular needs (e.g. to allow the cells to function like in the *in vivo* situation, as much as possible, to enhance test result utility) and match the natural behaviour of the cells as much as possible, including the maximum acceptable levels of serum components, such as immunoglobulins (which may have inhibitory effects), endotoxins (indicative of bacterial contamination, but are also powerful cell mitogens), and haemoglobin (indicative of haemolysis during clotting).

Animal sera are a potential source of microbiological contaminants, notably mycoplasma, bovine viruses, and possibly the agent which causes Bovine Spongiform Encephalopathy (BSE). Consideration should be given to impact of contamination with the most common viruses in bovine herds such as BVDV and BPV agents (WHO 2010). Suppliers use a variety of techniques, including filtration, irradiation and heat-inactivation, to reduce microbial contamination. Nevertheless, it is wise, and for some applications, obligatory, to specify sourcing of serum from countries where there is a low risk of infection, and, in the case of bovine sera, from not too old animals to reduce any potential risks (Festen 2007). The use of human serum is restricted to specialised applications as it carries additional risks, such as the potential presence of human pathogenic viruses. Its use must be subject to the strictest quality controls, including documentation to demonstrate origin and viral safety (Coecke *et al.*, 2005).

Variability between serum batches can lead to experimental variability and reduce inter-laboratory reproducibility, representing a major cost associated with cell culture (Usta *et al.*, 2014). Notably, the use of serum can possibly lead to unexpected or undesired outcomes. In this regard, a study has shown that serum can inhibit transforming growth factor (TGF)-β1-induced chondrogenesis in fibroblast-like type-B synoviocytes (Bilgen *et al.*, 2007). In another study, FBS, compared to autologous (human) serum, was found to induce a more differentiated and less stable transcriptional profile in human bone marrow mesenchymal stem cells, particularly at late passages, as shown by analysis of genome-wide microarray analysis (Shahdadfar *et al.*, 2005).

Moreover, EU legislation, through the Directive 2010/63/EU on protection of animals used for scientific purposes, offers a certain amount of protection for foetal forms of mammals. Considering that blood harvesting from foetal calves can cause pain and distress in these animals, as previously reported (Jochems *et al.*, n.d.; van der Valk *et al.*, 2004), and

considering Articles 1 and 3 of the Directive, harvesting of FBS from live bovine fetuses in the last third of their development for scientific purposes is a procedure under the Directive.

In 2008 the ECVAM Scientific Advisory Committee (ESAC) stated that "*for methods forwarded to ECVAM for validation/prevalidation where [the use of non-animal alternatives to serum] is not fulfilled a justification for future use must be provided, including measures taken to seek non-animal alternatives to [FBS]*" (ESAC, 2008) The drawbacks of using FBS and recommendation to replace it with available chemically defined serum free media is also mentioned in the 2005 GCCP guidance issued by EURL ECVAM (Coecke *et al.*, 2005).

Additionally, FBS is a by-product of the meat industry and for this reason is dependent on several external and generally unpredictable factors, such as weather conditions (e.g., changes of climate with drought or flood) or variations in beef consumption (e.g., switch to poultry). All these factors may make FBS supply unpredictable (Brindley *et al.*, 2012).

For all these reasons, serum-free media, (animal) protein-free media and chemically-defined media has gained interest over the years in order to avoid many of the disadvantages associated with the use of serum, as recently commented (Pamies *et al.*, 2016).

4.3.2 Serum-free media and serum replacements

Serum-free media are thought to circumvent the batch-to-batch variability issues associated with serum and offer better reproducibility and the potential for selective culture and differentiation of specific cell types (Geraghty *et al.*, 2014). Nevertheless, serum-free compositions may still need to be validated and monitored similarly to serum containing media as they are often not completely chemically defined. For instance, serum free media can include other poorly defined supplements such as pituitary extracts, chick embryo extracts, bovine milk fractions or bovine colostrum. Furthermore, some so-called 'defined' media contain complex serum replacement mixtures including chemically undefined agents. Notably B27 and its alternative NS21 used in the culture of neural cells contain bovine serum albumin and transferrin which can exhibit batch-to-batch variation in biological activity (Chen *et al.*, 2008). Therefore, it may be useful to carry out pre-use testing on new batches of reagents which could demonstrate variability that cannot be foreseen from manufacturers' information. Another example of an essential component prone to batch-to-batch variability is the so-called 'basement membrane extract', purified from Engelbreth-Holm-Swarm (EHS) mouse sarcoma cells and marketed under various trade names.

Serum-free medium formulations play a critical role for the culture of stem cells, such as human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs), which currently hold great promise for toxicology and regulatory testing and biomedical research. Both hESCs and hiPSCs are often maintained on inactivated mouse or human embryonic fibroblasts or under feeder-free conditions (using extracellular matrices) in chemically defined, serum-free media, in order to avoid the presence of undefined or unknown serum components (which may compromise the differentiation towards desired cell lineages) and the risk of contaminations from pathogens (e.g., mycoplasma, viruses, and prions) (Pistollato *et al.*, 2012; Yamasaki *et al.*, 2014).

4.4 The use of antibiotics in cell culture

Routine culture of cell lines under GCCP (Coecke *et al.*, 2005; Geraghty *et al.*, 2014; Stacey and Davis, 2007) should not require the use of antibiotics and it can never be relied on as a substitute for effective aseptic techniques. However, its use is still widespread due to established routine procedures in many laboratories. Antibiotics are agents that may arrest or

disrupt fundamental aspects of cell biology, and, while they are effective against prokaryotic cells (i.e. bacteria), they are also capable of causing toxic effects in animal cells. Not surprisingly, antifungal agents, being directed at higher order, eukaryotic micro-organisms, are likely to be more toxic to animal cell cultures. Given these obvious contra-indications, the use of antibiotics in cell and tissue culture should be focused in two areas: a) protection of materials at high risk of contamination such as tissues, organs and primary cultures in cases where sterility cannot be guaranteed; and b) the positive selection of recombinant cell clones based on the expression of antibiotic resistance genes (Coecke *et al.*, 2005).

4.5 Additional media components

Some media components are heat labile (e.g., L-glutamine), sensitive to light (e.g., retinoic acid) or have a limited half-life in diluted state or at high ionic strength, such as in prepared media (e.g., epidermal and fibroblast growth factors). These issues are best addressed by preparing a small volume of media necessary to cover the period of stability of the most sensitive component and discarding bottles after a set time period. Appropriate size aliquots of those labile components may be frozen by an appropriate method for long-term storage. In this respect, stock solutions with low concentrations of protein aqueous growth factors may require the addition of albumin or other excipients, to prevent adsorption to plastic and to increase stability in the frozen state. Stabilised forms of glutamine and retinoic acid are also available to avoid these issues.

In case culture media or other reagents have to be sterilised via heat or filtration, the impact of the procedure should be assessed and recorded. For example, heat sterilisation may result in degradation (or denaturation) of one or more of the components and filtration can remove individual and/or essential components (e.g. Fe^{2+} or Fe^{3+} iron products that enhance growth of mammalian eucaryotic cells in serum-free cultures).

4.6 Dedicated media to particular cell lines

Different cell types or tissues need to be cultured in culture media containing various components at different concentrations to allow optimal growth. Although certain cell lines may be grown in media with the same composition, sharing media increases the risk of cross-contamination. Therefore, each cell line should be cultured with separate dedicated media, which must not be shared with other cell lines.

5 Test Systems

In vitro method cell and culture-based test systems are routinely used by all industries and regulatory bodies in toxicity testing, safety assessment, and risk evaluation. The greatest use of *in vitro* test systems, however, is for elucidating mechanisms of toxicity and/or demonstrating the biological process involved, when exposing test systems to toxicants of various kinds. With the advances in genetics and genetic screening approaches, routine *in vitro* methods include already the use of either genetically altered cells, stem cells, stem-cell-derived models and organ-on-chip models and other complex and sophisticated systems.

5.1 GCCP

Good Cell Culture Practice identifies a set of core principles of best practice for working with simple but also with more complex cell and tissue culture systems (Annex 1 and Annex 2). Failure to comply with GCCP can have serious consequences for individual researchers and their employers, which include invalidation of important data sets, ethically compromised research, loss of crucial cultures due to microbiological contamination, failed patent applications and injury or infection amongst laboratory workers. These could clearly have serious consequences for the institutions involved, including the threat to scientific reputation, legal and regulatory compliance and cost of wasted resources.

The principles of the first version of GCCP published in 2005 (Coecke *et al.*, 2005) remain highly relevant to cell culture practice for *in vitro* methods today (see Annex 1), and aims at a far broader set of applications, including research, manufacture of medicines, and laboratory based GLP testing. In providing the detailed and specific principles of best practice for the handling and management of cell cultures systems, GCCP is a vital component of GIVIMP.

A new and updated version of GCCP called "GCCP2.0" is under development through collaboration between European, Japanese and North American experts and incorporates more recent developments in cell culture, and in particular addresses the new technological developments of human pluripotent stem cell lines and complex 3D culture systems. During a workshop in 2015 (see Annex 2) new key elements for GCCP 2.0 were identified.

5.2 Cell and tissue sourcing

Complex biological substrates such as cell and tissue cultures may have passed through a number of stages of development before they are qualified for an *in vitro* method and it is important that this track is documented, thus assuring traceability to original source materials, source cell banks, absence of contamination by major classes of biological agents (e.g. mycoplasma, bacteria, fungi and cytopathic viruses), genetic identity/consistency and stability of desired functionality. See Annex 1, GCCP principle 3 and Table 1~~Table 1~~ in section 1.2 for examples of document requirements concerning the origins of cells and tissues and see paragraph 5.5 below.

Proposed sources of test systems should be qualified by the user to assure they are appropriately traceable, have been quality controlled for key features (see paragraph 5.7 below) and in addition, the user should check that there is solid ethical provenance (e.g. hPSCreg registry[16]) and safety assessment for the cells. In addition, intellectual property

[16] http://www.hpscreg.eu/

Formatted: Font: Times New Roman, 12 pt

Field Code Changed

should be checked to ensure it does not impact on the use of the cell line and data. For more detailed information on these issues see (Stacey *et al.*, 2016).

In the case of human tissues and primary cells, there will also be a requirement to assure donor consent and to manage sensitive personal data e.g. according to the EU Data Protection Directive. Human tissue is precious and difficult to secure. The broad range of issues in securing tissues for testing were addressed at the 32nd Workshop of the European Centre for Validation of Alternative Methods (Anderson *et al.*, 1998) and where tissues cannot be sourced via a qualified tissue bank, there should be an agreed testing method in place with clinical contacts regarding all aspects of harvesting, preparation, labelling, storage and transfer (for an example see (Stacey and Hartung, 2006)). It is also important to assess the risks of viral contamination of primary cells and tissues. Approaches to risk assessment are described in (Stacey and Hartung, 2006). Tissues should be obtained from tissue banks holding only materials from screened donors and this will significantly assist in managing viral safety issues.

Availability of cell lines from a certified source (established cell banks with a high quality standard; reputable culture collections, commercial provider), that usually provide extensive documentation on the origins and characterisation of the test system, should be assured[17]. Moreover (or alternatively), master and working cell banks should be established to guarantee a cell supply of constant quality and records kept of the original source.

If test systems used in *in vitro* studies are genetically modified the Directive 2009/41/EC (EU, 2009) is applicable. This Directive lays down common measures for the contained use of genetically modified micro-organisms (GMMs), aimed at protecting human health and the environment. A notification has to be sent to the competent authorities before any use commences in the premises. A risk assessment of the GMMs used has to be performed. The Annexes to the Directive detail the criteria for assessing the risks of GMMs to health and the environment, as well as the protective measures for each of the four levels of containment. The Directive lays down the minimal standards applicable to the contained use of GMMs. Individual European member States are also permitted to take more stringent measures.

5.3 Handling and maintenance of the test system

During routine handling and maintenance, growth and survival characteristics of the cell system (such as cell viability, doubling time, etc.) and subculturing details (e.g. date of subculture, subculture intervals, seeding density, passage number, etc.) should be recorded and documented in the study report, since they are required for the complete traceability of results. The documentation provided by the test system supplier should be taken into account together with the historical data, when available, and used to establish acceptance criteria. See Table 1~~Table 1~~ for examples of document requirements concerning the handling, maintenance and storage of cells and tissues.

Different cell lines have different growth rates which may depend on several environmental factors. Whether cells grow and divide in a monolayer or in suspension, they usually follow the same characteristic growth pattern in which four different phases can be recognized: lag, log (or exponential), stationary (or plateau) and decline. Growth during exponential growth or log phase is usually fairly constant and reproducible for a given set of growth conditions (ATCC, 2014). Each cell line will show different cell proliferation kinetics during the log

[17] http://wiki.toxbank.net/w/images/1/18/ToxBank_D4_6_final_10_04_13.pdf

Formatted: Font: Times New Roman, 12 pt

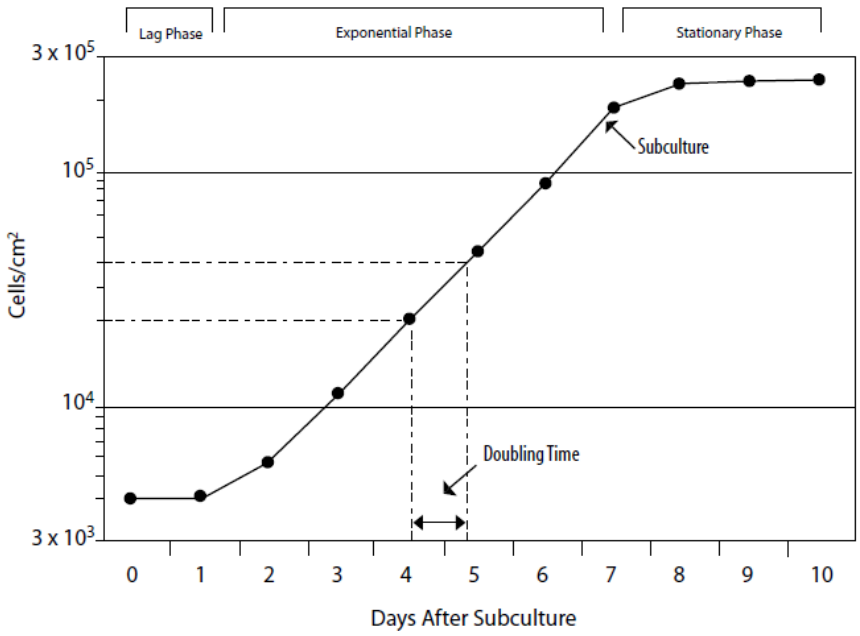
Field Code Changed

phase and it is therefore the optimal phase for determining the population doubling time (ECACC, 2010).

Many dividing primary human cell cultures have a split ratio of one in two (1:2), while continuous cell lines have much higher splitting rates. In order to ensure viability, genetic stability, and phenotypic stability, cell lines need to be maintained in the exponential phase, i.e. they need to be subcultured before a monolayer becomes 100% confluent or before a suspension reaches its maximum recommended cell density.

Many cell lines can be subcultured based on a rough estimate of cell density, and this is the usual practice unless stated otherwise in the cell maintenance protocol. Some cell lines require a fixed seeding density and subculturing scheme and counting the number of cells is required (Wilson *et al.*, 2015). Most commonly cell counting is performed using the Bürker Türk or Neubauer counting chambers. When automated cell counters are used, their correct functioning would need to be demonstrated (Cadena-Herrera *et al.*, 2015; Gunetti *et al.*, 2012; Phelan and Lawler, 2001).

Figure 3: Growth curve for cells grown in culture. Cells should be subcultured while still in the exponential phase (ATCC, 2014)



Each test facility should develop SOPs, where details are provided about how to thaw, handle, count, maintain and store cell lines. For example, the procedure to univocally assign progressive passage numbers as well as the selected assay(s) to determine the cell stock viability should be established.

5.4 Cryopreservation

Improved technologies that allow cryopreservation of *in vitro* cell and tissue cultures at different stages of differentiation, and their long-term storage has introduced new or more standardised *in vitro* test system into the pipeline of potential *in vitro* methods to be used in human safety assessment. Controlled-rate and slow freezing, also known as slow programmable freezing have been used all over the world for freezing down cell and tissue cultures to better preserve it for eventual thawing, before it is frozen, or cryopreserved, in liquid nitrogen. New methods are constantly being investigated due to the inherent toxicity of many cryoprotectants.

As described in GCCP Principle 1 'Establishment and maintenance of a sufficient understanding of the *in vitro* system and of the relevant factors which could affect it' (Coecke *et al.*, 2005) it is important to prepare preserved banks of cells intended for use, to assure that reliable stocks can be obtained for testing, which are at a consistent passage level from the original 'seed stock'. This is in order to avoid the effects of changes or cross-contamination which may occur if cell lines are maintained indefinitely. Standard cryopreservation methods using DMSO (10%) and serum (20%) as cryoprotectants, combined with a slow cooling rate (e.g. -1^{o}C/min), will usually be successful for most cell lines. However, it is important to check the viability of preserved stocks in case of freezing failure and also to try to assure consistency between individual vials in a cell bank regarding cell number, viability and desired functionality. It is important to note that viability measurements made immediately post-thaw can give misleadingly high values as many cells can be lost during the 24 h recovery phase post thawing. Therefore, it may be important to understand what losses occur at this stage.

5.5 Cell line identity and genetic aberrations

Ideally, there should be traceability to the original provider of the cell line and the related documentation. However, a frequent problem in the use of cell culture is the use of cell lines which have become cross-contaminated, misidentified (see ICCLAC database of cross contaminated or misidentified cell lines[18]), mixed-up, or underwent genomic instability (Allen *et al.*, 2016; Frattini *et al.*, 2015; Kleensang *et al.*, 2016; Vogel, 2010). This is not always detectable by cell morphology only. Establishing an early stock (or retention of a sample of original tissue) which is DNA fingerprinted will provide an important reference for future cell banks and for other centres. Short Tandem Repeat (STR) profiling is typically applied and has considerable background qualification for use in human samples (ISCBI, 2009). STR analysis can be performed in most laboratories that have the capabilities to execute molecular techniques. It is an easy, low cost and reliable method for the authentication of human cell lines. An ANSI US standard exists for this technique with cell lines. For non-human samples, STR methods might be available but may need to be qualified for specificity to individual cell lines (ASN-0002: Authentication of Human Cell Lines: Standardization of STR Profiling can be found at http://webstore.ansi.org). Other techniques have also been used including isoenzyme analysis but probably the most common and well-qualified technique is cytochrome oxidase gene sequencing (Ono *et al.*, 2007). However, the field of genetic analysis is progressing rapidly and interested parties should maintain knowledge of current best scientific practice in this area as next generation sequencing begins to become a routine tool.

Field Code Changed

Field Code Changed

[18] http://iclac.org/databases/cross-contaminations/

Field Code Changed

Genetic instability is inherent in cell cultures and it is wise to minimise the number of passages over which cells are maintained (typically p15-20). Although passage number alone is not a reliable parameter to ensure good cell functioning, it is good practice to define a limit for the maximum number of passages, possibly in combination with defined performance characterstics. At that limit, new cultures should be restarted from a working cell bank. The use of cells at higher passage numbers must be justified and their integrity and functionality demonstrated. Cultures at passage numbers beyond which it is known that the cell line functionality is maintained should not be permitted for use. Where cells are known to be extremely unstable, some form of genetic testing, such as karyology or molecular analysis like single nucleotide polymorphism arrays (aSNP) or comparative genomic hybridisation (aCGH) may need to be performed. In particular, this applies to recombinant cell lines including those maintained with antibiotic selection.

There are special issues for stem cells. Stem cell lines may contain a mixture of diploid and aneuploid cells, which may be unavoidable, but genetic testing (see above) can be used to screen for progressive change (e.g. between master and working cell banks) which could impact on the suitability of the cell culture. Human iPSC lines should also be tested for absence of ectopic expression of reprogramming genes and where produced by non-integrating vectors, for elimination of the vector.

5.6 Contaminants screening: sterility, mycoplasma, virus

Standard sterility tests are published and may be used for cell stocks and cultures. However, it is important to bear in mind that these are usually based on inoculation of broth cultures which may not support the growth of all contaminating micro-organisms. Alternative molecular methods such as identification by PCR and DNA sequencing of ribosomal RNA may be used.

It is absolutely mandatory to quarantine (see 3.5) and to confirm to be negative for mycoplasma and free of bacteria, yeast and fungi of all cultures passaged in a particular laboratory environment.

Viruses may arise as contaminants of cell cultures via the original donor used to produce the cell line or feeder cells and other biological reagents used in cell culture. They may cause cytopathic effects, in which case the culture should be discarded, or they may have no effect and become diluted out when fresh uncontaminated reagents are used. In certain cases they may establish persistent infections, although this is believed to be rare. Whatever the outcome, their presence and influence on cell biology may be significant as amongst other effects they may modify transcription factor networks and alter the cells' biology. To assure lab worker safety, some organisations require testing of all human cell lines for serious human pathogens such as human immunodeficiency virus (HIV) and Hepatitis B&C or evidence that the donors did not have these pathogens. However, such testing clearly does not cover more common human infections, and human pathogens may also be carried by cells from other species. Cell cultures should therefore always be handled and disposed of as if they are potentially infectious (Stacey, (2007) chapter 31: Risk assessment of animal cell cultures, In : Medicines from Animal Cells (Stacey and Davis, 2007). Cell line testing may be initiated if there are special hazards associated with the work or with the cells. Workers should always follow local rules for performing cell culture work, maintain their competence in aseptic processing, as well as carry out regular and careful inspection of cells for any unusual effects or morphologies that might indicate infection. It is an important part of a robust testing regime for contamination to have a procedure for managing positive results,

whether this means immediate discard or quarantine until a means of action can be decided along with detection of the root cause supplementary testing (Stacey, 2011).

5.7 Quality Control

It is important that certain key go/no-go points are established during the preparation and use of the test system for an *in vitro* method.

Typically, for use of cell lines, appropriate integrity checks should be applied at the sourcing of new cell lines or cell bank preparation, to assure stable passage and expansion for use and reproducible starting cells used to provide the test substrate, and, finally, to ensure acceptable and reproducible performance against controls in formal tests. Table 4~~Table 4~~ shows an example of measures at different stages.

