EURL ECVAM
Test Submission Template (TST)

TABLE OF CONTENTS

SUBMISSION OF CONFIDENTIAL INFORMATION 4
LIST OF ABBREVIATIONS 4
SUBMITTER'S REQUEST TO EURL ECVAM 4

1. INFORMATION ON VALIDATION MODULES 5

1.1 MODULE 1: TEST DEFINITION 5
1.1.1 Human health, environmental or other biological effects addressed by the test method 5
1.1.2 Scientific basis – biological and/or mechanistic relevance 6
1.1.3 Intended purpose of the test method 9
1.1.4 Evidence demonstrating the need for the test method 12
1.1.5 Technical specifications 13
1.1.6 Test items used for developing and optimising the test method (protocol and PM) 23
1.1.7 Cost and time estimates per test item 24
1.1.8 Occurrence of non-qualified tests 24
1.1.9 Known limitations and drawbacks of the test methods 28
1.1.10 Intellectual Property Rights (IPRs) 28
1.1.11 History of test method development 28
1.1.12 Quality system(s) of the developing laboratory 30
1.2 MODULE 2: WITHIN-LABORATORY REPRODUCIBILITY (WLR) 30
1.2.1 Rationale for the selection of the test items used for assessing WLR 30
1.2.2 Assessment of within-laboratory reproducibility of experimental data 30
1.2.3 Identification and discussion of outlying values 32
1.2.4 Quality system(s) of the testing laboratory 32
1.3 MODULE 3: TRANSFERABILITY (TF) 33
1.3.1 Rationale for the selection of the test items used for assessing the TF 33
1.3.2 Training required for transferring the test method 33
1.3.3 Obstacles pertaining to transferability that are specific to the test method 34
1.3.4 Organisation of the transfer phase 34
1.3.5 Assessment of the transferability to another laboratory 34
1.3.6 Quality system(s) of the other laboratory 34
1.4 MODULE 4: BETWEEN-LABORATORY REPRODUCIBILITY (BLR) 34
1.4.1 Rationale for the selection of test items used for assessing BLR 34
1.4.2 Assessment of reproducibility 35
1.4.3 Identification and discussion of outlying values 36
1.4.4 Quality system(s) of the testing laboratories 36
1.5 MODULE 5: PREDICTIVE CAPACITY (PC) 37
1.5.1 Rationale for the selection of test items used for assessing PC 37
1.5.2 Assessment of the predictive capacity of the test method 37
1.5.3 Identification and discussion of false predictions 43
1.6 MODULE 6: APPLICABILITY DOMAIN (AD) 43
1.6.1 Applicability of the test method identified through testing 43
1.6.2 Limitations of the test method identified through testing 43
*Table 9. Differences between the gastric fluid SOP used in the round robin study and the current one shown in Attachment 1a ..... **Error! Bookmark not defined.***
1.6.3 Suggested Applicability Domain (AD) of the test method 44
1.7 MODULE 7: PERFORMANCE STANDARDS (PS) 45
1.7.1 Suggestions of essential test method components for the evaluation of similarity 45
1.7.2 Suggestions of possible reference test items 45
1.7.3 Suggestions of defined accuracy and reliability values 45

2. ESSENTIAL INFORMATION FOR A SPECIFIC VALIDATION PROCESS 47

2.1 EXTERNAL VALIDATION STUDIES (NOT COORDINATED BY EURL ECVAM) 47
2.1.1 Study organisation and management (Project Plan) 47
2.1.2 Study objective and goals 48
2.1.3 Summary of the study results 48
2.1.4 List of test items used in the validation study 49

2.1.5 Study conclusions *49*
2.1.6 Recommendations *50*
2.2 ESSENTIAL INFORMATION FOR A RETROSPECTIVE VALIDATION STUDY 52
2.2.1 Information/Data Collection *52*
2.2.2 Protocols identified *52*
2.2.3 Data analysis and Prediction Models *53*
2.3 ESSENTIAL INFORMATION FOR A VALIDATION STUDY BASED ON PERFORMANCE STANDARDS (FOR SIMILAR OR UPDATED TEST METHODS) 53
2.3.1 Performance standards based on the validated reference test method(s) *53*
2.3.2 Essential test method components (used for similarity assessment) *53*
3. ADDITIONAL INFORMATION 54
3.1 ADDITIONAL INFORMATION 54
3.2 LIST OF REFERENCES 54
4. ATTACHMENTS 60

Submission of confidential information

It should be noted that during the evaluation phase of test method submissions, EURL ECVAM will share information contained in the TST with its Advisory Structure, e.g. the EURL ECVAM Scientific Advisory Committee (ESAC), and to the extent possible, with European regulatory authorities. Whoever gets access to the TST will be bound to treat all information submitted to EURL ECVAM in a confidential manner. Nevertheless, EURL ECVAM procedures involve that Background Review Documents on the evaluation of the test method and its outcome are compiled and made publicly available which might imply disclosing information/data submitted with the TST together with those generated during validation.

Therefore, if you consider some of the information provided in the TST as confidential (e.g. confidential business information (CBI)), please clearly identify below those paragraphs where confidential information has been entered. Briefly describe (not more than 100 words per paragraph) why this information is considered confidential. Please specify if such confidential information may be made publicly available.

[…]

List of abbreviations

Please list the abbreviation used in the submission

[…]

Submitter's request to EURL ECVAM

If you seek to enter the EURL ECVAM validation process with this test submission, please indicate at what stage in the validation process you think the test method should enter. For a description of the validation process and the different types of validation, see the Explanatory Note on the TST. Otherwise, please specify any other type of request to EURL ECVAM by filling the appropriate text box.

	Please tick where appropriate
Prevalidation	
Prospective validation	
Retrospective validation	
Validation based on PS	
Peer-review	
Other	Please specify your exact request to EURL ECVAM: Assessment of available data, improvement and validation of an existing protocol for which information and data is already available

1. INFORMATION ON VALIDATION MODULES

The EURL ECVAM modular approach to validation breaks down the information required for establishing the scientific validity of a test method into independent modules allowing greater flexibility with regard to the time of procuring information during a validation study.

1.1 MODULE 1: TEST DEFINITION

1.1.1 Human health, environmental or other biological effects addressed by the test method

Please describe which human health, environmental or other biological effects the test method will address, e.g. whether it can be used to predict repeated-dose toxicity in humans, to predict fish chronic toxicity, to measure bioaccumulation, to be used in quality control of immunobiologicals etc.

The proposed test method is an in vitro physico-chemical test that measures metal releases in artificial gastric fluid. This type of test, commonly referred to as a bioelution test, does not directly address a biological effect, but rather provides an in vitro estimate of oral bioaccessibility[1] of metals from a variety of materials (e.g. metal containing substances, complex metal-containing materials like UVCBs, alloys, pigments). This test can be used to compare the metal releases from reference and target substances[2].

In vivo oral bioaccessibility is the first step leading to oral absorption and is thus related to the systemic bioavailability[3] of metal ions. The basic premise is that the systemic toxicity effects of metal-containing materials are related to the bioavailability of the metal ion after systemic absorption. Thus, once it is established that a relationship can be defined for a particular metal between in vitro bioaccessibility in gastric fluid and in vivo oral bioavailability (based on metal-containing soils or metal compounds), this is also relevant to other materials containing that same metal (e.g. alloys, pure metals) that are also absorbed systemically through the oral route. The bioelution test data in vitro can therefore be used comparatively to predict the oral metal bioavailability from target materials (substances or mixtures) relative to that of reference materials (substances or mixtures), as long as the in vitro and in vivo speciation of the dissolved metal ions are the same.

This information can be applied in several different ways. Here, we discuss two main applications of oral bioaccessibility data (for simplicity we will refer to them as 1-Grouping and 2-Alloy classification.

1-Grouping. In this application, bioelution data are used to group substances of a given metal and predict if they will have similar or different oral bioaccessibility. Relative oral bioaccessibility is considered to be predictive of relative oral bioavailability. Bioaccessibility data can contribute (in a weight of evidence approach) to bridging and read-across to fulfil human health data requirements for some systemic effects and/or classification under

[1] **In vivo oral bioaccessibility** is defined as the fraction of a substance that is released in the gastrointestinal tract, and becomes available for absorption. This can be estimated in vitro by measuring the dissolved quantity of a metal ion released under surrogate physiological conditions (**in vitro bioaccessibility**) in bioelution tests.

[2] **Reference** substances are those substances, which are well characterised and for which toxicological data and/or harmonized classifications exist. **Target** substances are the substances lacking toxicological data for one or more health endpoints but have sufficient physico-chemical data.

[3] **In vivo bioavailability** is defined as the extent to which a substance is taken up by an organism and is available for metabolism and interaction at target organ/sites (e.g. kidney, skin).

chemical management programmes. Grouping of substances containing the same metal can help prioritize and limit the number of animal toxicity tests conducted.

2- Alloy classification. When bioelution tests are applied to alloys and their ingredient metals, the data that these tests generate can be used for the calculation of the bioaccessible concentration of a metal ingredient in the alloy. In this way, this test can allow for a refinement of the bulk concentration of a metal ingredient in a mixture by considering the fraction of that concentration that becomes bioaccessible for gastrointestinal absorption prior to being taken up into systemic circulation and becoming bioavailable.

These applications apply to micron size powder and massive forms of metal-containing materials but may not apply to nanoforms[4].

1.1.2 Scientific basis – biological and/or mechanistic relevance

Please describe the relationship between the test method and the effect of interest. This may include a reference to a) the mechanistic relevance (e.g. the mechanism of action) and/or b) the biological relevance (e.g. how well the test method models the target organ) and/or c) an empirically observed (correlative) relationship to the effect of interest.

The direct or indirect systemic toxicity effects of most metals and metalloid-containing inorganic materials are associated with the bioavailability of released metal ions and closely related to their chemical speciation in the solution of interest (e.g. Cassarett and Doull 2013; Hedberg and Wallinder 2015). In other words, it is the bioavailable metal ions that drive the systemic toxicity of metal-containing materials after oral ingestion of micron size powders or massive forms. Some illustrative examples are lead and cadmium. In both cases the toxicity effects are systemic (e.g. neurotoxicity, kidney toxicity) and in both cases metals in blood or urine data (biomonitoring) are often used to refine the relevant exposure when predicting toxicity or assessing risk.

Health effects secondary to systemically bioavailable metals (e.g. reproductive effects) can manifest after exposures via dermal, inhalation or oral routes of exposure, as long as these exposures result in bioavailable metal levels above the threshold values for these effects. In general, the dermal absorption of inorganic metals and metal compounds is very low <1-2% and will not contribute significantly to systemic health effects (see section 1.1.3). Inhalation exposure can contribute to systemic health effects via direct absorption of metal ions from particles deposited in the alveolar-pulmonary region of the lung or via gastrointestinal (oral) absorption of metal ions from particles deposited in other respiratory tract areas (tracheobronchial and nose). Therefore, the oral route of exposure covers not just the ingestion of metals but also a significant fraction of the inhalation of metal particulates.

Yet, the simple presence of a metal in a material does not necessarily impart to that substance the biological properties associated with the fully water-soluble metal ion or the metal by itself. When the bioavailability of the released metals (e.g. ions) at relevant exposure levels differs, their toxic effect differs as well. That is why e.g. compounds of the same metal (and valence) that are at the extremes of the bioavailability range often have different toxicological profiles and associated classifications (e.g. nickel sulphate a highly Ni(II) bioavailable compound is classified as Repro 1B, while nickel oxide having a much lower Ni(II)

[4] Nanoforms as defined by the EU Commission in 2011 (e.g. a natural, incidental or manufactured material containing particles, in an unbound state or as an aggregate or as an agglomerate and where, for 50 % or more of the particles in the number size distribution, one or more external dimensions is in the size range 1 nm - 100 nm).

bioavailability at relevant exposure levels is not). Another factor that affects toxicity of metal compounds is the metal ion valence. For example, compounds of chromium differ in carcinogenic potential and classification depending on their valence (III or VI); thus, assessing the valence of any bioavailable Cr ion is of high importance.

Information on bioavailability can be derived from in vivo sources such as toxicokinetic or toxicological test data. ***For the two applications described in this document absolute bioavailability[5] information is not required***. Rather, bioavailability estimates from reference and target substances are compared e.g. for the purposes of grouping and read across (Application 1-Grouping). In situations where the ***relative bioavailability*** of a target metal or mineral substance is not known or where it is not desirable to determine this in vivo (e.g. animal welfare considerations), and data exist from other substances (reference) of known bioavailability and toxicity, the amount of released metals (ions, complexes) "available for absorption" may be quantified using in vitro bioelution test methods. Bioelution tests measure the degree to which a substance (e.g. metal ions) is released in simulated physiological fluids. (e.g. gastric fluid as one of the oral fluids). The value resulting from bioelution testing (e.g. µg metal /L of fluid or µg metal/g of sample) is termed as ***absolute bioaccessibility***, and is defined as the fraction of a substance that can be released into solution under surrogate physiological conditions and therefore is potentially available for oral absorption into systemic circulation. For compounds, mixtures and complex substances of a metal or metalloid, ***their relative bioaccessibility and bioavailability*** can have an influence when assessing their relative hazard. A more thorough description of various definitions for absolute and relative bioaccessibility and bioavailability can be found on pages iii-v of Attachment 18.

Since the basic premise is that the systemic toxicity effects of metal-containing materials are related to the bioavailability of the metal ion after systemic absorption, to use bioaccessibility data for application-1 Grouping, it is necessary to previously establish if a relationship can be defined between in vitro bioaccessibility in oral fluids and relative in vivo bioavailability (e.g. from soil of metal compound studies). Once that relationship has been established for a particular metal, then this relationship is also relevant for other applications, like application 2-Alloy classification.

Bioelution assays can be used as a tool to establish categories of metal/mineral substances (i.e. application 1-Grouping) and are part of the weight-of-evidence approaches applied to perform read-across (extrapolation of known data from one substance to another substance based on the assumption that the two substances will cause similar biological responses) and grouping. Until recently, water solubility was often used as a surrogate for in vivo oral bioavailability and a starting point for read across for a number of metal compounds. However, recent toxicity studies have shown that this approach could in some cases be either

[5] In vivo **bioavailability** data can be '**absolute'** when AUC (area under the curve) data are available for target (e.g. soil) and reference (usually a water soluble metal compound) substances. These values can be converted to '**relative'** bioavailability values (RBA), by considering that the AUC for the reference metal compound is 100%. Similarly, when the in vitro **bioaccessibility** results are reported as the amount of the sample that is released in gastric fluid or as the percent of the metal content in the sample that is released in gastric fluid (IVBA, as it is done in the soil studies), these are also '**absolute'** values. The water soluble compounds of the metal (often used as reference in animal studies) will have a percent release that may closely correspond to 100%; in this case the 'absolute' values obtained for other compounds of this metal also correspond to the '**relative'** bioaccessibility values. If the water soluble reference metal compounds have <100% bioaccessible metal, then bioaccessibility values of other compounds relative to those of the water soluble compound can be calculated.

too precautionary or not conservative enough (OECD 2014). In addition, many studies have demonstrated that the oral bioavailability of metals correlates better with bioaccessibility in synthetic oral fluids (e.g. gastric) than in water (Attachment 18). As indicated for the nickel example above, differences in bioavailability can also be considered to justify differences in classification.

The scientific basis of application 2-Alloy classification is as follows: under EU CLP, alloys are considered as mixtures and classified based on their content of classified metal (metalloid) ingredients (i.e. the concentration of classified ingredient is compared to its classification cut-off value, such as GCL or SCL[6]). Yet, alloys (composed of more than one metal ingredient) have metallic bonds that can affect metal releases in such ways that the bulk concentration of the metal ingredients is not reflected in the bioavailability or bioaccessibility of that metal from the alloy, as evidenced by the unique physico-chemical properties of alloys. Since the systemic toxicity effects of alloys are related to the bioavailable metal ions, it follows that the bioavailability of metals from alloys is not expected to match that of a simple mixture. In vitro bioaccessibility data from alloy and pure ingredients can be used to calculate the relative bioaccessible concentration of the metal in the alloy, thus refining the "effective" ingredient concentration prior to comparison with GCL or SCL.

Bioaccessibility data can thus be considered as a refinement over simple water solubility (application 1) or bulk concentration (application 2).

The focus here is on oral bioaccessibility and systemic bioavailability through the oral route of exposure.

