TECHNICAL REPORT

ISO/TR 79

First edition
2015-04-15

Reference materials - Examples of reference materials for qualitative properties

Matériaux de référence - Exemples de matériaux de référence pour les propriétés qualitatives

Reference number
ISO/TR 79:2015(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/REMCO, *Committee on reference materials*.

Introduction

In 2007, ISO/REMCO created an ad hoc group (AHG) to investigate the need for guidance on the production of Reference Material (RM) certified for a qualitative property. AHG 01 carried out a gap analysis, contacting 12 organizations and bodies using qualitative RMs and reviewed 13 documents referring to qualitative RMs. Based on this gap analysis, ISO/REMCO decided in 2008 to create a working group (WG) and to entrust it with the drafting of an ISO document.

Due to the limited information submitted in the following years to WG13, the drafting of internationally harmonized guidance turned out to be impossible. Instead, it was decided to focus on an ISO Technical Report (TR), which summarizes the state of the art of the production of qualitative RMs. This TR lists examples of RMs which are either certified for a qualitative property or which can be considered as in-house RMs characterized for a qualitative property. Therefore, many of the RM examples listed here are based on the principles elaborated in ISO Guide 35[1] and ISO Guide 80.[2] The examples represent the experience gathered by various organizations and bodies and their interpretation of qualitative properties, but did not undergo a consensus building process.

In this TR, the following six RM examples are presented:

a) the certification of RMs for their DNA sequences by an ISO Guide 34[3] accredited reference material producer (RMP) (Clause 2);

b) the in-house characterization of organic chemicals as RMs for identification purposes by a laboratory (Clause 3);

c) the identification of a RM biospecimen by an ISO/IEC 17020[4] accredited tissue bank (Clause 4);

d) the development of a reference material for darnel seed identity (Clause 5);

e) the classification and between-sample homogeneity testing of a freshwater cultured pearl (Clause 6);

f) European Pharmacopoeia reference standards for qualitative analysis (Clause 7).

The lack of international standardization in the area of qualitative properties has been recognized by several groups. This includes WG 2 of the Joint Committee for Guides in Metrology (JCGM), officially responsible for the International Vocabulary of Metrology (VIM),[5] which investigates updating and expanding the VIM to cover also qualitative properties. As these discussions are on-going, the terminology used in the various examples presented in this TR may differ, e.g. some groups refer to qualitative properties as nominal properties. Likewise, no agreement has yet been made on international level if the term measurement is limited to quantitative properties or may as well be used for qualitative properties. To foster the readability of this TR, the term *qualitative property* has been given preference and the term *measurement* has been restricted to its use in conjunction with quantitative properties, following the recommendations expressed by the majority of ISO/REMCO delegates during their 37th annual meeting in 2014.

Due to the lack of common guidance on the production of RMs for qualitative properties, the approaches and understanding of terms properly defined for quantitative properties (e.g. homogeneity and traceability) are differently interpreted and applied for qualitative properties by the various organizations and bodies which contributed to this collection of examples. Likewise, the border between qualitative and quantitative properties is differently interpreted. Ordinal properties are perceived by some groups to be restricted to quantitative properties, while others suggest distinguishing between quantitative and qualitative order.

As the predominant aim of this TR is to contribute to the on-going discussion, these differences were on purpose maintained.

During the writing of ISO/TR 79, ISO/REMCO WG 13 identified a number of discussion items, which could not yet be answered with consensus, but which are considered to be crucial in case further efforts will be made to transform this TR into a Guide.

— The expression of confidence related to identification is discussed and in the majority of the cases, no uncertainty is estimated, although experts agree that the probability for a wrong identification forms also part of the result. The identity of an object does not have an uncertainty; however, the assessment of the identity of an object is related to the possibility for misclassification. Ways to estimate the uncertainty of qualitative analysis are especially suggested in the area of DNA sequencing.[6] At the same time, several areas require an assessment uncertainty equaling zero, like e.g. the classification of blood group values.[7]

— Forward/backward DNA sequencing is considered by many experts as an orthogonal method or method free of parameters influencing the result. At the same time, the question is asked what makes DNA sequencing specific.

— Heterogeneity of materials used as RM for qualitative property identification does not necessarily ruin the intended use. Ways are needed to check to which extend for instance inhomogeneity can be accepted.

— A working group at AOAC International developed internationally harmonized guidelines for the validation of qualitative binary chemistry methods.[8] The Guidelines for Validation of Qualitative Binary Chemistry Methods approach the question from the view point of the method. The question whether the RM used in presence/absence testing needs to be certified for a quantitative or qualitative property has not been discussed in this working group so far.

Reference materials - Examples of reference materials for qualitative properties

1 Scope

This Technical Report summarizes the state of the art of the production and certification or characterization of qualitative property reference materials (RMs).

The need for guidance documents for the production of RMs certified for qualitative properties was recognized by many experts. At the same time, the available information was found to be too immature to develop an internationally accepted guidance document. Additionally, the lack of an international vocabulary for terms and definitions for qualitative properties made it more difficult for the experts from various testing areas to communicate with each other.

ISO/TR 79 summarizes the available expertise. It aims to contribute to the on-going discussion on nominal properties and the production of such RMs. The investigation of nominal properties is referred to differently in various specialized areas (examination, classification, identification, testing, observation, etc.). ISO/TR 79 tries to foster the future development of an internationally harmonized guidance document.

2 Reference materials certified for their DNA sequence

2.1 General

The following is a compilation of the certification approaches applied for three reference materials which were certified for their DNA sequence by the Joint Research Centre of the European Commission, Institute for Reference Materials and Reference Materials (IRMM, Geel, BE).

2.2 Selected examples of certified reference materials

CRM ERM-AD427[9] is composed of plasmid DNA certified to contain certain DNA fragments. It is used for the quantification of Genetically Modified Organisms (GMOs) and the calibration of a defined quantitative Polymerase Chain Reaction (PCR) method. The Certified Reference Material (CRM) contains a plasmid carrying two defined 2'-deoxyribonucleic acid (DNA) fragments. The plasmid calibrant is certified by DNA sequencing for containing two specific DNA targets per plasmid. The number ratio between the two targets is equal to 1, allowing the use as calibrant for relative real-time PCR measurements. The DNA sequence identity has been confirmed by dye terminator cycle sequencing with a negligible error probability for the sequence identification.

CRM IRMM-448[10] is composed of genomic DNA extracted from a microorganism and certified for its DNA identity (with the PCR region of interest verified by DNA sequencing). IRMM-448 is used as positive control in a defined qualitative PCR method for food testing. The CRM consists of a purified and freeze-dried genomic DNA (gDNA) of *Campylobacter jejuni* (NCTC11351). The identity of the gDNA was confirmed by DNA sequence analysis of the ceuE gene, supporting the harmonization and validation of PCR methods by their use as taxonomic controls in PCR reactions. An indicative value for the mass of freeze-dried gDNA is given.

CRM IRMM/IFCC-490[11] is composed of plasmid DNA certified for its DNA sequence (whole sequence). IRMM/IFCC-490 is intended to be used as positive control in quantitative PCR in the area of genetic testing. The CRM consists of purified plasmid DNA (pDNA) pUC18 containing a specific fragment of the human Factor II (prothrombin) gene sequence. It is intended to support the validation and the harmonization of PCR-based methods used for the detection of the mutation in the human prothrombin gene. In all cases, PCR amplification is followed either by restriction enzyme digestion, hybridization

protocols, single-strand conformation polymorphism analysis, melting curve analysis, denaturating gradient gel analysis or sequencing.

2.3 Certification approaches applied

2.3.1 Processing of the materials

2.3.1.1 Processing can be carried out fulfilling the requirements laid down in ISO Guide 34.[3] Detailed knowledge and understanding of the individual processing steps are for the production of qualitative RMs of outmost importance as the knowledge about the raw material used is one way to ensure the desired identity. Likewise, meaningful processing controls are important.

An example is the use of extracted genomic DNA for the RM preparation. Genomic DNA from the desired microorganism strain can be obtained from culture collection centres together with a certificate of analysis confirming the identity of the material. The issuing of the certificate for genomic DNA of a particular microorganism strain is typically based on biochemical-, morphological- and microbiological information. Likewise, this approach of controlled origin can be applied for DNA fragments which will be selected for cloning.

2.3.1.2 Verification of the desired DNA sequence of fragments includes techniques such as:

a) purification of DNA fragment prior to ligation of the amplification product with e.g. PCR fragment purification kit;

b) control of the correct length of the DNA amplicons by agarose gel electrophoresis by comparison to a DNA molecular mass ladder;

c) melting temperature investigation of the qPCR products obtained with defined primers and probes to further confirm the identity of the DNA sequence;

d) assessment of the DNA sequence identity by DNA sequence analysis of the fragment inserted into the vector.

2.3.1.3 For cloning, the conditions should be selected in such a way that the cloning of other fragments than the desired one is prohibited. The following should be considered.

a) The synthetic vectors used for cloning should be a high copy vector from the same incompatibility group. As a result, the transformed bacterial clones can only bear one single plasmid and not a mixture of different plasmids.

b) By repeated plating, it should be ensured that only a single colony is cultivated.

2.3.1.4 The outcome of cloning can be checked by:

a) endonuclease restriction analysis to control correct cloning and to check the orientation of the DNA insert;

b) testing of the resulting plasmid by qualitative PCR for the presence of the insert;

c) assessment of the DNA sequence identity by DNA sequence analysis of either the whole DNA sequence or the DNA fragment of interest.

2.3.1.5 In the case of different DNA fragments inserted into a plasmid in a desired ratio, methods should be employed to verify this ratio.

a) Digital PCR experiments can confirm the expected ratio between two target sequences.

b) Real-time PCR method targeting one of the fragments and using the second fragment as normalizer can be applied provided that amplification efficiencies are equivalent.

2.3.1.6 The DNA sequence identity needs to be confirmed. This confirmation can either concern the whole DNA sequence (e.g. in case of a plasmid) or one or more parts of the DNA (e.g. the PCR target region in genomic DNA from microorganisms used for identification of strains or different DNA fragments inserted into a plasmid in a desired ratio).

Part of the identity confirmation is carried out during the processing (2.3.1 *Processing of the materials*). Depending on the processing steps, repeated application of confirmation method should be considered (e.g. DNA sequencing at various crucial processing steps and if possible in the final preparation).

To ensure that DNA sequencing data have a low probability of incorrect reads, forward and backwards sequencing (Sanger-based sequencing) should be applied. The probability for incorrect reads can be estimated according to Reference [6].

2.3.1.7 In any case, the following should be taken into account with respect to homology searches of the found DNA sequence:

a) if published sequence data were judged to be trustworthy, a homology search of the found DNA sequence (either whole DNA sequence or DNA fragment of relevance) should be applied;

b) a comparison of the found DNA sequence with e.g. data provided by a culture collection centre should be made.

2.3.2 Purity assessment

2.3.2.1 First step

As a first step in the DNA purity investigation, a list of potential DNA contaminants is made. Their impact on the later use of the CRM is evaluated and ways to check for impacting DNA contaminants are elaborated.

2.3.2.2 Plasmid-based CRMs

2.3.2.2.1 Remaining traces of genomic DNA from host bacterial cell or traces of RNA molecules:

It should be checked if such traces would affect the identity of the plasmid and its suitability for real-time PCR measurements, for instance if a high homology is given. If the sequence identity of the genomic DNA of the bacteria and the DNA fragments of interest is low, it is reasonable to conclude that they would not have an impact. However, as remaining traces may induce a bias in the UV absorbance-based DNA quantification of the plasmid solution, the user should be made aware that this could lead to erroneous estimation of the absolute number of plasmid copies.