Formatted: Font: Times New Roman, 12 pt

Table 4: Applicability of Integrity Checks on Cultures

Characteristic	Source *	Early stocks of preserved vials**	Cell banks***	Routine in process testing of stock cultures****
Morphology	X	X	X	X
Viability	X	X	X	X*****
Identity	X	X	X	
Doubling time	X	X	X	X******
Mycoplasma	X	X	X	X
Viruses	(X donor)		(X master bank)	
Bacteria and Fungi			X	X*******
Function/phenotype		X	X	
Genetic stability			X	(X)
Absence of reprogramming vectors (iPSC lines)		X	X	

*This may be provided in writing but ideally with evidence of test results or a qualified service provider test result. These tests should also be performed on the cells arriving in the lab as soon as samples can be obtained.

** A small number of vials frozen as soon after arrival of a new culture to avoid loss in case of a lab accident.

***Ideally master and working banks should be established (Coecke *et al.*, 2005) but testing may be focused on the master stock with more routine checks applied to working cell banks e.g. mycoplasma and viability.

****Stock cultures of cell lines maintained for use in *in vitro* methods should be tested routinely e.g. all lines in culture tested once per month.

*****Viability testing at passage will also be helpful to ensure consistent seeding of fresh cultures and assays for more reliable maintenance of stock cultures and reproducibility of cell-based *in vitro* methods. For this, the assays described under 6.2.1.1 can be applied.

******For diploid cultures, passage number is roughly equal to the number of population doublings (or population doubling level) since the culture was started (ATCC, 2014).

*******To avoid development of low grade contamination, sterility testing may be desirable for long term cultures. These may also be sustained as separate replicate sets of flasks to provide backup cultures in case of contamination.

Where primary cell cultures and tissues are used, variation in properties between individual donors must be considered, and each new batch should be qualified or controlled for key functionality. Special care should be taken to note any unusual observations in case of contamination or viral cytopathic effects or transformation, and all primary cell cultures should ideally be cryopreserved and screened for mycoplasma. Human and animal tissues and primary cells used for testing will also need to be appropriately documented, As part of QC for tissues, their differentiated state should also be documented, which may require a range of assays, which may include, morphology, histochemistry, cell markers, specific tissue function and cell-cell/matrix interactions (Stacey and Hartung, 2006). For primary cells prepared from tissues stored as banks of cryopreserved vials of cells, similar QC approaches can be used as adopted for banks of continuous cell lines (see Table 4~~Table 4~~).

Formatted: Font: Times New Roman, 12 pt

5.8 Biomarkers and functional tests to confirm the required cell function state

It is important to recognise that cell quality can vary during passaging, and in particular the time point in the growth curve at which cells are harvested may affect performance. In general, cells are best harvested in the logarithmic phase of growth. Accordingly, each culture used to set up an *in vitro* method should be subject to a key control regime measuring or indicating functionality. Acceptability criteria should be defined for functional tests and biomarkers that indicate the correct cell state. These may for example include: neuronal activity, competency of biochemical transformation, response to reference bioactive compounds, response to reference items in the particular *in vitro* method the cells are to be used for etc. In this way, each culture can be controlled, and consistency in *in vitro* methods is supported. Additionally, key markers which are associated with poor performance may be identified for future improvement For example, expression of self-renewal genes (e.g. Oct4, Nanog, Sox2) in stem cell cultures is crucial to the functionality of the cell population (further examples for stem cells are laid out in Pistollato *et al.*, 2014 and Stacey *et al.*, in press).

5.9 Special issues for microbial strains

Microbial strains are used in many *in vitro* methods, and all the generic aspects for *in vitro* methods given here are also applicable. However, there may be special requirements needed to be applied to assure selection of certain mutant or recombinant strains (as for some recombinant cell lines). Moreover, functional tests (biochemical, cell biological, etc.) may need to be necessary to assure correct identity of each species and/or strain. Preservation of cultures is usually achieved by freeze-drying, for which storage conditions and their stability may vary for different organisms. Some strains may require cryopreservation and/or more careful testing for genetic integrity (including maintenance of plasmids) and relevant function (e.g. sensitivity to UV-light or antibiotic resistance) on recovery.

5.10 Qualification of reference strains

Where a common cell line or organism is recommended for a particular *in vitro* method it may be available from numerous sources. The selected cell line should be qualified for use in a way which can assure consistent function in the method. The idea of reference cell banks to act as central sources for all users of cell substrates has been developed by WHO (WHO, 2010) and formal or regulatory *in vitro* methods may quote catalogue references from culture

collections for both microorganisms and cell lines. However, culture collections may not necessarily check the performance of such strains using reference testing methods and as such, this still will need to be conducted by the *in vitro* method developer or user. If an original and quality controlled source of a cell line is not available, it is wise to obtain cells from more than one source in order to compare their performance and authenticity, as cross contamination of cell lines is very common. Moreover, cell lines with different histories of use in test facilities have shown to result in different phenotypic characteristics.

6 Test and reference items

This chapter describes the characterisation and preparation of the test item (e.g. a chemical to be characterized for potential hazard) and of relevant reference items for the *in vitro* test. Furthermore, details will be given as to how test item stability must be monitored, how the *in vitro* test environment can affect the test item and how the biokinetics of the test item in the *in vitro* method should thus be assessed. These aspects are important to ensure reproducibility among laboratories and certainty that the outcome of the *in vitro* test is indeed related to the test item. Additionally, it is described which control and additional items should be applied in general to verify correct function of the *in vitro* test, such as negative and positive controls (OECD, 2004c).

A distinction is made between aspects that are important in the development phase of an *in vitro* method, and those that matter when the *in vitro* method is used routinely for regulatory purposes. For instance, the use of test and control items may differ between test development and routine use.

The stability of the test item (or its representative preparation) and whether it is in contact with the reactive part of the test system in a known or assessable concentration should be determined when applying the *in vitro* method, in order to collect robust data from the *in vitro* method.

6.1 Test item

6.1.1 Considerations during the development of the method

Test items are not used in the phase of *in vitro* method development, i.e. before the method performance has been validated. During the phase of test development, chemicals or products with well-known characteristics are evaluated to assess the relevance of the method and to amass results which will be used to set the acceptance criteria (see 6.4). Nevertheless, during the development phase, it is important to determine:

- The different natures of test items for which the method is intended (define if the *in vitro* method can be used for liquids, solids, certain powders, mixtures, multi-component test chemicals, certain preparations, suspensions, nanoparticles, emulsions, etc. (OECD, 2000)). Moreover, the process of preparation of these test items should be explored.
- The process of preparation of the test item, if needed before the test.
- Which chemicals are suitable as reference and control items.

6.1.1.1 Nature of test items for which the method is suitable

The nature of the substances for which the developed *in vitro* method is suitable for may be characterised, but not necessarily limited to, by describing it by using a number of different properties (see below). The list given below is not exhaustive and may need to be developed depending on the nature of the test. It is important to note that solubility is a highly important yet often neglected characteristic and is therefore described in more detail in a separate section (6.1.1.2):

- State: solid, liquid, gas and all in between-states such as aerosol, dust, or viscous liquid (see OECD TG 114 for determination of the viscosity of liquids); depending on its state, the substance could require preparation steps before the test (see also 6.1.1.3~~6.1.1.3~~) or a specific administration mode in the method, such as dry dispersion with pressurised air, nebulisation of a liquid formulation, or spark generation.

Formatted: Font: Times New Roman, 12 pt

- Appearance: nominal size, morphology, size distribution, aggregation and agglomeration phenomena and surface characteristics (surface area, surface charge, surface chemistry) are essential characteristics to know the nature of a certain nanomaterial (OECD series on the Safety of Manufactured Nanomaterials, n°36).
- Colour: some test items interfere due to this characteristic with the endpoint detection method.
- Physicochemical characteristics:
  - pH for test item in solution (see OECD TG 122 for pH determination) and pK_a (pKa indicates to what extent the test item may become ionised at the pH of the test system). Changes of pH can also affect the test item in other ways than its ionisation (see OECD TG 111 for sensitivity of hydrolysis to the pH, for example).
  - Osmolality
  - Volatility
  - Solubility: see 6.1.1.2
  - Dissociation constants in water (see OECD TG 112): dissociation is the reversible splitting into two or more species which may be ionic. The dissociation governs the form of the substance in the test system, which in turn determines its behaviour and transport and which thus may affect the adsorption of the substance to culture dishes or the penetration into cells or adsorption onto proteins in solution or resulting in a aspecific aggregation behaviour.
  - Lipophilicity (see OECD TG 123 and 107 for determination of the partition coefficient = K_{ow}).
  - Homogeneity and conditions of stable homogeneity
  - Sensitivity to photolysis (OECD TG 316 is meant for environmental fate, thus less suitable for this purpose, but can be used as a basis)
    - Composition and purity: chemical purity/contaminants, microbiological contaminants (including e.g. cell walls of decomposed microorganisms), biological purity (e.g. of cells lines or test microorganisms, or complex protein mixtures (vaccines)), composition of complexes (vegetal extracts, products of fermentation, etc.). In case of a mixed solution, the list of ingredients with percentages of each component can be relevant to describe the composition. For each component, information like molecular weight, chemical formula, CAS number, etc., is useful. Complex substances could require different information. For example, substances of Unknown or Variable composition, Complex reaction products or Biological materials (UVCBs) cannot be sufficiently identified by their chemical composition, because the number of constituents is relatively large and/or the composition is, to a significant part, unknown and/or the variability of composition is relatively large or poorly predictable. The composition could then be defined by the manufacturing process description[19].
- Conditions of stability: the limits of temperature, pressure, and humidity to maintain stability of the test item (to be compared with the *in vitro* method conditions).
- Microbiological status: requiring aseptic conditions or not.

In general, the limits on test item suitability are determined so that reproducible and definable interactions between test item and test system can be guaranteed (see chapter 6.2).

[19] https://echa.europa.eu/documents/10162/13587/10_sb_siduvcb_d1_lrws_20120203_en.pdf

Field Code Changed

6.1.1.2 Solubility

The solubility defines how much of the test item (molecular and ionized forms) that can be maximally dispersed in the solvent to be used for the test (=thermodynamic solubility). The rate of dissolution (in practical terms: fraction of dissolved test item after a fixed time) is called kinetic solubility (Jouyban and Fakhree, 2012). This section is focused on thermodynamic solubility.

The reason why solubility is an important property to know of a test item is that, above this limit, the substance precipitates and the effective concentration in the test medium is lower than the nominal concentration. Precipitates may also affect read-outs of the *in vitro* method and lead to impaired reproducibility within and between labs. It is thus important to ensure that all test items are properly dissolved for use in the *in vitro* method. The highest test concentration has to be below the solubility limit.

Different ways to determine the solubility of a substance in a defined solvent are available. While computational methods for solubility predictions in different types of solvent or matrices are available (Bergström et al., 2002; Persson et al., 2013), it remains difficult to predict the solubility in the medium used for the *in vitro* method. Most frequently, solubility of a compound is determined visually, by determining whether a clear solution is formed after adding the substance of interest. However, while the visual test is very simple to apply, it gives only a rough impression of the solubility. In most cases however, it is sufficient for simple checking of solubility where reliability can be enhanced by use of microscopy to detect solid particulates or liquid droplet suspension (indicative of insolubility). However, it is a rather subjective operator-dependent judgement. Reliability can also be improved in a relatively easy way by centrifugation, particularly for detection of precipitation in medium dilutions, where foaming may obscure visual observation. Note that solubility is affected by the composition of the substance (presence of impurities) and by the test conditions (temperature, incubation time, possible adsorption to the test vessel or to medium constituents (e.g. albumin)). OECD TG 105 can be used for the determination of the aqueous solubility of pure substances which are stable in water and not volatile, and OECD TG 116 can be used for fat solubility determination (fat solubility is the mass fraction of substance which forms a homogeneous phase with a liquid fat (oil) without giving rise to chemical reactions). Nephelometric, UV-spectroscopic and HPLC methods can also be used to determine solubility (Hoelke *et al.*, 2009).

Nephelometry facilitates solubility determination, particularly suited to serial measurement (e.g., ranges of chemicals and/or concentrations) allowing systematic and precise evaluation of turbidity due to dispersed precipitation, independent of matrix composition. However, the measurement is relative, requiring a definition of threshold turbidity for insolubility based on expedient practice with detection limit dependent on instrument sensitivity. Moreover, even nephelometry may not detect chemicals such as transparent immiscible liquids for which visual inspection, enhanced by experienced microscope observation, remains a reliable approach.

HPLC and UV spectrophotometry provides a quantitative determination of concentration with the use standard curves. While both methods are valid for solutions prepared in solvent, they may not be valid for preparations in biological media, which contain many components that often interfere with solubility detection. Cell culture media cannot be injected into HPLC columns and their multiple components will likely obscure the compound of interest through their inherent UV absorbance over a range of wavelengths. This necessitates pre-purification and extraction steps for quantifying test item concentration in media via HPLC.

Test item chemicals are generally dissolved in solvent (e.g., DMSO, Ethanol) to create a stock solution at a predetermined target concentration (e.g., 50mg/mL or 100mM). The test item should have a relatively high solubility in the solvent of choice and the solvent should not interfere with the test item (e.g. inactivate the compound). For example, the commonly used solvent DMSO can reduce the effects of the platinum complexes (Hall *et al.*, 2014). In addition, the solvent should not affect cell health or the phenotype of the cells in the assay when diluted in media. Furthermore if the stock solution is diluted in media up to a concentration exceeding the solubility of the substance, the test item may precipitate.

Regarding the sample preparation procedure, the following issues are key for reproducible results:

- Optimal time for dissolution in solvent: Does the drug dissolve immediately (1 min vortex) in the solvent or does it require additional treatment (sonication and warming) and time (e.g., 2 hours or overnight)?
- Solubility in media upon dilution: Including incubation to mimic assay conditions may be relevant: i.e., monitoring stability on incubation (e.g. at assay temperature and in the presence of CO_2) over a time period (e.g. 24 hours).
- Visual inspection of solubility: Unless sample material is expensive or available quantities are limited, stock solutions for visual inspection of solubility should be prepared with a minimum weight of 25mg, and in a minimum volume of 0.5 ml.

The solubility in assay medium may be higher than in pure water, due to adsorption to medium contents such as proteins. In addition, the higher temperature applied in the *in vitro* method (37°C) than in OECD TG 105 (20°C) will give rise to a higher solubility in the assay. However, assay media typically have a rather high ionic strength and an inherently a complex composition. This makes it difficult to predict test item solubility upon dilution from stock solution in various media solvents. Therefore, it is also necessary to determine the solubility of the final test concentrations in the assay medium under assay conditions. In case of inorganic substances, the anion and cation part of the test item may precipitate with other cations and anions present in the culture medium, if the solubility of these newly combined salts is exceeded. It is therefore recommended to visually monitor the test system for precipitation, and to verify whether the anions and cations present in the medium can form low-solubility salts with the test item.

Test items insoluble within an acceptable concentration range should be considered incompatible with the *in vitro* method. A remedy may be to increase the solvent concentration in the *in vitro* method. However, the tolerable solvent concentrations will depend on the solvent and the test system used: As a general rule, the final solvent concentrations should be as low as possible to avoid any potential interference with the *in vitro* method. The concentration of solvent present in the final test concentration should be considered in the test controls.

As for nanomaterials, special issues on measuring solubility and dispersion characteristics may arise. For these materials, the specific guidance documents are best followed, which are continuously being developed (Scenhir, 2015). Any toxicity testing using *in vitro* methods should pay special attention to the agglomeration/aggregation behaviour, and the insoluble/ partially-soluble nature of nanomaterials (Scenhir, 2015). Possibilities for dis-agglomeration and re-aggregation of nanomaterials should also be considered: some properties of nanomaterials may change due to interaction with the surrounding media.

6.1.1.3 Test item preparation

Test items may have to go through various steps of preparation, such as dissolution, dilution, extraction by wetting or centrifugation before being suitable for use in the *in vitro* method. These steps must preserve the characteristics of the test item. The purpose of each step of the preparation has to be explained, and the critical limits step/procedure should be determined. The impact on the test item stability, homogeneity and integrity should be assessed.

Once prepared as a dilution in a solvent, chemical stability in solution is crucial, since it determines the concentration of the work solution to be applied to the test system. The concentration in stock solutions and exposure medium has to be measured over time, to check for stability and avoid aberration in cell treatment. Relevant examples of compounds hydrolysed in aqueous solutions can be found in (Crean *et al.*, 2015; Pomponio *et al.*, 2015).

For more complex test items there are existing guidelines to aid this process: e.g. ISO 10993-12 gives conditions of the extraction to obtain a representative extract of medical devices depending on their composition, and The OECD series on the Safety of Manufactured Nanomaterials, n°36, gives advice on how to prepare and characterise a nanomaterial solution.

6.1.2 Considerations for the final user of the validated method

The type of test item for which the developed *in vitro* method is suitable has to be defined in the description of the validated method. Test items at the user stage should be characterised as required by the GLP principles (see OECD n°1, chapter 6.2): "each test and reference item should be appropriately identified (..); the identity, including batch number, purity, composition, concentrations, or other characteristics to appropriately define each batch of the test or reference items should be known (...).The stability of test and reference items under storage and test conditions should be known (…). If the test item is administered or applied in a vehicle, the homogeneity, concentration and stability of the test item in that vehicle should be determined (…)."

Interactions with the test system should be predicted and a way of assessment of such interactions should be implemented (cf section 6.2). Important examples are adsorption to protein, precipitation with carbonate in the medium and hydrolysis by aqueous environment.

6.2 Interaction between test item and test system

A two-way interaction has to be assumed between the test item and the test system. In one respect, the test system can affect the test item (in analogy to pharmacokinetics in *in vivo* models; detailed in 6.4). In another respect, the test item can affect the test system in specific ways (alteration of a readout, in accordance with the design and intended application of the test system; (see Sections 5 and 8) or in non-intended ways, by interfering with the overall performance of the biological model on which the test system is based, or by disturbing a correct readout of the test endpoint. There are endless possibilities for artefacts to be created in this way. As not all of these can be controlled automatically, experienced operators and personnel interpreting the test data are required to detect potential problems. Problem detection is also facilitated by regular inclusion of consistency controls and plausibility considerations (e.g. do compounds with similar structure or similar mode of action behave similarly?; can effects be reversed; does another test system for the same biological process give similar results?; are findings consistent with biological expectations concerning concentration and timing of effect?; etc.). For more detailed discussion, three elements of an *in vitro* test will be considered separately (Aschner *et al.*, 2016; Schmidt *et al.*, 2016): the test

system (biological model), the test endpoint, and the analytical endpoint (Leist *et al.*, 2010). These issues are relevant both in the development stage of an *in vitro* method and in the application stage.

6.2.1 Interference with the test system

Test items can disturb the test system, especially if it is based on living cells, as they are highly responsive to changes in their environment. The most frequent and serious disturbance is general cytotoxicity often leading to cell death.

6.2.1.1 Cytotoxicity

If the test endpoint is not cytotoxicity, then cytotoxicity triggered by test items is a serious confounder and needs to be controlled for. Indeed, changing cell numbers *in vitro* is known to affect observed effect concentrations (Gülden *et al.*, 2015, 2001). This is particularly important in repeated treatments (Kramer *et al.*, 2015).

Measurement of cytotoxicity should be done using the same conditions as used for the specific test endpoint (i.e. in identical samples, ideally during the same run, or even better on the same plate), so as to obtain reliable and relevant cytotoxicity data. Alternative approaches use measurements in parallel cultures. Viability controls in related, but not identical, culture conditions (different plate format, different cell preparation, etc.) should be avoided. The definition of cytotoxicity is not trivial. Often 100% cytotoxicity is referred to as 100% cell death, however, for a cytotoxicity assay with a metabolic activity endpoint, a 100% cytotoxicity just means that the metabolic activity is 0% compared to the healthy control cells. Therefore, the choice of method used for cytotoxicity determination, but also the interpretation of the results needs careful consideration (see Table 5~~Table 5~~). A single endpoint is usually not sufficient to be fully conclusive. A combination of cell counting and a population measure (e.g. resazurin reduction), or a combination of a viability measurement (calcein staining; dye exclusion, neutral red uptake) and a cell death measurement (propidium iodide uptake; LDH-release; annexin V staining) provides a greater level of certainty. Importantly, positive and negative controls for the viability assay have to be included and need to be considered for normalisation of viability data.

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

An important point to consider is the timing of cell death. For instance, a test item may trigger definitive cytotoxicity (e.g. > 70% of the cell population dying within16 h). If the specific test endpoint (and cytotoxicity) is measured after e.g. 4 h, then cytotoxicity will not be measurable by the method of choice in the test, although the cascade of cell death may have been triggered. There are no established rules on how to deal with this (relatively frequent) situation. One solution is to follow up on test results by alternative tests for the same endpoint, or by using the same test with a changed incubation scheme (e.g. prolonged incubation). This is particularly important, if data are used for risk assessment and far-reaching regulatory decisions.