The following artificial fluids have been considered for the oral route: saliva, gastric and intestinal, separate or in sequence, with variations in the composition of these fluids. In general, the gastric fluid (pH 1.5) has been found to result in higher metal releases from soils, metal compounds and alloys and can therefore be used to provide a conservative estimate of metal bioaccessibility through the oral route.

The human stomach pH varies considerably (from 1.2−6.4) among individuals due to prandial state, type, and volume of food consumed (Mudie et al. 2010; Kalantzi et al. 2006). The typical pH range for fasted gastric fluid is 1.2−2.7, and has been reported with a median value of pH 2.4 (Axson et al. 2015). The pH for fed gastric fluid can range from 3−6.4 depending on the stage of digestion (e.g. Kalantzi et al. 2006; Jantratid et al. 2009). Therefore, gastric fluid at pH 1.5 represents a worst case fasted state.

The absolute percent gastric bioaccessibility of a metal is generally higher than the absolute metal bioavailability since not all bioaccessible metals are absorbed before they are eliminated from the gastrointestinal tract. In addition, bioavailability depends on a number of factors including dose, metal speciation, metal uptake, fasting conditions, etc. The data generated by bioelution tests is <u>not</u> intended to predict absolute bioavailability but rather it is intended to compare the bioaccessibility of a given metal between different metal-containing materials (i.e. determine the relative bioaccessibility) in application 1-Grouping and to calculate the bioaccessible concentration of metal ingredients in alloys in application 2-Alloy classification. That is why we focus on estimating relative metal bioaccessibility.

[6] GCL: generic concentration limit; SCL: specific concentration limit.

There are some metals/metalloids for which high releases may be observed at neutral pH like Ag, Mo, Sb, Si, W and Hg (e.g. Herting et al. 2014; Mörsdorf et al. 2015); in these cases, bioelution tests in gastric fluid may not correspond to a worst case scenario and tests in artificial intestinal or saliva fluids would be required to predict relative systemic bioaccessibility and bioavailability.

Only the gastric fluid is the subject of the present validation submission. The materials to be considered are metals, inorganic metal compounds, or complex metal-containing materials (like alloys, UVCBs, pigments) in micron-size powder or massive forms.

1.1.3 Intended purpose of the test method

Please describe the intended purpose (i.e. practical use) of the test method (e.g. regulatory safety testing under REACH, for cosmetics, for pharmaceuticals, or non-regulatory applications).

The purpose of the bioelution test itself is to generate metal release data in simulated gastric fluid (µg metal /L of fluid, µg metal/g of sample, µg metal /g of metal in sample) from reference and target substances (or complex mixtures or alloys), all containing the same metal. The data from reference and target substances can have a variety of regulatory applications in hazard and risk assessment. For example, bioelution test data in gastric fluid has been used to estimate the exposure to metal-containing soils when performing risk characterisation of contaminated sites. In these cases, the relative metal bioaccessibility is considered to provide a reasonable estimate of relative oral bioavailability (Juhasz et al. 2016). Measured metal release data are also used in the context of the Toy Directive (BS EN 71-3 2013) but in this case, absolute metal releases from toy parts are compared to reference values.

Here we only discuss and provide justification for two main applications of bioelution test data. In one application (1-Grouping) bioaccessibility data from two metal-containing substances are compared directly. In the second application (2-Alloy classification) bioaccessibility data are used to calculate the bioaccessible concentration of a metal ingredient in an alloy.

For both applications, the information generated from the gastric fluid bioelution test is relevant for the oral route of exposure and the systemic effects of metal-containing materials. The systemic toxicity of metals is directly linked to the bioavailability of released metals (e.g. ions, complexes) as indicated in 1.1.2.

While the results are directly relevant to the oral route, they are also relevant to the fraction of any inhaled material that is swallowed and absorbed through the gastrointestinal tract. Inhalation and dermal exposures predominate in the workplace, while oral and dermal exposure can be more common for consumers and the general public. Importantly, the absorption of metals and the systemic bioavailability of metals can be very different for each of these routes of exposure. For example, dermal absorption of inorganic metals is generally very low (≤1-2%) for water soluble compounds and even lower for the more insoluble ones (≤0.1-0.2%) (HERAG 2008).

Table 21: Dermal absorption data for metals and inorganic metal compounds

Metal/compound	Test system	Results	References
Data as extracted and concluded upon in the various existing EU RA reports:			
Zinc oxide / Zinc sulphate	*in vitro*, porcine skin	2 % from liquid media 0.2 % from dust exposure (EU RAR assessment, Rapporteur: The Netherlands)	Grötsch (1999)
Cadmium metal, Cadmium oxide	(analogy)	< 1 % (EU RAR assessment, Rapporteur: Belgium)	EU RAR (2004)
Nickel metal, Nickel sulphate, Nickel chloride, Nickel nitrate, Nickel acetate	*in vivo*, human skin, tape stripping	0.2 % (EU RAR assessment, Rapporteur: Denmark)	Hostýnek et al. (2001a) Hostýnek et al. (2001b)
Nickel sulphate, Nickel chloride, Nickel nitrate, Nickel acetate	*in vitro*, human skin	2 % (EU RAR assessment, Rapporteur Denmark) 1 % when material bound to stratum corneum is discounted	Tanojo et al. (2001)
Diantimony trioxide	*in vitro*, human skin	0 - 0.1 %	Roper & Stupart (2006)
Copper compounds (not specified)	*in vitro* (unspecified)	0.3% soluble/insoluble Cu compounds (VRA Copper)	Roper (2003) Cage (2003)
Lead oxide	*in vitro*, human skin	0 - 0.1 % (VRA Lead)	Toner & Roper (2004)
Additional (non-exhaustive compilation) data made available from metal industries participating in HERAG:			
Zinc oxide	*in vitro*, porcine skin	< 0.1%	Gamer et al. (2006)
Aluminium chlorohydrate (^{26}Al-labelled)	*in vivo*, two human volunteers	0.012 % uptake (industry data)	Priest (2004), citing from Flarend et at. (2001)
Cobalt metal	*in vitro* (Franz diffusion cell, human skin)	Absorption not given as a percentage of the applied dose but as a steady-state flow of (0.0123 ± 0.0054) µg cm^{-2} h^{-1} with a lag time of (1.55 ± 0.71) h. Significant absorption only took place, when the metal was oxidised to Co^{2+} by stirring in artificial sweat for 30 minutes.	Filon et al. (2004)
Titanium dioxide	*in vitro*, porcine skin	< 0.1%	Gamer et al. (2006)

Table 1. Dermal absorption studies with a variety of metal compounds (corresponds to Table 21 of HERAG 2008)

Thus, while dermal exposure is quite relevant for local skin effects, it does not contribute significantly to the systemic bioavailability of most inorganic metals. Regarding inhalation exposure, a fraction of the inhaled material, the one with larger aerodynamic diameter and greater metal mass, will be deposited in the trachea-bronchial and extrathoracic regions of the respiratory tract. These particles are systemically absorbed through the gastrointestinal tract after they are swallowed. Thus, the oral route is relevant for the fraction of the inhaled dose that is not deposited in the alveolar region of the respiratory tract. The fluids that are considered representative of the respiratory tract are the interstitial fluid (Gamble, neutral pH) and the lysosomal fluid (pH 4-5). While all the particles in the alveolar region come into contact with interstitial fluid, only those particles that are taken up by cells (e.g. macrophages) will come in contact with lysosomal fluid. Metal release in lysosomal fluid can, for some metals, be higher than in gastric fluid due to the presence of additional complexing agents.

One study demonstrates that the in vitro lysosomal fluid can overestimate the intracellular dissolution of metal particles; cultured cells observed 24 h post exposure to nickel particles still showed a full loading of particles, while lysosomal fluid results predicted complete particle dissolution (Latvala et al. 2016).

Altogether, bioelution tests in gastric fluid can be used to assess the potential systemic bioavailability of metals after oral ingestion and for the fraction of the dose that is deposited in tracheo-bronchial and extrathoracic regions after inhalation exposure and ends up in the gastrointestinal tract. **Gastric fluid bioelution tests are not relevant to assess local respiratory tract effects after inhalation.**

The test method will assess the differences in metal release from various metal-containing materials (e.g. comparing target and reference materials) in simulated gastric fluid and as such provide an estimate of relative metal in vivo bioaccessibility for the two substances tested under the same conditions (provided the same speciation-valence prevails in the dissolved metal). The test method should always be used in a comparative way. In other words, the test as proposed below is not intended to provide information on the absolute bioavailability or bioaccessibility of metal-containing materials but rather to determine their relative bioaccessibility; i.e., comparison of released metal ions in gastric fluid that may become available for oral uptake and that eventually determines their relative systemic bioavailability.

In **application 1-Grouping**, relative bioaccessibility data are used to support grouping and read-across to fulfil Chemical Management programmes data requirements for systemic endpoints, reducing the need for additional animal toxicity studies. It can also be used in a weight-of-evidence approach to read-across classifications for systemic endpoints. This application is suitable for grouping and read across of metal-containing substances and mixtures (bridging). It should be noted that for this application, relative bioaccessibility data will not be used in isolation but will complement existing information on phys-chem (e.g. water solubility), toxicokinetics, and/or toxicity data for reference and target substances or mixtures in a weight of evidence approach. While water solubility has been used as a surrogate for in vivo oral bioavailability and a starting point for read across for a number of metal compounds, this approach has been found to be both too precautionary or not conservative enough based on recent toxicity studies (Henderson et al. 2012; OECD 2014). In addition, many studies have demonstrated that the oral bioavailability of metals correlates better with bioaccessibility in synthetic oral fluids (e.g. gastric) than in water (Attachment 18).

Application 2-Alloy classification. For some metal-containing mixtures such as alloys, matrix effects affect the extent to which metals can be released: the amount of released metal per gram of metal can either be reduced (in most instances) or enhanced (in a few cases) compared to that of the pure metal ingredients. For these reasons, the bioaccessible concentration of a metal in an alloy in gastric fluid is expected to provide a refinement of its bulk concentration. For example, recent bioaccessibility data in gastric fluid has shown how different the release of nickel from SS316L alloy (69% Fe, 16% Cr, 11% Ni) and from a simple mixture of powders containing 69% Fe, 16% Cr and 11% Ni is. The difference in Ni release was >100-fold, indicating a strong alloy matrix effect that limits the Ni release compared to a simple mixture of ingredients with the same bulk Ni concentration (ECTX 2017, see Attachment 20). This approach has been proposed to be used when determining the

hazard classification of alloys for systemic endpoints according to the EU CLP[7]. **This approach is not relevant for classifications involving local effects** (e.g. dermal sensitisation).

The proposed bioelution test in artificial gastric fluid is to be applied to micron size powder and massive forms of metal-containing materials but not to nanomaterials (see footnote 4).

1.1.4 Evidence demonstrating the need for the test method

Please summarise the need for the test method in relation to existing methods in the context of regulatory testing (relevant test guidelines and legislation) and the 3Rs. If applicable, describe whether the test method represents an improvement compared to an existing method. Possible improvements include a) better information (e.g. higher accuracy or addressing a mechanism of action), b) effectiveness in terms of throughput (e.g. amenable to high-throughput testing), c) cost.

Chemical management programmes require substances and mixtures to comply with a minimum set of data requirements, while minimizing animal testing. Grouping and read across (**application 1-Grouping**) is one of the tools to fulfil this double objective and has been extensively used under EU REACH.

- For example, in 2010, an average of 12 compounds for each of 27 metals (metalloids) were registered. If this number is multiplied by the number of human health endpoints for which data is required (e.g. ~10) we get an estimate of the extent of potential animal tests required: 3,200 tests. Industry did not undertake such extensive animal testing for its metal compounds. Instead, to reduce animal testing, industry applied read-across and grouping based on a weight-of-evidence approach that included bioelution test results (i.e. bioaccessibility data).

The EU has a key objective to ensure high quality data for chemicals and in this context is now checking the validity of the read across (e.g. RAAF 2017). **Having a recognised test method for the generation of bioaccessibility data will greatly contribute to the robustness of these assessments while greatly diminishing the need for animal testing.** This is also true for non EU jurisdictions[8]

In the case of **application 2-Alloy classification**, the current mixtures approach to the classification of alloys based on the bulk concentration of their classified metal ingredients does not work. It leads to situations in which an alloy that releases several-fold less metal in bioelution tests than predicted based on bulk concentration and release from the pure ingredient, is classified as hazardous for systemic effects, promoting the deselection of safe materials. The mixtures approach is also known to lead to underestimation of risk due to more metal release than predicted based on bulk concentration and release from the pure ingredient, potentially resulting in under-classification. Validated bioelution test methods could possibly lead to refinements in the classifications of alloys. The EU Commission has indicated that for bioelution test data to be considered in a regulatory application related to the classification of alloys, the data needs to be generated by an internationally recognized protocol that can

[7] The classification of alloys (as mixtures) under EU CLP is a self-classification by the companies placing them on the market. The use of the bioaccessible concentration for the refinement of the classification of alloys under the CLP is currently under consideration.

[8] For example, bioelution -based read across has been incorporated by Australia NICNAS in their IMAP chemical management programme.

guarantee the reproducibility and reliability of the results. **Without this protocol, there is no possibility to incorporate bioaccessibility data into the current EU CLP. The need for this protocol is global when UN GHS is considered.**

At present, there are many bioelution methods that look at metal releases in synthetic fluids from various materials, including soils (e.g. ASTM D5517 2007; BS EN71-3 2013; BARGE UBM 2016), but **there are no internationally recognised (e.g. EU or OECD) bioelution test methods specifically designed to assess metal releases relevant to systemic health effects after oral exposure**. [*While an OECD test guidance note exists for environmental hazard assessment of metal-containing materials, this method is not suitable for human health effects after oral exposure (OECD 2001)*].

The current proposal is to focus first on the validation of a fluid relevant to the oral route of exposure (i.e. gastric fluid). A test method is needed that will allow the generation of consistent and reliable metal release information in a fluid representative of the gastric compartment. This test method and the data it generates will contribute (in a weight- of-evidence approach) to improved grouping and read-across for hazard identification (systemic toxicity endpoints) of metal-containing materials (substances and mixtures) relevant to application 1 (Grouping). In addition to reducing the number of animal tests needed (e.g. to estimate bioavailability and/or toxicity when filling data requirements of registered substances under EU REACH), bioelution tests have lower costs compared to in vivo tests, are simpler in design, highly reproducible, and throughput is higher than for other tests (e.g. toxicokinetics), so testing can be done more quickly.

A validated gastric test method could eventually lead to more appropriate classifications of alloys for systemic toxicity endpoints secondary to oral absorption (application 2-Alloy classification).

1.1.5 Technical specifications

1.1.5.1 Protocol(s) of the test method

Please append the protocol(s) as used to generate the data submitted in this TST as Attachment 1a. Furthermore please attach the protocol(s) in the DB-ALM protocol[9] *format as Attachment 1b. Please specify how the protocol(s) relate to any existing DB-ALM protocol, if applicable, i.e. whether it is identical or differing from the DB-ALM protocol. In case of deviations, please outline the major components and/or steps that differ.*

Bioelution Testing of Metals, Inorganic Metal Compounds, and Complex Metal-Containing Materials: Simulated Gastric Fluid (November 2018, see Attachment1arev).

The protocol(s) should contain a list and description of the materials needed, a description of what is measured and how (specification of parameters/endpoints measured), as well as a (preliminary) prediction model (PM), if applicable. In addition, a brief description of the following key elements of the protocol(s) should be provided in the following paragraphs:

a) Brief description of the test system

Briefly specify the test system used (e.g. tissue model, specific cells grown to confluence, etc.).

The test system is the whole of the chosen test vessel (usually 250 mL Erlenmeyer flask) with the appropriate volume of test medium (a minimum of 50 mL). Test medium

[9] The so-called "DB-ALM" protocols are ECVAM's reporting format for the dissemination of a test method protocol via the ECVAM database web service on alternative methods http://ecvam-dbalm.jrc.ec.europa.eu/

corresponds to simulated gastric fluid (0.07N HCl), pH adjusted to pH 1.5 (37 °C ± 1 °C) in compliance with the ASTM 5517 standard.

b) Parameters and endpoints measured

Please specify the parameter(s) (e.g. optical density) and endpoint(s) (e.g. cell viability, EC_{50}) measured to make predictions and describe how this parameter(s) is/are measured.

The concentrations of released elements of interest are measured with an appropriate (validated) measurement method. In order to guarantee the basic quality (use of a validated measuring method, involvement of appropriately trained staff, traceability of the measurements, data processing and archiving,…) the sample analyses are preferably performed in a laboratory that works according to a standardised quality system (e.g. GLP, GMP, ISO17025, ...).