2.3.2.2.2 Presence of other plasmids than the one containing the DNA fragment(s) of interest:

a) In order to prevent that other plasmids than the one containing the DNA fragment(s) of interest are present, the synthetic vectors used for cloning should be high copy vectors from the same incompatibility group. As a result, the transformed bacterial clones can only bear one single plasmid. It can therefore be concluded that each single bacterium extracted from one colony contains only one type of plasmid.

b) Presence of other plasmids in higher abundance can be tested by enzymatic restriction, with conditions allowing a full digestion. However, one has to bear in mind that it is very difficult to prove that all plasmids were indeed fully digested, as traces of undigested plasmids will not be visible after gel electrophoresis and ethidium bromide staining.

c) As additional proof of purity, the plasmid DNA can be sequenced completely to verify that the desired sequence was correctly cloned. To confirm the purity, the results should not reveal the presence of a mixed population of plasmids. The sequencing technique should be carefully selected in view of the question to be answered; nowadays, a Next Generation Sequencing (NGS) analysis could be applied, provided that settings are chosen in such way that disturbing impurities are detected.

d) End-point PCR followed by agarose gel electrophoresis can be used to further investigate the purity. The qualitative PCR should be targeted at the DNA sequence(s) which is (are) considered to be either a problem for the later use of the CRM and/or at the DNA sequence which could be expected as contaminant. However, the ability of this method to investigate the purity of the material is limited and often only a purity of 90 % can be proven. Likewise, digital end-point PCR can be used to check for contamination.

2.3.3 Genomic DNA CRMs

Purity of the gDNA from one cell culture:

a) In order to ensure that the gDNA has the desired identity, the origin of the material should be controlled and documented as well as possible. Genomic DNA from the desired microorganism strain should be obtained from culture collection centres together with a certificate of analysis confirming the identity of the material. The issuing of the certificate for genomic DNA of a particular microorganism strain is typically based on biochemical-, morphological- and microbiological information.

b) As an additional proof of purity, the gDNA could be sequenced. This could be done on the whole genome (if feasible) or limited to part of the sequence of interest (i.e. the target region of a PCR).

2.3.4 Characterization (leading to certified values)

Characterization and value assignment for qualitative DNA CRMs rely to a large extent on proper control of processing and the outcome of the purity assessment. In several cases, the characterization could be classified as confirmation, which confirm that the intended material was processed.

Two possible confirmation techniques are as follows.

a) Control of the restriction pattern of the plasmid DNA samples can be done by gel electrophoresis. Gel electrophoresis can be carried out on the gDNA or on the PCR amplicons generated after performance of a specific PCR.

b) The whole DNA should preferably be sequenced by two independent laboratories/instruments or protocols, using preferably forward and backward sequencing (Sanger-based sequencing). Double stranded DNA should be sequenced on both strands in order to ensure an accurate determination by a twofold-coverage of the generated sequences, as well as a complete characterization of the molecular composition of the double stranded DNA.

In the case that a whole sequencing is not possible or considered not to be of interest, sequencing can e.g. be limited to amplicons generated after performance of a specific PCR. The amplicons need to be purified and cloned into a plasmid, which after purification can be sequenced.

The uncertainty related to the sequence determination by Sanger-based sequencing techniques can often be considered as negligible, as the probability to report a wrong base (error probability) is very low. For further guidance, see Reference [6].

2.3.5 Additional characterization (leading to non-certified values)

In various cases, the DNA amount is useful additional material information. Quantification can be done by fluorometry. However, in the absence of reference materials certified for their DNA amount, the resulting value is an estimation of the DNA quantity.

2.3.6 Suitability study

A suitability study is carried out to ensure that the material is fit for purpose and can be used for its intended use. Intrinsically, the suitability study should be designed in such a way that existing commutability problems would be discovered.

As the examples used in this document concern CRMs certified for their sequence and intended to be used for PCR measurements, the intended use for PCR is elaborated here.

The correct behaviour of a qualitative CRM for its intended purpose (i.e. as PCR calibrant or positive control in PCR) needs to be verified. Ideally, the performance can be tested by laboratories which are potential (later) users of the CRM.

If the type of DNA from analysed samples and calibrants differ (i.e. gDNA and plasmids), commutability problems should be investigated. The calibrant should behave in the analytical process in the same way as the material under investigation.

Suitability of the CRM might not only depend on certified parameters (DNA sequence) but also on non-certified parameters (i.e. the DNA amount). This should be taken into account in the suitability study as well as in the homogeneity study.

Positive outcome of a suitability study proves that the material is fit for its intended purpose.

2.3.7 Homogeneity study

A CRM certified for its DNA sequence is certified for identity. The sequence identify defines the structure. The homogeneity of the material is therefore determined by the purity of the material.

However, additional homogeneity studies should verify the homogeneity in view of the intended use of the CRM.

a) This could e.g. concern the DNA amount, justifying to measure the mass of double stranded DNA in a representative number of vials by fluorescence. The obtained data can be evaluated like a homogeneity study (e.g. as outlined in ISO Guide 35[1]), proving that the mass of DNA in each vial is not significant different from each other (95 % confidence level).

b) For gDNA, the intactness of the DNA might be of relevance and a number of vials could be analysed by gel electrophoresis. A single band of DNA confirms a reasonable intactness of the gDNA in each vial.

c) PCR amplification in different vials should be checked, when the CRM is intended for PCR. For real-time PCR, this could include the monitoring of the crossing point threshold (Cp) and/or the melting temperature. This would contribute to the confirmation of the sequence identity (within the DNA sequence defined by the primers).

Uncertainties related to homogeneity studies should not be above a level at which the random selection of a vial would have an impact on the performance. A CRM certified for its DNA sequence and intended to be used as positive control in PCR applications needs, e.g. to contain in each unit enough of the DNA so that the PCR result is not negative. Impurities may influence the suitability of the material if they interfere with the PCR but are not relevant if they e.g. concern not targeted DNA and/or background DNA added to enhance the stability of the targeted DNA in the CRM.

2.3.8 Minimum sample intake

A CRM certified for its DNA sequence is often a pure material. Therefore, classical studies designed to investigate the minimum sample intake[12] are not applicable.

Instead the suitability and homogeneity studies, in which the CRM was used for its intended purpose, can be used to demonstrate that a certain amount of the CRM is sufficient to obtain the expected results. A minimum sample intake for a given purpose can be recommended. The establishment of a minimum sample intake in this case often resembles the typical sample intake area, rather than the minimum (possible) sample intake.

2.3.9 Short-term and long-term stability

A CRM can be certified for its DNA sequence. The sequence identify is unlikely to change, even DNA degradation would strictly speaking not impact the identity.

However, during short-term stability studies the transport stability and during long-term stability studies the to-be expected storage stability in view of the intended use of the CRM should be investigated. Based

on the outcome, transport conditions and shelf-lives for specified storage conditions, not influencing the intended use should be selected.

Uncertainties related to stability studies should ensure that the material can still be used without impact on the performance after a certain time and/or exposure to specified temperatures.

2.3.10 Identification probability

No uncertainty concept exits for identity. The probability to misidentify for instance the DNA sequence of a material is often very low. To demonstrate this, the probability should be calculated that two sequencing procedures, independently performed in different laboratories, generated identical results by forward and backward sequencing.

However, one has to be aware that the DNA sequence information is used to 'identify' the material. This identification can be based on established classification schemes, which in itself can be subject to changes. It therefore should be mentioned which classification scheme has been applied at which time.

The probability for wrong identification in some specific areas can be rather high. e.g. if the DNA sequences are very similar, if crossings occur and if the availability of DNA sequence information is very limited. However, it should be carefully discriminated between the probability of a wrong outcome of DNA sequencing and the probability of a wrong identification because of a classification scheme which was set up on limited DNA information.

2.3.11 Documentary and metrological traceability

The certified identity is often based on a documentary traceability to a document linking the DNA sequence to a certain identity. Additionally, a metrological traceability exists for the confirmation step (such as dye terminator cycle sequencing or PCR measurements) carried out on the DNA sequence.

Beside the metrological traceability, the traceability (trackability) to a classification scheme is of interest.

3 Organic reference materials for qualitative analysis

3.1 General

Chemical measurement, whether it is quantitative or qualitative, requires reference materials for which the analyte of interest has been correctly identified. Laboratories are encouraged to purchase appropriately certified reference materials from an accredited reference material producer (RMP), providing confidence that the identity of their product has been correctly assigned. Due to the lack of reference materials certified for identity, laboratories are often forced to purchase chemicals from reputable chemical suppliers or simply use a sample "found on the shelf" and put their faith in the manufacturer's labelling and/or associated paperwork. Regardless of the source of the material, it would be considered prudent to perform in-house cross checks before embarking upon a testing program with a potentially incorrectly labelled reference material. Most modern day analytical laboratories will have access to mass selective detectors, linked to either liquid chromatography (LC-MS) or gas chromatography (GC-MS) equipment, which provide a fingerprint of the compound in hand. Specialized laboratories may also have access to infrared (IR) spectrometers, which can provide further structural information. Demonstrated equivalence with literature spectra is generally considered sufficient for structural confirmation. In cases where literature data are limited or non-existent, the analyst would need to exercise a lot more care when interpreting the structural information. The analyst also needs to appreciate the limitations of these techniques, particularly when considerations such as regio- and stereochemistry are important. In such cases, nuclear magnetic resonance (NMR) spectroscopy is recommended, as this technique provides a definitive structural analysis. On the downside, this introduces the need to access an NMR spectrometer and suitably qualified staff to interpret the spectra, which may be difficult for many laboratories. Fortunately, many university chemistry departments now offer this service.

3.2 Verification of the identification of a marker for steroid abuse

The following example serves to warn laboratories of the dangers of purchasing chemicals from commercial suppliers for use as a reference material (RM) for qualitative analysis, without performing adequate cross checks to confirm the identity.

7 β-Hydroxydihydroepiandrosterone (7β -hydroxyDHEA) **1**, a marker for steroid abuse, is required by the sports doping control community. Being a human metabolite of DHEA, sufficient quantities of the 7 β-hydroxy metabolite can only be realized through synthesis. Having exhausted the original batch, National Measurement Institute Australia purchased 7 β-hydroxyDHEA from a commercial company specializing in the provision of steroids. Unfortunately, the compound received was not the desired 7 β-hydroxyDHEA, but a compound of similar structure. Suspicions were first raised when the solubility characteristics of the supplied material did not match that exhibited by the originally certified material. IR analysis, while confirming the presence of a ketone group (strong peak at 1 734 cm^{-1}) and at least one hydroxyl group (broad peak at 3 283 cm^{-1}), did not match the IR spectrum of the certified sample which displayed a strong carbonyl peak at 1 726 cm^{-1} and a relatively narrow hydroxyl band at 3 447 to 3 474 cm^{-1}. A GC-MS based co-elution study (Figure 1) confirmed beyond all doubt that the purchased sample had been incorrectly assigned, although it is worth noting that separation of the two compounds required significant optimization of the chromatography conditions.