Table 5: Viability testing of cell cultures

Endpoint	Assay	Mechanism and comments
1. Structural cell damage (non-invasive)	Evaluation of overall cell shape, cytoplasmic structure, flatness and outline	Screening assay that covers many forms of damage with high sensitivity, if observer is experienced. May be automated and rendered quantitative to some extent by high

	properties on a good phase contrast light microscope	content imaging. **Advantages:** high throughput, non-invasive, repeatable on same well over time. **Disadvantages:** No clear prediction model (only qualitative data, no exact cell death definition), no stand alone approach; requires extensive experience of operator.
	LDH-release test	Cells with intact membrane retain their content of LDH enzyme; LDH is released when cell membranes rupture (non-viable cells), and the enzyme can then be measured in the supernatant. To give fully quantitative data, the assay requires normalisation to the total LDH content of a culture well. It can to some extent be repeated for the same culture at different time points. **Advantages:** Measurement of a definite/unambiguous cell death endpoint; can be combined with cell function assays. Allows cells to be used for other purposes, if only supernatant is sampled. **Disadvantages:** Information only for cell populations. Normalisation necessary (extra wells for controls). Frequently high background LDH levels are observed (e.g. from serum components; signal/noise ratio can be bad in some culture media or with some cell types). Problems with long-term assays involving medium changes.
2. Structural cell damage (invasive)	Membrane penetration by dyes to detect 'cytotoxicity' (e.g. naphthalene black, trypan blue, propidium iodide, ethidium bromide, EH-1)	Dyes are selected so that they stain non-viable cells, but do not enter viable cells with an intact cell membrane. Some of the dyes stain the entire cell (e.g. trypan blue), others stain the nucleus/DNA (e.g. propidium iodide). Dyes that only stain dead cells usually need a combination with a method that stains/identifies all cells (such as phase contrast for trypan blue, or a nuclear counterstain (H-33342, acridine orange, SYTO-13) for fluorescent dyes. **Advantages:** Rapid and usually easy to interpret. Gives information on the single cell level. High throughput and absolute quantification are possible (high content imaging). **Disadvantages:** May overestimate viability since apoptotic cells continue to have intact membranes and may appear viable. Some dyes (e.g. trypan blue, H-33342) are cytotoxic, so that the evaluation has to be performed rapidly.
	Retention of dyes within intact cells to detect 'viability' (e.g. fluorescein diacetate or calcein-AM)	After dye exposure, viable cells fluoresce when observed under UV light. The lipid-soluble dyes are transformed by cellular enzymes (esterases) into lipid-insoluble fluorescent compounds that cannot escape from cells with intact membranes. Thus, cells can be observed under a microscope (single cell analysis) or with a fluorescent plate reader (population analysis). The dyes are often used in combination with a cytotoxicity stain (e.g. propidium iodide). **Advantages:** Rapid and usually easy to interpret. Gives

		information on the single cell level (including morphological information on the cell shape). High throughput and absolute quantification are possible (high content imaging, fluorescent plate reader or FACS). **Disadvantages:** Some cells leak the dyes; some cells actively export the dyes through P-gp activity. Many fluorescent dyes are prone to photo-bleaching, and some may be sensitive to their local environment (pH etc).
	Evaluation of programmed cell death/apoptosis markers	As programmed cell death is a universal cell biological process based on defined biochemical pathways and organelle changes, the activation of cell-death-associated pathways is often used as surrogate marker for cell death. An example for such a pathway is the activation of caspases (detectable in populations by enzymatic analysis or in single cells by staining) or the activation of endonucleases (detectable on population level as DNA-fragmentation). Moreover, a typical type of chromatin condensation (detectable by DNA stains) and the display of phosphatidylserine on the outside of the plasma membrane (detectable by annexin staining) is highly correlated with apoptotic death. **Advantages:** Adds mechanistic information to cytotoxicity data. Several endpoints are easy to quantify and useful for high through-put measurements. **Disadvantages:** Not all types of cell death may be detected by a given endpoint. Needs to be combined with a general cytotoxicity test. Some endpoints are prone to artefacts (annexin staining) and some staining techniques (TUNEL, caspase-3) lead to an un-intentional selection of subpopulations. Caspase activity measurement does not easily yield a prediction model for the extent of cell death.
3. Cell growth	Cell counting	For some cell populations, continued growth is a defining feature, and thus impaired growth needs to be considered as a reduction of viability. Notably, impaired growth/proliferation is not necessarily correlated with cell death; it is thus rather a functional viability endpoint than a cytotoxicity measure. A special case for growth is the increase in cell size without proliferation. This feature is e.g. seen for the extension of neurites by neurons. The gold standard analytical endpoint for the growth/proliferation endpoint is counting (or morphometry). There are many ways of counting cells, either as single particles (e.g. by FACS or HCI) or by assessing a biochemical parameter correlated to cell number (e.g. DNA content). **Advantages:** growth can be a sensitive parameter of cell well-being. **Disadvantages:** growth is a functional endpoint, not necessarily linked to cytotoxicity; artefacts for growth endpoints may arise from inhomogeneous growth of subpopulations: moreover, growth may hide ongoing cell

		death, and thus needs careful control in combination with cytotoxicity assays.
	BrdU or EdU incorporation	Measures new DNA synthesis based on incorporation of the easily detectable nucleoside analogs BrdU (or EdU) into DNA. BrdU can be detected e.g. by fluorescent-labelled antibodies in permeabilised cells. Alternatively, radiolabelled thymidine can be used. **Advantages**: Measurement on single cell level. Easy to quantify and use at high throughput. **Disadvantages**: BrdU/EdU can be cytotoxic; no information available on how often one given cell has divided. High cost and effort compared to counting.
	Staining of cellular components that are proportional to overall cell mass (proteins by e.g. sulforhodamine B or crystal violet; DNA by Hoechst H-33342)	These assays evaluate a surrogate measure of overall cell mass and assume that it correlates with total cell number. In non-proliferating cells, or with continuous ongoing proliferation, the endpoints are also frequently used as indicators of cytotoxicity, as dead cells often detach from plates and reduce the overall cell mass. **Advantages**: Simple and cheap; lots of historical data **Disadvantages:** Mostly not a single cell measure but only population level. Protein staining is only a surrogate endpoint of real cell number. For DNA quantification with Hoechst 33342: fluorescent probe penetration, bleaching, and cytotoxicity are issues to be considered. .
4. Cellular metabolism	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium (MTT) assay, or similar tetrazolium dye reduction assays from multiple suppliers	Biochemical activity (mostly mitochondrial metabolism; production of reducing equivalents like NAD(P)H) in viable cells causes reduction of the tetrazolium dye. The resultant formazan is extracted and measured spectrophotometrically. The rate of formation of formazan corresponds to the function of essential cellular processes like respiration. **Advantages:** High throughput, easy, robust, low cost. Used in several ISO standards and OECD test guidelines. High sensitivity. Can be used for tissue constructs. **Disadvantages:** Measures amount of viable cells (only indirect measure of cell death), and needs control for contribution of proliferation. Cells with reduced mitochondrial function may appear non-viable. Inhibition of cell metabolism by the test item causes low values in the assay which is not necessarily related to cell viability. Some test items interfere with the assay e.g. by reducing the dye why interference testing is recommended. Measurement usually not on single cell level. Some cell cultures need long time to reduce sufficient amount of dye (no sharp time point for viability definition). Assessment of kinetic of the reduction may be necessary to ensure proper selection of incubation time with a tetrazolium dye (to avoid reaching plateau of OD).

	Resazurin reduction assay (sometimes called Alamar blue)	Similar to tetrazolium reduction assays. Fluorescent/absorbent resorufin is formed from resazurin through mitochondrial metabolism of viable cells. **Advantages:** Many tests can be performed rapidly in multi-well dishes. Cells can be tested repeatedly (non-invasive measurement). High sensitivity. **Disadvantages:** Cells with reduced mitochondrial function may appear non-viable. Some test items interfere with the assay (e.g. superoxide also reduces the dye) why interference testing is recommended. Measurement only on population level. Some cell cultures need a long time to reduce sufficient resazurin (no sharp time point for viability definition).
	Mitochondrial depolarisation assays (based on fluorescent indicator dyes)	Many organelle functions are used as endpoints of general cell health. Most frequently, mitochondrial function is assessed (see MTT, resazurin). One mitochondrial test on the single cell level is the measurement of mitochondrial membrane potential by addition of potential sensing fluorescent dyes like JC-1, TMRE, MitoTracker, etc. Quantification is by HCI or FACS **Advantages**: fast, cheap, high throughput; single cell information. **Disadvantages**: as for MTT (measures cell function, not cytotoxicity). Artefacts by test items that affect mitochondria specifically. Artefacts by test items that affect plasma membrane potential. Artefacts due to bleaching, quenching and unquenching, and due to shape changes and clustering of mitochondria.
	Neutral red assay (ISO 10993)	A cell oranelle function assay assessing lysosomal function. Active cells accumulate the red dye in lysosomes and the dye incorporation is measured by spectrophotometric analysis. **Advantages:** Low cost. Used in several ISO standards and OECD test guidelines. Historic data base. **Disadvantages:** Normalisation required for quantitative measurement, e.g. with protein content or number of cells. Gives usually information only at the population level. Not suited for tissue constructs and certain cell lines.
	ATP assays	Measurement of the total ATP content in a cell population. Dying cells fail to produce ATP, have an increased ATP consumption, and may lose ATP through perforations of the plasma membrane. For the test, cell lysates are prepared, and the ATP content is assesses by a luminometric assay. **Advantages:** fast, high throughput **Disadvantages:** no single cell data, expensive, requires a luminometer, as MTT: measurement of viable cell mass, not a direct measure of cytotoxicity. Artefacts as for other mitochondrial tests.

6.2.1.2 Functional disturbances

While the strict definition of cytotoxicity refers to cell death, a wider interpretation also includes adverse effects on cells that alter their functionality but do not lead to cell death (within the observation period). For instance, protein synthesis may be impaired, or mitochondrial function altered. Cytostasis, where dividing cells do not die but cease dividing, is another example of delayed cell death which can impact the endpoint measures. This can affect the specific endpoint of a test system (e.g. luciferase reporter assay), without being relevant for the intended test objective. Test items with such properties can lead to erroneous readouts.

6.2.1.3 Disturbed differentiation state or gene/protein expression

A special case of artefacts caused by test items is the change of biological properties of the test without overt cytotoxicity. The most common example is an altered differentiation of cells or an altered composition of cell sub-populations. For instance, a test item might alter cell differentiation state in a migration assay, and this alteration might lead to altered migration. The item would be wrongly classified as modifying cell migration. Another example would be measurement of monocyte function (e.g. cytokine release) in a whole blood assay. If a test item leads to platelet degranulation, it might influence the overall endpoint of the *in vitro* method without affecting the monocyte response as such.

6.2.1.4 Altered communication/adhesion properties

Another special case of artefacts can be generated by interference of the test item with cell adhesion or communication. This is listed here separately, as it would not normally be detected by cytotoxicity assays, but it would strongly alter the behaviour of the test system (biological model) in the test situation. An example is binding of test item to molecules used for the coating of culture dishes. This would then alter readouts such as migration or neurite growth, without really affecting such processes within the cells (and without necessarily being relevant *in vivo*).

6.2.2 Interference with *in vitro* method endpoint

For pure test items, most of the unwanted interactions with the *in vitro* method are covered by considering undesired interactions with either the test system or the *in vitro* method endpoint. The majority of interferences with the *in vitro* method endpoint will be related to cytotoxicity (immediate or delayed cell death or functional impairment), as covered in 6.2.1.1. For test items containing impurities or non-inert additional substances in their formulation (see 6.2.4) the situation can become more complex: this highlights the need to have clear specification for the test item, as stated above. For instance, impurities (e.g. detergents or solvents) may alter skin or blood-brain-barrier (BBB) permeability (without being cytotoxic) and thus result in incorrect data on the pure test item of interest, if the *in vitro* test assesses e.g. skin or BBB permeation capacity. In other cases the test item is a finished product. Then potential impurities and contaminations are part of the product and their effect on the response is important and has to be evaluated.

6.2.3 Interference with the analytical endpoint

Interference of the test item with the *in vitro* method endpoint means that the test item disturbs the normal measurement results. This can be controlled for by performing the *in vitro* method using adequate positive, negative, blank or vehicle controls. If the endpoints are

of analytical nature, the controls can also be spiked with test item to verify that the test item do not in any way hinder the normal function of the test system or interfere with the readout.

Examples of such kind of interference include:

Fluorescence/absorbance-based methods: disturbance by test items that fluoresce or absorb at the evaluation wavelength, or test items that quench fluorescence.
Enzymatic assays: alteration of enzyme function, of co-factor, or of other limiting reagents by test item; display of enzymatic activity (or chemical reactivity) by test item itself.
Resazurin/ or MTT reduction: strongly reducing agents directly reduce resazurin/ (or MTT) non-enzymatically. Compounds that trigger the release of superoxide can trigger reduction of resazurin by superoxide. This results in erroneous cytotoxicity data.
Another relevant example of this kind of interference is provided by the interactions between test reagents and nanomaterials in colorimetric assays for cytotoxicity (such as sulforhodamine B dye, or MTT used in the viability assays) (Scenhir, 2015). Moreover, some nanomaterials may themselves disperse/absorb light and therefore interfere with the measurements in colorimetric assays. Some of these problems might be overcome by either adding appropriate controls or modifying existing protocols: as an example removal of nanomaterials via centrifugation before reading the assay can reduce the variations in data generated for the same nanomaterials (Scenhir, 2015).

6.2.4 Consideration of interferences not coming from the active ingredient

With test items that are not pure, the interferences with the test may come from impurities or from ingredients of the formulation. Particularly difficult cases arise when such additional chemicals are inactive alone, but synergize somehow with the effect of the test item.

This can also occur for the solvent of the test item. Frequently, a solvent concentration that does not affect the standard endpoint of a test as such (e.g. 0.1% DMSO) may still alter the effect of a test item on the test system (e.g. in the case of DMSO: through the antioxidant properties of DMSO; or through its effect on cell membranes; or through other activities including cell differentiation).

In case of a test item consisting of a natural mixture (e.g. essential oils), it should be considered to test the mixture as well as the known pure substances present, since the other ingredients of the mixture can change the overall effect of the test item. The different kinetics of the ingredients must then be considered; however, although ingredients that are not absorbed *in vivo* will not be able to have an effect on the test item systemic toxicity, they may affect test item toxicity *in vitro*.

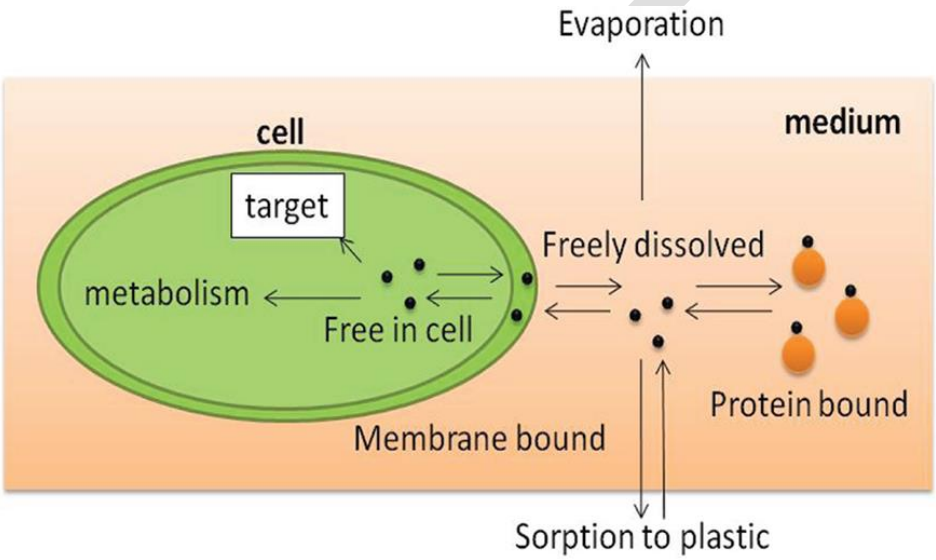
6.3 Biokinetics / dose extrapolation

Just like the biokinetics *in vivo* are about what the body of the organism does to the test item, the biokinetics *in vitro* concerns what the *in vitro* test environment does to the test item. A central issue in biokinetics is that generally only the freely dissolved molecules of a chemical can pass membrane barriers and reach a target inside a cell. Thus, in an *in vitro* system, the freely dissolved concentration of the test item in the medium or in the cell (as close to the target as possible) is the central parameter. Different processes result in a freely dissolved concentration that is not the same as the nominal concentration, (i.e. the added concentration). These processes are described in e.g. (Heringa *et al.*, 2006) and (Groothuis *et al.*, 2015), and were one of the main topic investigated by the FP7 EU Project Predict-IV, aimed to improve the predictivity of *in vitro* assays for unwanted effects of drugs after repeated dosing integrating biokinetics and biodynamic data. As one of the project outputs, a

step-wise strategy was applied to measure and model cell exposure levels over time of a selected number of drugs in the developed *in vitro* assays. The strategy and the major obtained results are described in (Kramer *et al.*, 2015).

Figure 3 schematically depicts the different processes, which can affect a xenobiotic *in vitro* bioavailability. Also the processes that change the identity of the test item, affecting its stability, are described briefly in the following.

Figure 4: Schematic representation of the processes that can cause the final target concentration to be different than the nominal concentration in an *in vitro* test (Kramer *et al.*, 2012)



6.3.1 Kinetic processes

6.3.1.1 Evaporation / plastic binding / precipitation

In vitro systems are often open, with a small gap between the well plate and the lid, to allow air circulation for provision of oxygen for the cells. This air circulation allows volatile substances to evaporate into the air of the incubator. This may decrease the concentration in the medium in the test system, but can also contaminate medium in e.g. blank wells, as the substance can dissolve from the air into the medium of other wells present in the well plate or even the incubator. An example showing the effect of evaporation on test results can be found in Tanneberger et al 2010 (Tanneberger *et al.*, 2010).

Lipophilic substances tend to bind to the plastic the cell culture plates are made of, although differences exist among the types of plastic used. The adsorption to polycarbonate is limited, but in organ-on-a-chip devices made of Polydimethylsiloxane (PDMS), there will be partitioning between the PDMS and the medium. PDMS is even used as an extraction material for-solid phase microextraction (SPME) (Heringa and Hermens, 2003), it is therefore not suitable for *in vitro* test devices for testing of chemical substances. Glass is a better material to avoid binding to some extent, but very lipophilic substances are known also to bind to glass. Silanized glass can decrease this binding even further. Using glass has other

practical downsides in *in vitro* tests. Examples where considerable binding to plastic was measured are the one of Kramer *et al.* (Kramer *et al.*, 2012). In this study, it is also shown how the addition of serum to medium decreases the binding to plastic. Other examples are reviewed in (Kramer *et al.*, 2015), reporting results of the Predict-IV project on cyclosporine A, amiodarone and chlorpromazine. The addition of serum to medium decreases the binding to plastic, but likely also the uptake into the cells (Pomponio *et al.*, 2015).

Sorption of the test item to cell-attachment matrices (e.g. collagen or matrigel layer used with hepatocytes in culture) is a specific aspect of interaction with the test device, although the relationship between a test item's lipophilicity and binding to is not as clear cut as it is for binding to plastic labware. The possible physical sequestration of test items, can lead to overestimating intracellular concentrations (Kramer *et al.*, 2015).

6.3.1.2 Chemical degradation

The aqueous environment of the medium in an *in vitro* test enables spontaneous hydrolysis (i.e. without the aid of an enzyme) of substances with structures sensitive to this chemical reaction. During the time the test system, e.g. the well plate, is outside of the incubator, light will reach the medium and photolysis can take place for light sensitive substances. Therefore, information on hydrolysis and photolysis sensitivity is necessary before a substance is tested in an *in vitro* method (see 6.1.1.1). More generally, each test facility should have adequate test item characterisation procedures in place to identify if the test item characteristics are compatible with the *in vitro* method.

6.3.1.3 Metabolism/metabolic stability

Some cell types have metabolic capacity, meaning that they contain significant levels of enzymes that convert the test substance to another substance. Especially cells originating from liver, intestine and lung are known to possess metabolic capacity, in decreasing order. In test systems with such cells, especially from these tissues, the concentration of the test item may decrease because of this metabolism, and the concentration of metabolites will increase. When a positive hazard response is obtained in such a cell system, it may thus either be caused by the test item itself, or its metabolite(s). The time profile of the response can reveal which is the main causative agent: when there is a lag time for the response (compared to the positive control or other reference items), it could be that a metabolite is responsible for the response. A good example is described in (Pomponio *et al.*, 2015)

6.3.1.4 Protein binding

Serum is often added to cell culture medium to supplement it with important factors required for cell proliferation and maintenance. Serum-free medium is available and used, but not all cell types thrive in such culture conditions. Serum contains proteins, including albumin, which has non-specific binding sites, to which most organic substances tend to bind. As proteins are large molecules that do not transfer across a membrane, the binding to a protein renders a test item unavailable for cellular uptake, thus unable to reach any target inside the cell. An example of the effect of serum protein binding can be found in Heringa *et al.*, (Heringa *et al.*, 2004) and in Pomponio *et al.*, (Pomponio *et al.*, 2015). On the other hand, serum proteins can also make some test items more accessible or more stable. E.g., for medical devices a medium with serum is preferred for extraction because of its ability to support cellular growth as well as to extract both polar and non-polar substances. In addition, protein binding also occurs *in vivo*. Gülden *et al.* 2005 therefore developed an *in vitro-in vivo* extrapolation method extrapolate nominal effective *in vitro* concentrations equivalent *in vivo*

plasma concentrations by accounting for the differences in protein concentrations (Gülden and Seibert, 2003).

6.3.1.5 Cell membrane absorption

Cell membranes are composed of fatty acids, thus provide a lipid environment in which lipophilic substances will like to absorb. These absorbed molecules are then also not available for a target inside the cell. An example showing the effect of membrane sorption can be found in Gülden *et al.* (Gülden *et al.*, 2001) and in Bellwon *et al.*, (Bellwon *et al.*, 2015).