Examples of universal and commonly used method for measuring dissolved element concentrations (and thus recommended) are:

• ICP-MS (inductively coupled plasma-mass spectrometry); well documented in ISO 17294-1:2004 Water quality - Application of inductively coupled plasma mass spectrometry (ICP-MS) - Part 1: General guidelines and ISO 17294-2:2016 Water quality - Application of inductively coupled plasma mass spectrometry (ICP-MS) - Part 2: Determination of selected elements including uranium isotopes.

• ICP-AES (inductively coupled plasma-atomic emission spectrometry); well documented in ISO 11885:2007 Water quality - Determination of selected elements by inductively coupled plasma optical emission spectrometry (ICP-OES)

However, other validated analysis methods can also be applied (such as AAS (graphite furnace or flame -atomic absorption spectrophotometry).

Unless otherwise agreed, only 1 measurement result is reported for each analytical sample and element (in µg/L or mg/L).

Recent measurements of Ni (high & medium level) and Co (low level) in gastric fluid by GAAF or AAS-flame and ICP-MS have given very comparable results; mean values were within 10% of each other. Please see Table 2 below.

Sample	Co (µg/L)		Ni (µg/L)	
	GAAF	ICP-MS	GAAF	ICP-MS
1	< LOD (0.3)	0.05	<LOD (1.0)	0.30
2	<LOD (0.3)	0.05	<LOD (1.0)	0.33
3	0.5-<LOQ (0.6)	n.m.	36620*	32496
4	0.5-<LOQ (0.6)	n.m.	36180*	31954
5	5.55	5.89	130.0	123.3
6	5.50	5.96	111.4	122.4

*AAS-flame; n.m. not measured. Sample 1 and 2: gastric fluid medium (negative control); samples 3 and 4: 0.2 g/L Ni powder (high Ni release); samples 5 and 6: SS316L powder (medium Ni release, low Co release)

Table 2. Measurements of Ni (high & medium level) and Co (low level) in gastric fluid by GAAF or AAS-flame and ICP-MS.

In addition to the measurement results of the test samples, a reporting limit is always provided for each element of interest (limit of quantification, LOQ or at minimum the limit of detection LOD). Other analytical performance characteristics of the analysis (measuring range, repeatability, reproducibility, accuracy, measurement uncertainty) should be reported.

c) Quality criteria applied to the test system

Please specify the quality criteria applied to the test system (e.g. appropriate stratification and barrier function of a reconstructed human epidermis in each lot/batch).

The test system consists of vessel + gastric fluid test. A test can be assumed valid (i.e. satisfies quality criteria of test system) if:

-The values of pH and temperature (T) of the test medium at the start of the test fall within the set boundary conditions: i.e. pH 1.5 ± 0.1 and temperature at 37 °C ± 1 °C.
-The values of pH and T of the negative control at the end of the test fall within the set boundary conditions: i.e. pH 1.5 ± 0.1 and temperature at 37 °C ± 1 °C.

d) Positive control, negative control, benchmarks

Please indicate all concurrent controls used and specify if they are used as acceptance criteria for the run (a run consists of one or more test items tested concurrently with a positive and a negative control). Please include also the acceptance range for the control responses and, where available, any historical data used to establish the acceptance range.

Positive Control(s)
The positive control(s) should consist of a test item(s) well known to elicit a positive response in the test method.

Positive control: The positive control can easily be prepared, for example, by spiking certain amounts of the Certified Standard Metal solution into the test medium. Its purpose is to assure that the gastric test produces repeatable results. It provides evidence that the test system is responsive under the actual conditions of the assay. The positive control is run in parallel with the test materials of that metal or element. The positive control is independent of the application of the test data. The positive control does not need to match the physical form of the test materials. The positive control is predicted to give reproducible results over time. The positive control can also be used to assess the recovery of the metal ions through the test (mixing with test medium, incubation, filtering, sampling, etc). This solution shall be run at the time when the SOP for one or several metals of interest is established in a lab. Using this solution, high and low metal ion concentration would be added to the test medium at time = 0 and run through the protocol. Recovery of 100 ± 10% would be considered acceptable.

Negative Control(s)
*The negative control(s) can consist of treatment with the vehicle used and/or a test item known **not** to elicit a positive response in the test method.*

The **negative control** vessel contains only the test solution (e.g. artificial gastric fluid). The negative control measurements of the elements of interest after the incubation period should generally be below < 20 x SD of the negative controls for

the lab (based on 10 or more replicates). Higher values would indicate possible contaminations of the test system, glassware, etc.

Benchmarks *(if applicable)*
Benchmarks consist of test item(s) that produce a midrange response in the test method.

Proficiency materials are the materials used to assess the transferability of the protocol to a new laboratory. Proficiency materials are well characterized and defined materials (e.g. regarding elemental composition and particle size) that may or may not include the metal of interest but need to match the physical form of the test materials (e.g. samples of powder metals or metal compounds, embedded and non-embedded massive forms of alloys). These materials are tested initially by the lab when establishing the SOP and are run periodically afterwards (at least once a year). Proficiency materials are independent of the application of the test data. A physical repository and Shewart charts with test results will be made available. Initially, the database will have at least 10 measurements for each proficiency material collected by at least 2 different labs.

A possible list of proficiency chemicals is provided in Attachment 19rev. For the these materials, measurements of the elements of interest should have a recovery of ±20% of reported values[10].

e) Acceptance criteria applied to the results

Please specify the acceptance criteria for the experimental data.

Any visual observations of e.g. precipitation, colouration, decolouration that differ between replicate test vessels or are consistent for all three replicates, should be recorded as they can help explain the bioelution results.
-
The test results are considered acceptable if the following criteria are met:

- **A LOQ is available for each metal of interest (or at a minimum the LOD)**
- **the between test material vessels' Coefficient of Variation (CV) is ≤ 40% for each metal of interest in ≤ LOQ – 10x LOQ range, or CV ≤20% for measurements in 10x LOQ – end of measuring range,**
- **accurate metal measurements (± 10% of reported values) are obtained from a positive control solution that contains the metal of interest**

When the coefficients of variation values of test materials exceed the recommended values (leading to data rejection), the proficiency material should be run. If the values for the proficiency materials are met but not for the test material, a thorough review of the sample characteristics should be conducted before it is attempted to repeat the study; in some instances, the large variability in metal release could be linked to the intrinsic nature of the sample. Some materials (e.g. UVCBs) are very complex and can, despite being from the same batch, have a wide range in the composition and also in particle size which will result in significant variations within a sample and between "identical" samples. Large variations will therefore provide important information on the test material but this information needs to be combined with proper physical-chemical

[10] A publicly available database will be set up that will include bioelution measurements for proficiency materials. These measurements will be collected using the gastric SOP, at two or more labs. New measurements will continue to be added after a proper curating process.

characterization of the test material, e.g. SEM-EDS, XPS, IR, BET, size distribution, etc.
The average negative control measurements that exceed the LOQ should be subtracted from test sample values. Contamination is suspected when negative control values are > 20 x SD of the mean negative control value for the lab (based on 10 or more replicates). In the latter case, a thorough review of all the test data should be conducted (including ratio of test sample value to negative control value) before deciding whether the study should be repeated.

1.1.5.2 Data analysis and Prediction Model (PM)

Please specify how the raw data are processed and analysed. Please describe the (preliminary) procedure to translate the test method results into predictions of effects on human health or the environment. This procedure ideally should be a (preliminary) PM.

The raw data generated by the gastric bioelution test are reported as mass of soluble metals that is released per volume of solution (µg/L), at a given loading (mass of sample added per volume of solution) and time point (2 h).

The results can also be reported as metal mass released/mass of sample (e.g. µg metal/g of sample) by taking the loading into account. When the metal content of the sample is known, it is possible to calculate the µg metal released /g of metal in the sample or the percentage of metal content that is released (at a given loading and time point). For samples of metals and alloys (zero valence), the release of soluble metals can also be expressed per unit of surface area.

The relationship between gastric bioaccessibility and bioavailability needs to be considered when translating bioelution test data into predictions of similar bioavailability. Once this has been established for a particular metal, based on oral ingestion of some metal-containing substances (e.g. compounds or soils), this relationship is valid for other substances containing that same metal (e.g. alloys).

Attachment 18 describes the relationships between oral bioaccessibility (including gastric fluid) and relative oral bioavailability of metals and metalloids. Studies that looked at the ability of oral bioaccessibility to predict in vivo bioavailability of metals from soils tend to report the information as % in vitro bioaccessibility (IVBA) and % in vivo relative bioavailability (RBA). From knowledge of the loading (mass sample/L) and measured metal mass released per liter of solution (e.g. mg As/L), the % IVBA can be calculated [e.g. (mg As/L) / (mg sample/L) x 100 = % As IVBA]. The RBA requires knowledge of the bioavailability of a metal from soils (or other substances-mixtures) and the bioavailability of that metal from a water soluble compound, both tested in animal studies (e.g. area under the curve). In general, it is difficult to find the absolute bioavailability of metals from soils reported in the published manuscripts; the RBA values are usually provided.

As an example of how data from soil studies are presented, Diamond et al. 2016 conducted a meta-analysis looking at the relationship between the in vitro bioaccessibility (IVBA) and the in vivo relative bioavailability (RBA) of arsenic from soils. The samples were authentic site samples; IVBA was assessed with the 0.4 M glycine/pH 1.5 assay; and bioavailability was assessed in either C57BL/6 mice or juvenile swines. Two types of regression analyses were performed on the data. A

preliminary regression model relied on ordinary least squares linear regression, while the final regression modelling used weighted least squares regression. This approach allowed incorporation of uncertainty in individual IVBA and RBA estimates into the regression model. Figure 1 below corresponds to Figure 1 of Diamond et al. (2016). The solid line represents the weighted regression model and the dotted lines are the 95% prediction limits for the weighted model. The authors conclude that the data from IVBA can predict RBA of arsenic in soil.

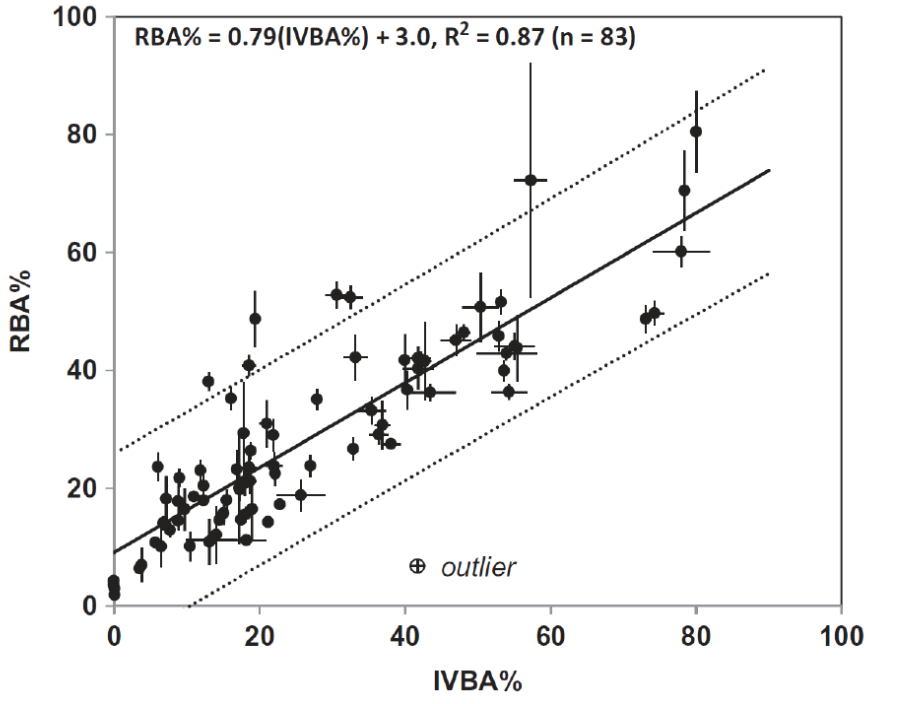


Figure 1. Percent RBA as a function of %IVBA for arsenic (corresponds to Figure 1 of Diamond et al. 2016).

Strong predictive relationships between RBA and IVBA emerged from the review for Pb and As, with fewer available data for Sb, Cd, Co, Ni, and Hg, and no data for Be and Se. For Cd, Co, Pb and Ni, higher IVBA are achieved with lower pH conditions and this effect of pH outweighs other components of the extraction fluid that are intended to create a more "physiological" environment. For As, solubility from soil appears to have a lower dependence on pH, although for some soils, low pH is associated with higher IVBA. Arsenic results from the low pH extraction methods are more closely correlated with RBA when studied using paired soils. In many cases the IVBA values are higher than the RBA ones and in other cases they are equal or lower. **In all cases studied, a relationship between IVBA and RBA was established** (see Attachment 18 for a detailed report). **The data is considered robust enough to support applications 1 (Grouping) and 2 (Alloy classification).**

Application 1-Grouping.

When reference and target samples are tested at equal loadings and times in the same solutions, metal releases per volume of solution (e.g. µg metal/L) can be directly compared (e.g. for grouping and read-across), and no further modification of the data is

needed. In some cases, when very high or very low releases are expected, different loadings may be applied to the reference and target samples. In this case, each loading can be considered to express the results as mass of released metals / mass of sample. Metal release per mass of sample from reference and target substances and mixtures can then be compared. With knowledge of the concentration of a metal in a sample, it is also possible to calculate the percent of metal content that is released.

When applying bioelution test data from for grouping and read-across, relative bioaccessibility data is considered in a weight-of-evidence approach to assess if target and reference substances are predicted to have similar or different oral systemic toxicities.

Example. Data from Henderson et al. (2012) can be used to illustrate how gastric bioaccessible data from metal compounds can be used for grouping and read-across in a weight-of-evidence approach. Table 3 shows 'absolute' bioaccessibility values (µg Ni/g sample, column 3) and 'absolute' values as percent of Ni content released (column 4)[11] in gastric fluid for different nickel-containing compounds of variable water solubility. Direct comparison of µg Ni/g sample would allow grouping of the first ten compounds together; with the category of green Ni oxides showing lower metal release per gram of sample. Likewise, comparison of % Ni released of Ni content allows discrimination into 2 main groups of compounds: high solubility (>80% of content) and low solubility (<80% of content), as indicated by the shading in Table 3. The low solubility group could also be subdivided into less water soluble sulphides and oxides releasing 5-30% of content, and green Ni oxides releasing < 0.5% of content.

Test substance	Sample ID	Gastric bioaccessibility- 2 h (µg Ni/g sample)[a]	Gastric bioaccessibility- 2 h (% Ni content released)[b]
Ni sulfate hexahydrate[d]	N58	201,670	90.55
Ni chloride hexahydrate	N98	215,632	89.85
Ni acetate tetrahydrate	N103	209,810	88.50
Ni hydroxycarbonate	N109	408,690	84.30
Ni sulfamate tetrahydrate	N104	147,783	83.40
Ni fluoride tetrahydrate	N111	264,930	82.35
Ni oxide (black)	N105	222,328	29.60
Ni hydroxide	N106	142,693	26.30
Ni subsulfide[d]	N18	158,185	22.65
Ni sulfide	N97	57,485	9.75
Ni oxide (green)[d]	N9, N46[e]	2540	0.33
Ni oxide (green)[d]	N112	<25	<LOQ[f]

NC, not completed.
[a] Bioaccessibility reported as µg Ni released per gram of sample. Reported values are mean values from duplicate experiments.
[b] Bioaccessibility reported as the percent of available Ni content released in synthetic gastric fluid after 2 h. Reported values are me
Different shading indicates a possible pre-grouping of the samples based solely on % Ni content released in gastric fluid after 2 h.

Table 3. Gastric bioaccessibility data for nickel compounds (corresponds to Table 4 of Henderson et al. 2012).

Yet, bioaccessibility is only one of the elements of read-across. Table 4 below compares the percent of Ni content released in gastric and intestinal fluids to the in vivo absorption in rats. The results are consistent with the gastric bioaccessibility data as they show that oral absorption of water soluble Ni compounds is much higher than that of insoluble ones.

[11] Because the absolute bioaccessibility of water soluble compounds is close to 100%, the absolute values also reflect the relative values when using water soluble compounds as a reference.