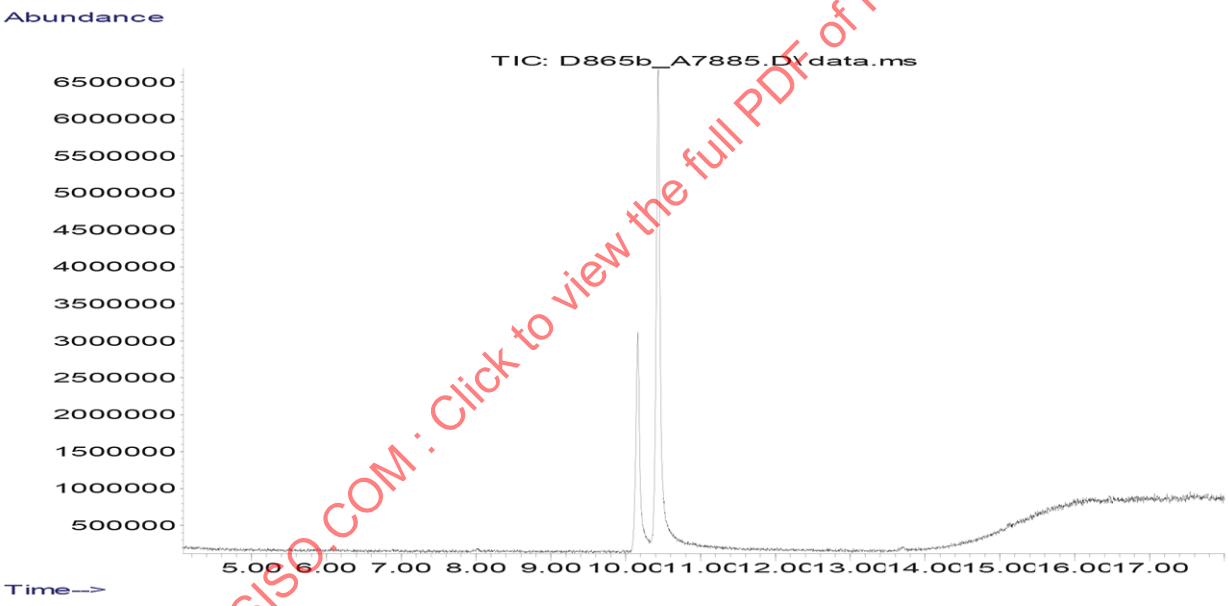


Figure 1 — GC-MS based co-elution study of a commercial unknown (r.t. = 10,1 min) and 7 β-hydroxyDHEA (r.t. = 10,4 min)

The respective mass spectra (Figure 2) of the two compounds are clearly different, although key peaks are common to each. Both compounds afford a weak molecular ion (304 *m/z*), and a base peak at 286 *m/z* representing the loss of water, while peaks at 271 *m/z* and 268 *m/z* represent the loss of a methyl group and a second water molecule respectively, in line with the proposed structure. A similar scenario was observed for the tris-TMS derivative of each compound. The observation of a vanishingly small molecular ion at 520 *m/z* in each case, and key fragments common to both compounds, provided further evidence that the requisite functionality was present on both steroids. The possibility that 7α-hydroxyDHEA had been provided was ruled out, again by direct comparison with a certified reference material which displayed almost identical IR and MS spectra to the desired 7 β diastereomer.

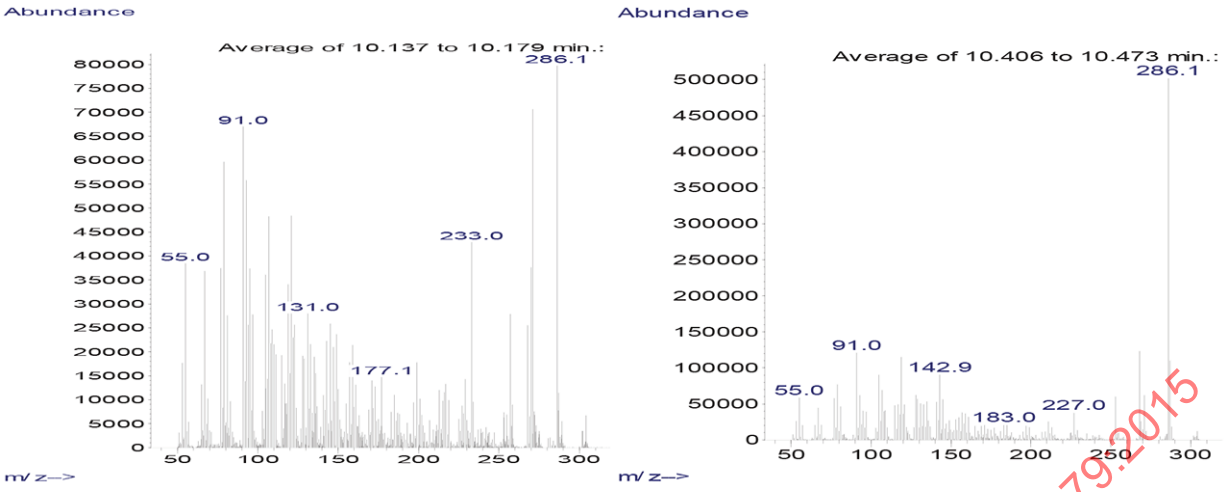


Figure 2 — Electron ionization mass spectra of the unknown (r.t. = 10,1 min) and 7 β-hydroxyDHEA (r.t. = 10,4 min)

^{1}H NMR analysis (Figure 3) confirmed beyond all doubt that the commercial material was not 7 β-hydroxyDHEA, and proved necessary to convince the manufacturer that they had indeed provided an incorrectly assigned material. Comparison with certified samples of DHEA (Figure 3 top) and 7 β-hydroxyDHEA (Figure 3 middle) clearly shows that while the commercial material (Figure 3 bottom) has a related structure, the absence of the H3 α proton is beyond question. For the original batch of certified 7b-hydroxyDHEA, the structure was confirmed by direct comparison with a certified sample of DHEA. The multiplicity of H3α of DHEA at 3,5 ppm is retained in 7 β-hydroxyDHEA, confirming retention of stereochemistry at C3. The introduction of the second hydroxyl group has the effect of changing the multiplicity of the olefinic H6 proton from a doublet in DHEA to a singlet, confirming substitution of one of the H7 methylene protons and the assigned regiochemistry. Consideration of the coupling constant (J = 8,5 Hz) of the H7 α doublet (pseudo axial) and dihedral angles to H8 (axial) confirm the β stereochemistry of the hydroxyl group. Supporting evidence came from direct comparison with 3 β-acetoxy-17, 17-ethylenedioxyandrost-5-en-7 β-ol **2** and 3 β-acetoxy-androst-5-en-7 β-ol-17-one **3** reported in the literature.[13]

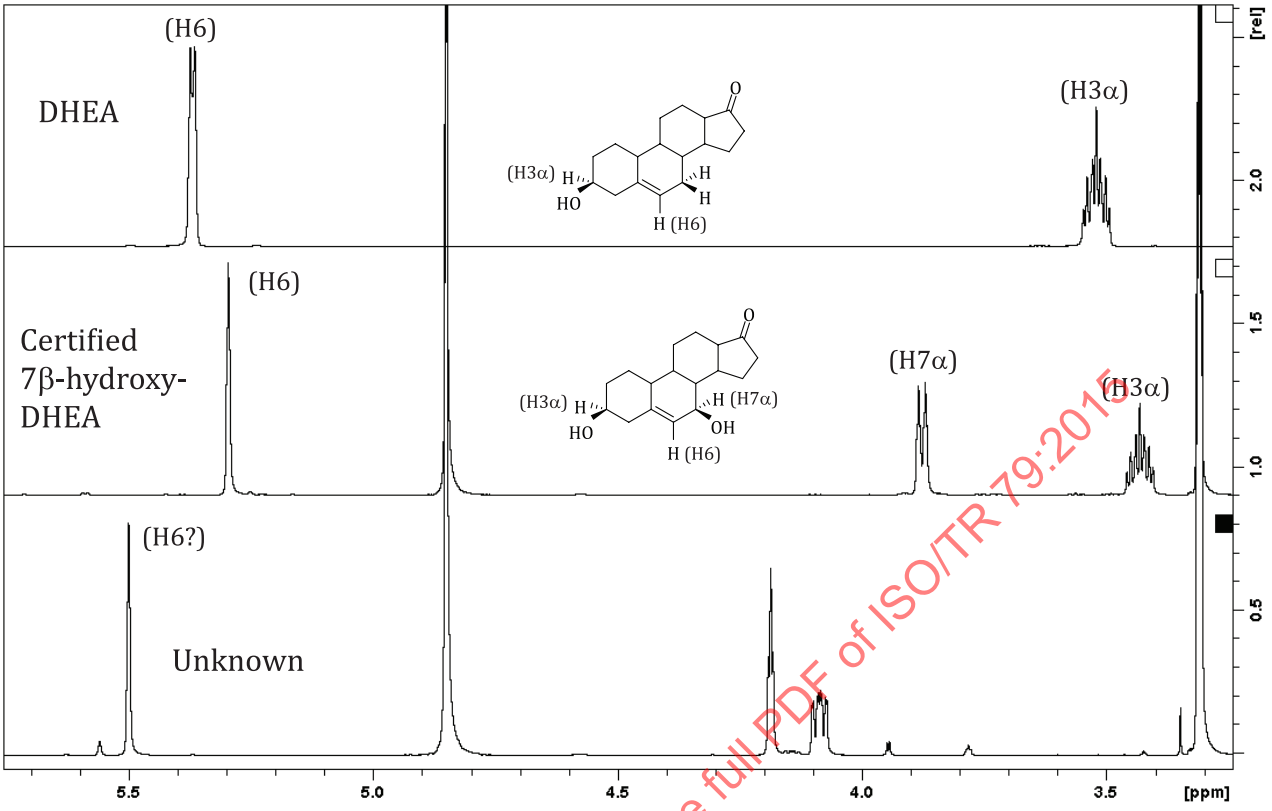
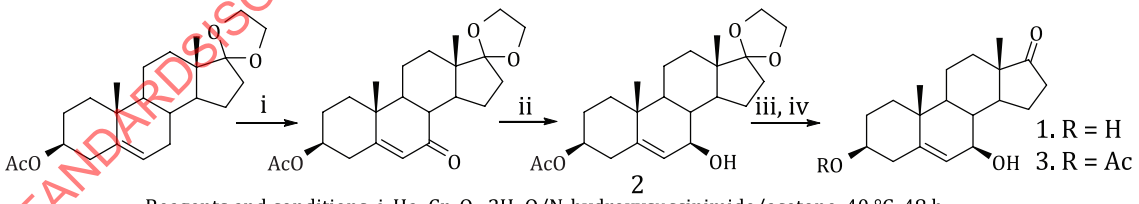


Figure 3 — ^{1}H NMR of DHEA (top), 7 β-hydroxyDHEA (middle), and the unknown (bottom)

3.3 The added value of knowing the synthetic procedure

To a large extent, this example also demonstrates the value of having knowledge of the synthetic procedure employed to afford the compound in hand. The originally certified material was synthesized at a local university chemistry department following literature precedents,[13][14][15][16] ensuring key chemical transformations such as the regiospecific oxidation at C7 and subsequent stereoselective reduction of the resulting ketone, yielded the desired compound (Figure 4).



Reagents and conditions: i. $Ha_2Cr_2O_7.2H_2O$/N-hydroxysuccinimide/acetone, 40 °C, 48 h;
ii. $NaBH_4/CeCl_3.7H_2O$/MeOH/THF, R.T., 1 h; iii. TsOH/acetone/water, iv. 15%KOH/MeOH

Figure 4 — Literature synthesis of 7 β-hydroxyDHEA

In summary, laboratories using commercial, or "off the shelf", samples for use in their qualitative analysis will need to make a judgment call as to what level of structural information is required. Demonstrated co-elution using a suitable chromatographic technique, with an authentic sample, preferably certified by an accredited body, and identical spectroscopic properties (usually mass) will generally be considered sufficient to confirm the identity of the sample in hand. In the event that a certified authentic sample is not available, comparison of spectroscopic data with literature precedents will be required. When using this approach, the analyst will need to take into consideration variation in spectroscopic data arising from instrument variation, sample quality and concentration. Furthermore, these techniques can be limited in situations in which the analyst needs to distinguish more closely related compounds. For example, it is highly unlikely that MS and IR spectroscopy would be able to distinguish the 7α -

and 7 β-hydroxy diastereomers when used in isolation. In the example presented here, both MS and IR spectroscopy were sufficient to distinguish the commercial sample from a certified reference material of 7 β-hydroxyDHEA. In such cases where regio- and stereochemistry are important factors in the structural assignment, ^{1}H NMR should be the method of choice. Caution also needs to be exercised when interpreting spectroscopic data from first principles in the absence of literature precedents. Both the MS and IR spectra of the unknown commercial sample displayed characteristics attributable to 7 β-hydroxyDHEA, which in the absence of other information could easily be interpreted to "fit" the proposed structure, leading to an incorrect assignment. The same could also be said for the ^{1}H NMR analysis, although direct comparison with compounds of similar structure immediately identified a problem in this case.