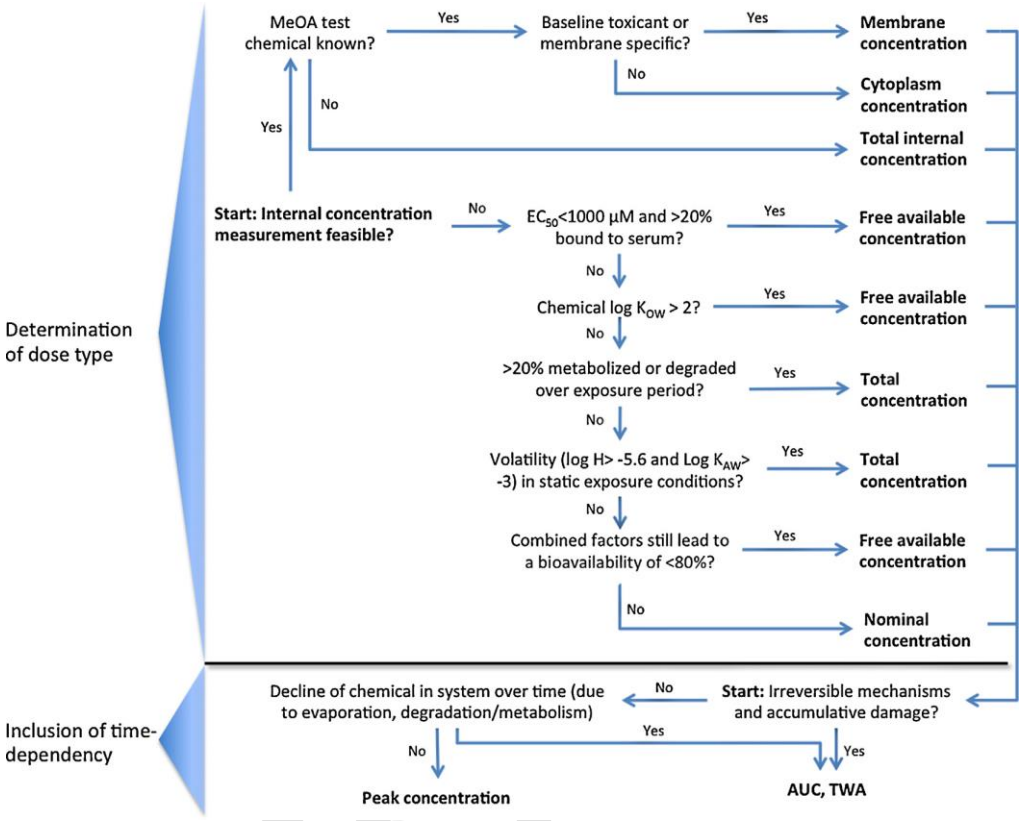
6.3.2 Measurement of free concentration /passive dosing

Clearly, several processes can influence how much of the added test substance will actually reach the target and this relates to its saturation concentration. If test results are based on the added, or nominal, concentrations, considerable variation between laboratories may be obtained. Furthermore, *in vitro* processes are included in these results (e.g. an EC_{50}), rendering these unfit for extrapolation to *in vivo* (see IVIVE below and (Kramer *et al.*, 2015)). For example, if there is considerable evaporation, the EC_{50} *in vitro* will appear to be much higher than it will be in the same tissue *in vivo*, as there is no evaporation *in vivo*. Thus, in order to obtain pure EC_{50} values, that relate target concentrations to responses, these target concentrations should be measured. As the precise concentration at the target site inside the cell is too difficult to measure, the best approximation should be measured: the free concentration in the cell or in the medium. The free concentration in the cell is often still difficult to measure, therefore the free concentration in the medium (similar to the cellular concentration for membrane permeable substances), or the total concentration in the cells (often for metals) are usually measured. Further information can then be added by calculations that take physicochemical and biochemical properties (e.g. transporter substrates) of the substances into account.

Methods with which the free concentration can be measured have been reviewed in Heringa *et al.* 2003 (Heringa and Hermens, 2003). This review also describes how negligible depletion-solid phase extraction (negligible depletion SPME) should be applied to measure free concentrations. This method is very suitable for *in vitro* tests, as it is suitable for small volumes. Examples of its application in *in vitro* tests are (Heringa *et al.*, 2004), (Broeders *et al.*, 2011), and (Kramer *et al.*, 2012).

Measuring the free concentrations does require extra effort and resources in the conduct of the *in vitro* test, as e.g. a chemical analysis method is necessary. This effort can be saved in some instances, depending on the properties of the test item: in case of very hydrophilic, non-volatile substances that hardly bind to serum proteins, there will hardly be any losses and the nominal concentration will be very similar to the free concentration. Groothuis *et al.* (Groothuis *et al.*, 2015) provide a decision scheme on which concentration should/can be used as dose metric.

Figure 5: Flow chart to aid in choosing an appropriate dose metric for a specific *in vitro* toxicity test (Groothuis *et al.*, 2015)



First, a choice should be made for dose type based on the characteristics of the chemical and available knowledge. Then, the metric can be integrated or averaged in case of time-dependent exposure and irreversible mechanisms, or steady reduction over time. Peak concentration is defined here as the maximum concentration reached during the exposure period. BK/TD may be applied to model partitioning and assess concentration changes over time. The chart has been compiled by (Groothuis *et al.*, 2015) using literature data (Austin *et al.*, 2002; Gülden *et al.*, 2010; Gülden and Seibert, 2003; Knöbel *et al.*, 2012; OECD, 2011, 2006a, 2006b; Reinert *et al.*, 2002; Riedl and Altenburger, 2007).

To avoid the effort of measuring free concentrations in every sample, passive dosing can be applied. In this method, a disk or ring of absorbent material, which is loaded with the test substance, is added to the sample. After a time of equilibration, the free concentration will have become proportionate to the concentration in the disk or ring, governed by the partition coefficient between water and the disk or ring material. If this partition coefficient has been predetermined, and if the amount of substance in the ring or disk by far exceeds the amount to be dissolved in the medium, then the free concentration in each sample can be easily calculated, and does not need to be measured. A more detailed description of the method can be found in Smith *et al.* (Smith *et al.*, 2010). This reference, as well as Smith *et al.* (Smith *et al.*, 2013), give examples of how passive dosing can be applied to *in vitro* tests.

In vitro to in vivo extrapolation (IVIVE) refers to the qualitative or quantitative transposition of experimental results or observations made *in vitro* to predict phenomena *in vivo*, on full living organisms. When the response of the *in vitro* test is plotted against the free

concentration (or the nominal concentration only in case it can be demonstrated/estimated this approximates the free concentration), toxicity parameters such as the EC_{50} or a benchmark concentration (BMC) can be derived from the obtained curve. This *in vitro* toxicity parameter can be used as point of departure (PoD) for *in vitro* test circumstances and directly applicable to *in vivo* extrapolations (Leist, 2014, Blaauboer 2012). The corresponding *in vitro* concentrations can be converted into relevant plasma concentrations by taking the protein and lipid concentrations in plasma and cell culture medium into account (Bosgra and Westerhout, 2015; Zimmer *et al.*, 2014). In a final step, this concentration can be used as input for physiologically based pharmacokinetic (PBPK) models to estimate the dose that would result in the respective plasma concentration in man. This way an external benchmark dose (BMD) can be obtained. PBPK models describe the kinetic processes *in vivo*, relating external doses to tissue concentrations in time. For these models, some physical-chemical properties of the test substance need to be known, as well as some kinetic parameters such as the fraction absorbed, rate of metabolism, tissue partition coefficients, protein binding coefficients and urinary excretion rate. An example of how this can be performed is described in Louisse *et al.* (Louisse *et al.*, 2010). Good modelling practices for PBPK models have been described by Loizou *et al.* (Loizou *et al.*, 2008). The recommendations from a joint EPAA - EURL ECVAM on how PBTK modelling platforms and parameter estimation tools could enable animal-free risk assessment are reported in Bessems *et al.*, (Bessems *et al.*, 2014).

6.4 Reference and control items

The inclusion of relevant reference and control items, and the setting of acceptance criteria on the basis of historical data, is essential for regulatory applicability of *in vitro* methods. By including the correct reference and control items, the data set obtained from the *in vitro* method will demonstrate the correct functioning of the test system and the method used for analysis and therefore the validity of the executed experiments.

Reference items can be one or more items where a specific readout and well-known response is expected (OECD, 2004a). The reference item(s) is used to provide a basis for comparison with the test item or to validate the response of the test system to the test item i.e., provide a known measurable or observable response. Notable requirements for the reference chemicals include having a well-defined chemical structure and purity and availability from commercial sources without prohibitive costs, hazards or disposal considerations. Reference items should be specific to the endpoint being measured. Documentation should be available to justify the selection of the reference items. While these reference chemicals are meant to represent the minimum to evaluate the performance of *in vitro* methods they can be substituted with other chemicals from the same class or substances possessing similar activity.

Reference item(s) should be tested for batch-to-batch variability and be appropriately characterised (e.g., purity, stability) and identified (e.g., CAS number) (GLP Principles, No.1). Solubility, stability, and purity need to be established for each reference item used, and acceptance criteria based on historical data need to be developed. The continuous monitoring of the reference items, e.g. in the format of a control chart, is important to prove that the *in vitro* method continues to perform within the limits, and is consistent over time.

Control items are used to control the proper performance of the test system (OECD, 2004a). Monitoring and recording performance against negative and positive control items may constitute sufficient proof for the responsiveness of a given test system. Non-response of the test system to the negative control and response to the positive control, within the acceptance criteria, show that the test system is "reactive" and behaves as expected. For positive,

negative, and vehicle control items (in those cases where the negative control is different from the solvent control), it may or may not be necessary to determine concentration and homogeneity, since it may be sufficient to provide evidence for the correct, expected response of the test system to them. Such evaluation may consist of documented evidence that the response of the respective test systems to these positive, negative, and/or vehicle control items does not deviate from the historical control values recorded in the test facility, which should furthermore be comparable to published reference values. Guidance on how to compile and use historical data can be found in literature. Hayashi (Hayashi *et al.*, 2011) describes the compilation and use of historical data specifically for genotoxicity data, but this approach can also be applied in a broader context. A more general approach is described by Yoshimura (Yoshimura and Matsumoto, 1994).

The positive control item is one that has a consistent and predictive effect in the *in vitro* test system. It is often used to assess aspects such as the characteristics of the test system and whether the *in vitro* methods gives reproducible results for this positive control over time in accordance with historical data. Moreover, a positive control-induced endpoint value that meets the predefined acceptance criteria, assures that when a test item has no effect on the *in vitro* method, the negative result is not caused by an error during the testing phase. Therefore, the results from the control test items are of utmost importance to show that a valid run has been performed when test data for each unknown test item are submitted to regulatory authorities.

The response to reference items may be used to characterise the test system. Therefore, particular attention should be given to the definition of appropriate acceptance criteria for the response to each selected reference item. Acceptance criteria should be available when the method can be considered as fully developed and subsequently enter the validation process As such, acceptance criteria, should be available prior to the start of the *in vitro* study when used for regulatory purposes.

Acceptance criteria for reference items, positive, negative and vehicle control items (e.g., purity, known response, cytotoxicity) should be developed and detailed in the SOP.

Records of receipt, storage, preparation and use should be available to allow for a full reconstruction of the history and use of each reference item.

6.5 Use of proficiency chemicals

While reference chemicals are used to validate the *in vitro* method, proficiency chemicals are used to test the suitability of a laboratory to carry out the *in vitro* method and produce acceptable results.

Transfer of a new method to another laboratory is part of the development and validation process before a new *in-vitro* test can be accepted by regulators. The developer of the *in vitro* method should provide a list of proficiency chemicals. Proficiency chemicals should be fully characterised in the new *in vitro* method. By comparing the results of testing the selected proficiency chemicals by both laboratories, the inter-laboratory transfer of the new *in vitro* method can be evaluated

Proficiency chemicals selected for validation of the assay should cover a range of responses within the *in vitro* method's dynamic range. In addition, a good proficiency chemical set should include representatives of several chemical classes within the applicability domain of the *in vitro* method [chemical classes for which the test is intended]. The number of proficiency chemicals should be chosen in such a way that a new laboratory can be confident that their results will be acceptable and robust. Since this greatly depends on the properties of

2878 the method, some methods may require 5 proficiency chemicals, while for others up to 20
2879 compounds should be tested. On average, a list of proficiency chemicals usually contains
2880 around 10 materials when the method is transferred to an OECD test guideline. In this way,
2881 new laboratories undertaking the *in vitro* method can demonstrate their proficiency.

2882

7 Standard Operating Procedures (SOPs)

According to the Principles of GLP, documented procedures which describe how to perform tests or activities normally not specified in detail in study plans or test guidelines are defined as standard operating procedures (SOPs). SOPs are a set of written documents describing routine or repetitive activities (e.g. *in vitro* methods and complementary procedures) that facilitate consistency in the quality and integrity of a product or end-result. SOPs are required by GLP. Outside the GLP environment, terms such as testing methods, instructions, worksheets, and laboratory operating procedures are often used. In principle SOPs are unique to an organisation, as they are part of an integrated quality assurance process of a facility.

The aim of a SOP is to ensure that procedures are carried out in a reproducible way by qualified personnel. Therefore SOPs need to be written with sufficiently detailed but clear work instructions and be kept simple, to minimise the risk for misinterpretation.

An *in vitro* method will be supported and documented by a number of different SOPs, forms, templates and worksheets. Besides the description of the main test procedure, also SOPs for supporting procedures (e.g. the handling of cell cultures, waste handling, cleaning procedures, operating and calibration instructions for the equipment, record keeping, reporting, archival, quality assurance procedures, etc.) need to be available and referred to. To avoid lengthy documents, the instructions are preferably divided into a series of short SOPs. SOPs should be readily available to personnel in each working area.

7.1 *In vitro* method SOP(s) development

The development of an *in vitro* method for regulatory testing purposes by using complex animal or human cells and tissues as test systems is a difficult and time-consuming task. In the initial stages of the development, the procedure will undergo many changes and each step needs to be described in laboratory records, which will crystallise into a laboratory procedure or a SOP along the test development process. During this period, also the historical data set of the model compounds will be collected. This dataset will be needed to define the critical and relevant end-parameters, the control and reference items and associated acceptance criteria.

Once the method is sufficiently developed and all parameters are defined, the *in vitro* method SOP(s) is/are sufficiently "routine" and standardised to be used for an in house validation process during which the *in vitro* method is checked for repeatability (accuracy & precision), selectivity, sensitivity, and stability assessment over time. Likewise, its robustness is assessed (i.e. the influence of critical (external) parameters on the outcome parameters), as it is important to secure the test performances in different laboratory environments, albeit within defined boundaries.

Figure 6: Evolution of a Standard Operating Procedure

SOP Evolution
No routine => No SOP => no reproducibility
Method becomes routine => Draft SOP - Describe the *in vitro* method procedure - Historical data of reference items are generated in a controlled way
SOP Version 01 - Advanced *in vitro* method description - Acceptance criteria for valid/invalid experiments - Lists of needed equipment, reagents, consumables and reference items - Calculation of results
SOP Version xx - Further optimised procedure - Acceptance criteria for valid/invalid experiments - Calculation of results for test items - Data recording Forms, Data Calculation forms - SOP is robust

During the in house validation process, weaknesses can come to light that demand for adaptation (optimisation) and which might also trigger the re-initiation of a new validation cycle. Optimisation of the SOP should be performed by following a formal procedure. It is critical that any parameter(s) to be changed should be chosen prior to the optimisation process, including the steps to be performed: it is recommended to define these steps in amendment in the validation study plan, so that every passage is recorded in a proper way. Also, the historical data should be annotated to allow tracking, comparison and measurement of the acquired optimisation.

Upon a satisfactory completion of the validation process, the method development can be finalised and the final set of SOPs associated with the *in vitro* method will be available.

In addition, the test developer should be aware that if the established *in vitro* method makes use of complex instrumentation and software, including data analysis and computer models and if developed in-house (e.g. excel data analysis templates), this software will also require documentation and full validation with SOPs for correct use, prior to transferring the *in vitro* method to the validation bodies.

Once an *in vitro* method has been validated and published, e.g. in the format of an OECD Test Guideline (TG), the end-users will, from the published method, need to develop their own set of SOPs which are applicable and integrated into their organisation to assure the correct execution of the *in vitro* method within their facility environment.

7.2 Preparing Standard Operating Procedures

As indicated above, the guidance how to correctly perform the *in vitro* method is given in a set of SOPs, covering how to execute the *in vitro* test but also SOPs referring to general supporting procedures e.g. test system handling, solubility assessment, cytotoxicity

measurement, equipment maintenance, calibration and cleaning; handling of test and reference items; record keeping, reporting, storage, and retrieval, etc. The reason for not having all these steps and processes described in one single SOP, but a set of methods/procedures referring to other specific SOPs is to have available and easy-to-handle documents, to be consulted by the personnel involved. However, to avoid deviating procedures over time, it is advisable to include cross-references between these SOPs rather than duplicating information as it might otherwise not be updated in all documents where it appears.

SOPs should be written in the active voice and concisely explain a procedure in a step-by-step, easy-to-read format. The information presented should be unambiguous and not overly complicated. The document should not be wordy, redundant, or overly lengthy but simple and short. The inclusion of a flow chart to illustrate the process can help to make it clear.

A SOP is best written by the people actually performing the work on a daily basis. The finalised SOP needs to be reviewed and approved by laboratory management. SOPs are not static documents and need to be systematically reviewed on a periodic basis and adapted whenever something changes (products/ equipment / facility…), in order to be possible to fully adhere to. As soon as a new version is approved, all concerned personnel need to be informed before the new version comes into force. From that date onwards the obsolete version needs to be made unavailable and adequately archived. To allow and control this, all SOPs need to have a unique identifier (Title / version number / approval date). It is also recommended to keep the revision history of the document in the document. SOPs should be formally authorised by test facility management.

SOPs may be written according to a pre-set template and may contain (adapted from "EPA (USA) "Guidance for Preparing Standard Operating Procedures (SOPs) – EPA QA/G-6 /2007"): (EPA, 2007) ;

- Title page or page header (including: a unique name which describes the essence of the SOP; the name of the author or authors and the signatures of reviewers and approvers, an identification code -if applicable-, the revision number, the date of approval and effective date, and number of total pages in case it is not included in the subsequent pages).
- Change log (list summarising the changes from each previous edition of the SOP).
- Table of contents in particular necessary for longer SOPs. In some GLP test facilities short SOPs (e.g. 5 pages) do not always have a table of contents.
- Purpose/objective – (identifying the intended use of the procedure).
- Field of Application/Scope (identifying when the procedure is to be followed, and explaining who the intended user of the document is and what the method does and does not cover).
- Summary of main steps (can be illustrated by a flow chart, if desired).
- Definitions (defining any words, phrases, or acronyms used in the SOP).
- Personnel qualifications/responsibilities (identifying any special qualifications users should have such as certification or training experience and/or any individual or positions having responsibility for the activity being described)..
- Reference to all related SOPs and other documents (forms, templates, worksheets etc.) to be used together with the SOP (not necessarily as a separate section)
- Health & safety warnings (indicating the possible risks to address and the personal protection equipment and containment equipment to use).
- Procedure (including: a list of the equipment, reagents, consumables and reference materials, with their quality requirements; pre-analytical preparation: sample collection,

handling and preservation; data processing; evaluation of results against acceptance criteria; post-analytical activities: reporting of results; sample and chemicals discarding).

- Criteria, checklists, or other standards that are to be applied during the procedure.
- Records management (specifically, e.g., as forms to be used and locations of files).
- *In vitro* method acceptance criteria section - describe any control steps and provisions for review or oversight prior to acceptance of the results.
- Reference Section - cite all references that have been consulted during the authoring of the SOP.
- In addition to the first page, all the subsequent pages of a SOP should include the title, the identification code (if applicable), the revision number, the page number (and the total number of pages).

8 Performance of the method

To assess *in vitro* method performance one needs to define the selection of reference and control items as these are used to check the performance of the method. Of critical importance is also how acceptance criteria (often based on historical data) are developed and defined for the method regarding its performance.

The number of replicates for each condition (e.g. chemical concentration) should be specified. During method development the number of replicates must be chosen using appropriate statistical methods. For example, a statistical power analysis (Crawley, 2015) can be used to calculate the desirable number of replicates to detect a defined difference between treatments with pre-set levels of confidence (Krzywinski and Altman, 2013). However, one should be aware that this number can be too high to be useful in practice. Alternatively the statistical power is provided for the chosen number of replicates.

Additionally, when multiple concentrations of a test item are tested, the mathematical curve fitting model (e.g. dose-response curve) can be computed with increasing number of replicates. The lowest number of replicates that gives satisfactory variability of the parameter of interest (e.g. IC_{50} within acceptable limits) can be used in future studies (Assay Guidance Manual, HTS Assay Validation, Section 5.2 assay guidance (Iversen *et al.*, 2004)). Apart from these statistical considerations, sometimes practicalities such as cost and availability of replicates may also play a role in the selection process. However the impact of reducing replicates should always be subjected to careful analysis and corresponding power should be given.

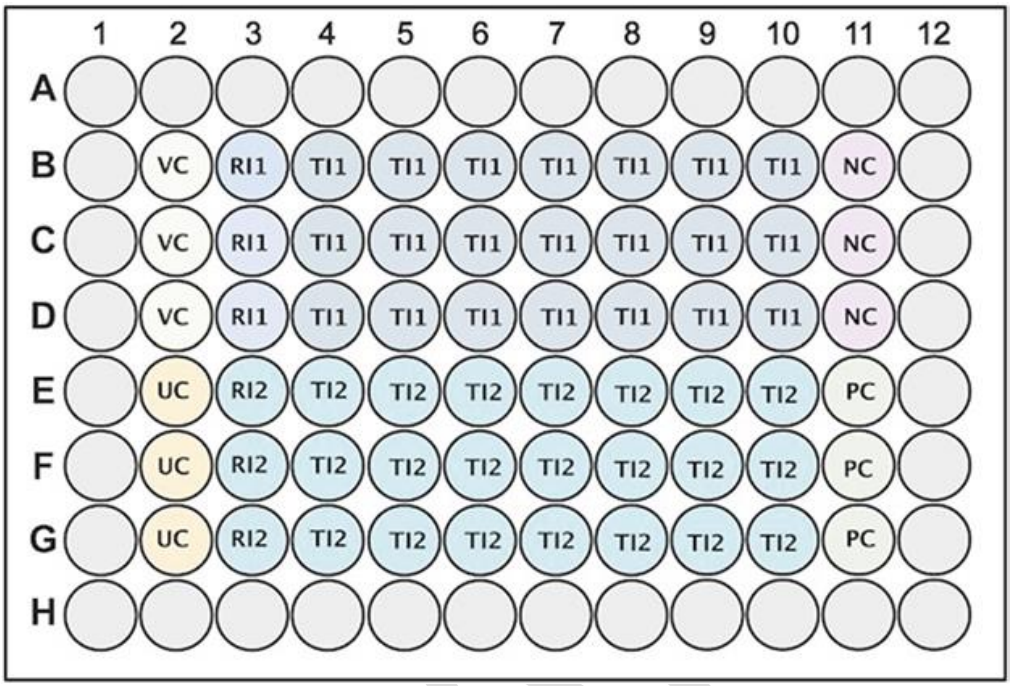
Similarly, the number of independent experiments needs to be evaluated. For instance, *in vitro* methods with a high degree of inter-experimental variability, such as those using primary tissues, may need a higher number of independent experiments compared to *in vitro* methods employing continuous cell lines.