Oral bioaccessibility and in vivo absorption and relative bioavailability of nickel compounds

Test Substance	Gastric Bioaccessibility after 2 hours (% Ni content released)	Intestinal Bioaccessibility after 24 hours (% Ni content released)	Absorbed Fraction after 24 hours (%)[a]	Relative Bioavailability after 24 hours (%)[a]
Ni sulphate	90.55	58.00	11.12	100
Ni chloride	89.85	38.45	9.8	88
Ni oxide (black)	29.6	0.32	0.04	0.36
Ni oxide (green)	9.75	0.12	0.01	0.09
Ni subsulphide	22.65	0.25	0.47	4.2
Ni sulphide	9.75	0.18	2.12	19

[a] Ishimatu et al. (1995).

Table 4. Oral bioaccessibility and in vivo absorption data for nickel compounds (corresponds to Table 1 of Attachment 2 in Attachment 18).

As an example, using these information in a weight-of-evidence approach together with other phys-chem, toxicokinetic and acute oral toxicity data, Ni sulfamate was grouped with the water soluble Ni sulfate and Ni chloride compounds and read across from these compounds a reproductive toxicity classification (Henderson et al. 2012).

Application 2-Alloy classification.

When data on released amounts of metals/ mass of sample from alloys and reference metal ingredients are considered, the bioaccessible concentration (BC) of the metal in the mixture can be calculated using the following equation.

$$BC\% = \frac{\mu g\ metal\ ion\ released\ (measured\ in\ extract)/g\ alloy\ sample\ tested}{\mu g\ metal\ ion\ released\ (measured\ in\ extract)/g\ pure\ metal\ sample\ tested}\ x\ 100$$

When applying bioelution test data from alloys to refine a metal's bulk concentration, the bioaccessible concentration is calculated and compared to the bulk concentration in order to assess the presence of a matrix effect.

Example: data from Hillwalker and Anderson (2014) paper is used to illustrate the ways in which data from gastric bioelution tests are reported for metals and alloys and how to calculate the bioaccessible concentration of a metal (Mn) in an alloy (carbon steel) using these data. The bulk concentration of Mn in carbon steel is 0.456%. Table 5 presents metal release data (metal bioaccessibility) from one alloy (carbon steel) in terms of µg metal/mL fluid (e.g. average 0.84 µg Mn/mL).

Supplementary Information

Table S2. Metal Concentration in Biofluid Extract (µg metal in extract / mL fluid)

Alloy / [a]grade	Biofluid	Cobalt	Chromium	Copper	Iron	Manganese	Nickel	Zinc
Carbon steel NIST 14g	gastric	<0.0054	<0.0038	<0.0055	186	1.0	<0.015	<0.016
					158	0.79		
					157	0.79		
					170	0.89		
					147	0.75		

Table 5. Metal concentration in gastric fluid extract (corresponds to Table S2 of Hillwalker and Anderson 2014).

Data on Mn bioaccessibility from carbon steel in terms of µg metal/g sample, corresponding to a loading of 2 g sample/L fluid is reported in the Table 6 (e.g. average 420 µg Mn/g carbon steel, calculated as 0.84 µg Mn/mL / 2 g/1000 mL= 420 µg Mn/g).

Supplementary Information
Table S3 Metal Concentration in Material (µg metal in extract / g alloy material); n = 2 to 5

Alloy / [a]grade	Biofluid	Cobalt Avg±SD	Chromium Avg±SD	Copper Avg±SD	Iron Avg±SD	Manganese Avg±SD	Nickel Avg±SD	Zinc Avg±SD
Carbon steel NIST 14g	gastric	<2.7	60 ± 7	<2.8	8.2 ± 0.7	420 ± 52	<7.5	<8.0

Table 6. Metal concentration in samples (corresponds to Table S3 of Hillwalker and Anderson 2014).

Table 7 presents data on metal bioaccessibility for the pure Mn in terms of µg Mn/mL fluid at a loading of 2 g/L. An average release of 1467 µg Mn/mL corresponds to an average release of 733,333 µg/g of Mn[o] sample (calculated as 1467 µg Mn/mL / 2 g pure Mn/1000 mL= 733,333 µg Mn/g pure Mn).

Table S2. Metal Concentration in Biofluid Extract (µg metal in extract / mL fluid)

Alloy / [a]grade	Biofluid	Cobalt	Chromium	Copper	Iron	Manganese	Nickel	Zinc
Mn metal					135	1300		
	gastric	<0.0054	<0.0038	<0.0055	135	1570	<0.015	<0.016
					126	1530		

Table 7. Manganese bioaccessibility in gastric fluid extract (corresponds to Table S2 of Hillwalker and Anderson 2014).

This information can be used to calculate the bioaccessible concentration of Mn in the carbon steel alloy as follows:

$$BC\% = \frac{\mu g\ Mn\ ion\ released\ (measured\ in\ extract)/g\ \ carbon\ steel\ tested}{\mu g\ Mn\ ion\ released\ (measured\ in\ extract)/g\ pure\ Mn\ metal\ tested}\ x\ 100$$

$$BC\% = \frac{420}{733{,}333}\ x\ 100 = 0.057\%$$

The bioaccessible concentration indicates that carbon steel alloy behaves as a mixture containing 0.057% of Mn metal in it. Since the content of Mn in the alloy is 0.456%, there is a matrix effect in the alloy that decreases Mn release (~8-fold) compared to what could be expected from a simple mixture. This example illustrates how the bioaccessible concentration of a metal can differ from its bulk concentration due to alloy matrix effects.

1.1.5.3 Expertise required for performing the test method protocol

Please describe the level of expertise and demonstrated proficiency required by the study personnel for performing the test method protocol.

The test method protocol is easily applicable in a standard laboratory environment as long as a well-defined SOP (Attachment 1a) is available; no special expertise is required to test powders (defined as being <1 mm diameter) and unmodified massive samples (defined as

being ≥1 mm in diameter, as sheets, pellets, foils, etc.). The proposed protocol is not directly applicable to nano materials.

To test massive forms, special equipment and some experience is required when epoxy-embedded samples are prepared and used (see guidance on sample preparation included in Attachment 1arev).
One must always be very aware of (cross) contaminations. Contaminations by easily soluble metal compounds or present as trace elements in chemicals, or in non-acid cleaned test vessels can adversely affect the results of sparingly or insoluble test items

The concentrations of released elements of interest are measured with an appropriate (validated) measurement method. In order to guarantee the basic quality (use of a validated measuring method, involvement of appropriately trained staff, traceability of the measurements, data processing and archiving,…) the sample analyses are preferably performed in a laboratory that works according to a standardised quality system (e.g. GLP, GMP, ISO17025, ...).).

Accordingly, a test report is prepared for each test material and shall include (but is not limited to) the following information:

- **Information on the lab/study**: Identification of the study, the sponsors, the test facility, and study director
- **Description of the test material**

- Name of the test material, including identifiers such as CAS/EINECS numbers when relevant
- Batch number and/or sampling date of the test material
- Expiration date of the test material when applicable
- Details of supplier(s)/manufacturer/importer (e.g. address, contact information)
- Information available on composition (e.g., Certificate of Analysis), at least all substances > 0.1 wt% to assess purity of the sample and nature of impurities (% or ppm)
- Information available on method of manufacture (particularly for powder forms) indicating if atomization (water or air), chemical or grinding method was used
- Information on processing, preparation, surface finish, transport and storage, if available (e.g. heat treatment/fabrication history; metallurgical microstructure particularly for massive forms)
- Information on specific surface area, surface charge (particles), surface composition, and other surface properties (e.g. surface oxidation for both massive and powder forms; surface finish particularly for massive forms), if available
- Particle size distribution (powders only) using parameters such as d0.1, d0.5, and d0.9. Report method of assessment and if values are based on volume or number distribution and in which media (e.g. composition, ionic strength, pH) the size distribution was measured (e.g. air, water)
- If surface charge information is available, include the medium used to measure the surface charge (e.g. composition, ionic strength, pH)- Information on morphology (shape of the particles) and metallurgical structure
- Information on particle agglomeration behaviour in solution
- Specific surface area in m^2/g (powders only)
- Surface area (massive only)

- If the sample (powder) was sieved, information about the original sample could also be provided

- Description of the test system and loading: the following information should be provided

- Test material weight, final loading and final volume
- pH and temperature at the start and the end of the test
- Detailed descriptions of the test apparatus and procedure, including analytical instruments used
- Surface area to volume ratio for massive materials

- **Results from the analyses of metal concentrations**

- Date of test
- LOQ for each metal of interest (or at a minimum the LOD)
- metal release (µg/L or mg/L) (measured element concentrations in the samples, including negative and positive controls, test material samples, and/or proficiency material samples)
- metal release per g test material (calculated mg/g)
- percent of total element released (calculated %)
- element release per unit surface area for metals and alloys (calculated mg/m2)
- description of quality assurance program applied

- **Deviations from this Standard Operating Procedure, if any and reason(s) for such**
- **Any circumstances that may have affected the results**
- **Reference to the records and all raw data necessary to reconstruct the study**

1.1.6 Test items used for developing and optimising the test method (protocol and PM)

Please indicate the number and basic physical/chemical properties of the test items used to develop and optimise the test method. Please append the full list of these test items in the form of a table, including their CAS numbers and basic physical/chemical properties and acquired data, as Attachment 2.

A synthetic gastric fluid similar to that used in the development and optimization stages, as well as in the validation study (as discussed in other sections of the TST) of the present method, has been used extensively. One example is the Comité Européen de Normalisation standard, Safety of Toys (BS EN 71-3 2013), which has been adopted in the United States as ASTM D5517 (2007; Standard Method for Determining the Solubility of Metals in Art Materials). The ASTM D5517 method served as the basis for the work further developed specifically for metals by Angeline Kirby Memorial Health Center (Wilkes-Barre, PA, USA). Its application to assess the bioaccessibility of various cobalt and nickel compounds was published by Stopford et al. (2003) and Henderson et al. (2012), respectively.

Variations of the ASTM D5517 method have been used to measure Cu, Cr, Fe, Mn, Mo, Si, Ti, Zn, Zr releases from compounds or alloys (Hedberg et al. 2012, 2016; Mörsdorf et al. 2015; Herting et al. 2014; Jiang et al. 2012; Hedberg and Midander 2014; Midander et al. 2010; Hillwalker and Anderson 2014). In addition, several unpublished datasets for other metals or metal-containing samples (Rh, Ru, Pt, Pd, Ag, Au, Re, Cu, In, Ge, Cd, Pb, Si, Al, Zr, Sb, V, Sn) were generated during this time period, further contributing to the development of the protocol. A list of the types of test items used in the development of the method is provided in Attachment 2. The method and simulated gastric fluid composition in the validation study were selected based on the outcome of these previous exercises; in general, those previously published by Stopford et al. (2003) and the KTH Laboratories (Hedberg et

al. 2010a; Midander et al. 2007a, b; Herting et al. 2006; Midander 2006) served as the basis of developing the SOP for the validation study.

1.1.7 Cost and time estimates per test item

Please give an estimate of the testing cost per test item, considering that the laboratory is equipped with all necessary standard equipments and not considering labour cost. Please indicate as well an estimation of the time required to complete data acquisition for a run and specify how many chemicals can be included in a typical run.

Testing of powders and/or unmodified or epoxy-embedded massives (directly applicable as received from the sponsor[12]):
-estimation of the cost per test material (including 2 loadings per test material, with triplicate samples per loading, with two measurements per sample); one negative control sample with two measurements; and one positive control sample with two measurements: Average 950€ (test vessels, chemicals for cleaning, testing, sampling filters, syringes, sample tubes, labels, analyses, waste disposal)
-estimation of the time required to complete data acquisition for a run: 5-7 days
-Typically up to five test items can be included in a typical run

1.1.8 Occurrence of non-qualified tests

On the basis of your experience/historical data please provide:
a) an estimate (e.g. in percentage) of the frequency in occurrence of non-qualified tests (i.e. tests which do not meet the acceptance criteria),
b) an average of the number of tests which usually have to be performed to acquire the requested number of qualified tests (as described in the protocol or prediction model).

The test results are considered acceptable if the following criteria are met:

- A LOQ is available for each metal of interest (or at a minimum the LOD)
- the between test material vessels' Coefficient of Variation (CV) is ≤ 40% for each metal of interest in ≤ LOQ – 10x LOQ range, or CV ≤20% for measurements in 10x LOQ – end of measuring range,
- accurate metal measurements (± 10% of reported values) are obtained from a positive control solution that contains the metal of interest

Notes:
When the coefficients of variation values of test materials exceed the recommended values (leading to data rejection), the proficiency material should be run. If the values for the proficiency materials are met but not for test material, a thorough review of the sample characteristics should be conducted before it is attempted to repeat the study; in some instances, the large variability in metal release could be linked to the intrinsic nature of the sample. Some materials (e.g. ores and concentrates) are very complex and can, despite being from the same batch, have a wide range in the composition and also in particle size which will result in significant variations within a sample and between "identical" samples. Large variations will therefore provide important information on the test material but this information needs to be combined with proper physical-chemical characterization of the test material, e.g. SEM-EDS, XPS, IR, BET, size distribution, etc.

[12] If the embedding of massive forms in epoxy needs to be conducted at the receiving lab prior to bioelution testing, the testing cost per test material would be 1750€ (epoxy resin, sanding and polishing consumables, test vessels, chemicals for cleaning, testing, sampling filters, syringes, sample tubes, labels, analyses, waste disposal).

The average negative control measurements that exceed the LOQ should be subtracted from test sample values. Contamination is suspected when negative control values are > 20 x SD of the mean negative control value for the lab (based on 10 or more replicates). In the latter case, a thorough review of all the test data should be conducted (including ratio of test sample value to negative control value) before deciding whether the study should be repeated..

To assess the occurrence of non-qualified tests (tests not meeting the acceptance criteria defined above), data from two published manuscripts were examined. Hillwalker and Anderson (2014) noted: "Recoveries of fortification samples and [instrument] check standards were from 85 to 110% and 93 to 109%, respectively." This suggests that the data acceptance criterion for the positive control (±10% or between 90-110% of value) was likely met. The second study, Henderson et al. (2014) did not include a positive control.

Hillwalker and Anderson (2014) reports bioelution data in gastric fluid (2h) for 7 metals (Co, Cr, Cu, Fe, Mn, Ni and Zn) released from 7 different materials (5 alloys and 2 pure metals). Triplicate test samples were included and mean values ± SD were reported. Method LOD values for the metals were provided as µg/L and µg/g sample. LOQ values were not reported. From the reported LOD values, LOQs were calculated as 2 x LODs. Negative control values were not reported in the manuscript and thus the possible presence of contamination could not be evaluated. A total of 63 measurements were reported; 36 were < LOD and 27 > LOD. Of the values > LOD, 8 values were in LOQ – 10x LOQ range and 19 values were in 10x LOQ – end of range. To calculate the %CVs, the SD was divided by the mean value and multiplied by 100. Of all the 63 measurements, 6 (9.5%) had CVs that did not meet the acceptance criteria; of the 27 values > LOD, 6 (22%) had CVs that did not meet the acceptance criteria.

Henderson et al. 2014 reports bioelution data in gastric fluid (2h) for 7 metals (Co, Cr, Cu, Fe, Ni, Pb and Zn) released from 6 different materials (2 compounds, 2 alloys, 1 concentrate, and 1 pure metal), for a total of 10 metal-sample combinations. [*The data corresponding to the release of Cr from Inconel in gastric fluid was not part of the round-robin analysis since 1 lab had 2/3 measurements < LOD and another lab had 1/3 measurements < LOD. Thus, the data for this sample is not included in the analyses below*.] For the other 9 metal-sample combinations, triplicate test samples were included, mean values +/- SD were reported. Independent data are available from 5 different labs, with the exception of Lab C that has data for only 7 metal-sample combinations. LOD values for the metals were provided as µg/g sample. LOQ values were not reported. From the reported LOD values, LOQs were calculated as 2 x LODs. To calculate the %CVs, the SD was divided by the mean value and multiplied by 100.

A lab-specific analysis is described below:

- Lab A had a total of 9 metal-sample means reported, representing 27 individual values: no individual values < LOD; 1 mean value was in ≤LOQ – 10x LOQ range and 8 mean values were in 10x LOQ – end of range. None of the mean values (**0%**) had CVs that were above the listed criteria.
- Lab B had a total of 9 metal-sample means reported, representing 27 individual values: no individual values < LOD; 1 mean value was in ≤LOQ – 10x LOQ range and 8 mean values were in 10x LOQ – end of range. None of the mean values (**0%**) had CVs that were above the listed criteria.