4 Biospecimens of human origin, certified for qualitative properties

4.1 General

Biospecimens play a critical role in understanding and improving human health conditions. These specimens are acquired, validated, processed, stored and distributed to academic or private research end-users by professional biobanks.[17] Identification and validation of clinically relevant biomarkers, including potential prognostic, predictive and diagnostic markers, and drug targets, is based on analysis of biospecimens, such as serum, urine, solid tissue or cell lines.

Biospecimens must have appropriate annotations to ensure their usefulness and the reliability of research results based on their analysis, especially when it comes to clinical validation. These annotations contain patient-related clinical / biological and specimen-related processing information, also referred to as preanalytical information. Relevant annotations often concern paired biospecimen-related information, obtained on a different type of specimen than the one being certified for a nominal value. Recently, efforts were made to define which preanalytical variables are essential and should be attached to certain types of collected biospecimens.[18][19][20] Accurate clinical annotation and/or accurate biological or anatomopathological annotations, as well as precise pre-analytical records are required.

4.2 Selected human biospecimen examples of reference materials

A. Acute *Chlamydia trachomatis* infection serum

The bacterial sexually transmitted infection (STI) pathogen *C. trachomatis* can cause serious reproductive complications such as pelvic inflammatory disease, ectopic pregnancies and tubal infertility. Serological diagnostic tests are important in STI screening programs and infertility clinics. Assessment of performance of a serological assay, for acute infection diagnosis, is based on analysis of certified serum samples, collected from patients with acute infection.

B. Lung adenocarcinoma frozen tissue-extracted DNA

Lung cancer is the leading cause of cancer-related mortality worldwide. Comprehensive genomic analysis of lung adenocarcinoma samples is expected to facilitate identification of new therapeutic targets. Tumor whole genome sequencing, study of mutation patterns and tumour DNA rearrangements is based on analysis of certified DNA samples, collected from tumours from patients with adenocarcinoma.

C. Parkinson's disease urine

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disease of the central nervous system (CNS), but also the enteric nervous system (ENS) in almost all cases, affecting mainly dopaminergic neurons. The two hallmarks in PD are: 1) the loss of dopaminergic neurons in the substantia nigra with loss of dopamine in the striatum; and 2) the presence of intracellular inclusion bodies in the soma (Lewy bodies) or in neuritis (Lewy neuritis). Diagnosis is based on clinical symptoms. Because PD diagnosis is hindered by heterogeneous clinical presentations, variable progression rates, and different clinical subtypes, it is often established rather late. Thus, there is a need for efficient tools to diagnose the disease earlier (before appearance of motor symptoms) and neuroprotectants. Diagnostic biomarkers would

decrease the cost, time, and effort it takes to ensure diagnosis and different types of biospecimens are studied for discovery of such biomarkers.

D. Pancreatic cancer cell line

Pancreatic adenocarcinoma (PA) is an aggressive disease with very poor prognosis. The poor prognosis of PA is due in part to a lack of molecular information on disease development. Pancreatic cancer cell lines are a useful tool to investigate the PA-underlying molecular events.

4.3 Certification approaches applied

4.3.1 General

In the analysis of the certification approach on the above examples of human biospecimen CRM, it is shown that the RM producer (RMP), a biobank in this specific case, can be certified for the core biobanking activities, and accredited for a number of assays, including immunological, molecular biology and pathology assays. Typical processing activities for biospecimen CRMs include e.g. centrifugation, extraction, purification and freeze-drying. These activities are carried out by the RMP, while the testing methods (e.g. nucleic acid quantification, purity investigations, antibody and protein identification and quantification) are often performed by a subcontractor.

4.3.2 Processing of the materials

A.C. For serum and urine specimens, the following pre-analytical variables should be documented: type of blood collection tube; pre-centrifugation delay and temperature; centrifugation conditions (speed, time, temperature, brake); post-centrifugation delay and temperature; storage container and temperature. Freeze–thaw cycles should be documented.

B. For snap frozen tumour tissue specimens, the following pre-analytical variables should be documented: type of collection (e.g. biopsy, surgery, autopsy); warm and cold ischemia times; storage container and temperature.[19][20]

D. The original tissue specimen from which the cell line has been derived should be documented with data on type of collection (e.g. biopsy, surgery, autopsy), warm and cold ischemia times, transport medium composition. The primary culture record should contain information on the method of tissue dissociation (trypsin, collagenase, primary explants culture), composition of culture medium, batch numbers, passages, seeding details (cell concentration, split ratio, volume per flask), composition of freezing medium, freezing protocol. The cell culture aliquots that are used for the different characterization assays described below should be traceable to the passage / flask ID. The DNA extraction for short tandem repeat (STR) profiling should be done on cells collected during their logarithmic growth phase and have > 90 % viability, as assessed by trypan blue staining or flow cytometry methods, in order to avoid DNA extraction from apoptotic cells.

4.3.3 Purity assessment

A. Serum samples are used for the development and validation of species specific serological diagnostic kits. The interpretation of results from serum antibodies is often difficult because of potential cross reactions against common chlamydial antigens, and because of widespread contact with chlamydial species, such as *Chlamydophila pneumonia*. Therefore, it is important to assess the purity of a serum sample relative to antibodies specific of other chlamydial species. For this purpose, serological assays based on recombinant antigens from the different species can be used.[21]

B. The degree of purity of tumour specimens corresponds to the percentage of tumour in the specimen. This percentage should be defined by consensus between three certified pathologists.

C. Absence of microbial contamination is assessed by culture and/or absence of leucocytes, nitrites and haemoglobin (Combur urine stick).

D. Purity of a cell line requires verification of the absence of contamination of other cell lines and/or the absence of microbial contaminants.

Absence of contamination by a different cell line can be assessed by STR profiling on 8 different STR loci and the amelogenin locus; this profiling can reveal differences of only one or two alleles[22]). Identification of the species of origin can be done by isoenzyme analysis, or species-specific antibodies,[23] if necessary (if multi-species cultures take place in the same biobank laboratory). Testing for absence of fungi, yeast and bacterial contamination is done by visual inspection, by culturing a sample of the cell medium in nutrient broth or agar, or by polymerase chain reaction (PCR) methods (16S, 18SrRNA genes). Mycoplasma contamination is assessed by Hoescht 33258 staining or by PCR (commercial kits exist). Viruses' detection is performed by PCR. The laboratory performing all these assays should be accredited.

4.3.4 Characterization

Gender, age and ethnicity of the donor should always be documented. Medication taken by the donor by the time of specimen collection should always be documented too.

A. In order to certify that a serum sample corresponds to acute infection, direct diagnosis of a paired urogenital specimen should be performed by an accredited laboratory. *C. trachomatis* can be directly diagnosed by culture, direct antigen detection tests or nucleic-acid based tests with or without amplification. Nucleic Acid Amplifications Tests (NAATs) have replaced culture as the reference method for the laboratory diagnosis of *C. trachomatis*. Amplification and detection of nucleic acids from *C. trachomatis* in urine specimens or in urethral /cervical swabs, based on technologies such as PCR, ligase chain reaction, strand displacement amplification or transcription mediated amplification can be used for nominal value assignment as "acute infection" to a urogenital, and by extension, to a paired serum specimen. Stringent negative controls should be included in the NAATs in order to avoid false-positive results. Positive results by one NAAT should be confirmed using another NAAT. Alternatively, the same NAAT targeting another gene can be used.[24] If only one NAAT is available, and has *x* % specificity, then the uncertainty of the material characterization will be equal to (100 - *x*)%.

Extraction of circulating DNA from the serum sample can be performed in order to confirm the correspondence of the serum to the patient by DNA fingerprinting on the serum- and the paired urine-extracted DNA, by a DNA fingerprinting method.[22][25]

Screening of the serum for HIV, HBV, HCV, syphilis and parvovirus B19 should be performed, for biosafety reasons and the related records should be available.

B. Characterization/classification has historically been based on histological and immunohistochemical analysis of tumour tissues. Now, molecular characterization of tumour samples is also required. Lung adenocarcinoma, for example, can be subdivided into several subtypes, based on the presence of *EGFR, KRAS, BRAF* and *HER2* mutations and *ALK* translocations. In order to certify a tissue sample as an adenocarcinoma, clinical indications and peripheral localization of the tumour will be documented, but the characterization is essentially histological. The histological definition of an adenocarcinoma is "a malignant epithelial tumor with glandular differentiation or mucin production, showing acinar, papillary, bronchioloalveolar or solid with mucin growth patterns or a mixture of these patterns"[26]. The anatomic location of the tumour (right or left lung; superior, middle or inferior lobe; segment) will be documented after macroscopic examination. Microscopic examination of a mirror FFPE (formalin fixed paraffin embedded) sample allows a certified and experienced pathologist to classify the tumour as an adenocarcinoma (WHO 8140/3) and furthermore, to define the exact histological type

of adenocarcinoma. The histological classification of adenocarcinomas, according to WHO[26] is the following (ICD-O codes should be used):

Adenocarcinoma /8140/3

Adenocarcinoma, mixed subtype /8255/3

Acinar adenocarcinoma /8550/3

Papillary adenocarcinoma /8260/3

Bronchioalvelolar carcinoma /8250/3

Nonmucinous /8252/3

Mucinous /8253/3

Mixed nonmucinous and mucinous or indeterminate /n8254/3

Solid adenocarcinoma with mucin production /8230/3

Fetal adenocarcinoma /8333/3

Mucinous (colloid) carcinoma /8480/3

Mucinous cystadenocarcinoma /8470/3

Signet ring adenocarcinoma /8490/3

Clear cell adenocarcinoma /8310/3

The adenocarcinoma should be staged according to the international TNM system for pathological (pTNM) and clinical (cTNM) classification.[27][28] Histological grading of pulmonary adenocarcinomas is based on conventional histological criteria and three grades are used: well (grade 1); moderate (grade 2); and poorly differentiated (grade 3) tumours.

Complementary immunohistochemical analysis, on a mirror FFPE sample, for the following antigens should be performed (standard profile is indicated in parentheses, although profile depends on the exact type of adenocarcinoma): CK7 (+); TTF1 (+); **napsin A(+)**; CK5/6 (-); **p40 (-)**; p63 (-); chromograninA (-); CD56 (-); synaptophysin (-). The antigens in bold correspond to the minimum panel to be tested.

Analysis and documentation of *EGFR, K-ras* oncogene, *p53* gene and *c-erbB2* oncogene mutations can complement the molecular characterization of a pulmonary adenocarcinoma.

Cytogenetic analysis (the mean chromosome number is near the triploid range[29] and CGH analysis can also complement the characterization of the adenocarcinoma sample (frequently observed imbalances are deletions on chromosomes 3p, 4q, 5q, 6q, 8p, 9, 13q, and gains on 5p, 8q, 20q).