Statistical methods can be very useful in the process of optimising new *in vitro* methods. Newly developed *in vitro* methods can include steps that may be interpreted differently and result in significantly different readouts for the different interpretations. To obtain an *in vitro* method that leads to accurate, reliable and robust readouts, the results of several combinations of any changes in the *in vitro* method would have to be assessed. Given the big number of variables that may or may not be changed, it is nearly impossible to approach all these combinations experimentally. In such cases, factorial design of experiments may be of added value. These are efficient at evaluating the effects and possible interactions of several factors (independent variables). A statistical approach predicting the effect of changes in testing method steps on the observed readout (known also as method robustness assessment) would allow for the development of an efficient *in vitro* method design, since the experimental robustness check can be based on a much smaller subset of combinations (Box *et al.*, 2005; Groten *et al.*, 1997).

8.1 Plate layout

The plate layout should be such that cross-contamination (e.g. between test items) can be controlled for (replicates). It should also allow for cross-plate comparison by using appropriate reference and control items. An example of an experimental 96-well plate layout using reference and control items is shown in Figure 1~~Figure 1~~ (Coecke *et al.*, 2014).

Figure 7: Example of plate layout. PC positive control, RI reference item, NC negative control, UC untreated control, VC vehicle control, TI test item



The example plate layout (Figure 6) minimises potential edge effects (difference between outer and inner wells due to evaporation). For strategies to assess edge effects and drift see Section **Error! Reference source not found.**~~8.7~~. Another way to assess plate drift is to include vehicle controls (VC) on both the left and right side of the plate. Left and right VCs should not differ more than a certain percentage for the plate to be accepted. For example, in the test in appendices, the following criteria is used: a test meets acceptance criteria if the left and the right mean of the VCs do not differ by more than 15% from the mean of all VCs (National Institutes of Health, 2001).

Comparison across plates, evaluation that different plates provide comparable data, should also be taken into account for the plate design.

The inclusion of relevant reference and control items, and setting of acceptance criteria on the basis of historical data, is essential for regulatory applicability of *in vitro* methods and should be considered when developers decide on their plate lay-out. By including the correct reference and control items, the data set obtained from the *in vitro* method will demonstrate the correct functioning of the test system and the method used for analysis and therefore the validity of the experiments executed.

In addition, certain reference chemicals may be volatile (e.g. solvents) or may contaminate neighbouring wells by capillary action, the wicking effect (Sullivan, 2001) and this may need to be taken into account in designing plate layouts. For instance, the commonly used cell lysis surfactant Triton X can affect cell viability in neighbouring wells and should be used at low concentrations or separated from cell-containing wells by placing wells containing media or buffer in-between.

8.2 Data analysis

When data need to be transformed by formulas for normalisation, computer scripts and/or any model equations fit to the data, this should be documented in the SOP (OECD, 2014). Formulas for normalisation (checked for accuracy) should be documented, validated (when implemented in electronic format) and disclosed along with a description and justification of the controls used in the calculation. Computer scripts used to process raw data (e.g. Excel spreadsheets, scripts, macros etc.) should be validated and fully documented.

When a model equation is fitted to the data (as in the case of dose-response curves, standard curves or other models), the equations and reasoning behind their choice need to be stated. For example, when fitting a dose-response curve, the type of the equation fitted to the data should be stated (e.g., four parameter logistic curve), along with any constraints (e.g. top constrained to 100% in normalised data) and weightings (e.g. by inverse data uncertainty) applied (Motulsky and Christopoulos, 2004). Furthermore, the software name and version used to fit the equations should be listed/stated, as well as the confidence interval of the measurements of interest (e.g. IC50 values) and the relevant goodness of fit parameters (R-square, sum of squares etc.) stated. Also see section **Error! Reference source not found.**~~8.8~~ of the current document and section 4.3 of OECD 211 (OECD, 2014) (OECD, 2014) as the same data will need to be reported in the late stages of method development during the assessment of method performance.

8.3 Data-intensive *in vitro* methods

The last decennia brought a paradigm shift in toxicity testing of chemical compounds, relying more on less expensive and higher throughput high-content screening *in vitro* methods. They allow the processing of hundreds or thousands of compounds simultaneously enabling the identification of mechanisms of action, and ultimately facilitating the development of predictive models for adverse health effects in humans. Furthermore, image analysis and genomics-based *in vitro* method read-outs are getting more popular for *in vitro* method developers due to the data rich information obtained with such methods.

The utility of "big data" for regulatory safety assessment has been discussed recently, for example omics (ECETOC 2013) or High Throughput screening (Judson 2013). These data could be used in various contexts, such as supporting evidence for read-across, defining categories or to allow the design of Integrated Testing Strategies (ITS). Still, most applications have focused on screening and prioritisation as in the EPA ToxCast program (Judson 2010).

Although some technologies have been extensively used for more than a decade, e.g. microarrays, debate is still ongoing about the reproducibility of experiments and the comparability of results at different sites and on different platforms. Consensus is still to be achieved concerning best practices in many critical aspects such as the experimental design and protocols for sample preparation and handling, data processing, statistical analysis and interpretation. One of the important challenges to be addressed for regulatory acceptance is related to quality assurance (Bouhifd *et al.*, 2015). The maintenance of high standards is essential for ensuring the reproducibility, reliability, acceptance, and proper application of the results generated. A certain level of standardisation is also needed since "big data" are generated using diverse technological platforms and various biochemical, analytical and computational methods, producing different data types and formats.

8.4 Acceptance criteria

Because the *in vitro* tests for the prediction of human health effects are intended to predict a quantitative endpoint response predictive of the degree of human hazard, it is essential that the test performs consistently over time and across laboratories. The positive control response plays a major role in determining that consistency (**Error! Reference source not found.**~~Figure 10~~). To do that, the positive control needs to induce a known change in the endpoint measure(s) that falls at a point within the dynamic range of the test. That way, increased and decreased magnitudes of response can be measured.

In many cases, the acceptable range for the positive control response is based on a statistical analysis of the historical range for the assay (firstly produced at the developer's laboratory, then supplemented by data from validation studies, and finally produced in the laboratory performing the *in vitro* method). For data, which follows an approximately normal distribution, the mean and standard deviation are calculated and the acceptance range is set at for instance plus or minus two standard deviations from the mean response. For example, the Bovine Corneal Opacity and Permeability (BCOP) *in vitro* method using 100% ethanol as the positive control (selected because it induces both increased opacity and fluorescein passage), the mean published *in vitro* score (opacity + 15×permeability) was 51.6 ± 6.2 (mean ± standard deviation SD), which would set the acceptable range (± 2SD) to be 39.2 to 64.0 (n=1171 trials) (Harbell *et al.*, 2014).

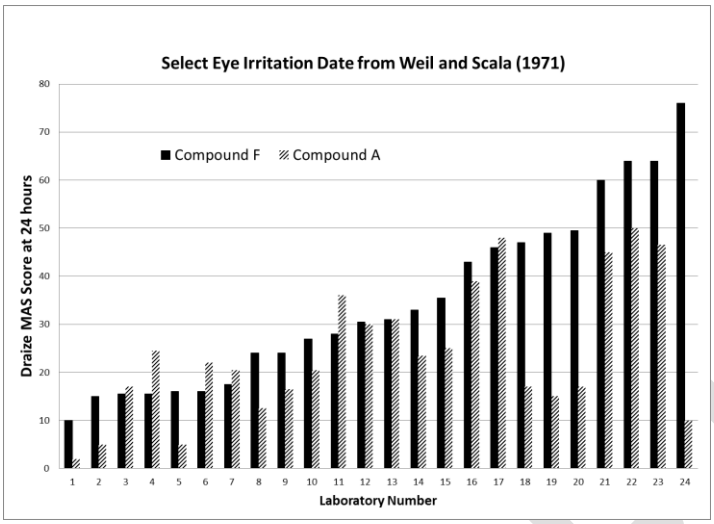
Development of the positive control should begin as early in the *in vitro* method development process as is practical (Hartung *et al.*, 2004). Positive control performance can help identify dependent variables that impact assay consistency. It is essential that the positive control should be incorporated into the assay every time it is performed. The positive control is particularly important during the *in vitro* method development stage where data are being generated to define the prediction model for translating the *in vitro* endpoint data into a prediction of human (animal) health effects. A robust positive control historical record will facilitate technology transfer to another test facility as part of the *in vitro* method transfer process.

The challenge of evaluating *in vitro* method performance over time and across laboratories is not unique but is often not addressed with *in vivo* test system. *In vitro* test systems have a large number of dependent variables. The operator is wholly responsible for maintaining test system homeostasis and consistency.

The importance of positive controls is exemplified by a Draize eye irritation study "Study of intra- and inter-laboratory variability in the results of rabbit eye and skin irritation tests" in (Weil and Scala, 1971). Figure 8~~Figure 8~~ shows the 24-hour Draize Maximum Average Scores (MAS) for two of the test materials: Compound F (46% aqueous triethanolamine lauryl sulfate) and Compound A (95% ethanol). The data are arrayed according to the MAS reported for Compound F and A. Without a positive control, there is no means to compare data from one laboratory to another or one laboratory over time.

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Figure 8: Sample data from the Draize Eye Irritation test on two chemicals at 24 hours after instillation



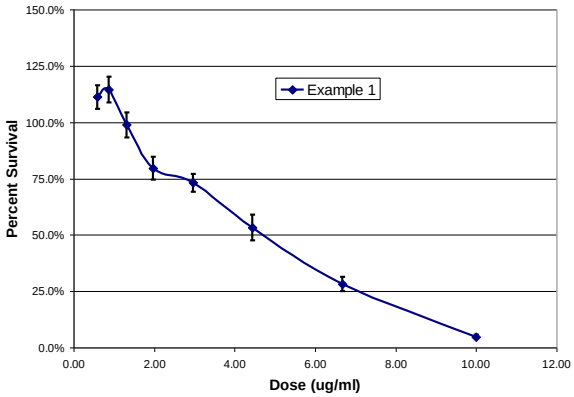
The data are arrayed in order of increasing maximum average score (MAS) for compound F. The corresponding MAS for compound A is paired with the MAS for compound F from that laboratory.

Figure 9~~Figure 10~~ shows three concentration response curves from a keratinocyte-based cytotoxicity assay (neutral red uptake endpoint) treated with sodium lauryl sulfate. The difference between these curves shows the information that can be gained from a concurrent positive control. They also illustrate the importance of having multiple concentration points in the active range so as to narrowly define the 50% relative survival point. The 50% survival point (concentration) for the positive control compound is often the basis for establishing the acceptance criteria for a dilution-based cytotoxicity assay. This approach allows increased and decreased sensitivity to be readily identified.

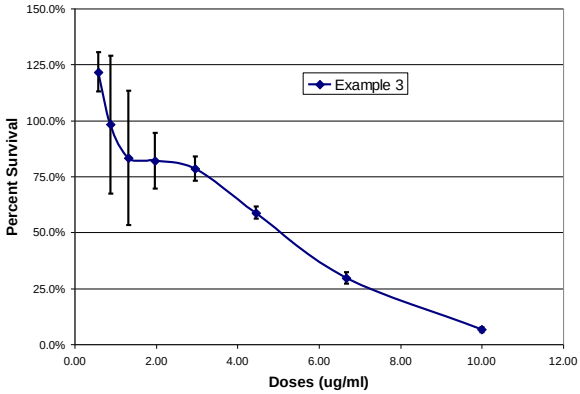
Formatted: Font: Times New Roman, 12 pt, Font color: Text 1

Figure 9: A 96-well based cytotoxicity assay using human keratinocytes

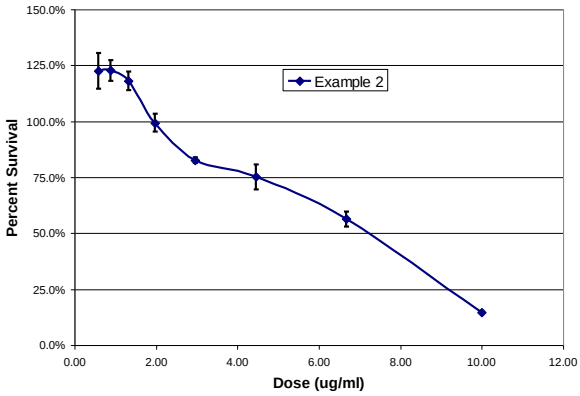
A. An example of normal positive control concentration response curve for this assay.



B. This is an example of poor cell seeding leading to inconsistent neutral red uptake in replicate wells.



C. An increase in cell number per well has induced a right shift of the SDS cytotoxicity curve.



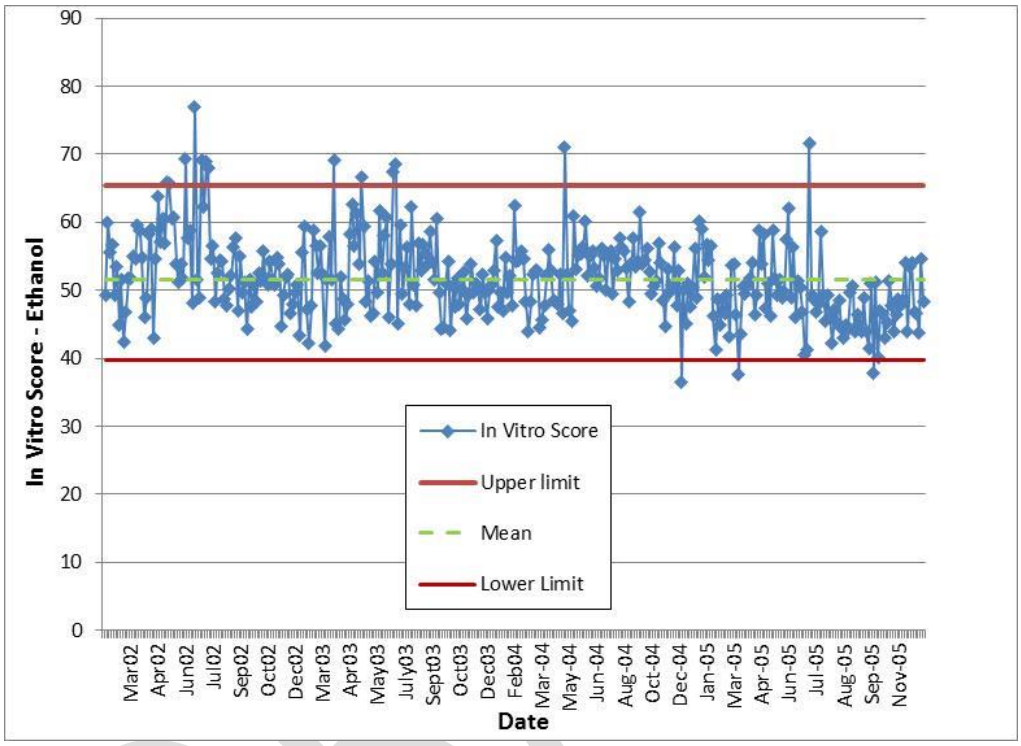
In selecting a positive control, it is desirable but not always possible to select a single positive control to address all endpoints or exposure conditions. In the genetic toxicity *in vitro* methods, such as the bacterial reverse mutation (Ames) assay, two positive controls are used for each bacterial strain to address direct mutagenic activity and metabolic activation of a promutagen with rat liver S9 (Zeiger *et al.*, 1988). In the BCOP, one positive control is used for the liquids exposure testing method and another for the solids exposure testing method (OECD, 2009).

Finally, the importance of testing the positive control concurrently with the unknown test materials each time the *in vitro* method is run is illustrated in Figure 10~~Figure 10~~ (Ulrey et al., 2005). This Figure shows the quality control chart for the BCOP using the ethanol control for each test performed over a period of 2.5 years. The acceptable upper limit is shown with a light line between 60 and 70. There is a cluster of points that extend above the upper limit (in mid-2002). These were failed assays which were repeated. The basis of the failures was not immediately clear since the eyes looked quite normal upon arrival in the laboratory. However, the patterns became clear and the cause was traced to improper handling of the eyes in the abattoir. Without the concurrent positive control data, it would not have been possible to identify the problem and prevent inappropriate data from being reported. Isolated tissues or tissue constructs as test systems can be difficult to properly evaluate visually and so

Formatted: Font: Times New Roman, 12 pt, Font color: Text 1, Do not check spelling or grammar

the functional test provided by the concurrent controls is often the only way to measure their integrity.

Figure 10: A quality control chart for the ethanol positive control for the BCOP showing the acceptable range and deviations above and below (March 2002 to November 2005- courtesy of IIVS)



In the same vein, establishing an expected range for the negative control is important to assure that the test system performs normally and to determine the detection limit of the assay.

Finally, it is also important to establish the cut-off value of the acceptance criteria for quantitative endpoints, i.e. whether it will be accepted not less to or higher than a specified value, including the number of significant figures. Having a result of 50.4, and an acceptance criteria of >50 is not acceptable as the number of significant figures are not comparable (using an acceptance criteria of ≥ 50 would be acceptable), however the preferred approach is to specify the same number of significant figures both for the acceptance criteria and the measured result e.g. acceptance criteria of > 50.0 (one significant figure) and a measured result of 50.4 (would be accepted).

8.5 Dynamic range/range of application

The response of the instrument and the *in vitro* method with regard to the readouts of interest should be known, and should be evaluated over a specified concentration range. An important characteristic of the method performance, for both quantitative and qualitative *in vitro* methods is often referred to as Limit Of Detection (LOD). It determines the lowest actual concentration or signal that can be consistently detected with acceptable precision, but not necessarily quantified. In case of normally distributed data, the LOD is often determined as

the concentrations at the average response + 3* SD of the negative control range, as this gives only 1% chance of a false positive.

For quantitative measurements, the boundaries of the dynamic range are determined by the lowest and highest analyte concentrations that generate results that are reliably produced by an *in vitro* method without dilution of the sample. To ensure the ability of the method to yield test results that are directly proportional to the analyte concentration (cell number, enzyme activity) the linearity of the method should be ensured within the given range. Subsequently, to facilitate efficient method transfer, the linear regression parameters (correlation coefficient, slope, y-intercept as well as residual sum of squares) should be submitted along with a plot of the data. When the upper limit is exceeded, samples fall outside of the linear range, they may need to be diluted (when the upper limit is exceeded), if possible. In case the samples give a result below the lower limit of the linear range, it may be useful to change to a different apparatus with a higher sensitivity or adapt the sample preparation to obtain higher concentrations, if possible.

The lower limit of linearity is frequently referred to as the lower limit of quantification (LLOQ) and the upper limit of linearity as the upper limit of quantification (ULOQ). The upper limit of linearity may be restricted by the highest available concentration in a sample or by the saturation of the signal generated by the instrument. The lower limit of quantification may also be based on the negative control range.

For certain assays/methodologies, equations other than the linear can be fit as a standard curve, provided that the user is operating within the range of the assay/equipment (see 4.1 and above). However, it is recommended that the simplest model that adequately describes the concentration-response relationship is used. Selection of weighting and use of a complex regression equation should be justified. (Burd, 2010; EMEA, 2011; FDA, 2001; Viswanathan *et al.*, 2007).

8.6 Signal intensity

The signal to noise (S/N) ratio is frequently applied for methods, which exhibit background noise (observed as the variation of the blanks) as baseline. It is calculated by comparing measured signals from samples with positive control item with those of blank samples. Note: concentrations of positive controls should be chosen in a low to medium range (giving a clear signal, but rather in the lower range of the linear range). While the S/N ratio is useful for tests, in which the blank (negative control) and the positive control exhibit comparable levels of variation, measures that take into account the levels of variability of both the blank and the sample may be more appropriate for biochemical and cell biology research.
They include:

- Signal windows (SW) (Sittampalam, 1997): these allow for the calculation of the separation between the sample (positive control) and the blank (negative control/solvent) including the variability of both measurements:

$$SW = \frac{Mean_{sample} - Mean_{control} - 3*(SD_{sample} + SD_{control})}{SD_{sample}}$$

A common SW acceptance criterion is SW≥2 for biological assays.

- Z-factors (Zhang *et al.*, 1999): they represent both the assay signal dynamic range and the data variation associated with the signal measurements, and therefore are suitable

for assay quality assessment. Z-factors ≥0.4 indicate acceptable separation between the signal and the blank.

$$Z = 1 - \frac{3*(SD_{sample} + SD_{control})}{Mean_{sample} - Mean_{control}}$$

8.7 Signal variability and plate uniformity assessment

The Assay Guidance Manual (HTS Assay Validation, Section 3 (Iversen *et al.*, 2004)) proposes a signal variability and plate uniformity assessment when developing new *in vitro* methods. This assessment serves to quantify the variability of assay output for the minimum (base, background), middle (close to EC_{50}) and maximum (highest/maximal) assay readings. By employing an interleaved format one can calculate the coefficient of variation of the min, mid and max signals, Z-factors and signal windows. Exemplary acceptance criteria have been proposed as coefficient variation CV≤20%, Z-factor≥0.4 and SW≥2. In addition, the data from this assessment can be used to derive the number of required replicates, using power analysis (Crawley, 2015; Krzywinski and Altman, 2013).

Furthermore, seeding density variation, plate edge effects and drift can also be examined in this experimental setup using heat maps (e.g. created with Excel) or graphical plots.

Edge effects are differences in growth in outer wells compared to inner wells. They are often due to uneven evaporation rates or plate stacking and can be a source of variation, as outer wells can often present as outliers compared to inner wells. Edge effects can be detected by using a heat map of the readouts from the plate uniformity assessment, or by plotting the signal readouts as a function of well position.