- Lab C had a total of 7 metal-sample means reported, representing 21 individual values: no individual values < LOD; all 7 mean values were in 10x LOQ – end of range. One of the 7 mean values (**14%**) had a CV that was slightly above the listed criteria. For the other 2 metal-sample combinations no values were reported.
- Lab D had a total of 9 metal-sample means reported, representing 27 individual values: no individual values < LOD; all 9 mean values were in 10x LOQ – end of range. One of the 9 mean values (**11%**) had a CV that was slightly above the listed criteria.
- Lab E had a total of 9 metal-sample means reported, representing 27 individual values: no individual values < LOD; 2 mean values were in LOQ – 10x LOQ range and 7 mean values were in 10x LOQ – end of range. Three of the 9 mean values (**33%**) had CVs that were above the listed criteria.

In total, 5 of 43 values (**12%**) had CVs that did not meet the acceptance criteria based on the Henderson et al. 2014 data.

Negative control values were not reported in the manuscripts. However, negative control values were included in a few of the lab reports for the round robin exercise reported in Henderson et al. 2014. An analysis based on these limited data is shown below:

- Lab A reported 3 negative controls/each for Fe, Pb and Zn and 6 negative controls/each for Co, Cu and Ni (total of 27 negative controls). Of these values, 26 were lower than 20 x SD of the corresponding metal negative control and one value was reported as outlier (cannot be evaluated).
- Lab B reported 2 negative controls/each for Co, Cu, Pb, Zn and Fe and 4 negative control for Ni (total of 14 negative controls). Of the 14 negative controls, 5 were higher than 20 x SD of the corresponding metal negative control (2 Co, 2 Ni and 1 Zn); only 3 of these values were higher than 21 x SD of the negative control . The 5 values > 20 x SD of the negative controls corresponded at most to 0.6% of the test sample values and were not considered to invalidate the results from these assays.
- Lab D. Release data for negative control samples were only recorded as <250 µg/g sample" for all relevant measurements except one, Ni from Inconel alloy (recorded as <25 µg/g sample); no additional data was provided.
- Lab E reported 2 negative controls/each for Fe, Pb, and Zn and 4 negative controls/each for Co, Cu and Ni (total of 18 negative controls). Three of these values (1 Co, 2 Zn) were higher than 20 x SD of corresponding metal negative control . One of the 3 values > 20 x SD of the negative control corresponded to 0.3% of the test sample values and were not considered to invalidate the results from these assays; two (duplicate negative controls) of the negative controls, however, corresponded to 50% of the test sample value and would warrant a repeat under the current SOP acceptance criteria.

Overall, of the 58 negative control measurements reported, 8 (**14%**) were > 20 x SD of the negative controls; based on present data, and only 2/58 (**3.5%**) would have warranted a repeat of the test.

While the information provided by these two studies is not very extensive (particularly regarding the positive control), the limited positive control data, CV and negative control analyses suggests that the proposed data acceptance criteria is reasonable.

The variability in metal measurements from epoxy embedded massive samples can be gleaned from the studies reported in Table 8A and B below. Table 8A includes results from triplicate measurements of two different alloy samples and two different metals (Ni and Cu). Results show good repeatability with one exception (CV 14%). Table 8B includes the results for Cu release from alloy CW602N when tested in the same lab (CIMM, Chile)[13] in 2010 and in 2015, and in a second lab (ECTX, Belgium) in 2018. When comparing results at the same loading (2 g/L) for the two labs, results show good reproducibility with values of 203 and 262 mg Cu/g massive alloy sample. In all cases 3 independently prepared samples of the material were tested.

A. Repeatability

	Samples and metals measured (µg /L) at a loading of 2 g/L				
	CW602N alloy 1h-Cu	CW602N alloy 2h-Cu	CW602N alloy 4h-Cu	SS316L 2h-Co	SS316L 2h-Ni
Replicate 1	177	385	966	0.126	2.94
Replicate 2	175	385	995	0.122	3.1
Replicate 3	224	447	1092	0.133	3.28
Mean	192	406	1018	0.127	3.11
CV%	14*	9	7	5	5
Reference	ECTX X02a-146			ECTX X02a-081	

*Does not pass the repeatability < 10% criterion

B. Reproducibility

	Mean 2 h copper values for alloy CW602N from 5 assays conducted at different loadings and at 2 different labs over an 8-year period				
Year of study	2010	2010	2015	2015	2018
Loading (g/L)	0.2	2	0.25	3.9	2
Metal release: µg copper/g alloy	217	262	462	229	203
CV%	5	3	19*	7	9
Reference	CIMM, 2010	CIMM, 2010	CIMM, 2015	CIMM, 2015	ECTX X02a-146

*Does not pass the repeatability < 10% criterion

[15] Personal communication from S. Baken of EU Copper Alliance.

Table 8. Metal releases from embedded massive samples in gastric fluid.

1.1.9 Known limitations and drawbacks of the test methods

Please specify any known limitations of the test method beyond those addressed by the Applicability Domain (Module 6). Examples of possible limitations and drawbacks:

- *The test method requires expensive equipment and/or is based on an expensive test system*
- *Equipment / test system needed is no longer commercially available*
- *Etc.*

To test powders and massive forms without further modification, few limitations and drawbacks of the test method are known if the sponsor can provide test-ready material. The method is applicable with standard laboratory equipment. Large vessels (up to 1 L) may be needed to accommodate testing of massive forms.

However, if the test materials need to be prepared (epoxy embedding and polishing to prepare massive samples) additional specialised equipment is required (e.g. polishing machine). This step is described in the Guidance Note provided with the SOP.

One important limitation is that nanomaterials as defined by EU Commission (2011) are not covered. Furthermore, to avoid/reduce contamination, working in a clean environment is necessary

1.1.10 Intellectual Property Rights (IPRs)

Please state whether any component of the test method (e.g. protocol, test system, equipment) is patented, copyright protected, trade-marked or registered. Please specify who is holding the IPRs.

To our knowledge, no component of the test method is copyright protected

1.1.11 History of test method development

Please provide any information on the process of developing the test method that might be of relevance for its validation.

In the last 10 years there have been a variety of bioelution tests in simulated gastric fluid. These tests have been employed in different applications (e.g. the ASTM 05517 protocol has been applied to assess the migration of metals from arts materials; EN 71-3 is a test method developed to assess metal release from swallowed toy parts and it is specified in the Toys Directive). In 2003, Stopford et al. applied a variation of the ASTM 05517 test to the releases of cobalt from various cobalt-containing substances. Since that time, many metal commodities adopted this protocol and applied it to their metal-containing materials (e.g. Henderson et al. 2012 for nickel, Midander et al. 2010 for FeCr alloys, Hedberg et al. 2012 for TiO_2 and Zr, Hedberg et al. 2010b for Sb substances; Hedberg et al. 2016 for Inconel).

In 2014, a round robin was conducted in which different metal compounds and alloys were tested at 5 different laboratories using a variation of the currently proposed protocol. Some of the metal-containing substances already tested using this method are those registered, and still to be registered in 2018, under EU REACH (including metals, metal compounds, UVCBs ranging from ores and concentrates, intermediates in metal production to pigments). Metal-containing complex mixtures such as alloys constitute a huge number of mixtures which have been required to be self-classified under EU CLP since 1 June 2015.

In parallel to the application of bioelution tests during the registration or classification of various metal compounds and complex substances, bioelution tests for the oral route of

exposure (including gastric compartment) have been applied to the assessments of human exposures to metal compounds and minerals in soils (e.g. Ruby et al. 1999). This research has been conducted by independent, government, and scientific research groups (e.g. Wragg et al. 2011). USEPA Validated Test Method 1340 (one of SW-846 guidance methods) assesses the bioaccessibility assay for Pb in soil. Validation of an In Vitro Bioaccessibility Test Method for Estimation of Bioavailability of arsenic from Soil and Sediment (ESTCP Project ER-200916 2012; Brattin et al. 2013). The UBM (BARGE 2016) met the criteria on repeatability (median relative standard deviation value <10%) and the regression statistics (slope 0.8 to 1.2 and r-square > 0.6) for As, Cd, and Pb. The proposed test protocol is similar to BARGE's gastric test. The learnings from the development and validation of the OECD N° 29 Guidance note have also been helpful in building up the experience for the development of the SOP and the validation for the current bioelution gastric test.

A summary of the composition of the various artificial oral fluids is shown in Table 9 below (see Attachment 18).

Table 6. Composition and parameters utilized in common in vitro bioaccessibility assays

Method / phase	ASTM D5517-4	BARGE UBM	DIN	EN 71-3	IVG	PBET	RBALP	SBRC
		Saliva–gastric		"Maceration"				
		Saliva (pH 6.5±0.5):		for oily or waxy samples				
In vitro parameters								
Composition (g l−1)		0.896 g KCl, 0.888 g NaH2PO4, 0.2 g KSCN, 0.57 g Na2SO4, 0.298 g NaCl, 1.8 ml of 1 M NaOH, 0.2 g urea, 0.145 g amylase, 0.05 g mucin, 0.015 g uric acid		water				
pH		1.2		not specified				
Soil/ solution ratio		0.6:22.5		1:25				
Extraction time (h)		1		1 min, follow by gastric				
	Gastric	Gastric phase	Gastric	Gastric	Gastric	Gastric	Gastric	Gastric
In vitro parameters								
Composition (g l−1)	0.7N HCl	2.752 g NaCl, 0.266 g NaH2PO4, 0.824 g KCl, 0.4 g CaCl2, 0.306 g NH4Cl, 8.3 ml of 37 % HCl, 0.65 g glucose, 0.02 mg glucuronic acid, 0.085 g urea, 0.33 g glucosaminehydrochloride, 1 g bovine serum albumin, 3 g mucin, 1 g pepsin	1 g pepsin, 3 g mucin, 2.9 g NaCl, 0.7 g KCl, 0.27 g KH2PO4	0.7N HCl	10 g pepsin, 8.77 g NaCl	1.25 g pepsin, 0.5 g sodium malate, 0.5 g sodium citrate, 420 µl lactic acid, 500 µl acetic acid	0.4M Glycine	0.4M Glycine
pH	1.0 to 1.5	1.2	2.0	1.0 to 1.5	1.8	2.5	1.5	1.5
Soil/ solution ratio	1:50	0.6:22.5	1:50	1:50	1:150	1:100	1:100	1:100
Extraction time (h)	1shaken, 1 stationary	1	2	1shaken, 1 stationary	1	1	1	1
		Gastric–intestinal	Intestinal		Intestinal	Intestinal	Intestinal	Intestinal
		Duodenal phase (pH 7.4±0.2)						
In vitro parameters								
Composition (g l−1)		7.012 g NaCl, 5.607 g NaHCO3, 0.08 g KH2PO4, 0.564 mg KCl, 0.05 g MgCl2, 0.18 ml of 37 % HCl, 0.1 g urea, 0.2 g CaCl2, 1 g bovine serum albumin, 3 g pancreatin, 0.5 g lipase	9.0 g bile, 9.0 g pancreatin, 0.3 g trypsin, 0.3 g urea, 0.3 g KCl, 0.5 g CaCl2, 0.2 g MgCl2		3.5 g bile, 0.35 g pancreatin	1.75 g bile, 0.5 g pancreatin		1.75 g bile, 0.5 g pancreatin
pH		6.5	7.5		5.5	7.0		7.0
Soil/ solution ratio		0.6:58:5	1:100		1:150	1:100		1:100
Extraction time (h)		4	6		1	4		4
		Bile phase (pH 8.0±0.2)						
In vitro parameters								
Composition (g l−1)		5.259 g NaCl, 5.785 g NaHCO3, 0.376 g KCl, 0.18 ml of 37 % HCl, 0.25 g urea, 0.222 CaCl2, 1.8 g bovine serum albumin, 6 g bile						
pH		6.5						
Soil/ solution ratio		0.6:58:5						
Extraction time (h)		4						

ASTM: International Association for Testing Materials
BARGE UBM: Bioaccessibility Research Group Europe, Unified BARGE Method
DIN: Deutsches Institut fur Normunge (2000)
EN 71-3: European Committee for Standarization
IVG: In Vitro Gastrointestinal (Rodriquez 1999)
PBET: Physiologically-Based Extraction Test (Ruby 1996)
RBALP: Relative Bioavailability Leaching Procedure (Drexler & Brattin, 2007)
SBRC: Solubility Bioavailability Research Consortium (Ruby 1999)

Table 9. Summary of examples of various oral fluid compositions (corresponds to Table 6 of Attachment 18).

The composition of gastric fluid employed in a variety of other studies is generally uniform, with HCl acid (0.07N) and pH in 1-2 range being the most common fluid composition (Brandon et al. 2006; Brattin et al. 2013; Denys et al. 2012; Drexler and Brattin 2007; Juhasz et al. 2009, 2014; Maltby et al. 1991; Molina et al. 2013; OSWER 2007; Ruby et al. 1999; Schroder et al. 2004; Walsh et al. 2015). The pH is generally reported as 1.5 (typical interval found: 1.2 to 1.8, with a pH of 1.5-1.6 being more biologically relevant).

1.1.12 Quality system(s) of the developing laboratory

Please state the quality system(s) in place, if any, in the laboratory that developed the test method (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

The original gastric fluid method ASTM 5517 was adapted from the European Toy Safety Directive EN 71-3:1994 in 1996 and last revised in 2014. ASTM' role is to develop voluntary consensus technical standards. We do not have information on the quality system of this laboratory but supporting data has been filed as ASTM international headquarters and may be obtained by requesting Research Report RR: D01-1121. The test method was further modified by e.g. changing the loading specified in ASTM 5517 (20 g/l) to 0.2 g/l during the round robin of a gastric fluid protocol almost identical to the current test (Henderson et al. 2014). The quality systems for participating labs are described in section 1.2.4.

1.2 MODULE 2: WITHIN-LABORATORY REPRODUCIBILITY (WLR)

1.2.1 Rationale for the selection of the test items used for assessing WLR

Please describe the criteria applied for selecting the test items used for assessing WLR. Please specify the extent to which these test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible, please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. Please append the full list of these items in the form of a table, including their CAS number and relevant properties, as Attachment 3.

Six different inorganic metal-containing materials were selected for inclusion in the inter-laboratory study: Co oxide, Co powder, Cu concentrate, Inconel alloy, leaded brass alloy, and Ni sulfate hexahydrate. This range of materials was identified to represent the various types of metals and materials to which this method is intended to be applied. Specifically, very water-soluble substances (Ni sulfate hexahydrate), insoluble substances (Co oxide), pure metals (Co powder), alloys (Inconel alloy and leaded brass), and ore concentrates (Cu concentrate). The variety of physical-chemical characteristics of these samples impact the amount of released metals in the synthetic fluid, and therefore demonstrate a range of in vitro bioaccessibilities, and potential in vivo bioaccessibilities and systemic bioavailabilities. The complex materials (alloys, ore concentrate) also provide examples of materials containing more than one metal that can potentially be released. In addition, they provide examples of complex matrix materials, i.e. those that do not behave as simple mixtures. In matrix materials, the ingredients (mixture components) can interact with each other in a way that results in a material with altered bioelution, and associated physico-chemical and toxicological properties, compared to those of the different pure ingredients alone.

A list of the six test materials used in the study is provided in Attachment 3. All test materials were powders with a median particle size <60 µm in diameter representing a size range relevant for oral exposures. Laboratories were supplied with 100 g of each test material from the same original batch and samples were tested as received without further grinding or other manipulation to alter particle size.

1.2.2 Assessment of within-laboratory reproducibility of experimental data

Please provide an assessment of the within-laboratory reproducibility of experimental data, i.e. the agreement among results obtained from testing the same test items over time using the same protocol in one laboratory. Please specify possible sources of variability. Please append the data in the form of tables and/or figures, as Attachment 4.

Results from each of the five laboratories that participated in the study were coded in random order for confidentiality. Because the study included a single particle loading for all samples, we found that, for some materials, the metal releases were very low, often close to or below the limit of detection (LOD) for one or more labs. Prior to statistical analyses, the amounts of released metals that were not reported by the laboratories or were below the respective LOD were excluded. In addition, any fluid/time point/lab dataset with two or more labs reporting results below the limit of detection (<LOD) were excluded from the inter-laboratory validation. The within-lab reproducibility was then assessed by two independent third-party biostatisticians.

First, Dr. Peter-Th. Wilrich, professor at Freie Universität Berlin (Berlin, Germany) conducted statistical analysis of the measurement results based on ISO 5725-2 (1994). According to this method, measurement results obtained in an inter-laboratory study were inspected for within-lab consistency by plotting Mandel's k statistics and for outliers by application of the Cochran test. Mandel's k statistics are the within-laboratory standard deviations divided by the square root of the mean of the within-laboratory variances. The Cochran test for the detection of an outlying within-laboratory standard deviation uses the largest within-laboratory variance divided by the sum of the within-laboratory variances as test statistic, i.e. the maximum of the squares of the k statistics multiplied by the number of laboratories.