The DNA sample, extracted from the adenocarcinoma tissue, will be characterized by quantification by either spectrophotometry or spectrofluorimetry. Its purity (absence of protein contamination) will be assessed by the spectrophotometric ratio A260/A280 (expected to be > 1,8). Its integrity will be assessed by agarose gel electrophoresis (expected molecular weight > 20 kb). The extraction method will be mentioned and the composition of elution buffer will be specified. The correspondence of the DNA to the patient of origin can be confirmed by DNA fingerprinting comparison between the adenocarcinoma tissue–extracted and paired whole blood-extracted DNA.[22]

C. Diagnosis of PD is only based on clinical symptoms (UK Brain Bank criteria) and these should be documented in a clinical report form (CRF). The three main symptoms are akinesia (slow and difficulty with movements), muscular hypertonia (rigidity, known as “extrapyramidal”) and tremors (regularly in 4 to 7 cycles per second when still). Moreover, up to 83 % of PD patients develop dementia during the course of their disease. In addition, other symptoms such as olfactory dysfunction, pain, cramps, constipation, pins and needles, orthostatic hypotension, frequent urination, visual dysfunction,

sweating, excessive production of saliva, anxiety, REM sleep behaviour disorder (RBD), depression, irritability and manias can persist.[30]

Tomodensitrometry (TDM) and magnetic resonance imaging (MRI) are not used systematically to diagnose idiopathic PD when the clinical picture is typical. Likewise, olfactory tests and transcranial sonography (TCS) are not used routinely in diagnosis. However, when the clinical picture is unclear, it is necessary to eliminate other causes of parkinsonian syndrome (iatrogene, vascular, toxic causes,...). For this purpose, complementary examinations can be applied, but these are not systematic (DaTSCAN to differentiate essential tremor from PD, TDM or MRI when vascular parkinsonian syndromes are suspected for example).

To determine the stage of the disease and to follow its course, neurologists use a rating scale. The most commonly used scale is the Unified Parkinson's Disease Rating Scale (UPDRS). Scoring to the first 4 sections of UPDRS should be documented:

— Part I: evaluation of mentation, behaviour, and mood;

— Part II: self-evaluation of the activities of daily life including speech, swallowing, handwriting, dressing, hygiene, falling, salivating, turning in bed, walking, cutting food;

— Part III: clinician-scored motor evaluation;

— Part IV: Hoehn and Yahr staging of severity of Parkinson's disease. (H&Y is a scale of 0 to 5; 0 being a stage without signs of disease and 5, a stage with total loss of autonomy.)

A new improved version was developed by the Movement Disorder Society (MDS) and published in 2008. The MDS-UPDRS has four parts, namely, I: Non-motor Experiences of Daily Living ; II : Motor Experiences of Daily Living ; III : Motor Examination ; IV Motor Complications. Moreover, 20 questions are completed by the patient/caregiver.[31]

All the above data items are part of the characterization and should be reported. The Parkinson subtype should also be reported: postural instability gait difficulty (PIGD), tremor dominant (TD) or indeterminate. The Parkinson cognitive status should be reported: normal, MCI or dementia. Cognitive markers closest to biospecimen collection should be documented: MMSE (Mini Mental State Examination) total score; CDR (Clinical Dementia Rating) overall; CDR sum of boxes.

Neuropsychological tests can be performed and documented as part of the characterization:

a) a word list test (e.g. CERAD 10-word learning test) with documentation of the sum learning trials raw score, the sum learning trials z-score, the delayed recall raw score and the delayed recall z-score;

b) a type story recall test with documentation of the sum immediate recall raw score, the sum immediate recall z-score, the delayed recall raw score and the delayed recall z-score;

c) a type visuoconstruction test, with documentation of the copy figures raw score, the copy figures z-score;

d) a type verbal fluency test;

e) an attention/executive function test with documentation of the Trail Making Tests TMTA and TMTB raw and z-scores.

If any of the following imaging has been performed within 12 months of biospecimen collection, the results should be documented: MRI, Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), fluorodeoxyglucose (FDG) PET.

If any genotyping data among the following are available, these should be reported: SNCA, LRRK2, VPS35, EIF4G1, PARK2, PINK1, DJ-1.

The concentration of the urine sample can be characterized by measuring its creatinin and its cystatin-C concentration. Urine samples can be characterized as of „high", „intermediate" or „low" concentration" based on these measures.

D. Characterization of pancreatic cell lines includes information on clinical history, *in vitro* and *in vivo* growth characteristics, phenotypic characteristics (adhesion, invasion, migration, tumorigenesis, all of which can be compared to a well-established and commercially available cell line, such as Capan-1), and genotypic characteristics.

More specifically:

The site of derivation of the initial tumour specimen, exact histological characterization, pTMN, cTNM classification, ICD-O classification, should be documented by a certified pathologist.

Adhesive ability of the cell line to different extracellular matrix proteins (fibronectin, collagen I, collagen IV, laminin) should be tested and documented. Documentation includes the % adherence and the length of time allowed for adhesion, but also technical details on the cell quantification method used, the cell culture conditions and extracellular matrix handling.

Invasive ability of the cell line, monitoring cell movement through Matrigel (a mixture of laminin, type IV collagen, entactin and heparin sulfate), should be tested and documented. Invasive ability is expressed as mean invaded cells per high-powered field, cells /chamber, cells/cm^2 or % of added cells having completely migrated into the bottom chamber. The length of time allowed for invasion, the cell culture conditions, the cell quantification method and the invasiveness assay or commercially available kit should be documented.

Levels of pro-angiogenic factors (COX-2, PGE2, VEGF, IL-1α, IL-8) secreted by the cell line is used as a surrogate measure to estimate its angiogenic potential. Pro-angiogenic factors are preferably expressed in absolute units (pg/ml, pg/mg protein, pg/106 cells) and again, the exact experimental conditions should be documented.

Tumorigenicity, which describes a cancer cell line's ability to produce tumours *in vivo*, is measured after injecting a suspension of pancreas cancer cells into the subcutaneous tissue of immunocompromised mice and allowing a tumour to grow. Tumorigenicity is expressed as tumour volume and/or mass (mm^3) having formed from a specified injected number of cells (e.g. 106 cells) and after a specified time period (20 to 50 days).

Genotypic analysis should include mutation analysis of the KRAS, TP53, CDKN2A and SMAD4 genes, which are frequently mutated in pancreatic cancers (http://pathology.jhu.edu/pancreas/geneticsweb/Profiles.htm).

Any cell line should be characterized by STR profiling with STR primers previously published.[22] It is advisable to submit any cell line to a recognized central cell bank (ATCC, DSMZ, ECACC, JCRB, RIKEN) and to submit the STR profiling data to a central database (Cell Line Data Base, http://bioinformatics.istge.it/hypercldb/, Cell Line Integrated Molecular Authentication database, http://bioinformatics.istge.it/clima/).[32] The same STR analysis should be used to connect the cell line to its original patient, by DNA fingerprinting on the cell line- and the paired whole blood-extracted DNA.[22],[25],[33]

Confirmation of the tissue of origin can be made by immunocytochemistry with tissue specific markers on fixed cells, e.g. carboxypeptidase B for pancreas specificity.

Finally, cell concentration and post-thaw cell viability should be reported.

4.3.5 Suitability study

A. If the downstream application is related to serum immunoglobulins, any serum sample, with documented standard pre-analytical code (SPREC) and stored at −80 °C is fit-for–purpose. Presence of anti-*C. trachomatis* IgG, IgA and/or IgM antibodies in the serum sample should be confirmed by a CE-marked or FDA-approved enzyme-linked immunosorbent assay (ELISA).

The pre-analytical quality of the serum can be assessed by measuring soluble CD40L.[34] Concentration of sCD40L should be higher than 6 ng/ml. Using this cut-off provides a sensitivity of 97,5 % in detection of serum samples which have been exposed to room temperature for at least 48 h.

B. The suitability of the DNA sample for downstream applications can be assessed by multiplex PCR[34] and by assessment of inhibition of amplification by the SPUD assay.[36]

C. The suitability of urine for downstream proteomic applications can be assessed by the absence of haemoglobin (ELISA).

D. The suitability of a cancer cell line is assessed by testing of the phenotypic characteristics described above.

4.3.6 Homogeneity study

The principles laid down in ISO Guide 35 [1] can be applied to homogeneity studies.

A.C. Homogeneity of serum or urine samples can be assessed by microparticle counting. Total microparticle counting can be performed by an impedance based method. Serum homogeneity testing, relative to anti-*C. trachomatis* antibodies, measured by immunosorbent assay should also be performed.

B. Homogeneity of DNA samples can be assessed by spectrophotometry.

D. Homogeneity of pancreatic cancer cell lines cannot be assessed. Pancreatic cancer cell lines should be considered as heterogeneous populations of cells due to potential for genetic drift and the presence of different tumour cell sub-types. Although it sounds difficult to certify a RM without homogeneity assessment, this reflects the actual clinical situation.[37] Homogeneity can still be assessed relative to attributes such as cell viability.

4.3.7 Stability study

The principles laid down in ISO Guide 35[1] can also be applied to stability studies.

A. Short and long term stability studies of serum, relative to anti-*C. trachomatis* IgG, IgA and IgM antibodies, measured by immunofluorescence and/or immunoenzymatic assays, should be performed.

B. Short and long term stability studies of DNA, relative to specific downstream applications (e.g. whole genome sequencing, methylation analysis) should be performed.

C. Short and long term stability studies of urine, relative to specific downstream applications (e.g. proteomics, metabolomics) should be performed.

D. Stability of a cancer cell line should be assessed relative to potential genetic drift over sequential passages. STR profile, caryotype (development of aneuploidy and heteroploidy), and gene mutations should be assessed regularly over passages from one flask to another.

4.4 Conclusion on biospecimens as reference material certified for qualitative properties

4.4.1 General

Different examples of samples (serum, urine) and derivatives (DNA extracted from tissue, cell line derived from tissue) were used to illustrate the need of accurately annotated CRMs, for use in research. The qualitative properties for which these materials can be certified is “C. trachomatis acute urogenital infection”, “lung adenocarcinoma”, “Parkinson’s Disease” or “pancreas cancer cell line”. However, we showed that in order for this qualitative property to be accurate, various clinical, biological, histopathological, immunological, cell biology and molecular biology data have to be collected and reported. Although the certified value is a qualitative property, some of the certification attributes, e.g. homogeneity and stability may need to be assessed by quantitative assays. The list of assays that need to be performed in order to have accurate and relevant qualitative property assignment may be long (depending on the material and the qualitative property) (Table 1). This is the reason why accreditation is required for biobanks to operate as RMPs.

Table 1 — Assays used for characterization of 4 examples of biospecimen types as reference materials certified for qualitative properties

Immunological	Cell biology/ Functional	Molecular biology	Histopathology	Imaging	Other
ELISA IgG, IgA, IgM	Cell counting	STR profiling	Histology	MRI	Isoenzyme analysis
HIV, HBV, HCV, syphilis, parvovirus B19 ELISA	Cell viability	NAAT microbial DNA	Immunohis-to-chemistry	TDM	Spectro-photom-etry
ELISA pro-angiogenic factors	Adhesive ability	PCR 16S. 18D rRNA	Immunocy-to-chemistry	SPECT	Spectro-fluorim-etry
ELISA sCD40L	Invasive ability	PCR Mycoplasma		PET	Proteomics
ELISA haemoglobin	Tumorigenicity	Genotyping			Metabolomics
		Caryotyping			
		SPUD assay			
		DNA methylation			

4.4.2 Documentary and metrological traceability

The certified qualitative clinical property of a biospecimen is based on clinical data and biological data. Documentary traceability of the origin of the specimen is ensured through a general quality management system. Additionally, metrological traceability should be ensured for those biological data where validated methods and CRMs are available.