The plate uniformity assessment can also detect drift effects, i.e. trends in signal from left-right or top-bottom. Drifts can be due to seeding density variation during the process of initial cell seeding in plates. For example, cells may be settling down in the master vessel which is used to store a cell suspension used to seed a particular plate. Additionally, using the same set of tips on a multichannel pipette while pipetting cells in media compositions prone to foaming, may compromise the accuracy of the seeding. Higher variability, which cannot be resolved via technique optimisation may require increased numbers of replicates, more drug concentrations used to calculate dose-response or a higher numbers of independent experiments (Iversen *et al.*, 2004).

8.8 Reliability of endpoint calculations

The coefficient of variation, signal windows and Z-factors are poorly suited to determine the variation and reproducibility for potency estimates in dose-response experiments. Comparing inflection point parameters such as IC_{50} or EC_{50} is further complicated by the fact that they are not normally distributed and require logarithmic transformation before averaging or statistical comparison (Christopoulos, 1998; Motulsky and Christopoulos, 2004).

8.8.1 Confidence intervals for EC_{50}s

While the parameters derived from non-linear regression analysis, like IC_{50} and EC_{50} values, are not normally distributed, their logarithms are (the potency estimates are log-normally distributed). Therefore, when multiple independent experiments are run, EC_{50} values from multiple runs can be combined by using their geometric mean, calculated by averaging the

3302 logEC_{50} values. The standard error of the mean of the logEC_{50} values can be further used to
3303 calculate 95%-confidence intervals for the EC_{50} of a number of independent experiments.
3304 Another way of quantifying dose-response curve reproducibility is by using the minimum
3305 significant ratio (MSR) (Eastwood *et al.*, 2006).

3306 ### 8.8.2 Minimum significant ratio

3307 The MSR is a parameter that can be used to quantify assay reproducibility and resolution (the
3308 smallest ratio between compound potency which can be detected in the *in vitro* method). The
3309 so-called 'replicate experiment study' (Assay Guidance Manual, HTS Assay Validation
3310 Section 4 (Iversen *et al.*, 2004)) uses the differences between the potency estimates for 20-30
3311 compounds in two independent assay runs to estimate the within-run assay variability over a
3312 range of potencies. The SD of the difference between logIC50s in run 1 and run 2 is
3313 computed. The $MSR=10^{2*SD}$ is computed and example acceptance criteria have been given as
3314 MSR<3 (meaning up to three-times potency difference between runs is judged acceptable).

3315 The control compound MSR can be calculated for controls where data for at least six runs is
3316 available and gives an estimate of between-run variability. The control compound MSR is
3317 calculated as $MSR=10^{2\sqrt{2}s}$, where s is the standard deviation of the $\log_{10}EC_{50}$ values across
3318 runs, assuming one EC_{50} result per run (Haas *et al.*, 2004).

3319 ### 8.8.3 Variability and outliers

3320 The degree of variability judged acceptable should be given for the critical parameters of the
3321 assay- CV, Z-factors, SW, MSR or 95%-CI for IC_{50} measurements.

3322 Criteria to detect/remove outliers should be stated and the reasoning behind should be given
3323 (Motulsky and Brown, 2006; Pincus, 1995).

3324 Variability should be determined for within runs (= repeatability), for within the laboratory (=
3325 intra-laboratory reproducibility), and for between laboratories (inter-laboratory
3326 reproducibility) during validation (Burd, 2010; EMEA, 2011; FDA, 2001; Viswanathan *et*
3327 *al.*, 2007).

3328 ### 8.8.4 Transferability: in between lab transfer

3329 The transferability of an *in vitro* method from the developer laboratory to a second laboratory
3330 is a crucial step for demonstrating the robustness of the established procedures and/or SOPs.
3331 This step is necessary to evaluate the practicability of the test and to identify possible sources
3332 of within- and between-laboratory variability. Moreover it provides also an estimation of the
3333 amount of training that will be necessary to successfully transfer the test to an inexperienced
3334 laboratory. A good inter-laboratory reproducibility proves transferability.

3335 ## 8.9 Accuracy, reliability and uncertainty

3336 The usefulness of an *in vitro* method depends on its accuracy and reliability to correctly
3337 classify chemicals according to its stated purpose (e.g., sensitivity, specificity, positive and
3338 negative predictivity, false positive and false negative rates). These values often are obtained
3339 from continues data and categorised accordingly (e.g. as strong, weak). In such case cut off
3340 values are used and their impact on the accuracy and reliability should be taken into account.
3341 The use of confidence bounds taking into account the distance to these cut off values should
3342 be considered.

There are a number of measures that can be used to assess method performance and the confidence of the test predictions (Gaddis and Gaddis, 1990), based on Table 6. Below is an example that has been used in the validation process of different *in vitro* methods that are now OECD test guideline methods.

Table 62: Possible outcomes of a test result of a chemical in a validation

		Gold Standard Test	
	Total number	Condition positive	Condition negative
Test outcome	Test outcome positive	True positive (TP)	False positive (FP)
	Test outcome negative	False negative (FN)	True Negative (TN)

The sensitivity is the ability of a test to reliably classify positive substances:

$$Sensitivity\ (\%)\ =\ 100\ \cdot \frac{TP}{(TP+FN)}$$

$$False\ negative\ rate\ (\beta)\ =\ 1-sensitivity = \frac{FN}{(TP+FN)}$$

The specificity is the ability to reliably classify negative substances:

$$Specificity\ (\%)\ =\ 100\ \cdot \frac{TN}{(TN+FP)}$$

$$False\ positive\ rate\ (\alpha)\ =\ 1-specificity = \frac{FP}{(TN+FP)}$$

The positive predictive value (PPV) is the proportion of correct positive responses among materials testing positive by an *in vitro* method:

$$PPV\ (\%)\ =\ 100 \cdot \frac{TP}{(TP+FP)}$$

The negative predictive value (NPV) is the proportion of correct negative responses among materials testing negative by an *in vitro* method:

$$NPV\ (\%)\ =\ 100\ \cdot \frac{TN}{(TN\ +\ FN)}$$

What level of sensitivity, specificity, etc. is acceptable is not standardised. Such levels are also very dependent on the list of chemicals with which they are determined, therefore strict boundaries in acceptable levels for these accuracy parameters are not realistic. Generally though, sensitivities below 75% should not be accepted.

A discussion on what is a gold standard for an *in vitro* test, is very important here, as this is where many validations of *in vitro* tests fail currently. Comparison to an *in vivo* test is

problematic, because of several reasons such as: (i) species differences e.g. in case of rat *in vivo* and human cells *in vitro*; (ii) *in vivo* tests may include more pathways to a certain effect (endpoint) than the *in vitro* test. In this case the *in vitro* test would detect fewer *in vivo* ("true") positives. The situation may also be reverse, e.g. due to compensations *in vivo* not existing *in vitro*; (iii) the test endpoints *in vivo* and *in vitro* may not be 100% identical; (iv) the *in vitro* test does not reflect *in vivo* toxicokinetics and metabolism, which can lead to fewer (e.g. in case of toxification by metabolic enzymes) or increased positives (e.g. in case of detoxification or an absorption barrier) being found by an *in vitro* test without any metabolic competence, compared to an *in vivo* test; (iv) lack of mechanistic understanding of the underlying biological processes resulting in a positive or negative effect *in vivo*. Hereby it should be mentioned that many routinely used animal-based methods correlate poorly to the effects observed in humans. It might happen, therefore, that an *in vitro* test with human tissues shows very poor sensitivity and specificity when compared to the animal test and would therefore seem to be a poor test. Comparison to human data as the golden standard could in that case show whether the *in vitro* test is really poor, or actually a better test than the animal test.

The chemicals used for determining accuracy and reliability should therefore be selected with care (only the ones working through the pathways included in the *in vitro* method) and this assessment should include toxicokinetic information on these chemicals.

EFSA's Scientific Committee defines uncertainty as referring to "all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources available for the assessment". Examples include:

- Possible limitations in the quality and representativeness of data.
- Comparing non-standardised data across countries or categories.
- Choosing one predictive modelling technique over another.
- Using default factors (such as the weight of an average adult).

Since 2013, EFSA's Scientific Committee has been developing guidance[20] on how to assess uncertainty in a structured and systematic way. The aim is to offer a tool-box of methodologies – both quantitative and qualitative – and related training for EFSA's Scientific Panels and staff as well as for other organisations (e.g. researchers, national authorities) that carry out scientific work on EFSA's behalf.

In addition, IPCS has published a guidance document on "Evaluating and expressing uncertainty in hazard characterization"(IPCS, 2014), which is focussed on animal test results. However, described concepts such as the benchmark dose (BMD), including its uncertainty characterisation, can also be used for *in vitro* assays.

[20] https://www.efsa.europa.eu/sites/default/files/160321DraftGDUncertaintyInScientificAssessment.pdf

Field Code Changed

9 Reporting of Results

In vitro methods must be fully documented by good recording and reporting practices and contain all pertinent details to allow subsequent and adequate analysis and reporting of results. For example, lot numbers, catalogue numbers, supplier details, and expiry dates for chemicals and reagents must be listed, as well as temperatures and times (e.g. storage of chemicals, incubation steps in the *in vitro* method), specific identification of critical equipment used and, perhaps most importantly, any deviations from SOPs. All this information must be directly and accurately recorded, signed and dated by the person performing the activity, as these recordings are important for the correct interpretation of the results obtained.

Experimental details and results should be easily located; a log page at the front of a notebook may help track the recordings and observations. Any reference to computer files containing data should also be catalogued in the notebook. Data files should always be backed-up in case of computer failure, corruption, or deletion.

Reporting requirements depend on the different development phases of the *in vitro* method. For regulatory use, requirements for reporting are described in the GLP Principles. Reporting adequate information and results of all developmental phases will increase the confidence in the *in vitro* method and would allow for general acceptability by regulatory authorities.

Classical use of *in vitro* method information such as results derived and reported from OECD *in vitro* test guideline methods (e.g. Ames, *in vitro* skin irritation, *in vitro* eye irritation) conducted at GLP test facilities are accepted by regulatory authorities of OECD member states and MAD-adhering economies. In contrast, results derived from non-standard *in vitro* methods and non-testing methods (QSARs, etc.) are not necessarily accepted by regulators but might be submitted as supporting information.

There is an increasing tendency towards more transparency when publishing work which may lead to better reproducibility of published data (Guidelines for Transparency and Openness Promotion (TOP) Open Science Framework[21]). The EU Competitiveness Council has also announced their target that all scientific publications resulting from publicly funded research should be publicly available by 2020[22].

Sharing of data in public repositories is also being encouraged and best principles regarding the publication of scientific data have also been addressed by others, such as the FAIR (Findable, Accessible, Interoperable, and Reusable) Guiding Principles for scientific data management and stewardship, by the Nature Publishing Group[23]. This initiative not only promotes more transparency and openness but also promotes the use of computer readable datasets and data mining so that computers have the ability to access the data autonomously, unaided by their human operators, which is core to the FAIR Principles.

In addition to the increasing openness and transparency, the publication of negative results is also gaining more ground e.g. Journal of Negative Results in BioMedicine is an open access, peer reviewed journal that provides a platform for the publication and discussion of non-confirmatory and "negative" data.

[21] https://osf.io/ud578/

[22] http://english.eu2016.nl/documents/press-releases/2016/05/27/all-european-scientific-articles-to-be-freely-accessible-by-2020

[23] http://www.nature.com/articles/sdata201618

Field Code Changed

Field Code Changed

Field Code Changed

The information provided by individual *in vitro* methods, as well as *in silico* predictions, can be combined, interpreted and used for regulatory decision making by means of Integrated Approaches to Testing and Assessment (IATA) (OECD 2016c). An IATA is an approach that integrates and weighs all relevant existing evidence and guides the targeted generation of new data, where required, to build up a hazard or risk assessment acceptable in regulatory decision-making. Ideally, IATA should be informed by mechanistic understanding of the underlying toxicokinetics and toxicodynamics. A framework for capturing the toxicodynamic information is provided by Adverse Outcome Pathways[24].

Structured integration of different data types can be performed at different levels, including raw data and summarised level data. Different levels of data integration can then be used including Boolean combinations of categorised results, scoring approaches, decision trees, deterministic and probabilistic approaches. As experience is gained, approaches to data integration can become standardised. Such approaches, called "defined approaches," can thus become core elements of IATA. A defined approach is a formalised decision-making approach consisting of a fixed data interpretation procedure used to interpret data from a defined set of information elements (OECD 2006a,b).

When submitting *in vitro* data to a receiving authority, the use of OECD endorsed templates (e.g. OECD Harmonised Template 201 – Intermediate Effects) is encouraged but is not yet obligatory. This is facilitated by the use of IUCLID (International Uniform ChemicaL Information Database), a software application used to record, store, maintain and exchange data on intrinsic and hazard properties of chemical substances.

9.1 Publishing

It is essential to have all the results reported in a uniform manner to facilitate their use in the IATA framework, where the same dataset can be used in many different ways. It is good practice to publish scientific results in a timely manner. The results will be used and re-used by other scientists, competitors, modellers or validation study statisticians. Moreover, for any systemic endpoint the prediction is/will be based on the results of many different studies, using different methods performed in different facilities.

Therefore, data sharing is encouraged by default, unless there is reason for confidentiality. It is recommended to not only publish the results, but also the method/SOP (e.g. Nature testing methods, JOVE, Testing method Exchange, Springer Testing methods, DB-ALM, JRC-QSAR DB) and using public data-sharing standards and repositories for raw data such as ISA-TAB (http://isa-tools.org/, Dryad, Figshare, and Scientific data (http://www.nature.com/sdata/). In the same vein, assay modifications and further developments should be published, while tests are still in a development or pre-validation phase. Such publications should include the changes leading to improvement, the rationale for them, and this should also entail information on which changes reduce assay performance, or that do not result in an improvement.

Field Code Changed

Field Code Changed

9.2 Reporting of method validation

Validation is at the interface between *in vitro* method development/optimisation and regulatory acceptance/international recognition and ensures a science-based and conscientious evaluation of *in vitro* methods and approaches (e.g. Integrated Testing

[24] http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm

Field Code Changed

Strategies, ITS), independent of specific interests, establishing their overall performance and fitness for a given purpose, i.e. their scientific validity[25].

The approach taken by the validation bodies may vary according to the needs of that body as to whether they will coordinate the validation or whether a validation study should be submitted to that body for assessment. In general, once a validation study has been completed, the validation report will be submitted to the validation body for assessment of the outcome of the validation. While the individual study raw data may not need to be reported to the validation body, the data analysis and validation report will be required. Completed validation studies will then undergo independent scientific peer review e.g. by the EURL ECVAM Scientific Advisory Committee (ESAC), whose outcome will also be made public.

On completion of the validation study assessment, the recommendations of the validation body plus the validation report will be made public, e.g. EURL-ECVAM publishes its EURL-ECVAM Recommendations on its website. Before finalising and publishing these recommendations, EURL-ECVAM, as a matter of routine, invites comments from the general public. The aim of an EURL-ECVAM Recommendation is to provide EURL-ECVAM views on the validity of the *in vitro* method in question, to advise on possible regulatory applicability, limitations and proper scientific use of the *in vitro* methods, and to suggest possible follow-up activities in view of addressing knowledge gaps.

The validation study project plan should outline the

- Organisations or individuals responsible for data collection
- The means of data collection, back-up and archiving
- Procedures for data collection (to be established in collaboration with the participating laboratories
- A consistent system of paper-based or electronic labelling of files and folders including provisions to clearly label final versions.

9.2.1 Reporting Templates

To assist in the reporting of data generated in the various laboratories participating in a validation study, templates are prepared either by the test submitter, the validation body or a combination of both. These reporting templates should be designed to support later analysis of the 'high-level' data of validation studies (e.g. by the statistician) and should be validated before being introduced into the validation process. For a fully GLP-compliant study where each laboratory is a test facility, the test facility's SOPs and report templates might need to be used and the validation body subsequently introduces specific templates for analysing the multi-study validation data.

In case of laboratories operating under GLP, there should be adequate measures put in place to assure the quality and reliability of the data being reported and recorded. However, in cases of laboratories that do not operate under quality schemes (e.g. university laboratories) it is highly advisable to reflect on simple procedures that allow monitoring the adequacy of data reporting, to ensure that the data used by the statistician for final analysis are correct and hence relevant. The simplest possible measure would be that data files are checked, at random, by either a laboratory member or a member of the validation management (the responsibility should be outlined in the validation project plan). Typical and avoidable

[25] https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance

Field Code Changed

mistakes are the incorrect handling (copy / pasting) of numerical values (e.g. wrong column in an Excel sheet with consequences on automated data analyses), accidental deletion of embedded formulas or mistakes when normalising values.

Collection of data generated by participating laboratories can be done in a secure environment by using web-based information exchange tools such as the European Commission's CIRCABC system.

9.3 Data reporting for regulatory purposes

Data captured in GLP studies and results derived thereof will play an important role in increasing the relevance of *in vitro* data in regulatory contexts. Consideration and ultimately acceptance of *in vitro* GLP data can be promoted by using a standardised data format. IATA and AOP knowledge, if properly captured and presented, leads to a better understanding of toxicity mechanisms, and ultimately the AOP knowledge - derived from testing several chemicals - may be extrapolated to predict the toxicity of all chemicals that trigger the same Molecular Initiating Event (MIE) or Key Event (KE). Until recently, the absence of a template to report Intermediate Effects (like MIEs and KEs) was a limiting factor

The OECD had already designed and published 114 OECD Harmonised Templates (OHTs) to report test results concerning:

- physical/chemical properties (e.g. boiling point, density, flammability, …),
- human toxicity (e.g. carcinogenicity, acute toxicity, …)
- environmental toxicity (e.g. aquatic toxicity, terrestrial toxicity, …)
- other properties describing degradation, accumulation etc.

These templates are geared towards results derived from classical (mostly OECD guideline) studies, focusing on apical endpoints, i.e. Adverse Outcomes.

However, reporting MIEs or KEs with such a classical OHT would tie them inseparably to the one Adverse Outcome the one template covers, which is undesirable, as the (*in-vitro, in-silico* mechanistic) information is then not easily accessible for building AOPs leading to other Adverse Outcomes: A Key Event can be relevant not only for one AOP, but several. Reporting the Intermediate Effect in an "AO-neutral" template makes the data available for all kinds of AOPs.

A new, AO-neutral OHT was therefore needed that would allow reporting observations from mechanistic (*in-vitro* and *in-silico*) tests, without immediately locking into one of several AOs the Intermediate Effect could lead to.

Knowing not only about results of animal tests (classical OHTs), but being able to cross-reference these test results with the intermediate effect observations (new OHT) has the potential to lead the way towards a less animal-centred hazard assessment.

The OECD therefore started an initiative to come up with a stable, stakeholder-endorsed OHT for reporting on "intermediate effects" being observed via *in vitro* assays and possibly other non-animal test methods (computational predictions etc.). The template was titled "OHT 201 - Intermediate effects".

OHT 201 was endorsed by the OECD Joint Meeting in 2015 and was finally published in August 2016, see http://www.oecd.org/ehs/templates/harmonised-templates-intermediate-effects.htm for more details.

The basic principle of OHT 201 is that:

- one or several objective observation(s) (= results from non-classical test methods)
- lead(s) to one subjective conclusion (= Intermediate Effect present, yes or no).

A properly filled in OHT 201 template therefore conveys a clear statement:

- Based on observations O_1, O_2, …O_n
- a certain chemical
- triggers/does not trigger
- a certain intermediate effect
- on a certain biological level
- at a certain effect concentration.

With OHT 201 being implemented in IUCLID, the ICT system used by industry to fulfil reporting obligations under more and more legislative programmes (e.g. REACH), the notion of Intermediate Effects (and implicitly AOPs and predictive toxicology) has started to get attention in the regulatory world. This is a first step towards the acceptance of results from alternative tests for regulatory purposes, with the ultimate goal of replacing *in-vivo*-centred Adverse Outcome observations with alternative-methods-centred IATA/AOP considerations as the basis for risk assessment.

For *in vitro* methods without a guideline, the Office of Pesticide Programs U.S. Environmental Protection Agency recommends following OECD Guidance Document 211 (OECD, 2014) for describing non-guideline *in vitro* methods (EPA, 2016).

10 Storage and retention of records and materials

As the ultimate goal is to develop an *in vitro* method which will be formally validated for its future use in a regulatory environment following a quality system (e.g. GLP), it is essential to have some knowledge of the regulatory requirements specifically relating to the storage and retention of data, records and materials as the *in vitro* method should be designed so as to be easily transferrable into a GLP facility.

Data, records and material (specimen) integrity arrangements must be in place throughout the *in vitro* method lifecycle to ensure that the accuracy and completeness of data, records and materials are retained. The lifecycle includes all phases in the life of the data, records and materials, from their initial creation or purchase through processing, use, retention, archival and retrieval, and eventual destruction. It is vital that formal records used to confirm the results and how they were obtained are held in a stable/secure form and location which is documented and traceable and for which there is a minimum storage period. Disposal after such storage periods should be recorded and a summary report of the stored data and the means of destruction prepared and held.

Retention arrangements (classified as either archive or backup in case of electronic data) must be designed to protect data, records and materials from deliberate or accidental changes, manipulations or deletions thus ensuring integrity throughout the retention period. Archiving is defined as the long term, permanent retention of completed data and relevant metadata, records or materials in its final form. Archived data, records or materials may need to be stored for many years and must be permanently locked such that no changes can be made without detection. In the case of paper records, storage must be in dry and fire- and light-proof conditions. In addition to this, they should be easily retrieved for regulatory inspections.

Storage, retention and archiving requirements vary considerably throughout the lifecycle of the *in vitro* method. In the early stages, e.g. method development, there are less formal requirements for storage and retention of records and materials than in the later stages and in general will follow the university/institute/company policy regarding the storage and retention of data, records or materials (where such a policy exists). The development phase should be used to define the raw data, preferably described in the *in vitro* method itself, and any data (e.g. metadata), records or materials, to be retained when used in a regulatory environment (e.g. GLP).

As the use of human cells and tissues is critical for the development and implementation of *in vitro* methods for regulatory use in human safety assessment, already in the early stages care has to be taken regarding their ownership, their identity and who can control their fate. A number of treaties, laws, and regulations help to guide the ethical collection of human-derived specimens (Clinchem, n.d.)(http://www.clinchem.org/content/56/11/1675.full).