A laboratory mean or a within-laboratory standard deviation was marked as a straggler if the outlier test result was significant at the 5% level, and marked as an outlier if the outlier test result was significant at the 1% significance level. Following ISO 5725-2 recommendations, outliers were discarded and stragglers retained unless other explanations for the outlying observations were found.

Repeatability standard deviation (sr; within-lab) was used as measurements of within laboratory precision. For the gastric fluid treatment condition, Ni from Ni compound measurements were the least variable within labs, with Pb from leaded brass alloy and Co from Co compound also demonstrating relatively low variability. Iron from the Inconel alloy, a dataset with the fewest bioaccessibility measures for the gastric fluid treatment, demonstrated some of the highest within-lab variability. The results of the repeatability analysis are presented in Attachment 4.

The second independent statistical analysis was conducted by an independent consulting firm, ToxStrategies, Inc. (Austin, Texas, USA). Here, relative standard deviation (RSD) analysis of the log concentration was also used to assess within-laboratory measurement variability. This approach examines sr measures individually, assessing the fluctuations in the data relative to the log mean. Expressed in percentage terms, the formula for RSD was: (sd/mean log concentration) * 100. RSD values and associated thresholds represent an attempt to define absolute levels of acceptable sample-to-sample result variability (repeatability, r) within-lab. Criteria for the analysis were that the RSD for for repeatability should be less than 10%. The results of the RSD analysis are presented in Attachment 4.

Possible sources of variability affecting within-laboratory repeatability included handling of particles in solution (aggregation, sedimentation, separation) issues and differences related to metastable corrosion processes inherent to the metal and alloy samples.

1.2.3 Identification and discussion of outlying values

Please identify and discuss any outlying values.

Chromium from the Inconel alloy was completely excluded from the statistical analyses as there were two labs reporting results below the limit of detection (2/3 and 1/3 <LOD). Following analysis using the Cochran test, only the results for iron (Fe) from in the Inconel alloy in one laboratory were discarded as an outlier for the gastric fluid test. Nickel from Ni compound in a different laboratory was identified as a straggler but retained.

1.2.4 Quality system(s) of the testing laboratory

Please state the quality system(s) in place, if any, in the testing laboratory (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

The SOP specified the following regarding quality systems required to be in place at each of the participating laboratories:
"A study should be subjected to a Quality Assurance Program which could guaranty integer and reproducible results.
Possible quality systems:
- ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories
-Good Laboratory practice as specified by Directive 1999/11/EEC of the commission of March 08, 1999, relating to the adjustment of the principles of GLP to technical progress according to the Directive 87/18/EEC of the council for the adaptation of legal and administrative regulations for the application at test with chemical test items (official journal of the European Union N° L77 p-8-21 of 23 March 1999."
In general, the studies were conducted following the spirit of GLP as made evident by the following:
-Each laboratory followed a SOP provided by the study coordinator
-Each laboratory was independent from all study sponsors/coordinator
-Each laboratory had a dedicated study director
-All equipment, reagents, and materials used by laboratories were of the quality stipulated in the SOP; equipment was cleaned according to the SOP
-Each laboratory generated a comprehensive report to the study coordinator and archived records in-house
-Each laboratory report underwent a quality assurance exercise. A detailed review and comparison between the SOP and the five laboratory reports was performed by an independent study coordinator. As part of this review, individual exchanges were held with the labs to address information gaps and confirm data when necessary

The quality systems employed by each of the five participating laboratories, as described in their corresponding study reports, are listed below:

Lab A report states, "In order to assure the quality of the readings of the ICP-MS (Inductively Coupled Plasma Mass Spectrometry) metals determinations, two quality control standards, QC1 and QC2 were used. QC1 was prepared based on the QCS-26 solution; QC2 was based on the QCS-27 solution"

Lab B report states, "The test phase was conducted at [LAB B] which works according to the principles of GLP (Good Laboratory Practices). However, in this phase of research (optimizing SOP) it was impossible to conduct this study according to the principles of GLP. Nevertheless, all raw data and calibrations of balances, pH meter, thermometer and shaker are

available and will be stored properly according to the principles of GLP. The study has been audited by the Quality Assurance manager who followed several critical phases in this study (i.e. weighing of the test item, setting up the experiments, some of the sampling, controlling the raw data transfer and reporting). The delegated analytical phase of this study has been performed according to the criteria of the ISO 17025 standard by the accredited [ANALYTICAL LABORATORY]. The results of the measurements of the matrix spiked quality control samples are added to their analytical reports..." This lab used ICP-MS or ICP-AES to measure metal releases.

Lab C report states, "The [LAB C] program has operated under GLPs (Good Laboratory Practices) or a Quality Assurance Program (QAPP) plan since 1999. The program has been audited numerous times, including several facility audits, GLP raw data audits, and GLP in-phase audits…. QC samples comprise at least 15% of each analytical run…". Furthermore, the reports states, "Laboratory QC consists of various blanks, and check standards, spikes, and certified reference materials. The data meet program data quality objectives. The blanks, all types, standards, check standards, are individually listed in the spreadsheet, as requested." This lab used ICP-MS or ICP-AES to measure metal releases.

Lab D report states that the metal analyses were performed according to the EPA methodology, EPA Method #200.7 (ICP-AES). This laboratory is accredited by state of PA to do environmental testing and it conducts its work according to the principles of GLP (Good Laboratory Practices).

Lab E has yearly certifications to perform AAS metal analyses and complies with standard calibration and quality control solutions in a state of the art facility.

1.3 MODULE 3: TRANSFERABILITY (TF)

1.3.1 Rationale for the selection of the test items used for assessing the TF

Please describe the criteria applied for selecting the test items used for assessing transferability. Please specify the extent to which the test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. Please append the full list of these test items in the form of a table, including their CAS number and relevant properties, as Attachment 5.

Proficiency materials used for assessing transferability must be:
-easily available in the appropriate form at a low cost
-well characterised
-stable under ambient conditions
-not extremely toxic so that no special measures need be taken beside the usual protective clothing, dust mask, glasses and gloves

A list of possible proficiency materials is provided in Attachment 19.

1.3.2 Training required for transferring the test method

Please provide an estimation of the amount of training that is necessary to establish the test method in a naïve laboratory (i.e. a laboratory which is familiar with the techniques involved but not with the test method). If available, please append the training protocol as Attachment 6

No special training is needed. The test could be performed based on the detailed instructions recorded in the test protocol by any laboratory with analytical capability and knowledge of metals' analysis.

1.3.3 Obstacles pertaining to transferability that are specific to the test method

Please provide a summary of expected obstacles or difficulties that may impact on the transferability of the test method, e.g. level of complexity of some procedures in the protocol(s), etc.

One should always work in a clean environment and pay attention to potential contamination. Although test loadings are relative high (up to two gram per litre), the level of element release can be very low (µg/L or sub µg/L) and be significantly affected by contamination. In addition, the use of acid cleaned vessels is mandatory. Particle separation is another step that needs to carefully follow the guidance provided in the protocol to ensure that released, solubilized metal is accurately measured.

1.3.4 Organisation of the transfer phase

Please explain how the transfer phase was organised, including the criteria applied to assess success of transfer (see 2.3.5). If available, please append the transfer protocol as Attachment 7.

The transfer of the method can best be done by carrying out a study with one or more proficiency materials possibly under supervision of an experienced person who can point out possible problems and concerns.

1.3.5 Assessment of the transferability to another laboratory

Please provide an assessment of the transferability of the submitted test method to another laboratory. If available, please append the transfer report as Attachment 8...

The use of proficiency materials with the SOP can be used for the transferability of the bioelution protocol. Acceptable results are within ±20% of target value.

1.3.6 Quality system(s) of the other laboratory

Please state the quality system(s) in place, if any, in the other laboratory (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

The concentrations of released elements of interest are measured with an appropriate (validated) measurement method. In order to guarantee the basic quality (use of a validated measuring method, involvement of appropriately trained staff, traceability of the measurements, data processing and archiving,…) the sample analyses are preferably performed in a laboratory that works according to a standardised quality system (e.g. GLP, GMP, ISO17025, ...).

1.4 MODULE 4: BETWEEN-LABORATORY REPRODUCIBILITY (BLR)

1.4.1 Rationale for the selection of test items used for assessing BLR

Please describe the criteria applied for selecting the test items used for assessing BLR. Please specify the extent to which these test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible, please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. Please append the full list of these test items in the form of a table, including their CAS number and relevant properties, as Attachment 9.

The same six inorganic metal-containing materials used for the assessment of within laboratory reproducibility were selected for inclusion in the inter-laboratory reproducibility study. A list of the six test materials is provided in Attachment 3.

1.4.2 Assessment of reproducibility

Please provide an assessment of the between-laboratory reproducibility of experimental data, i.e. the agreement among results obtained from testing the same test items using the same protocol in different laboratories. Usually at least three laboratories are requested to properly evaluate between-laboratory reproducibility. Please specify possible sources of variability. Please append the data in the form of tables and/or figures as Attachment 10.

Results from each of five laboratories in the study were coded in random order for confidentiality. Prior to statistical analyses, the amounts of released metals that were not reported by the laboratories or were below the respective limit of detection were excluded. In addition, any fluid/time point/lab dataset with two or more labs reporting results below the limit of detection (<LOD) were excluded from the inter-laboratory validation. The between-lab reproducibility was then assessed by two independent third-party biostatisticians.

First, Dr. Peter-Th. Wilrich, professor at Freie Universität Berlin (Berlin, Germany) conducted statistical analysis of the measurement results based on ISO 5725-2 (1994). According to this method, measurement results obtained in an inter-laboratory study were inspected for between-lab consistency by plotting Mandel's h statistics and for outliers by application of the Grubbs tests. Mandel's h statistics are the differences between the laboratory means and their mean divided by the standard deviation of the laboratory means.

Reproducibility standard deviation (sR; between-labs) was used as measurements of precision. The between labs reproducibility standard deviation (sR) was calculated as the square root of ($s_r^2+s_L^2$), where s_L is the between-laboratory standard deviation. In the gastric fluid treatment condition, Ni from Ni compound measurements were the least variable within and across labs, with Pb from leaded brass alloy and Co from Co compound also demonstrating relatively low variability for both measures. Iron from the Inconel alloy, a dataset with the fewest bioaccessibility measures for the gastric fluid treatment, demonstrated some of the highest variability for both measures. The results of the reproducibility analysis are presented in Attachment 4.

The ratio of the repeatability standard deviation and the reproducibility standard deviation (sR:sr) of the log-concentration was determined and used as an indicator of the (dis)agreement between the mean results of the laboratories. Ratios up to 3 were considered to represent good agreement, ratios between 3 and 6 to represent fair agreement, and >6 were considered to mean that agreement between the laboratories needed to be improved. All metals/substances had ratios below 6 except for one (Co from Co powder was 6.2), demonstrating good agreement between the variability in repeatability and reproducibility. The results of the sR:sr analysis are presented in Attachment 4.

The second independent statistical analysis was conducted by an independent consulting firm, ToxStrategies, Inc. (Austin, Texas, USA). Here, relative standard deviation (RSD) analysis of the log concentration was also used to assess between-laboratory measurement variability. This approach examines sr and sR measures individually, assessing the fluctuations in the data relative to the log mean. Expressed in percentage terms, the formula for RSD was: (sd/mean log concentration) * 100. RSD values and associated thresholds represent an attempt

to define absolute levels of acceptable lab-to-lab result variability (reproducibility, R). Criteria for the analysis were that the RSD for reproducibility should be less than 20%, and the RSD for repeatability should be less than 10%. The results of the RSD analysis are presented in Attachment 4.

Possible sources of variability affecting between-laboratory reproducibility include differences in sensitivity of analytical instrumentation and metal analysis techniques, surface abrasion, etc. A study aimed at evaluating analytical procedures among labs was conducted prior to initiating the inter-laboratory bioaccessibility study. Samples of interstitial fluid (only fluid type used in this pre-study analysis) spiked with known metal concentrations (negative control, Co, Cu, Ni, Pb, and Zn) were provided (in blind fashion) to each of the laboratories to determine the analytical concentrations. After eliminating outliers, the statistical analysis resulted in an sR:sr ratio of about 6, indicating a lack of harmonization among laboratories. In addition, methods of pH control, sample weighing and loading, agitation, and sample filtering may have varied between labs. The recommendations for the gastric fluid were also carried forward into the current SOP as described in section 1.6.2 (Table 11).

1.4.3 Identification and discussion of outlying values

Please identify and discuss any outlying values.

As reported in section 1.2.3, chromium (Cr) from the Inconel alloy was completely excluded as there were two or more labs reporting results below the limit of detection (<LOD). Following analysis using the Double high Grubbs test, only the values for Ni in the Inconel alloy in two laboratories were discarded as outliers for the gastric fluid test. Cobalt in Co compound in one laboratory was identified as a straggler in the Single high Grubbs test but retained.

1.4.4 Quality system(s) of the testing laboratories

Please state the quality system(s) in place, if any, in the testing laboratories (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

Please see section 1.2.4.

1.5 MODULE 5: PREDICTIVE CAPACITY (PC)

1.5.1 Rationale for the selection of test items used for assessing PC

Please describe the criteria applied for selecting the test items used for assessing PC.. Please specify the extent to which these test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible, please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. If some of these test items were also used to develop the prediction model, please indicate which. Please append the full list of these test items in the form of a table, including their CAS number and relevant properties, as Attachment 11.

.

The predictive capacity of the bioelution test for applications 1 and 2 can be surmised from studies assessing the relationship between gastric metal bioaccessibility and oral metal bioavailability (see section 1.5.2 and Attachment 18). Test items include simple metal compounds (e.g. nickel, cobalt) and complex samples like soils.

Application 1-Grouping is applicable to a variety of metal (metalloid) containing samples such as metal compounds, metals, complex metal-containing materials like UVCBs, alloys, pigments, soils, etc.

Application 2-Alloy classification requires testing of alloys and their pure metal or metalloid ingredients in powder or massive forms.

1.5.2 Assessment of the predictive capacity of the test method

Please provide all available information on the predictive capacity of the test method. Please describe the accuracy of the proposed test method with respect to its ability to measure or predict the effect of interest. The accuracy values (i.e., overall accuracy, sensitivity, specificity, positive and negative predictive values, false positive and negative proportion) of the proposed test should be compared to that obtained for the appropriate reference test method (if available) and/or other data or information, especially from the species of interest (if available). Please append the data in the form of tables and/or figures, as Attachment 12.

The predictive capacity of the bioelution test for applications 1 and 2 can be surmised from studies assessing the relationship between gastric metal bioaccessibility and oral metal bioavailability. Oral bioaccessibility test systems, including the ones based on the static approach to simulate the worst case gastric fasting exposures in humans, have been used to predict the relative bioavailability of several metals, mainly from an environmental compartment, i.e. soil. Alloy LLC (2017) conducted a review of the available information on the relationship between in vivo relative bioavailability (RBA) and in vitro bioaccessibility (IVBA)[14] of metals from soils (Attachment 18). The review was limited to As, Pb, Be, Se, Sb, Cd, Co, Ni, and Hg (i.e. focusing on metals for which exposure is associated with systemic effects via the oral route). There is a broad base of research to understand and predict the bioavailability of metals from soil, because this information forms part of the risk assessment process that dictates the need for remedial action at contaminated sites. Findings from the studies of soil (mixtures of metal compounds), including evaluation of in vitro extraction testing to predict bioavailability, informs how bioelution test data can predict (or provide a conservative estimate of) the bioavailability of metals from metal containing substances and mixtures. Of the 39 studies reviewed by Alloy LLC, 22 included information on both RBA as measured in animals (mouse, rat, and swine) and IVBA measured in laboratory extraction tests. RBA corresponds to the bioavailability of metals from soils relative to that from fully water

[14] Please see sections 1.1.1 and 1.1.2, as well as pages iii to v of Attachment 18 for definitions of bioaccessibility and bioavailability.

soluble metal compounds. RBA does not indicate absolute bioavailability of metals from soils; the absolute bioavailability of metals from soil (ABA) is always lower than the RBA. Across all studies and all metals evaluated, the data provide strong evidence that soil- and site-specific factors exert control on the bioavailability of metals from soil relative to bioavailability of soluble forms. Strong predictive relationships between RBA and IVBA emerged for lead and arsenic, with fewer available data for Sb, Cd, Co, Ni, and Hg, and no data for Be and Se. An example of the relationship between RBA and IVBA in stomach fluid, or stomach and intestinal fluids is shown in Figure 2 below for Sb.