5 Reference material for darnel seed identity

5.1 General

Darnel (*Lolium temulentum* L.) is a kind of worst weed. It belongs to genus *Lolium* which includes 8 species. The seeds of these 8 species are more or less similar to each other, so that it is difficult for inspectors to distinguish them. The plants of these species grow together with wheat, and their seeds often appear in wheat seed lots. Darnel seeds contain a toxin named temulentine, which leads to toxicoses in people and livestock, such as vertigo, vomiting, abdominal pain, diarrhoea, fever, and even coma, shiver, spasm, and dead owing to central nervous system paralysis. Hence, identifying darnel seeds in wheat lots and distinguishing them from other species in genus *Lolium* is a routine quarantine activity in the wheat trade. Identification by morphyological criteria is the most frequent method used for identifying darnel seeds; however, the conclusion should be confirmed by a judgement of the leading expert. This situation led to risks of erroneous judgment by the leading expert and time loss due to the tedious program. Therefore, a reference material (RM) was developed to identify darnel seeds (GSB 11-2232-2008).

5.2 Preparation of the reference materials

Darnel was planted in an isolated nursery field. During a 2-year period, Darnel seeds were collected. Intact seeds were selected to be the candidate reference material by the following rules:

a) the caryopsis shall be rectangle ellipse, clay brown, with length around 5 mm and width around 2,15 mm ;

b) the awn shall be about 10 mm long;

c) the length ratio of embryo to caryopsis shall be about 1:4;

d) it shall be difficult to peel the lemma from caryopsis.

The selected seeds were dried at a temperature of 35 °C for 7 days, and sealed carefully one by one in resins at room temperature to preserve their morphology characteristics.

5.3 Intended use

The RM is regarded as a calibrator for distinguishing darnel from wheat, the other species in genus *Lolium*, and mixtures thereof. It is also suitable for calibrating sensory equipment for several important characteristics of darnel seeds.

5.4 Characterization

5.4.1 Qualitative property to be characterized

The qualitative property of the RM is its identity, i.e. the materials are seeds of darnel. If the qualitative property value is true, the materials belong to the taxonomy class of darnel.

5.4.2 Principal of characterization

5.4.2.1 General

The qualitative property is concluded using classification or ranking of the object under study, such as what kind of species the object belongs to, which rank the object locates on and which grade the object is. The conclusion of this ranking or classification is to grade or classify the object under study into a rank or a class. For convenience, rank is regarded as class in a broad sense here, as well as ranking is regarded as classification. In practice, a qualitative property is assigned by comparing a series of characteristics of the object under study with the class boundary. Hence, class and its boundary need to be defined first.

5.4.2.2 Class and its boundary

In principle, a class is a region with boundary in a multidimensional space, and the number of space dimension is equal to the number of characteristics employed. Undoubtedly, the objects under study should be defined by the same characteristics as those define the class and its boundary. Otherwise, the comparison causing the qualitative property results becomes meaningless.

Obviously, when an object falls inside the boundary of a class, it belongs to the class; otherwise, it does not belong to the class. This is the identification method. That's to say, if the whole series of characteristic values of an object falls into the characteristic range of the class, the object belongs to the class.

5.4.2.3 Characteristic and its values

A characteristic forms a coordinate axis of the multidimensional space, it may be continuous or discrete. One specific characteristic value corresponds to a point on the axis. In actual circumstances, qualitative property results are practically defined by dimorphic, categorical, ordinal, and/or quantitative characteristics. Quantitative characteristics have numerical values per se, if the attribute values of the other type characteristics are assigned by numbers, the comparison becomes simple. However, the characteristics may be expressed by numbers, words, symbols, even graphs, there is no consensus way to assign all their values by numbers, for this reason, the way of assigning the characteristic values should be stated clearly.

An object has a value on a specific characteristic and a class has a value range (sometimes a value) on the same characteristic. If the object value falls into the class value range, the object is coincident with the class on this characteristic.

5.4.2.4 Origin of the axis

Comparison needs a cardinal point which is used for comparison, that's a point in the multidimensional space. The projection of the cardinal point on a characteristic axis is the origin of the axis. As the way assigning the characteristic values, there is no consensus way to assign the origin. For convenience, the central point of the target class is always selected to be the cardinal point which has the same distances to any point on the class boundary. Nevertheless, in some cases, it might be more convenient that one end point on the specific characteristic is selected to be the origin. Hence, the method of selecting the origin should be also stated.

5.4.2.5 Similarity

Similar things should be classified to be of the same class. Similarity is suggested to be applied on describing the location of an object in the multidimensional space, it indicates how close the object is to the stated cardinal point, and it also implies how close the object is to the class boundary or the class.

Many mathematical expressions have been used to describe the degree of closeness among two or more objects, such as correlation coefficient, correlation index, kappa, Fleiss's kappa, Cohen's kappa, mean square contingency, entropy, Gower general similarity coefficient, included angle cosine, matching factor, γmeasure.

Gower general similarity coefficient (so-called Gower Similarity here for convenience)[38] is recommended due to its suitable for both qualitative and quantitative data, it is given by

$$g = \frac{\sum_{i=1}^{n} w_i g_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

where

g is the Gower general similarity coefficient;

w_i is the weight of the i-th characteristic;

g_i is the similarity on the i-th characteristic, which is given by

$$g_i = 1 - |x_i - x_0| / r_i = 1 - d_i / r_i \ ; \tag{2}$$

where

x_i is the value of the target object on the i-th characteristic;

x_0 is the value of the origin on the i-th characteristic;

r_i is the radius of the scope or range of values that the upper class covers on i-th characteristic;

d_i is the distance between the object under study and the origin (object distance) on -ith characteristic.

When a characteristic value of an object is out of the target class range, the similarity on this characteristic, g_i is a negative number. To avoid this confusion on the meaning of similarity, an upper class radius instead of the target class radius is used. However, in practice, the upper class radii of different kinds of characteristics might be determined in different ways depending on the types of the scale of characteristic value, the five cases commonly used are given below.

— **Case A**: The characteristic value is quantitative, while both the target class and the actual upper class cover finite ranges. Taking the target class centre as the origin, the imagined upper class is constructed by expanding the target class boundary along both directions of the characteristic axis

symmetrically up to the farther boundary of the actual upper class, as shown in Figure 5. The radius is defined as a half range of the imagined upper class, and the similarity is a number between 0 (the imagined upper class boundary) to 1 (the origin), so that the similarity is normalized.

— **Case B**: The characteristic value is also quantitative, while the target class covers a finite range and the actual upper class covers an infinite range on one side, for example on the upper side, the imagined upper class range is established by expanding only one unit on the infinite side of the actual upper class. One unit is expanded symmetrically from the target class boundaries, when the characteristic value is out of the range of the target class boundaries ± 1 unit, the range of imagined upper class is set as the target class boundaries ± 2 units, as shown in Figure 6. The radius and similarity are generated in the same way as in Case A with the changed range of the imagined upper class.

— **Case C**: The characteristic value is still quantitative, while the target class covers a finite range and the actual upper class covers an infinite range on both sides, the imagined upper class range is constructed by doubling the target class range and distributed symmetrically on both side of the target class, as shown in Figure 7. The radius and similarity can be treated by the same approach with Case A and Case B.

— **Case D**: The characteristic value is categorical or ordinal, the value in the centre of target class is set as origin, whether it is of physical meaning or not, the imagined upper class range is then constructed by a similar way as that in Case A, As shown in Figure 8. The radius and similarity are calculated also by the same way as in Case A.

— **Case E**: The characteristic value is dimorphic that means the target class covers only one point or one range, the range of the target class is set as 1, the imagined upper class range is set as 2, as shown in Figure 9.

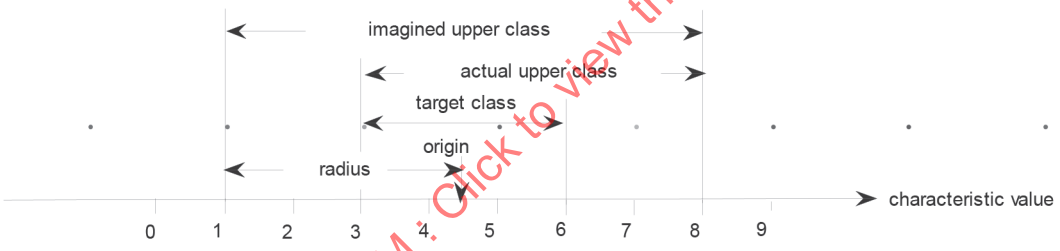


Figure 5 — Diagram of similarity on different types of characteristics, Case A: quantitative characteristic with finite range

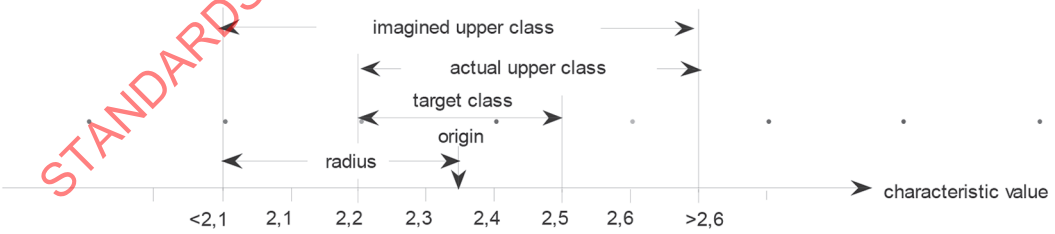


Figure 6 — Diagram of similarity on different types of characteristics, Case B: quantitative characteristic with infinite range on one side

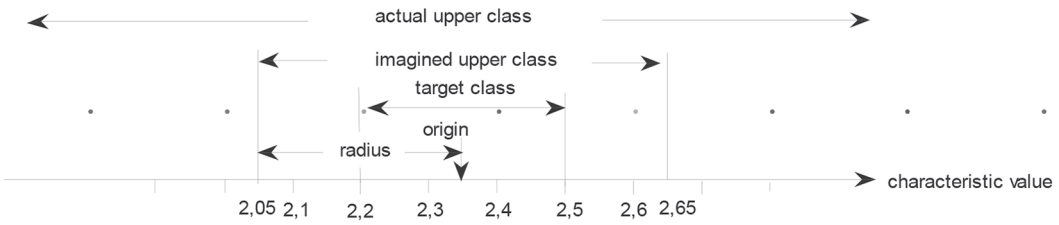


Figure 7 — Diagram of similarity on different types of characteristics, Case C: quantitative characteristic with finite range on both sides

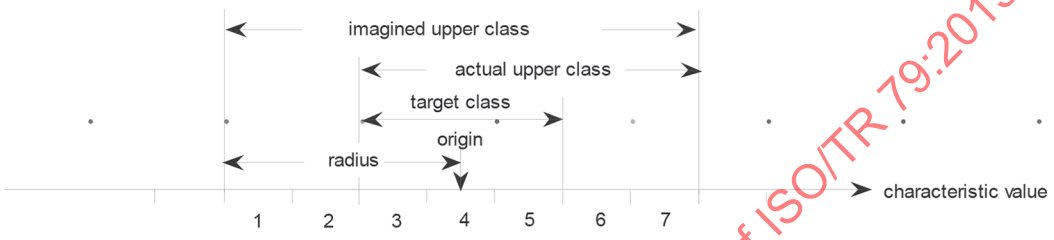


Figure 8 — Diagram of similarity on different types of characteristics, Case D: categorical or ordinal characteristic

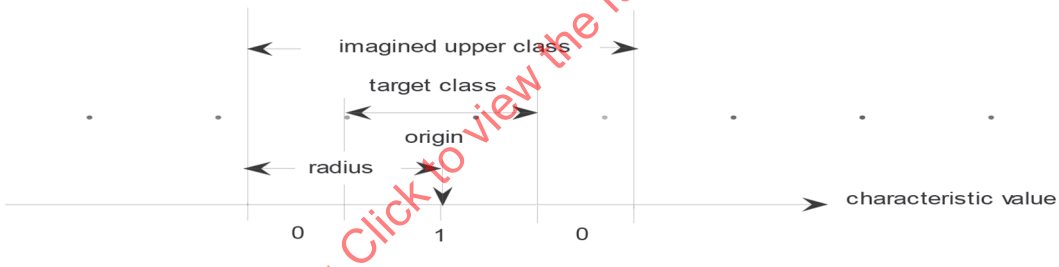


Figure 9 — Diagram of similarity on different types of characteristics, Case E: characteristics which target class only covers one point

5.4.2.6 Amount and weights of the characteristics

In reality, it is unnecessary and usually impossible to define the class by all its characteristics, and dominant characteristics are used. Obviously, the more characteristics are applied, the more exclusive the class is and the more precise the class boundary is. However, the bigger the number of characteristics used is, the more objects have the same value for some characteristics and the class boundaries between similar classes can become fuzzy. The risk of incorrect classification tends to rise. On the contrary, fewer characteristics make the classification easier, but when there are more similar classes to be classified, some classes cannot be distinguished by only a few characteristics, and the objects might not be distributed to the class in which they should be. Therefore, the number and types of characteristics should be reasonably selected. Not only should the class be sufficiently defined, but also the effectiveness of the number of classifying characteristics should be investigated.