10.1 Archiving of materials

An effective archiving system must also provide for the archiving of study samples and materials, e.g. slides, specimens, test items and reference material under suitable storage conditions (OECD, 2007a). Specimens and materials should be stored only as long as they are stable. The Principles of GLP state that: “a sample for analytical purposes from each batch of test item should be retained for all studies except short-term studies”. The same rules apply to these archives as apply to the paper based archive, i.e. access restrictions, retrieval and removal of items, etc.

The storage conditions should be optimal for these samples and often these archives will require dedicated storage facilities, e.g. low temperature storage such as -20°C, liquid nitrogen storage or storage of items under inert conditions. Where special storage equipment is required, the rules governing the control and maintenance of this equipment must be applied. Where computerised systems are used, these systems must also follow the facility's policy regarding the use of computerised system, including qualification and validation of said systems (OECD, 2016b).

Samples of test and reference items or specimen may however be discarded when the quality of the material no longer permits evaluation. Obviously, the storage conditions should be optimal for these samples. When samples of test and reference items or specimens are disposed of before the end of the required retention period, the reason for disposal should be justified and documented (e.g. the reason might be perishable specimens such as blood smears, freeze-dried preparations and wet tissues).

10.2 Document and record management

Good Document Practices should be followed for both paper records and electronic records in order to assure data integrity. These principles require that documentation has the characteristics of being Attributable, Legible, Contemporaneously recorded, Original and Accurate (sometimes referred to as ALCOA). These essential characteristics apply equally for both paper and electronic records (WHO, 2016).

When working under a quality system (e.g. ISO, GLP), the management of documents and records to be retained will be described in the quality system's documentation such as in the Quality Manual or various SOPs. In a regulatory environment, e.g. GLP, principles and national laws detail the archiving requirements (OECD, 2007a). Where no formal quality system exists the general concepts for storage of records still apply.

When storing electronic records it is critical that the relationships between data and metadata are maintained intact throughout the data life cycle, as the data alone cannot provide the necessary information to recreate a study.

Electronic records may be stored either as flat files, e.g. individual files, or in a relational database. Storing data in flat files is relatively trivial to implement, for example files may be stored on a local drive or on a network share, however the data stored in this way rarely contains all the relevant metadata, rendering the data incomplete. Due to the nature of flat files, the user will have full access to the file(s), which presents a greater data integrity risk since data can be moved, manipulated or even deleted without tracking.

Conversely, a relational database file structure is much more secure as it stores the data and metadata in different places but maintains the relationship between them. This makes it inherently more difficult to selectively delete, amend or recreate the original data and the metadata trail of actions, and most systems provide an audit trail to see who has done what and when. Such systems mean an upfront investment in software and time (training, etc.), and the life cycle of the relational database software must also be taken into consideration (e.g. upgrading, etc.). Once a system has been put in place it is difficult to migrate to another system and retain all metadata links, specifically the audit trail data. Migrating to a new system for whatever reason (e.g. current system becomes obsolete) will prove expensive in cost, time and retraining of personnel.

10.2.1 Documents and records to be retained

It is imperative that the historical data, paper-based or in the form of electronic data, are effectively managed so as to prevent any data integrity issues as this data may be requested when submitting the method for formal validation.

As compliance with the principles of GLP is required by law for safety studies on chemical products around the world, it is important that newly developed *in vitro* methods can be suitable to be performed in a GLP environment from the very beginning of their development, avoiding an adaption phase at the very end of the process (Coecke *et al.*, 2016). Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP (OECD, 2005b). Studies which support validations may or may not be subject to verification depending on compliance monitoring authorities' programmes[26].

In most cases electronic templates (e.g. Excel spreadsheets) will have been developed and validated for the processing of the study data. These electronic files must be populated with the acquired data from each laboratory and forwarded to the validation body upon conclusion of the study. It is recommended that also these electronic files are stored, at least until the conclusion of the complete validation process. The validation body also has responsibility for ensuring the integrity of the transfer of these files and their long term storage. It is highly recommended that the transfer of these files is not performed via email, but using a secure file transfer system (e.g. https or sftp or similar). The transfer integrity of the files may be controlled by creating a checksum for each file and checking the checksum at the end of the transfer steps.

FDA has previously advised that defining paper records as “raw data” (the so called typewriter rule) does not satisfy the predicate rules, that the industry has misinterpreted the 2003 21 CFR Part 11 Scope and Applications Guidance (FDA, 2003) and that "the printed paper copy of the chromatogram would not be considered a true copy" [27]. Although this comment was made about chromatographic data, the principles have much wider implications.

Many electronic records are important to retain in their dynamic (electronic) format, to enable interaction with the data. Data must be retained in a dynamic form where this is critical to its integrity or later verification. This should be justified based on risk.

10.3 Archiving and retention

In a regulatory environment, when studies are GLP compliant, the archiving retention time is usually defined in national legislation, however where there is no retention time specified, the OECD recommends that records and materials should be retained for at least three inspection cycles so that inspectors can evaluate the compliance of the test facility with the principles of GLP (OECD, 2007b).

Data is generated during the experimental phase of studies and during this phase the integrity of the data must be ensured until final archiving of the study. This data will usually be required for further analysis and as such will not be formally archived until the completion of the study. It is important that access to this data, both electronic and hard copies, is controlled until the final archiving upon completion of the study. It is recommended, where possible or

[26] http://www.oecd.org/env/ehs/testing/glp-frequently-asked-questions.htm

[27] http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm124787.htm

Field Code Changed

Field Code Changed

feasible, that the electronic data is set as read-only or that an audit trail is provided, detailing who did what and when.

10.3.1 Retrieval

Each facility should have in place procedures concerning the retrieval of archived records and materials. The procedures should detail who may retrieve records and materials, for how long and the return of records and materials to the archive. All steps mentioned above need to be documented and traceable.

In the case of electronic records, viewing the records without the possibility of alteration or deletion of the archived version does not constitute "retrieval" of a record. Most systems available nowadays support read-only access, without the possibility to change or delete the archived record.

10.3.2 Backup and restore

When archiving electronic documents, periodic backups should be performed of the electronic archives. These backups do not constitute archived records, however as they may be required to be restored in the case a system failure, the same rules regarding access to the archived electronic records should be applied to access to the backup(s). In general backups are foreseen for short term storage and not long term storage or archiving and therefore the long term readability of these archives is usually not an issue, however the restoration of the backups should also be checked on a regular basis.

Data generated during the experimental phase of the study should also be covered by the backup and restore policy of the facility.

11 References

Allen, M., Bjerke, M., Edlund, H., Nelander, S., Westermark, B., 2016. Origin of the U87MG glioma cell line: Good news and bad news. Sci. Transl. Med. 8, 354re3. doi:10.1126/scitranslmed.aaf6853

Anderson, R., O'Hare, M., Balls, M., Brady, M., Brahams, D., Burt, A., Chesné, C., Combes, R., Dennison, A., Garthoff, B., Hawksworth, G., Kalter, E., Lechat, A., Mayer, D., Rogiers, V., Sladowski, D., Southee, J., Trafford, J., van der Valk, J., van Zeller, A.-M., 1998. The Availability of Human Tissue for Biomedical Research: The Report and Recommendations of the ECVAM Workshop 32. Altern. Lab. Anim. 26, 763–77.

Andrews, P.W., Baker, D., Benvinisty, N., Miranda, B., Bruce, K., Brüstle, O., Choi, M., Choi, Y.-M., Crook, J.M., de Sousa, P.A., Dvorak, P., Freund, C., Firpo, M., Furue, M.K., Gokhale, P., Ha, H.-Y., Han, E., Haupt, S., Healy, L., Hei, D.J., Hovatta, O., Hunt, C., Hwang, S.-M., Inamdar, M.S., Isasi, R.M., Jaconi, M., Jekerle, V., Kamthorn, P., Kibbey, M.C., Knezevic, I., Knowles, B.B., Koo, S.-K., Laabi, Y., Leopoldo, L., Liu, P., Lomax, G.P., Loring, J.F., Ludwig, T.E., Montgomery, K., Mummery, C., Nagy, A., Nakamura, Y., Nakatsuji, N., Oh, S., Oh, S.-K., Otonkoski, T., Pera, M., Peschanski, M., Pranke, P., Rajala, K.M., Rao, M., Ruttachuk, R., Reubinoff, B., Ricco, L., Rooke, H., Sipp, D., Stacey, G.N., Suemori, H., Takahashi, T.A., Takada, K., Talib, S., Tannenbaum, S., Yuan, B.-Z., Zeng, F., Zhou, Q., 2015. Points to consider in the development of seed stocks of pluripotent stem cells for clinical applications: International Stem Cell Banking Initiative (ISCBI). Regen. Med. 10, 1–44. doi:10.2217/rme.14.93

Aschner, M., Ceccatelli, S., Daneshian, M., Fritsche, E., Hasiwa, N., Hartung, T., Hogberg, H.T., Leist, M., Li, A., Mundi, W.R., Padilla, S., Piersma, A.H., Bal-Price, A., Seiler, A., Westerink, R., Zimmer, B., Lein, P., 2016. Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use. ALTEX. doi:10.14573/altex.1604201

ATCC, 2014. Animal Cell Culture Guide, Atcc. doi:10.1093/chemse/bjt099

Austin, R.P., Barton, P., Cockroft, S.L., Wenlock, M.C., Riley, R.J., 2002. The influence of nonspecific microsomal binding on apparent intrinsic clearance, and its prediction from physicochemical properties. Drug Metab. Dispos. 30, 1497–503.

Bellwon, P., Truisi, G.L., Bois, F.Y., Wilmes, A., Schmidt, T., Savary, C.C., Parmentier, C., Hewitt, P.G., Schmal, O., Josse, R., Richert, L., Guillouzo, A., Mueller, S.O., Jennings, P., Testai, E., Dekant, W., 2015. Kinetics and dynamics of cyclosporine A in three hepatic cell culture systems. Toxicol. In Vitro 30, 62–78. doi:10.1016/j.tiv.2015.07.016

Bergström, C., Norinder, U., Luthman, K., Artursson, P., 2002. Experimental and computational screening models for prediction of aqueous drug solubility. Pharm. Res. 19, 182–8.

Bessems, J.G., Loizou, G., Krishnan, K., Clewell, H.J., Bernasconi, C., Bois, F., Coecke, S., Collnot, E.-M., Diembeck, W., Farcal, L.R., Geraets, L., Gundert-Remy, U., Kramer, N., Küsters, G., Leite, S.B., Pelkonen, O.R., Schröder, K., Testai, E., Wilk-Zasadna, I., Zaldívar-Comenges, J.-M., 2014. PBTK modelling platforms and parameter estimation tools to enable animal-free risk assessment: recommendations from a joint EPAA--EURL ECVAM ADME workshop. Regul. Toxicol. Pharmacol. 68, 119–39. doi:10.1016/j.yrtph.2013.11.008

Bilgen, B., Orsini, E., Aaron, R.K., Ciombor, D.M., 2007. FBS suppresses TGF-beta1-induced chondrogenesis in synoviocyte pellet cultures while dexamethasone and dynamic stimuli are beneficial. J. Tissue Eng. Regen. Med. 1, 436–42. doi:10.1002/term.56

Bosgra, S., Westerhout, J., 2015. Interpreting in vitro developmental toxicity test battery results: The consideration of toxicokinetics. Reprod. Toxicol. 55, 73–80. doi:10.1016/j.reprotox.2014.11.001

Bouhifd, M., Beger, R., Flynn, T., Guo, L., Harris, G., Hogberg, H., Kaddurah-Daouk, R., Kamp, H., Kleensang, A., Maertens, A., Odwin-DaCosta, S., Pamies, D., Robertson, D., Smirnova, L., Sun, J., Zhao, L., Hartung, T., 2015. Quality assurance of metabolomics. ALTEX 32, 319–26.

Box, G.E.., Hunter, J.., Hunter, W.G., 2005. Introduction to fractional factorial experimentation, in: Statistics for Experimenters. Wiley, New York.

Brindley, D.A., Davie, N.L., Culme-Seymour, E.J., Mason, C., Smith, D.W., Rowley, J.A., 2012. Peak serum: implications of serum supply for cell therapy manufacturing. Regen. Med. 7, 7–13. doi:10.2217/rme.11.112

Broeders, J.J.W., Blaauboer, B.J., Hermens, J.L.M., 2011. Development of a negligible depletion-solid phase microextraction method to determine the free concentration of chlorpromazine in aqueous samples containing albumin. J. Chromatogr. A 1218, 8529–35. doi:10.1016/j.chroma.2011.09.064

Burd, E.M., 2010. Validation of laboratory-developed molecular assays for infectious diseases. Clin. Microbiol. Rev. 23, 550–76. doi:10.1128/CMR.00074-09

Cadena-Herrera, D., Esparza-De Lara, J.E., Ramírez-Ibañez, N.D., López-Morales, C.A., Pérez, N.O., Flores-Ortiz, L.F., Medina-Rivero, E., 2015. Validation of three viable-cell counting methods: Manual, semi-automated, and automated. Biotechnol. Reports 7, 9–16. doi:10.1016/j.btre.2015.04.004

Chen, Y., Stevens, B., Chang, J., Milbrandt, J., Barres, B.A., Hell, J.W., 2008. NS21: Re-defined and modified supplement B27 for neuronal cultures. J. Neurosci. Methods 171, 239–247. doi:10.1016/j.jneumeth.2008.03.013

Christopoulos, A., 1998. Assessing the distribution of parameters in models of ligand-receptor interaction: to log or not to log. Trends Pharmacol. Sci. 19, 351–7.

Clinchem, n.d. Clinchem [WWW Document]. URL http://www.clinchem.org/content/56/11/1675.full

Coecke, S., Balls, M., Bowe, G., Davis, J., Gstraunthaler, G., Hartung, T., Hay, R., Merten, O.-W., Price, A., Schechtman, L., Stacey, G., Stokes, W., 2005. Guidance on good cell culture practice. a report of the second ECVAM task force on good cell culture practice. Altern. Lab. Anim. 33, 261–87.

Coecke, S., Bernasconi, C., Bowe, G., Bostroem, A.-C., Burton, J., Cole, T., Fortaner, S., Gouliarmou, V., Gray, A., Griesinger, C., Louhimies, S., Gyves, E.M., Joossens, E., Prinz, M.-J., Milcamps, A., Parissis, N., Wilk-Zasadna, I., Barroso, J., Desprez, B., Langezaal, I., Liska, R., Morath, S., Reina, V., Zorzoli, C., Zuang, V., 2016. Practical Aspects of Designing and Conducting Validation Studies Involving Multi-study Trials. Adv. Exp. Med. Biol. 856, 133–163. doi:10.1007/978-3-319-33826-2_5

Coecke, S., Bowe, G., Milcamps, A., Bernasconi, C., Bostroem, A.-C., Bories, G., Fortaner, S., Gineste, J.-M., Gouliarmou, V., Langezaal, I., Liska, R., Mendoza, E., Morath, S., Reina, V., Wilk-Zasadna, I., Whelan, M., 2014. Considerations in the development of in vitro toxicity testing methods intended for regulatory use, in: Bal-Price, A., Jennings, P. (Eds.), In Vitro Toxicology Systems, Methods in Pharmacology and Toxicology. Springer New York. doi:10.1007/978-1-4939-0521-8_25

Crawley, M.J., 2015. Chapter 1. Fundamentals, in: Statistics: An Introduction Using R, 2nd Edition. Wiley, pp. 8–10.

Crean, D., Bellwon, P., Aschauer, L., Limonciel, A., Moenks, K., Hewitt, P., Schmidt, T., Herrgen, K., Dekant, W., Lukas, A., Bois, F., Wilmes, A., Jennings, P., Leonard, M.O., 2015. Development of an in vitro renal epithelial disease state model for xenobiotic toxicity testing. Toxicol. In Vitro 30, 128–37. doi:10.1016/j.tiv.2014.11.015

Eastwood, B.J., Farmen, M.W., Iversen, P.W., Craft, T.J., Smallwood, J.K., Garbison, K.E., Delapp, N.W., Smith, G.F., 2006. The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays and estimation by replicate-experiment studies. J. Biomol. Screen. 11, 253–61. doi:10.1177/1087057105285611

ECACC, 2010. Fundamentals Techniques in Cell Culture.

Ekins, S., Olechno, J., Williams, A.J., 2013. Dispensing processes impact apparent biological activity as determined by computational and statistical analyses. PLoS One 8, e62325. doi:10.1371/journal.pone.0062325

EMA, 2014. Concept paper on review and update of EMA guidelines to implement best practice with regard to 3Rs (replacement, reduction and refinement) in regulatory testing of medicinal products.

EMEA, 2011. Guideline on bioanalytical method validation EMEA/CHMP/EWP/192217/2009.

EPA, 2016. Process for Evaluating & Implementing Alternative Approaches To Traditional in Vivo Acute Toxicity Studies for Fifra Regulatory Use 1–9.

EPA, 2007. Guidance for Preparing Standard Operating Procedures.

ESAC, 2008. ESAC statement on the use of FCS and other animal-derived supplements.

Esch, P.M., Moor, C., Schmid, B., Albertini, S., Hassler, S., Donzé, G., Saxer, H.P., 2010. Good Laboratory Practice (GLP) - Guidelines for the Development and Validation of Spreadsheets. Qual. Assur. J. 13, 41–56. doi:10.1002/qaj.466

EU, 2009. DIRECTIVE 2009/41/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the contained use of genetically modified micro-organisms.

FDA, 2003. Part 11, Electronic Records; Electronic Signatures — Scope and Application. FDA Guid. Ind. 12.

FDA, 2001. Guidance for Industry Bioanalytical Method Validation.

Frattini, A., Fabbri, M., Valli, R., De Paoli, E., Montalbano, G., Gribaldo, L., Pasquali, F., Maserati, E., 2015. High variability of genomic instability and gene expression profiling in different HeLa clones. Sci. Rep. 5, 15377. doi:10.1038/srep15377

Gaddis, G.M., Gaddis, M.L., 1990. Introduction to biostatistics: Part 3, Sensitivity, specificity, predictive value, and hypothesis testing. Ann. Emerg. Med. 19, 591–7. doi:10.1016/S0196-0644(05)82198-5

Geraghty, R.J., Capes-Davis, A., Davis, J.M., Downward, J., Freshney, R.I., Knezevic, I., Lovell-Badge, R., Masters, J.R.W., Meredith, J., Stacey, G.N., Thraves, P., Vias, M., 2014. Guidelines for the use of cell lines in biomedical research. Br. J. Cancer 111, 1021–46. doi:10.1038/bjc.2014.166

Grant, R.J., Roberts, K., Pointon, C., Hodgson, C., Womersley, L., Jones, D.C., Tang, E., 2009. Achieving accurate compound concentration in cell-based screening: validation of acoustic droplet ejection

technology. J. Biomol. Screen. 14, 452–9. doi:10.1177/1087057109336588
Groothuis, F.A., Heringa, M.B., Nicol, B., Hermens, J.L.M., Blaauboer, B.J., Kramer, N.I., 2015. Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. Toxicology 332, 30–40. doi:10.1016/j.tox.2013.08.012
Groten, J.P., Schoen, E.D., van Bladeren, P.J., Kuper, C.F., van Zorge, J.A., Feron, V.J., 1997. Subacute toxicity of a mixture of nine chemicals in rats: detecting interactive effects with a fractionated two-level factorial design. Fundam. Appl. Toxicol. 36, 15–29. doi:10.1006/faat.1996.2281
Gülden, M., Jess, A., Kammann, J., Maser, E., Seibert, H., 2010. Cytotoxic potency of H2O2 in cell cultures: impact of cell concentration and exposure time. Free Radic. Biol. Med. 49, 1298–305. doi:10.1016/j.freeradbiomed.2010.07.015
Gülden, M., Kähler, D., Seibert, H., 2015. Incipient cytotoxicity: A time-independent measure of cytotoxic potency in vitro. Toxicology 335, 35–45. doi:10.1016/j.tox.2015.07.002
Gülden, M., Mörchel, S., Seibert, H., 2001. Factors influencing nominal effective concentrations of chemical compounds in vitro: cell concentration. Toxicol. In Vitro 15, 233–43.
Gülden, M., Seibert, H., 2003. In vitro-in vivo extrapolation: estimation of human serum concentrations of chemicals equivalent to cytotoxic concentrations in vitro. Toxicology 189, 211–22.
Gunetti, M., Castiglia, S., Rustichelli, D., Mareschi, K., Sanavio, F., Muraro, M., Signorino, E., Castello, L., Ferrero, I., Fagioli, F., 2012. Validation of analytical methods in GMP: the disposable Fast Read 102® device, an alternative practical approach for cell counting. J. Transl. Med. 10, 112. doi:10.1186/1479-5876-10-112
Gupta, K., Rispin, A., Stitzel, K., Coecke, S., Harbell, J., 2005. Ensuring quality of in vitro alternative test methods: issues and answers. Regul. Toxicol. Pharmacol. 43, 219–24. doi:10.1016/j.yrtph.2005.03.010
Haas, J. V., Eastwood, B.J., Iversen, P.W., Weidner, J.R., 2004. Minimum Significant Ratio – A Statistic to Assess Assay Variability, Assay Guidance Manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences.
Hall, M.D., Telma, K.A., Chang, K.E., Lee, T.D., Madigan, J.P., Lloyd, J.R., Goldlust, I.S., Hoeschele, J.D., Gottesman, M.M., 2014. Say no to DMSO: Dimethylsulfoxide inactivates cisplatin, carboplatin, and other platinum complexes. Cancer Res. 74, 3913–3922. doi:10.1158/0008-5472.CAN-14-0247
Harbell, J., Raabe, H., J. Harbell and H. Raabe, 2014. In Vitro Methods for the Prediction of Ocular and Dermal Toxicity, in: Handbook of Toxicology, Third Edition. CRC Press, pp. 197–231. doi:doi:10.1201/b16632-6
Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V., 2004. A modular approach to the ECVAM principles on test validity. Altern. Lab. Anim. 32, 467–72.
Hayashi, M., Dearfield, K., Kasper, P., Lovell, D., Martus, H.-J., Thybaud, V., 2011. Compilation and use of genetic toxicity historical control data. Mutat. Res. 723, 87–90. doi:10.1016/j.mrgentox.2010.09.007
Heringa, M.B., de Wit-Bos, L., Bos, P.M.J., Hakkert, B., 2014. Do current EU regulations for the safety assessment of chemical substances pose legal barriers for the use of alternatives to animal testing?, RIVM Report.
Heringa, M.B., Hermens, J.L.M., 2003. Measurement of free concentrations using negligible depletion-solid phase microextraction (nd-SPME). TrAC Trends Anal. Chem. 22, 575–587. doi:10.1016/S0165-9936(03)01006-9
Heringa, M.B., Hogevonder, C., Busser, F., Hermens, J.L.M., 2006. Measurement of the free concentration of octylphenol in biological samples with negligible depletion-solid phase microextraction (nd-SPME): analysis of matrix effects. J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci. 834, 35–41. doi:10.1016/j.jchromb.2006.02.009
Heringa, M.B., Schreurs, R.H.M.M., Busser, F., van der Saag, P.T., van der Burg, B., Hermens, J.L.M., 2004. Toward more useful in vitro toxicity data with measured free concentrations. Environ. Sci. Technol. 38, 6263–70.
Hoelke, B., Gieringer, S., Arlt, M., Saal, C., 2009. Comparison of nephelometric, UV-spectroscopic, and HPLC methods for high-throughput determination of aqueous drug solubility in microtiter plates. Anal. Chem. 81, 3165–72. doi:10.1021/ac9000089
ISCBI, 2009. Consensus guidance for banking and supply of human embryonic stem cell lines for research purposes. Stem Cell Rev. 5, 301–14. doi:10.1007/s12015-009-9085-x
ISO, 2015. ISO 15519-2:2015 - Specifications for diagrams for process industry -- Part 2: Measurement and control.
ISO, 2005. ISO/IEC 17025:2005 - General requirements for the competence of testing and calibration laboratories.
ISPE, 2012. ISPE GAMP Good Practice Guide. A Risk-Based Approach to GxP Compliant Laboratory Computerized Systems, Second. ed. ISPE Publication.
Iversen, P.W., Beck, B., Chen, Y.-F., Dere, W., Devanarayan, V., Eastwood, B.J., Farmen, M.W., Iturria, S.J.,