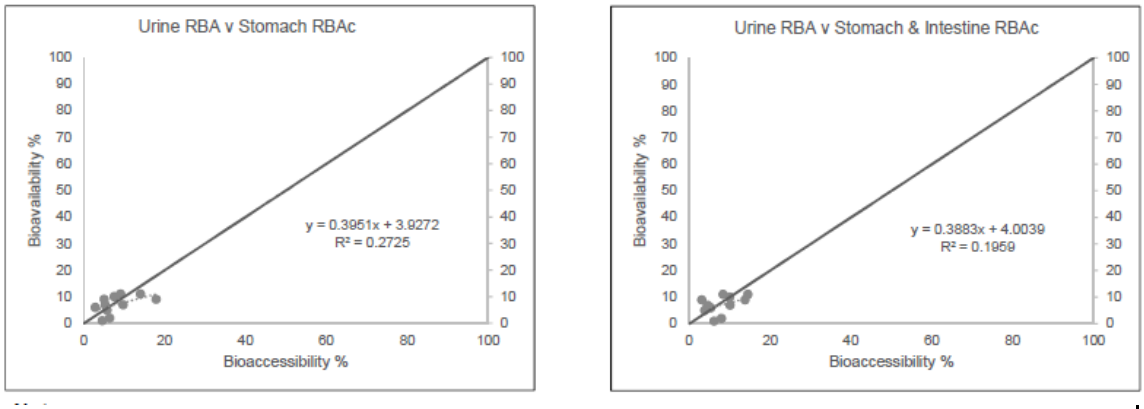


Figure 2. (corresponds to Figure 5a of Attachment 18).

In many cases the IVBA values are higher than the RBA ones and in other cases they are equal or lower. In all cases studied, a relationship between IVBA and RBA was established (see Attachment 18 for a detailed report).

For Cd, Co, Pb and Ni, simple extraction fluids (e.g. those that consist of a simple glycine-buffered system, or even an unbuffered HCl solution) resulted in higher measured IVBA than systems that include additional components to make the extraction fluid mimic physiological conditions more closely (Attachment 18). Thus, lower pH conditions outweigh other components of the extraction fluid that are intended to create a more "physiological" environment. For arsenic, solubility from soil appears to have a lower dependence on pH, although for some soils, low pH is associated with higher IVBA. Arsenic results from the low pH extraction methods are more closely correlated with RBA when studied using paired soils. Conversely, the IVBA of mercury is higher at higher pH. Figure 3 below corresponds to metal releases (Pb, Zn, Cd, Cu, Ni, As) from NIST standard reference material soils at different oral fluid pH (Koch et al. 2013).

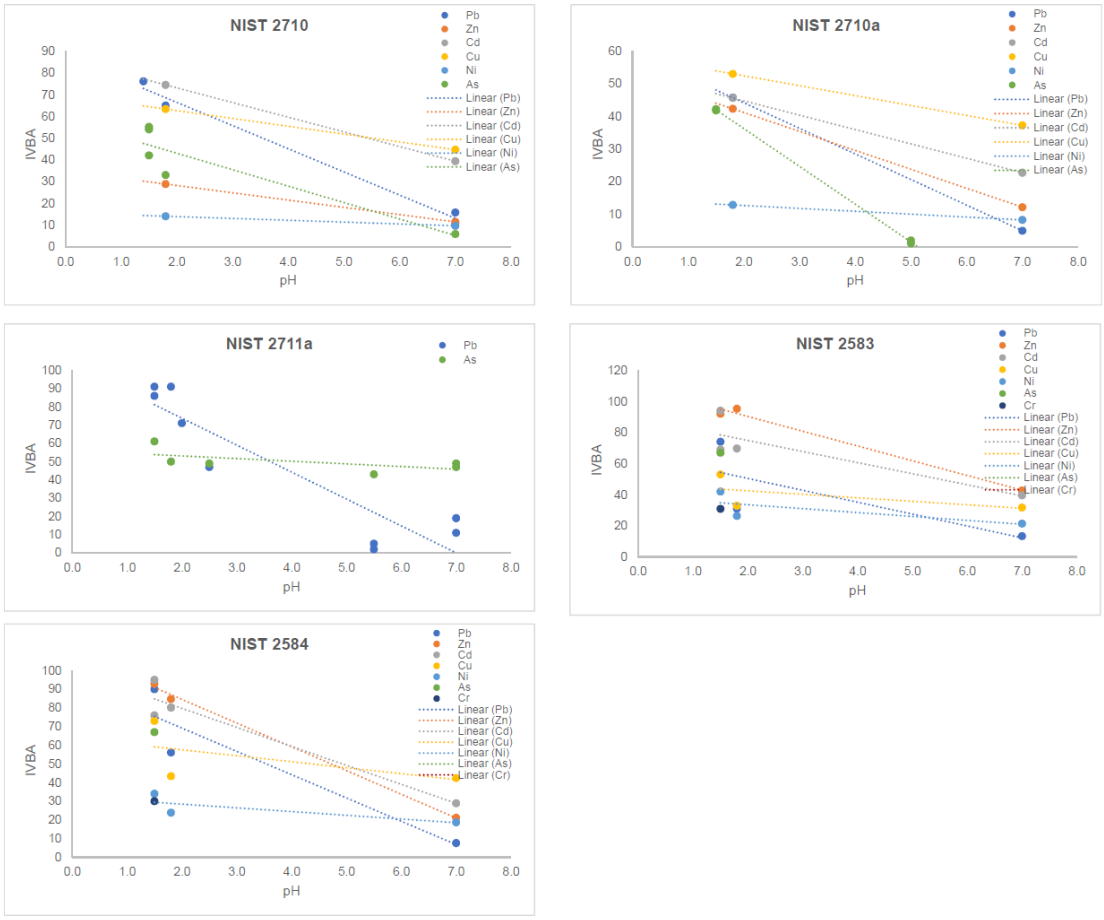


Figure 3. (corresponds to Figure 4 of Attachment 18).

For arsenic, Diamond et al. 2016 conducted a meta-analysis looking at the relationship between IVBA and RBA of arsenic from soils (see more details about this example under section 1.1.5.2). This approach allowed incorporation of uncertainty in individual IVBA and RBA estimates into the regression model (Figure 1). The solid line represents the weighted regression model and the dotted lines are the 95% prediction limits for the weighted model. The authors conclude that the data from IVBA can predict RBA of arsenic in soil.

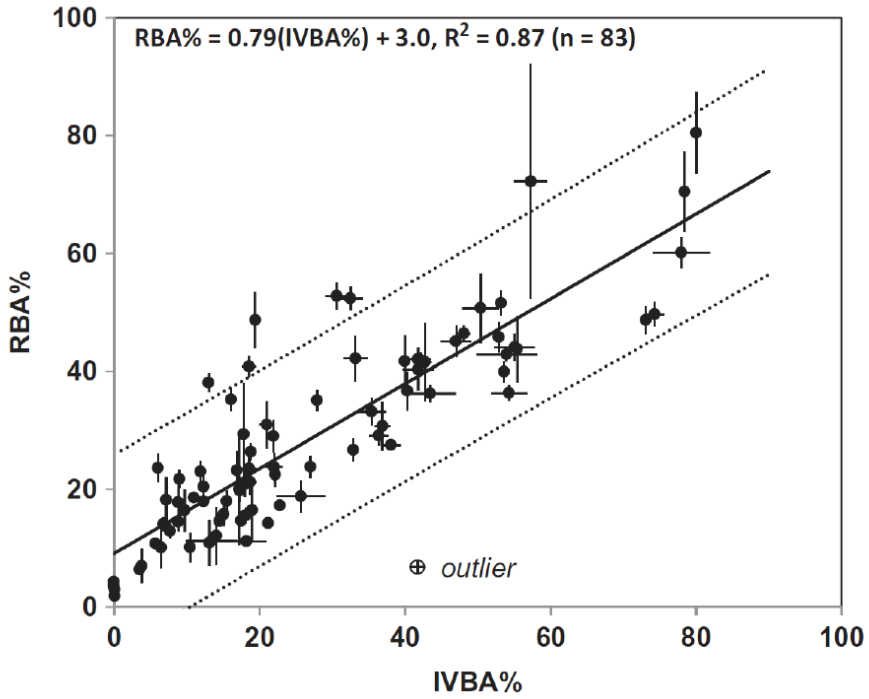


Figure 1. Percent RBA as a function of %IVBA for arsenic (corresponds to Fig 1 of Diamond et al. 2016).

In the above studies, bioelution at pH of 1.5 seems to yield the highest IVBA values independent of material or metal tested.

The review by Alloy LLC and other existing studies (e.g. Molina et al. 2013) showed that for several metals (As, Cd, Co, Cu, Mn, Ni, Pb, Sb, Zn) there was good statistical evidence to show that there was a relationship between the IVBA of the released metal in gastric fluid and the in vivo systemic RBA of the metal. In most instances, the RBA values were lower than the IVBA ones but there were also instances where the reverse was true. Importantly, in all instances a defined relationship between gastric bioaccessibility and in vivo bioavailability was established (see Attachment 18).

The review concludes that in general, IVBA methods provide an efficient and effective method for predicting the relative bioavailability of metals.

Application 1-Grouping.

In this application the bioaccessibility of metals from reference and target samples after 2 hours in gastric fluid are considered in a weight-of-evidence approach for grouping and read-across.

Example 1. Data from Henderson et al. 2012 can be used to illustrate how gastric bioaccessible data from metal compounds can be used for grouping and read-across in a weight-of-evidence approach. Please see this example description under Application 1-Grouping in section 1.1.5.2. Using this information in a weight-of-evidence approach with other phys-chem and oral toxicity data, a self-classification for reproductive toxicity was assigned to Ni sulfamate by grouping it with the water soluble Ni sulfate

and Ni chloride compounds and reading across their reproductive toxicity classification (Henderson et al. 2012).

Application 2-Alloy classification.

When data on released amounts of metals per mass of sample from alloys and reference metal ingredients are available, the bioaccessible concentration of the metal in the mixture can be calculated using the following equation.

$$BC\% = \frac{\mu g\ metal\ ion\ released\ (measured\ in\ extract)/g\ alloy\ sample\ tested}{\mu g\ metal\ ion\ released\ (measured\ in\ extract)/g\ pure\ metal\ sample\ tested}\ x\ 100$$

The bioaccessible concentration of a metal in an alloy can be compared to its bulk concentration to establish the presence or absence of a matrix effect that would warrant the application of %BC as a refinement to bulk concentration, for the purpose of alloy classification for systemic health effects.

Example 1. To demonstrate how the presence of a matrix effect in alloys can affect metal releases compared to those from simple mixtures, the following experiment was carried out by ECTX (2017). The release of nickel from a sample of SS316L alloy powder containing 10.7% of Ni (68.6% Fe, 16.4% Cr) (ECTX X02a-144) was compared to the release of nickel from a simple mixture containing 10.7% Ni, 68.6% Fe, and 16.4% Cr powders (ECTX X02a-130), both samples tested at the same sample loading of 2 g/L. Gastric fluid corresponded to an HCl solution at pH 1.5 and the duration of the test was 2 hours. The release from pure Ni metal powder was also reported (ECTX X02a-130). The results shown in Table 10 below demonstrate a 300-fold difference between the Ni releases from a simple mixture compared to release from the SS316L alloy (even if the concentration of Ni in both samples was the same). Using the release of pure Ni metal as a reference, the bioaccessible concentrations (BC) of Ni in the simple mixture (12%) and SS316 L (0.04%) were calculated[15]. The BC of Ni in the simple mixture (12%) coincides with its bulk concentration (11%) indicating that no matrix effect was present. By contrast, the BC of Ni in SS316L (0.04%) is ~300-fold lower than its bulk concentration (11%) and 300-fold lower than the BC of the simple mixture (12%). This illustrates how the alloy matrix effect can affect metal bioaccessibility. Details from this and other studies conducted by ECTX are described in Attachment 20.

[15] BC% calculations: simple mixture: 20,073 µg Ni/g sample/162,734 µg Ni/g pure Ni^o x 100 = 12%; for the SS316L: 61 µg Ni/g alloy /162,734 µg Ni/g pure Ni^o x 100 = 0.04%.

Sample	*Particle diameter*	*Mass loading (g/L)*	*Nickel loading (g Ni/L)*	*Nickel release (mg Ni/L)*	*Nickel release (mg Ni/g sample)*	*Bioaccessibl Concentratio of Ni in mixture (%)*
Simple mixture: *68.6% Fe, 16.4% Cr, 10.7% Ni*[a]	D_{90}= 2 µm	2	0.21	40.1	20.1	12.3
SS316L: *68.6% Fe, 16.4% Cr, 10.7% Ni*[b]	D_{90}= 6 µm	2	0.214	0.13	0.065	0.04**
Ni metal, *100% Ni*[a]	D_{90}= 2 µm	0.2	0.2	32.5	163*	

*a. ECTX X02a.130; b. ECTX X02a. 144; * previous results for this sample were 209 (rounded from 208.6) mg Ni/g sample (ECTX X02a. 144) and 182.6 mg Ni/g sample (ECTX X02a. 085). ** 0.03-0.04% when using previous test results for nickel metal samples as a reference.*

Table 10. Differences in metal release in gastric fluid from a simple mixture and an alloy containing the same bulk concentration of nickel (ECTX 2017).

While the particle size of the powders tested above was quite similar (D90 =2-6 µm), the surface area of the Ni metal powder in the simple mixture (2.58 m²/g) was different from that of the SS316L alloy powder (0.35 m²/g). If desired, the bioaccessible concentrations can be re-calculated after applying a correction for measured surface area of the samples.

Example 2: data from Hillwalker and Anderson (2014) paper can be used to illustrate the ways in which data from gastric bioelution tests are reported for metals and alloys. For example, the calculation of the bioaccessible concentration of a metal (Mn) in an alloy (carbon steel) was described under Application 2-Alloy classification in section 1.1.5.2.

The bioaccessible concentration indicates that carbon steel alloy behaves as a mixture containing 0.057% of Mn metal in it. Since the content of Mn in the alloy is 0.456%, there is a matrix effect in the alloy that decreases Mn release (~8-fold) compared to what could be expected without a matrix effect (e.g. simple mixture). This is another example that illustrates how the bioaccessible concentration of a metal can differ from its bulk concentration due to alloy matrix effects.

Other examples where the bioaccessible concentration of metals in alloys differ from their bulk concentrations exist. Mörsdorf and coworkers (2015) showed that the 2 h gastric fluid bioaccessible concentrations of Fe and Mo in a ferromolybdenum alloy (7% Fe and 26% Mo) were lower than their respective bulk concentrations (30.2% Fe and 59.8% Mo) by 2-4-fold (based on data from Figure 4 of Mörsdorf et al. 2015). Hedberg et al. 2016 showed that when tested for 168 hours in lysosomal fluid, Fe in Inconel had a bioaccessible concentration of <0.5% (calculated based on 2 different Fe metal reference samples) while its bulk concentration was 14.6% (data from Figure 4 of Hedberg et al. 2016). Stockmann Juvala et al. 2013 reported nickel releases from pure

Ni metal and SS316 L after 168 hours in lysosomal lung fluid of 0.87 and 0.00088 µg/g sample, respectively. From these data, the bioaccessible concentration of Ni in S316L powder (0.1%) can be calculated; the BC of Ni in SS316L was 100-fold lower than its bulk content (10.3%).

1.5.3 Identification and discussion of false predictions

Please identify and discuss any false predictions (e.g. false positives and false negatives).

[…]

1.6 MODULE 6: APPLICABILITY DOMAIN (AD)

1.6.1 Applicability of the test method identified through testing

Please identify the applicability of the test method on the basis of experimental evidence. In the case of chemicals, indicate the chemicals and/or chemical categories (e.g. based on functional groups and/or physicochemical properties) for which the test method makes reliable and relevant predictions.

On the basis of the experimental evidence, the test method is applicable to metals, inorganic metal compounds, or complex metal (metalloid)-containing materials (like alloys, UVCBs, pigments) in powder or massive form.

For both applications (1-Grouping, 2- Alloy classification), when powder materials (defined as particles < 1 mm diameter) are tested in gastric fluid, testing the subfraction of particles <100 µm in diameter is recommended because it is also relevant to the fraction of the inhalation exposure that is swallowed and becomes bioavailable through the oral route. Nanosize powders are not within the scope of current SOP.