A suitable set of characteristics for setting up the classifying criterion should be selected, the set ought to consist of a minimum number of characteristics and meanwhile these characteristics can sufficiently identify the class studied. A method named Minimum Dimension Analysis (MDA)[39] can be used for finding the suitable characteristics. The optimal characteristic should make the difference between classes significantly larger than that within the target class. It means that *either* the between-class variation should be as large as possible, and meanwhile the within-class variation should be as little as

possible; *or* the between-class variation should occupy a ratio large enough to cover the possible variation. This ratio, called the correlation ratio (λ), should be maximized. It is expressed as:

$$\lambda^2 = \frac{\sigma_b^2}{\sigma^2} = 1 - \frac{\sigma_w^2}{\sigma^2} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{3}$$

where

σ_b^2 is the between-class variation;

σ_w^2 is the within-class variation;

σ^2 is the total variation.

Squared correlation ratios are normalized by the following equation:

$$\lambda_{Ni}^2 = \frac{\lambda_i^2}{\sum_j^m \lambda_j^2} \tag{4}$$

where

λ_{Ni}^2 is the normalized squared correlation ratio on *i*-th characteristic;

λ_i^2 is the squared correlation ratio on *i*-th characteristic;

m is the total characteristic number.

When the sum of λ_{Ni}^2 of several major characteristics is close to 1, generally > 0,8 or > 0,9, these characteristics are the dominant characteristics that the MDA method is looking for.

When the dominant characteristics are determined, the corresponding correlation ratio should be recalculated by Formula (4). Then a new approximate multidimensional space can be constructed by the selected characteristics. The characteristic weight is the calculated normalized squared correlation ratio for the characteristic. The similarity of target class boundary can be calculated by using the new characteristic weights and the boundary values.

Even if the major characteristics could be found, the sufficiency of them is not always good, and the effect should be tested.

The similarity of the nearest point not belonging to the target class should be calculated, if it is below the similarity of target class boundary, the effect of classification is perfect, otherwise, extra characteristics need to be added and a new round of MDA should be carried out.

5.4.2.7 Variation on characteristics with unequal precision

Obviously, characteristic values are often coming from measurements with unequal precision. The division number which is equal to the imagined upper class range, could also be used to express the precision. For the characteristics which target class covers only one point (Case E in Figure 9), the division number is set as 2; for categorical and ordinal characteristics which target class covers a range (Case D in Figure 8) and the quantitative characteristics (Cases A to C in Figure 5 to Figure 7), the

division number is the range of imagined upper class. The object distance to the origin on a characteristic depends on the division number of the characteristic.

$$d_j/d_k = D_j/D_k = r_j/r_k \tag{5}$$

where

d is the object distance;

D is the division number;

r is the space radius;

j and k represent different divisions.

Then the variation of a group of object distances is proportional to the square of the division number.

$$\sigma^2(d_j)/\sigma^2(d_k) = (D_j/D_k)^2 = (r_j/r_k)^2 \tag{6}$$

For the characteristics which target class covers only one division, there is no difference detected on each object in the target class, the variation of these objects is used to be regarded as 0. In fact, these objects may be different; their variation is an infinitesimal in this case. If the target class is fractionized, the variation rises, while the positions of these objects did not move. To compare the correlation ratios, the variation should be defined on the same division. Considering the division of these types of characteristics is 2, and the target class occupies 1 division, the within-class variation on i-th characteristic can be defined as:

$$\sigma_{wi}^2 = (D_{Ti}/D_i)^2 \tag{7}$$

where

D_i is the division number of the upper class;

D_{Ti} is the division number occupied by the target class.

As defined in Formula (4), the similarity on each characteristic is independent of the division number. The between-class variation relies on the central point of every class, expressed in similarity as follows.

$$\sigma_{bi}^2 = \sum_j^n (s_{ij} - \overline{s_i})^2 \Big/ n, j = 1,2,\cdots,n \tag{8}$$

where

s_{ij} is the similarity of central point of j-th class on i-th characteristic;

$\overline{s_i}$ is the average of the similarities of these central points;

n is the number of classes to be classified in the upper class.

5.4.2.8 Characterization procedures

Following this principle, the qualitative property of the RMs could be characterized by the following procedures.

a) Characteristics: List all available characteristics defining the qualitative property, target class, upper class and their boundaries, assigned the characteristic values by numbers when the characteristic is discrete, and list division number of each characteristic.

b) Origin: Select the central point of the target class as the origin on each characteristic, if another way is employed, state it.

c) Characteristic similarity: Calculate the similarity of the target class boundary and similarity of each class central point for each characteristic.

d) Variation: Calculate the within-class variation by using division numbers of target class and upper class; calculate between-class variation by using similarities of central points of all classes to be classified.

e) MDA: Calculate the normalized squared correlation ratio for each characteristic, then determine the dominant characteristics by the sum of top squared correlation ratios, recalculate the normalized squared correlation ratio as the weight of each dominant characteristic by using only the selected characteristics.

f) Classifying criterion: Calculate the Gower Similarity of the target class boundary which will be used as the classifying criterion.

g) Efficiency check: Check the efficiency of classification by comparing Gower Similarity of the nearest point to the target class boundary, if it can be classify from the target class, the amount of dominant characteristics is enough for classification.

h) Characterization: Test the candidate RM and determine similarities on all dominant characteristics, calculate the Gower Similarity of the RM and determine the qualitative property of the RM.

5.4.3 Characterization of the reference material for darnel seed identity

Only the qualitative property to classify darnel and to distinguish it from the other species in genus *Lolium* was reported here as an example. The characteristics mainly employed in many references in the literature are listed in Table 2, and the characteristic values, including the central point values, are listed in Table 3.

Table 2 — Main morphological characteristics of the species in genus *Lolium*

No.	Name	Type	Range	Values
x1	caryopsis shape	Categorical	oblate ellipsoid — ellipsoid	1 = ellipsoid; 0 = oblate ellipsoid
x2	caryopsis colour	Ordinal	light brown — black brown	1 = light brown; 2 = yellow brown; 3 = clay brown; 4 = sepia; 5 = dark brown; 6 = black brown
x3	caryopsis length	Quantitative	3 mm — 8,5 mm	
x4	caryopsis width	Quantitative	1,1 mm — 3 mm	
x5	awn length	Quantitative	0 mm — 15 mm	
x6	tightness between lemma and caryopsis	Dimorphic	easy — difficult to peel	1 = difficult to peel; 0 = easy to peel
x7	length ratio of embryo to caryopsis	Ordinal	1:4 — 1:12	1 = 1:12; 2 = 1:11; 3 = 1:10; 4 = 1:4

Table 3 — Characteristic values of the species

Species[a]	x1	x2	x3	x4	x5	x6	x7
A	1	2—4(3)[b]	4—6(5)	1,8—2,5(2,15)	9—11(10)	1	4
B	1	1—5(3,5)	7—8,5(7,75)	2—2,7(2,35)	10—15(12,5)	1	4
C	0	3—6(3,5)	5—8(6,5)	2,5–3(2,75)	0—2,5(1,25)	1	4
D	0	3	3—5(4)	1,1–2(1,55)	0—5,5(2,75)	1	3
E	0	3	4,5—6,5(5,5)	1,5–2(1,75)	7—15(11)	1	1
F	1	3—5(4)	4,5—7(5,75)	1,3–1,5(1,4)	0—2(1)	1	2
G	0	3—4(3,5)	4—6(5)	1,3–1,5(1,4)	3—5(4)	1	2
H	0	3—4(3,5)	4—6(5)	1,3–1,5(1,4)	3—5(4)	1	2

[a] Species in genus *Lolium*: A = *Lolium temulentum* L.; B = *Lolium temulentum* var. longiaristum Parnell; C = *Lolium temulentum* var. arvense Bab.; D = *Lolium remotum* Schrank; E = *Lolium persicum* Boiss. Et Hohen; *F* = *Lolium perenne* L.; G = *Lolium multiflorum* Lam.; and H = *Lolium rigidum* Gaud.

[b] Number in the parenthesis: the central point value.

The similarity of the each class centre, the between-class variation, the within-class variation, and the squared correlation ratio were calculated. These are listed in Table 4. MDA was calculated. The order of the correlation ratios shows that the dominant characteristics are the awn length (x5), the length ratio of embryo to caryopsis (x7), the caryopsis length (x3), and the caryopsis shape (x1). The other characteristics contribute only about 13,5 % to the classification. These findings agree with expert opinions.

Table 4 — MDA calculation

Species	x1	x2	x3	x4	x5	x6	x7
A	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
B	100,0 %	83,3 %	21,4 %	81,0 %	75,0 %	100,0 %	100,0 %
C	100,0 %	83,3 %	57,1 %	42,9 %	12,5 %	100,0 %	100,0 %
D	100,0 %	100,0 %	71,4 %	42,9 %	27,5 %	100,0 %	71,4 %
E	100,0 %	100,0 %	85,7 %	61,9 %	90,0 %	100,0 %	14,3 %
F	100,0 %	66,7 %	78,6 %	28,6 %	10,0 %	100,0 %	42,9 %
G	0,0 %	83,3 %	100,0 %	28,6 %	40,0 %	100,0 %	42,9 %
H	0,0 %	83,3 %	100,0 %	28,6 %	40,0 %	100,0 %	42,9 %
R_i[a]	0 to 1	1 to 6	3,0 to 8,5	1,1 to 3,0	0 to 15	0 to 1	1 to 4
D_i	2	6	70	21	20	2	7
D_{Ti}	1	3	20	7	2	1	1
s_B[b]	100 %	66,7 %	71,4 %	66,7 %	90,0 %	100 %	100 %
σ^2_{wi}	0,250	0,250	0,082	0,111	0,010	0,250	0,020
σ^2_{bi}	0,188	0,012	0,065	0,063	0,105	0,000	0,097
λ_i^2	0,429	0,046	0,442	0,362	0,913	0,000	0,826
λ^2_{Ni}	14,2 %	1,5 %	14,6 %	12,0 %	30,3 %	0,0 %	27,4 %
Order	4	6	3	5	1	7	2

[a] Range of the upper class.

[b] Boundary similarity of the target class.

Figures 10 and 11 clearly demonstrate the classification effects obtained by using the first two and first three dominant characteristics.