Montrose, C., Moore, R.A., Weidner, J.R., Sittampalam, G.S., 2004. HTS Assay Validation, Assay Guidance Manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences.
Jochems, C.E.A., van der Valk, J.B.F., Stafleu, F.R., Baumans, V., n.d. The use of fetal bovine serum: ethical or scientific problem? Altern. Lab. Anim. 30, 219–27.
Jouyban, A., Fakhree, M.A., 2012. Experimental and Computational Methods Pertaining to Drug Solubility, in: Toxicity and Drug Testing. InTech. doi:10.5772/19206
Kleensang, A., Vantangoli, M.M., Odwin-DaCosta, S., Andersen, M.E., Boekelheide, K., Bouhifd, M., Fornace, A.J., Li, H.-H., Livi, C.B., Madnick, S., Maertens, A., Rosenberg, M., Yager, J.D., Zhaog, L., Hartung, T., 2016. Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function. Sci. Rep. 6, 28994. doi:10.1038/srep28994
Knöbel, M., Busser, F.J.M., Rico-Rico, A., Kramer, N.I., Hermens, J.L.M., Hafner, C., Tanneberger, K., Schirmer, K., Scholz, S., 2012. Predicting adult fish acute lethality with the zebrafish embryo: relevance of test duration, endpoints, compound properties, and exposure concentration analysis. Environ. Sci. Technol. 46, 9690–700. doi:10.1021/es301729q
Kramer, J.R., Schmidt, N.W., Mayle, K.M., Kamei, D.T., Wong, G.C.L., Deming, T.J., 2015. Reinventing Cell Penetrating Peptides Using Glycosylated Methionine Sulfonium Ion Sequences. ACS Cent. Sci. 1, 83–8. doi:10.1021/acscentsci.5b00054
Kramer, N.I., Krismartina, M., Rico-Rico, A., Blaauboer, B.J., Hermens, J.L.M., 2012. Quantifying processes determining the free concentration of phenanthrene in Basal cytotoxicity assays. Chem. Res. Toxicol. 25, 436–45. doi:10.1021/tx200479k
Krzywinski, M., Altman, N., 2013. Points of significance: Power and sample size. Nat. Methods 10, 1139–1140. doi:10.1038/nmeth.2738
Leist, M., Efremova, L., Karreman, C., 2010. Food for thought ... considerations and guidelines for basic test method descriptions in toxicology. ALTEX 27, 309–17.
Loizou, G., Spendiff, M., Barton, H.A., Bessems, J., Bois, F.Y., d'Yvoire, M.B., Buist, H., Clewell, H.J., Meek, B., Gundert-Remy, U., Goerlitz, G., Schmitt, W., 2008. Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps. Regul. Toxicol. Pharmacol. 50, 400–11. doi:10.1016/j.yrtph.2008.01.011
Louisse, J., de Jong, E., van de Sandt, J.J.M., Blaauboer, B.J., Woutersen, R.A., Piersma, A.H., Rietjens, I.M.C.M., Verwei, M., 2010. The use of in vitro toxicity data and physiologically based kinetic modeling to predict dose-response curves for in vivo developmental toxicity of glycol ethers in rat and man. Toxicol. Sci. 118, 470–84. doi:10.1093/toxsci/kfq270
Motulsky, H.J., Brown, R.E., 2006. Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics 7, 123. doi:10.1186/1471-2105-7-123
Motulsky, H.J., Christopoulos, A., 2004. Fitting models to biological data using linear and nonlinear regression.
National Institutes of Health, 2001. C & D of NIH Publication No: 01-4500. Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity 1–102.
OECD, 2016a. OECD principles No 18 - OECD Position Paper Regarding the Relationship between the OECD Principles of GLP and ISO/IEC 17025.
OECD, 2016b. OECD principles No 17 - Application of GLP Principles to Computerised Systems, OECD. doi:ENV/JM/MONO(2007)10
OECD, 2014. OECD Series on Testing and Assessment. Number 211. Guidance Document for describing non-guideline in vitro test methods. doi:ENV/JM/MONO(2014)35
OECD, 2011. OECD guideline for testing of chemicals. Section 2: effects on biotic systems. In: Guideline 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Organisation for Economic Cooperation and Development, Paris, France.
OECD, 2009. OECD Guidelines for the Testing of Chemicals, Section 4, Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants.
OECD, 2007a. OECD series on Principles of Good Laboratory Practice and Compliance Monitoring Number 15. Establishment and Control of Archives that Operate in Compliance with the Principles of GLP. Paris : OECD Publishing.
OECD, 2007b. OECD GLP 15 - Establishment and Control of Archives that Operate in Compliance with the Principles of GLP. Guideline. doi:ENV/JM/MONO(2007)10
OECD, 2006a. OECD Guideline for testing of chemicals. Draft proposal for a new guideline. In: Fish Embryo Toxicity (FET) Test. Organisation for Economic Cooperation and Development, Paris, France.
OECD, 2006b. OECD guideline for testing of chemicals. Section 2: Effects on biotic systems. In: Guideline 221: Lemna sp. Growth Inhibition Test. Organisation for Economic Cooperation and Development, Paris, France.
OECD, 2005a. OECD Series on Testing and Assessment. Number 34. Guidance Document on the Validation

and International Acceptance of New or Updated Test Methods for Hazard Assessment. ENV/JM/MONO(2005)14.
OECD, 2005b. OECD Guidance Document on the Validation and International Acceptance of new or updated test methods for Hazard Assessment No. 34. Guidance 33, 1–96. doi:ENV/JM/MONO(2005)14
OECD, 2004a. The Application of the Principles of GLP to in vitro Studies. OECD Publishing.
OECD, 2004b. OECD series on Principles of Good Laboratory Practice and Compliance Monitoring. Number 14. Advisory Document of the Working Group on Good Laboratory Practice: The Application of the Principles of GLP to in vitro Studies. ENV/JM/MONO(2004)26, OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. OECD Publishing. doi:10.1787/9789264084971-en
OECD, 2004c. OECD GLP 14 - Application of GLP to in vitro Studies. Guideline. doi:ENV/JM/MONO(2004)26
OECD, 2000. Guidance document on aquatic toxicity testing of difficult substances and mixtures. Environ. Heal. Saf. Publ. Ser. Test. Assess. - N°23 53. doi:ENV/JM/MONO(2007)10
OECD, 1999. OECD Series on Principles of GLP and Compliance Monitoring Number 4 (Revised) - QUALITY ASSURANCE AND GLP. Guideline 4. doi:ENV/JM/MONO(99)20
OECD, 1998a. OECD series on principles of good laboratory practice and compliance monitoring. Number 1. OECD principles on good laboratory practice. ENV/MC/CHEM(98)17. Guideline 33, 1–172.
OECD, 1998b. OECD Principles on Good Laboratory Practice.
OMCL, 2011. OMCL Network of the Council of Europe QUALITY ASSURANCE DOCUMENT. QUALIFICATION OF EQUIPMENT CORE DOCUMENT PA/PH/OMCL (08) 73 2R.
OMCL, 2009. Validation of Computerised Systems Annex 1: Validation of computerised calculation systems: example of validation of in-house software PA/PH/OMCL (08) 87 2R.
Ono, K., Satoh, M., Yoshida, T., Ozawa, Y., Kohara, A., Takeuchi, M., Mizusawa, H., Sawada, H., 2007. Species identification of animal cells by nested PCR targeted to mitochondrial DNA. In Vitro Cell. Dev. Biol. Anim. 43, 168–75. doi:10.1007/s11626-007-9033-5
Pamies, D., Bal-Price, A., Simeonov, A., Tagle, D., Allen, D., Gerhold, D., Yin, D., Pistollato, F., Inutsuka, T., Sullivan, K., Stacey, G., Salem, H., Leist, M., Daneshian, M., Vemuri, M.C., McFarland, R., Coecke, S., Fitzpatrick, S.C., Lakshmipathy, U., Mack, A., Wang, W.B., Daiju, Y., Sekino, Y., Kanda, Y., Smirnova, L., Hartung, T., 2016. Good Cell Culture Practice for stem cells and stem-cell-derived models. ALTEX. doi:10.14573/altex.1607121
Persson, L.C., Porter, C.J.H., Charman, W.N., Bergström, C.A.S., 2013. Computational prediction of drug solubility in lipid based formulation excipients. Pharm. Res. 30, 3225–3237. doi:10.1007/s11095-013-1083-7
Phelan, M., Lawler, G., 2001. Cell counting. Curr. Protoc. Cytom. 00:3A:A.3A.
Pincus, R., 1995. Barnett, V., and Lewis T.: Outliers in Statistical Data. 3rd edition. J. Wiley & Sons 1994, XVII. 582 pp. Biometrical J. 37, 256–256. doi:10.1002/bimj.4710370219
Pistollato, F., Bremer-Hoffmann, S., Healy, L., Young, L., Stacey, G., 2012. Standardization of pluripotent stem cell cultures for toxicity testing. Expert Opin. Drug Metab. Toxicol. 8, 239–57. doi:10.1517/17425255.2012.639763
Pistollato, F., Louisse, J., Scelfo, B., Mennecozzi, M., Accordi, B., Basso, G., Gaspar, J.A., Zagoura, D., Barilari, M., Palosaari, T., Sachinidis, A., Bremer-Hoffmann, S., 2014. Development of a pluripotent stem cell derived neuronal model to identify chemically induced pathway perturbations in relation to neurotoxicity: effects of CREB pathway inhibition. Toxicol. Appl. Pharmacol. 280, 378–88. doi:10.1016/j.taap.2014.08.007
Pomponio, G., Savary, C.C., Parmentier, C., Bois, F., Guillouzo, A., Romanelli, L., Richert, L., Di Consiglio, E., Testai, E., 2015. In vitro kinetics of amiodarone and its major metabolite in two human liver cell models after acute and repeated treatments. Toxicol Vitr. 30, 36–51. doi:10.1016/j.tiv.2014.12.012
Reinert, K.H., Giddings, J.M., Judd, L., 2002. Effects analysis of time-varying or repeated exposures in aquatic ecological risk assessment of agrochemicals. Environ. Toxicol. Chem. 21, 1977–1992. doi:10.1002/etc.5620210928
Riedl, J., Altenburger, R., 2007. Physicochemical substance properties as indicators for unreliable exposure in microplate-based bioassays. Chemosphere 67, 2210–20. doi:10.1016/j.chemosphere.2006.12.022
Rispin, A., Stitzel, K., Harbell, J., Klausner, M., 2006. Ensuring quality of in vitro alternative test methods: Current practice. Regul. Toxicol. Pharmacol. 45, 97–103. doi:10.1016/j.yrtph.2005.03.005
Scenhir, 2015. Final Opinion on the Guidance on the Determination of Potential Health Effects of Nanomaterials Used in Medical Devices. doi:10.2772/41391
Schmidt, B.Z., Lehmann, M., Gutbier, S., Nembo, E., Noel, S., Smirnova, L., Forsby, A., Hescheler, J., Avci, H.X., Hartung, T., Leist, M., Kobolák, J., Dinnyés, A., 2016. In vitro acute and developmental neurotoxicity screening: an overview of cellular platforms and high-throughput technical possibilities. Arch. Toxicol. doi:10.1007/s00204-016-1805-9

Shahdadfar, A., Frønsdal, K., Haug, T., Reinholt, F.P., Brinchmann, J.E., 2005. In vitro expansion of human mesenchymal stem cells: choice of serum is a determinant of cell proliferation, differentiation, gene expression, and transcriptome stability. Stem Cells 23, 1357–66. doi:10.1634/stemcells.2005-0094

Sittampalam, G.S., 1997. Design of Signal Windows in High Throughput Screening Assays for Drug Discovery. J. Biomol. Screen. 2, 159–169. doi:10.1177/108705719700200306

Smith, K.E.C., Heringa, M.B., Uytewaal, M., Mayer, P., 2013. The dosing determines mutagenicity of hydrophobic compounds in the Ames II assay with metabolic transformation: passive dosing versus solvent spiking. Mutat. Res. 750, 12–8. doi:10.1016/j.mrgentox.2012.07.006

Smith, K.E.C., Oostingh, G.J., Mayer, P., 2010. Passive dosing for producing defined and constant exposure of hydrophobic organic compounds during in vitro toxicity tests. Chem. Res. Toxicol. 23, 55–65. doi:10.1021/tx900274j

Stacey, G. (Glyn), Davis, J., 2007. Medicines from animal cell culture. Wiley.

Stacey, G., Coecke, S., Price, A., Healy, L., Jennings, P., Wilmes, A., Pinset, C., Sundstrom, M., Myatt, G., 2016. Ensuring the Quality of Stem Cell-Derived In Vitro Models for Toxicity Testing, in: Eskes, C., Whelan, M. (Eds.), Validation of Alternative Methods for Toxicity Testing. Springer, pp. 259–297.

Stacey, G.N., 2011. Cell culture contamination. Methods Mol. Biol. 731, 79–91. doi:10.1007/978-1-61779-080-5_7

Stacey, G.N., Hartung, T., 2006. The use of human tissue in drug discovery, scientific, ethical, legal and regulatory environments, Chapter 9., in: V, M.U. and S. (Ed.), Drug Testing In Vitro – Breakthroughs and Trends in Cell Culture Technology. Wiley-VCH, pp. 231–250.

Sullivan, B., 2001. Assay Development in High Density MicroWell® Plates: Use of Well Geometries, Format, Surface Modification and Optical Properties to Achieve Optimal Assay Performance. J. Assoc. Lab. Autom. 6, 47–52. doi:10.1016/S1535-5535(04)00125-X

Tanneberger, K., Rico-Rico, A., Kramer, N.I., Busser, F.J.M., Hermens, J.L.M., Schirmer, K., 2010. Effects of solvents and dosing procedure on chemical toxicity in cell-based in vitro assays. Environ. Sci. Technol. 44, 4775–81. doi:10.1021/es100045y

Ulrey, A.K., Curren, R.D., Harbell, J.W., Mun, G., A, H., Raabe, 2005. Applying Good Laboratory Practices (GLPs) to In Vitro Studies: One Laboratory's Perspective.

Usta, S.N., Scharer, C.D., Xu, J., Frey, T.K., Nash, R.J., 2014. Chemically defined serum-free and xeno-free media for multiple cell lineages. Ann. Transl. Med. 2, 97. doi:10.3978/j.issn.2305-5839.2014.09.05

van der Valk, J., Mellor, D., Brands, R., Fischer, R., Gruber, F., Gstraunthaler, G., Hellebrekers, L., Hyllner, J., Jonker, F.H., Prieto, P., Thalen, M., Baumans, V., 2004. The humane collection of fetal bovine serum and possibilities for serum-free cell and tissue culture. Toxicol. In Vitro 18, 1–12.

Viswanathan, C.T., Bansal, S., Booth, B., DeStefano, A.J., Rose, M.J., Sailstad, J., Shah, V.P., Skelly, J.P., Swann, P.G., Weiner, R., 2007. Quantitative bioanalytical methods validation and implementation: Best practices for chromatographic and ligand binding assays. Pharm. Res. doi:10.1007/s11095-007-9291-7

Vogel, G., 2010. Cell biology. To scientists' dismay, mixed-up cell lines strike again. Science 329, 1004. doi:10.1126/science.329.5995.1004

Weil, C.S., Scala, R.A., 1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicol. Appl. Pharmacol. 19, 276–360.

WHO, 2016. Annex 5 Guidance on good data and record management practices, in: WHO Technical Report Series, No. 996, 2016, Annex 5. pp. 165–210.

WHO, 2013. WHO Handbook: Quality practices in basic biomedical research, WHO. World Health Organization.

WHO, 2010. Recommendations for the evaluation of animal cell cultures as substrates for the manufacture of biological medicinal products and for the characterization of cell banks Table of contents, Who.

Wilson, H.K., Canfield, S.G., Hjortness, M.K., Palecek, S.P., Shusta, E. V, 2015. Exploring the effects of cell seeding density on the differentiation of human pluripotent stem cells to brain microvascular endothelial cells. Fluids Barriers CNS 12, 13. doi:10.1186/s12987-015-0007-9

Yamasaki, S., Taguchi, Y., Shimamoto, A., Mukasa, H., Tahara, H., Okamoto, T., 2014. Generation of human induced pluripotent stem (Ips) cells in serum- and feeder-free defined culture and TGF-B1 regulation of pluripotency. PLoS One 9, e87151. doi:10.1371/journal.pone.0087151

Yoshimura, I., Matsumoto, K., 1994. Notes on the use of historical controls. Environ. Health Perspect. 19–23.

Zeiger, E., Anderson, B., Haworth, S., Lawlor, T., Mortelmans, K., 1988. Salmonella mutagenicity tests: IV. Results from the testing of 300 chemicals. Environ. Mol. Mutagen. 11 Suppl 1, 1–157.

Zhang, J.J.-H., Chung, T., Oldenburg, K., 1999. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J. Biomol. Screen. 4, 67–73. doi:10.1177/108705719900400206

Zimmer, B., Pallocca, G., Dreser, N., Foerster, S., Waldmann, T., Westerhout, J., Julien, S., Krause, K.H., van Thriel, C., Hengstler, J.G., Sachinidis, A., Bosgra, S., Leist, M., 2014. Profiling of drugs and

environmental chemicals for functional impairment of neural crest migration in a novel stem cell-based test battery. Arch. Toxicol. 88, 1109–26. doi:10.1007/s00204-014-1231-9

List of Figures

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

List of Tables

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Annex 1 Good Cell Culture Practice (GCCP)

See pdf

Annex 2 Good Cell Culture Practice for stem cells and stem-cell-derived models

http://www.altex.ch/resources/Pamies_of_160823_v4.pdf

Field Code Changed

Annex 3 Experts participating at the GIVIMP meeting 24-25 February 2015

Maria Astridou	EFSA, Parma, Italy
Camilla Bernasconi	EURL ECVAM, Ispra, Italy
Gerard Bowe	EURL ECVAM, Ispra, Italy
Ann-Charlotte Bostroem	EURL ECVAM, Ispra, Italy
Sandra Coecke	EURL ECVAM, Ispra, Italy
Thomas Cole	EURL ECVAM, Ispra, Italy
Rita Cortvrindt	Vito, Mol, Belgium
Laureano Cuevas	Centro Nacional de Microbiologia, Majadahonda, Spain
Agnieszka Dudra	Bureau for Chemical Substances, Lodz, Poland
Eugene Elmore	University of California Irvine, Trabuco Canyon, Unites States
Chantra Eskes	SeCAM, Magliaso, Switzerland
Varvara Gouliarmou	EURL ECVAM, Ispra, Italy
Andrew Gray	United Kingdom GLP Monitoring Authority, The MHRA Medicines and Healthcare, London, United Kingdom
John Harbell	JHarbell Consulting LCC, Dallas, United States
Minne Heringa	RIVM, The National Institute for Public Health and the Environment, Centre for Safety of Substances and Products, Bilthoven, The Netherlands
Delyan Ivanov	University of Nottingham, Cancer Biology Queens Medical Centre, Nottingham, United Kingdom
Rob Jaspers	Health Care Inspectorate, The Netherlands
Ingrid Langezaal	EURL ECVAM, Ispra, Italy
Marcel Leist	University of Konstanz, Konstanz, Germany
Thomas Lucotte	ANSM, Saint-Denis, France
Marisa Meloni	Vitroscreen SRL, Milano, Italy
Maurits-Jan Prinz	European Commission, Directorate General Internal Market, Industry, Entrepreneurship and SMEs, Brussels, Belgium
Vittorio Reina,	EURL ECVAM, Ispra, Italy
Glyn N Stacey	United Kingdom Stem Cell Bank, Advanced Therapies Division, NIBSC, United Kingdom
Emanuela Testai	Istituto Superiore di Sanita, Rome, Italy
Frederique van Acker	TNO Triskelion, Zeist, The Netherlands
Iwona Wilk-Zasadna	Novartis Vaccines, Basel, Switzerland.
Maria-Chiara Zorzoli	EURL ECVAM, Ispra, Italy

Formatted: French (France)