— When the first two characteristics are used (the sum of contribution to the classification is 57,6 %), all the classes can be distinguished except A with B and G with H.

— When the first three characteristics are used (the sum of contribution to the classification is 72,3 %), only G and H could not be separated.

Things did not change when the first four characteristics were employed. Reviewing the characteristics selected, there is no difference between the species G and H for these characteristics, and other characteristics should be added.

It is very interesting that when the first five or six characteristics were used, the similarity of the target class boundary was less than that of the closest point from other classes. Consequently, redundancy characteristics may have lead to a drop in the classifying efficiency. It might be the reason for erroneous identifications.

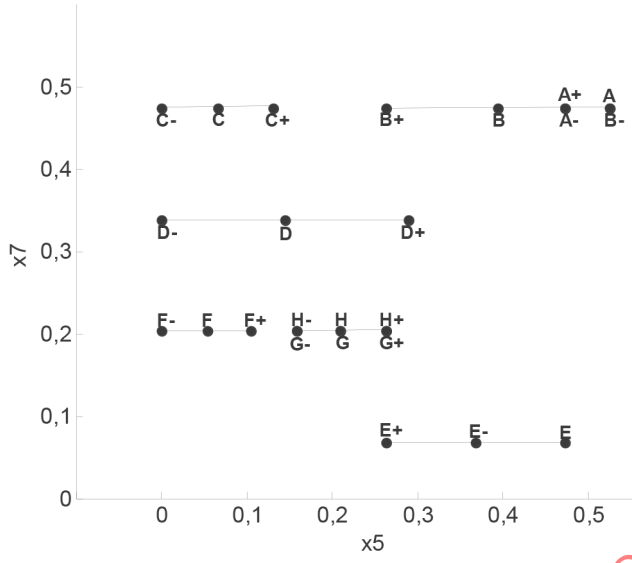


Figure 10 — Classification by the first two characteristics

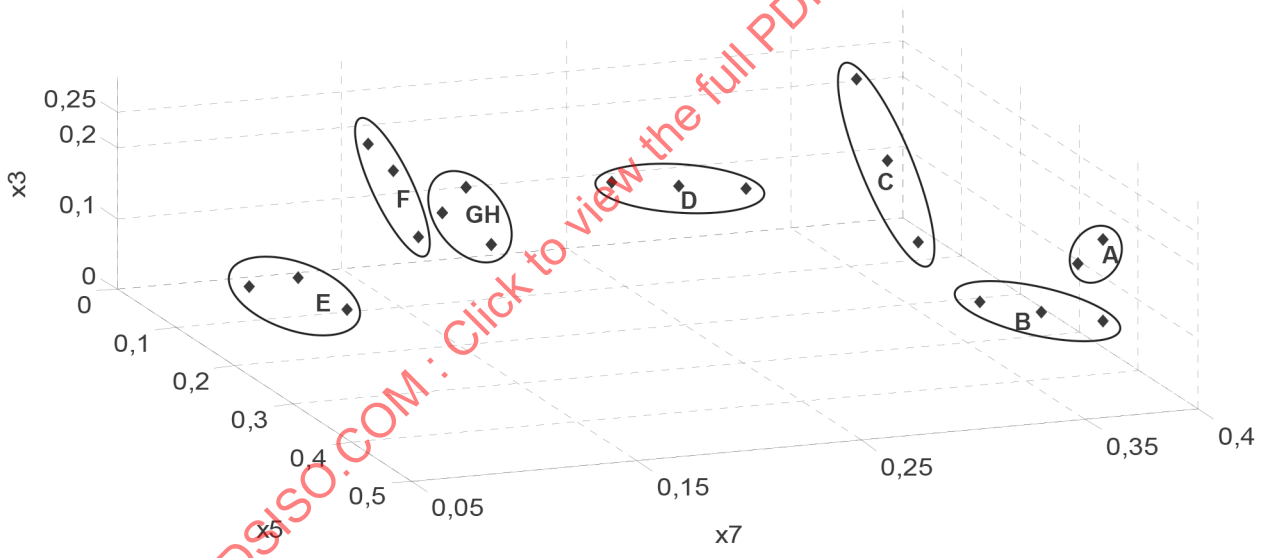


Figure 11 — Classification by the first three characteristics

The Gower coefficients of each species on the first three characteristics x5, x7 and x3 (including those for the central points, boundary points and the points closest to the target class centre) were calculated by using the corresponding normalized squared correlation ratios as the characteristic weights, as shown in Table 5. Classifying criteria, including a boundary similarity of the target class, 90,0 %, can be set with the three dominant characteristics, and the Gower Similarity of the most closest point is about 88,4 %, which is lower than the target class boundary. The classifying criteria are those that have the capacity to distinguish darnel seeds from other seeds in the genus *Lolium*.

Table 5 — Gower Similarity of the species

Species	Class centre	Point most closed to the target class centre	Point with the lowest values on each characteristic	Point with the highest values on each characteristic
A	100,0 %	100,0 %	90,0 %	90,0 %
B	73,6 %	88,4 %	88,4 %	58,8 %
C	54,7 %	58,1 %	58,1 %	51,2 %
D	53,0 %	70,3 %	35,7 %	70,3 %
E	60,5 %	60,5 %	52,1 %	37,9 %
F	36,3 %	36,3 %	33,6 %	33,3 %
G	53,2 %	53,2 %	51,6 %	43,3 %
H	53,2 %	53,2 %	51,6 %	43,3 %

5.5 Homogeneity and stability study

The qualitative property was quantitatively expressed by similarity, so the homogeneity and stability of the identity of the RM can be assessed by using the method listed in ISO Guide 35.[1]

Because the darnel seed identity RM in this study is characterized, and used one by one, it is unnecessary and impossible to assess its homogeneity.

The stability of the RM can be taken for granted because the seeds were sealed in resin, their morphological characteristics do not change for a long period.

5.6 Uncertainty

The uncertainty of qualitative property results also can be assessed by similarity. The RMs were sent to 5 experts to determine their Gower Similarities. The experts' results indicated that the Gower Similarity of each individual seed is identical.

However, there is a systematic error caused by the method, the Gower Similarity of an individual seed is supposed to conform to the rectangular distribution, the standard uncertainty is

$$u(g) = (1 - g) / \sqrt{3} \tag{9}$$

For example, the Gower Similarity of an individual seed is 95,8 %, the standard uncertainty of the qualitative property of the RM is about 2,4 %.

5.7 Expressing the qualitative property of the reference material

With the Gower Similarity and its uncertainty, the qualitative property can be expressed in the following form:

Qualitative property result, Gower Similarity ± Expanded Uncertainty (10)

For the example listed in Clause 5.6, the qualitative property of the darnel seed identity RM is

Darnel Seed, (96 ± 4) %

5.8 Traceability

The identity of the RM (darnel seed) is the realization of the darnel identity as defined by the taxonomy of plants. It is therefore traceable to the taxonomy definition via the similarity method given here. In

another words, the definition is the end point of the traceability chain, and any identification result for darnel seed can be made traceable to this RM and then to the end point, the definition of darnel identity, applying the similarity method.

6 Colour of freshwater cultured pearls

6.1 General

The grade of freshwater cultured pearl is determined according to the Chinese national standard GB 18781-2008 by its size, colour, shape, lustre, surface perfection, nacre thickness (if applicable), and matching attributes (if applicable). The pearl certificate provides the ranks of these properties.

For example, the colour attribute includes the white, red and purple series. Within the white series, there are 5 ranks: W1 - pure white; W2 - white with a bit yellow; W3 - white with yellow; W4 - white with obvious yellow; and W5 - white with deep yellow.

A set of CRMs (GSB 16-2521-2008) of Chinese natural freshwater cultured pearls was developed between 2008 and 2010. Five categories of conformity, comparing a colour test result to a colour rank are shown in Table 6; these categories were established by visual inspection carried out by humans.

Table 6 — Categories of conformity, comparing a pearl colour test result to a colour rank

Category	Too high (does not conform)	Slightly high	Matched	Slightly low	Too low (does not conform)
Category abbreviation	TH	SH	M	SL	TL
Category number/code k	1	2	3	4	5

6.2 Homogeneity testing

Variances of within- and between-sample qualitative and semi-quantitative test results applicable for homogeneity study of RMs are compared in.[40-42]

When a design of experiment for homogeneity study consists of testing M samples with n_m replicates for sample $m = 1, 2, \ldots, M$, a part of the test results n_{km} is related to category $k = 1, 2, \ldots, K$. As a rule, such a design of experiment requires an equal number of replicates n_m for every sample, providing the best estimates. However, some data may be "missed"; therefore, in general

$$n_m = \sum_{k=1}^{K} n_{km} \text{ and } \sum_{m=1}^{M} n_m = N \tag{11}$$

where N is the total number of test results, i.e. the size of the experiment. The number of the test results belonging to the k-th category is

$$n_k = \sum_{m=1}^{M} n_{km} \text{ and } \sum_{k=1}^{K} n_k = N \tag{12}$$

Let $p_k = \frac{n_{km}}{n_m}$ and $F_{km} = \sum_{i=1}^{k} p_{im}$ be the proportion and the cumulative relative frequency of data belonging to/(up to) the k-th category in the m-th sample.

Denote

$\pi_m = \frac{n_m}{N}$ as the total sample proportion ($\sum_{m=1}^{M} \pi_m = 1$),

$p_k = \frac{n_k}{N}$ as the total proportion of the test results belonging to the k-th category ($\sum_{k=1}^{K} p_k = 1$), and

$F_k = \sum_{i=1}^{k} p_i$ as the total cumulative relative frequency of test results belonging to the k-th category.

The hypothesis about the material homogeneity is tested using the following indicator:

$$I = \frac{V_{\text{BETWEEN}} / df_{\text{BETWEEN}}}{V_{\text{TOTAL}} / df_{\text{TOTAL}}} = \frac{N-1}{M-1} \cdot \frac{\sum_{k=1}^{K-1} \sum_{m=1}^{M} \pi_m \cdot (F_{km} - F_k)^2}{\sum_{k=1}^{K-1} F_k \cdot (1 - F_k)} \tag{13}$$

where

$$V_{\text{BETWEEN}} = \frac{4}{K-1} \cdot \sum_{k=1}^{K-1} \sum_{m=1}^{M} \pi_m \cdot (F_{km} - F_k)^2, \quad df_{\text{BETWEEN}} = M - 1 \tag{14}$$

are the between-sample variance and its number of degrees of freedom;

$$V_{\text{TOTAL}} = V_{\text{WITHIN}} + V_{\text{BETWEEN}} = \frac{4}{K-1}\left[\sum_{k=1}^{K-1} F_k \cdot (1 - F_k)\right], \quad df_{\text{TOTAL}} = N - 1 \tag{15}$$

are the total variance and its number of degrees of freedom; and

$$V_{\text{WITHIN}} = \frac{4}{K-1} \sum_{m=1}^{M} \left\{\pi_m \left[\sum_{k=1}^{K-1} F_{km} \cdot (1 - F_{km})\right]\right\}, \quad df_{\text{WITHIN}} = N - M \tag{16}$$

are the averaged within-sample variance and its number of degrees of freedom.

When $I < 1$, the hypothesis is not rejected. If $I > 3$ the hypothesis is rejected. When $1 < I < 3$, the homogeneity is questionable and obtaining additional data is necessary for the final decision.

The homogeneity study of a candidate RM of freshwater cultured pearls of pure white colour rank W1, for example, was performed with 10 replicate tests, i.e. $n_m = 10$ for every from $M = 5$ RM samples by one expert in the same conditions, so $\pi_m = \frac{1}{M} = \frac{1}{5}$. The test results by five categories $K = 5$ according to Table 6 are presented in Table 7.