Massive forms of metals (M^o) and alloys can also be tested as-received or epoxy-embedded; some bioelution tests conducted after 2012 have utilized epoxy-embedded massive forms of alloys (e.g. ECTX X02a-146; ECTX X02a-081). For massive forms of these substances, techniques such as BET cannot be used to assess the surface area. In this case, it is possible to embed the massive samples in epoxy to expose a limited and known surface area to the gastric solution, or expose a well-defined coupon of the same material calculating the geometric surface area (e.g. as it is done in BS EN1811 2011).

1.6.2 Limitations of the test method identified through testing

Please identify on the basis of experimental evidence the limitations of the test method. In the case of chemicals, indicate the chemicals and/or chemical categories (e.g. based on functional groups and/or physicochemical properties) for which the test method has been shown <u>not</u> to be applicable to.

One limitation that was identified during the round robin is the use of a single loading for all samples. The test method needs to incorporate a high (2 g/L) loading in addition to a low (0.2 g/L) loading so that materials that release low amounts of ions can be reliably assessed (measurements >LOD). This applies to both applications of the test method. Likewise, very soluble materials need to be tested at a lower loading to avoid underestimating the metal release.

For some particular metals such as silver (Ag), gastric fluid <u>alone</u> should not be used for the oral route as the presence of chloride can lead to precipitates of AgCl. In the case of metal

release from alloys, while concentration-dependent, silver reacts slowly with dilute hydrochloric acid and form chloride-rich corrosion products on the metal surface as well as AgCl(s) that precipitates out from solution. This layer of poorly soluble corrosion products reduces further metal release. These processes are expected to take place both in-vitro and in-vivo. In this case, the loading needs to be selected to avoid precipitation and sedimentation and be able to quantify the metal ion that remains in solution. In this case, conducting bioelution tests in saliva and/or intestinal fluid would be recommended for oral route assessments under applications 1 or 2.

Similarly, there are other elements (Mo, Sb, Si, W, and Hg) for which there is insufficient evidence to conclude that gastric fluid will yield the highest metal release values for the oral route. For these metals and metal compounds it is recommended to conduct bioelution tests in both gastric fluid (current SOP) and in saliva and/or intestinal fluid for oral route assessments under applications 1 or 2.

The main differences between the gastric protocol employed in the round robin (Henderson et al. 2014) and the one included in the current SOP are identified in Table 11

Differences in protocols	**Round Robin 2014**	**Current SOP (Attachment 1arev)**
Loading	One loading (0.2 g/L)	Two loadings (0.2 and 2 g/L)
pH measurement and adjustment	pH of initial gastric fluid solution was adjusted to 1.5; pH was adjusted again after adding the test material to the vessel; pH was measured at the end of the test	pH of initial gastric fluid solution was adjusted to 1.5;no further pH adjustment is carried out once fluid is added to the test material. pH is measured at the end of the test
Darkness during test	Not specified	Specified
Samples	Powders tested as < 100 µm subfraction	Massive (≥ 1 mm) and powder (≥ 0.1 µm - < 1mm) samples are tested. For powders, testing the subfraction < 100 µm is recommended when feasible.
Controls	Negative	Negative and Positive
Data Acceptance Criteria & Reporting	Not well defined	Described in great detail

Table 11. Differences between the gastric fluid SOP used in the round robin study and the current one shown in Attachment 1a.

1.6.3 Suggested Applicability Domain (AD) of the test method

Please define to the extent possible the (preliminary) AD of the test method on the basis of the information given in 1.6.1 and 1.6.2. Indicate whether the AD is considered to be wider than the one experimentally demonstrated and provide justification for this.

The test method is applicable to metals and metalloids (e.g. As, Si), inorganic metal compounds, or complex metal-containing materials (like alloys, UVCBs, pigments) in powder or massive forms. For the testing of materials in powder form (defined as particles < 1 mm diameter) in gastric fluid, testing the subfraction of particles <100 µm in diameter is

recommended because it is also relevant to the fraction of the inhalation exposure that is swallowed and becomes bioavailable through the oral route.

The test method is not meant to be applied to nanomaterials (where separation of particles from free ions may require different filtration or centrifugation methods), organometallic materials, or other, non-metal or non-mineral materials or matrices (e.g. plastic, rubber, glass, etc.) for which no metal release data has been generated to date.

If substances as placed on the market are expected to generate powders but do not meet the target particle size distribution for testing (e.g. <100 µm), the material should be altered to achieve the smaller particle size that can reasonably be expected to be generated, prior to testing. Some of the alloys and metallic substances placed on the market are often present in massive form. In some instances, it may be desirable to test a smaller particle size. In order to obtain the desired particle size, potential methodologies for sample preparation include crushing, grinding, milling, and/or sieving. Of these methods, sieving is preferred when sufficient material is available, as this method does not impart any physical or chemical changes to the sample (e.g. possible surface oxidation/passivation). Crushing, grinding, and milling operations can have a significant effect on metal release from some alloys and should only be considered when sieving is not feasible. Under these circumstances, it is important to test the metal release from the material both as-received and after being crushed/ground/milled in order to establish whether the crushing/grinding/milling operation has had an impact on the release properties of the test material (i.e., the generated particles have a chemical composition that is not too different from the real product). Characterisation of the surface oxides after crushing/grinding/milling operations, can be important to interpret the metal release results before and after crushing/grinding/milling, as this is the interface that governs the release process. Any such change of properties must be recorded and reported as well as the relative difference in surface area exposed to the test solution.

1.7 MODULE 7: PERFORMANCE STANDARDS (PS)

Performance Standards, based on a validated test method, can be used to evaluate the reliability and accuracy of proposed similar or updated test methods. The three elements of PS are: a) essential test method components, b) a minimum list of reference test items and c) the defined accuracy and reliability values.
Please consider this section as optional, since these standards are normally defined at the end of validation studies.

1.7.1 Suggestions of essential test method components for the evaluation of similarity

Please indicate the essential structural, functional and procedural components of a validated test method that should be retained in a proposed test method to be judged as similar.

Not applicable

1.7.2 Suggestions of possible reference test items

Please suggest test items which should be used to assess the accuracy and reliability of a similar test method. These test items should be a representative subset of those used to demonstrate the reliability and accuracy of the validated test method.

[…]

1.7.3 Suggestions of defined accuracy and reliability values

Please suggest the target accuracy and reliability values that should be achieved by the similar test method when evaluated using the reference test items.

[…]

2. ESSENTIAL INFORMATION FOR A SPECIFIC VALIDATION PROCESS

2.1 External Validation Studies (not coordinated by EURL ECVAM)

All validation studies, also those which are not coordinated by EURL ECVAM, should follow some good validation principles. The following sections of the TST (from 2.1.1 to 2.1.5) apply to all validation studies not coordinated by EURL ECVAM including external retrospective validation studies and external performance standards-based validation studies.

2.1.1 Study organisation and management (Project Plan)

For external validation studies (not coordinated by EURL ECVAM) a detailed project plan, describing the study organisation and management should be produced prior to the initiation of the study and provided as Attachment 13 (see Explanatory Note to the TST on pages 2-3). Please briefly summarise here below how the study was designed (e.g. number of participating laboratories, number of replicates within a single experiment, number of independent repetitions, etc.) and managed.

The method validation study was initiated in 2010 by a group of representatives from various metal commodities seeking to standardize fluid compositions and testing protocols for the basic bioelution methods being used across the industry. To investigate the within-lab repeatability and between-lab reproducibility of bioaccessibility testing in gastric fluid, five laboratories measured metal release from six metal-containing materials following one defined protocol issued by the study sponsors. The study was funded by five commodities groups as described in Attachment 13, with each organization having a key point person that participated in steering team and technical discussions.

Prior to conduct of the round robin, an independent third party (T. Brouwers of ECTX) was asked to collect and distribute the samples to the initial 4 participating labs. Eventually the ECTX lab agreed to also participate in the round robin at no cost; thus 5 labs were included.
-The metal commodity sponsors were represented by R Henderson (NiPERA, who was later replaced by A Oller), F Van Assche (International Zinc Association), K Delbeke (European Copper Institute), T Brock (Cobalt Institute) and V. Verougstraete (Eurometaux).
-R. Henderson (who by then had moved from NiPERA to ToxStrategies) was contracted as study coordinator to prepare the manuscript for publication. Prior to publication, R Henderson performed a Q&A of the reports from the laboratories and coordinated the statistical analyses.
-Two independent statistical analyses of the bioelution data generated by the round robin were conducted by P Wilrich (Institut fur Statistik und Okonometrie Freie Universitat Berlin) and ToxStrategies

Five independent laboratories participated in the study, each with a designated study director(s) as listed in Attachment 13. Samples of each of the six materials (pre-characterized for particle size, metal content, and surface area) were distributed to each of the five laboratories from two independent repositories (Particle Technologies Laboratory (Downers Grove, IL, USA) and Outotec (Finland)). Labs were asked to follow one Standard Operating Procedure (SOP; dated November 2010) provided and discussed prior to study initiation. All fluids and experimental set ups were prepared by each individual laboratory. In brief, test materials were added to simulated gastric fluid and extracted after 2 h standard conditions (e.g. pH, temperature). Following a filtration step, extracts were analyzed and the amounts of metals released into solution were reported. Laboratories measured the release of seven

different metals (Cr, Co, Cu, Fe, Ni, Pb and Zn) depending on the composition of the test materials. All extractions were prepared and analyzed in triplicate.

Each laboratory generated a comprehensive report, which underwent a QA exercise by the study coordinator prior to the publication of the results. A detailed review and comparison between the SOP and the 5 laboratory reports was performed. As part of this review, individual exchanges were held with the labs to address information gaps and confirm data when necessary. Finally, two independent biostatisticians were commissioned to perform statistical analysis of the within-lab repeatability and between-lab reproducibility. A detailed summary of the study organization and management is provided in Attachment 13.

2.1.2 Study objective and goals

Please specify the objective and goals of the study (e.g. evaluation of the transferability of the test method, evaluation of WLR and/or BLR, and/or assessment of PC) here below.

The objective and goal of the inter-laboratory study was to perform a cross-laboratory testing of different metal-containing materials in different simulated fluids that are relevant to characterizing key routes of human exposure (e.g. oral, inhalation. dermal), using defined protocols. Of direct relevance to the present submission was the testing in gastric fluid. Five laboratories measured the release of metal from six different metals and metal-containing materials in synthetic gastric fluid. The results of these bioelution analyses were evaluated by characterising within-laboratory repeatability (WLR) and between-laboratory reproducibility (BLR) measures. One of the fluids included was gastric fluid, relevant to oral ingestion and to the fraction of an inhaled dose that is swallowed and systemically absorbed from the gastrointestinal tract.

2.1.3 Summary of the study results

Please briefly summarise the study main results, from the information included in the other sections of the TST, in light of the objectives and goals set for the study (e.g. agreement of the results for WLR and BLR when experimental data have been generated using the same test items and adhering to the same protocol, accuracy of the test method to measure or predict the effect of interest).

Laboratory data for bioaccessibility after 2 h in synthetic gastric fluid were available for all but the Cr from Inconel alloy. In this treatment condition, Ni from Ni compound measurements were the least variable within and across labs, with Pb from leaded brass alloy and Co from Co compound also demonstrating relatively low variability for both measures. Iron from the Inconel alloy, a dataset with the fewest bioaccessibility measures for the gastric fluid treatment, demonstrated some of the highest variability for both measures. Metals measured included: Co, Cr, Cu, Fe, Ni, Pb, and Zn.

Analyses of repeatability measures using two different approaches (sR:sr ratios and RSD) show that the within-laboratory variability was generally satisfactory. The average repeatability standard deviation (sr) of the log-concentration varied slightly (between 0.009 and 0.083; average 0.047). These findings demonstrate good within-laboratory repeatability. With regards to between-laboratory reproducibility, good agreement between mean results of the laboratories was observed with three measurements (Co from Co compound, Ni from Inconel alloy, Ni from Ni compound; sR:sr <3) and fair agreement was observed with five measurements (Cu from Copper concentrate, Cu from Leaded brass alloy, Fe from Inconel alloy, Pb from Leaded brass alloy, Zn from Leaded brass alloy; sR:sr =3-6). Only one

measurement had a ratio slightly above the fair range (sR:sr = 6.2) and the average sR:sr ratio across all test materials was 3.4. Similar findings were observed when analyzing the data using relative standard deviation analysis (RSD). All measurements in gastric fluid were well below the 10% and 20% criteria for RSD for within-laboratory repeatability and between-laboratory reproducibility, respectively. The low within-lab variability was demonstrated by all RSD values being <4% and the average across all test materials = 1.2%. Between-laboratory reproducibility also has low variability, with all measurements have RSD values <12%, and an average across test materials of 4.3%. The results of the study are presented in Attachment 4.

2.1.4 List of test items used in the validation study

Please provide as Attachment 14, a table listing the test items used in the validation study and specify for each test item whether it has been used for a) test method development, b) WLR and/or BLR assessment and c) predictive capacity assessment.

Six different inorganic metal-containing materials were selected for inclusion in the inter-laboratory validation study and used to assess WLR and BLR: Co oxide, Co powder, Cu concentrate, Inconel alloy, leaded brass alloy, and Ni sulfate hexahydrate. These samples were tested as received without further grinding or other manipulation to alter particle size. The list of test items used in the validation study is provided in Attachment 3.

2.1.5 Study conclusions

Please provide your conclusions regarding the outcome of the study [e.g. is the test method easily transferable to other laboratories? Is the test method generating reproducible results within a single laboratory and between laboratories? Is the test method sufficiently accurate compared to the reference data (e.g. in vivo data)/ target accuracy and reliability values (for performance standards-based validation studies)?], in the light of the study objectives and in consideration of the extent to which such conclusions are supported by the study results.

In conclusion, analyses using two different approaches (sR:sr ratios and RSD) show that the within-laboratory repeatability and between-laboratory reproducibility were satisfactory regarding the bioaccessibility of metals in the synthetic gastric fluid, i.e., the test method generated reproducible results within a single laboratory and between laboratories. These results were obtained with powder samples tested as received. The possible impact of manipulation of the material to get an adequate sample size has not been assessed during the round robin. Similarly, repeatability and reproducibility associated with the testing of epoxy embedded massive forms of alloys was not assessed.

The objective of the test method is not to predict actual *in vivo* bioavailability values, but rather to allow comparisons of relative bioaccessibility of a given metal between test materials. The comparative nature of the approach is inherent to both proposed applications. As such, there are no conclusions regarding whether the method is "sufficiently accurate compared to the reference data." Rather, it is useful to compare the results of the current study to those of similar inter-laboratory validation studies of specific bioelution methods. In the study of Drexler and Brattin (2007) an *in vitro* relative bioaccessibility leaching procedure (RBALP) designed to mimic oral Pb exposure conditions was performed by three laboratories on 19 different test materials. The results of each lab were subsequently compared to *in vivo* relative bioavailability (RBA) measures. The authors reported that the intra- and inter-laboratory in vitro results were ''highly reproducible'' with a coefficient of variation (e.g. RSD) equal to 6% and 4%, respectively, and concluded that the RBALP method could reliably estimate Pb RBA *in vivo*. Another round-robin study looked at a different physiologically-based method for estimating the bioaccessibility of Pb, as well as Cd and As,

from soils (Wragg et al. 2011). The United BARGE Method (UBM) method, which includes synthetic saliva, gastric and intestinal fluids, was used to assess metal release from As, Cd, and Pb samples. Measurements from seven laboratories were compared to *in vivo* RBA data and the overall outcomes were evaluated based on a set of four benchmark criteria. Results of the UBM method were reported to have met the inter-laboratory criteria for As (RSD = 7.43% for stomach phase and 15.72% for stomach + intestine phase). However, compliances for the stomach phase only for Pb (RSD = 22.78%) and stomach plus intestine phases for Cd and Pb (RSD = 35.35% and 81.39%, respectively) were above the benchmark criteria (i.e., RSD ≤20%). The authors suggested that tighter control of gastric pH may be helpful and noted that a follow up inter-laboratory study would be needed.

Using the same RSD criteria the results of the current study appear to be in line with those of Wragg et al. (2011). In the context of some other studies of similar characteristics it is possible that the criteria used here (RSD ≤10% and ≤20% for intra- and interlaboratory variability, respectively) may be too stringent. An RSD of 30% or even 40% may be a more realistic cut-off for determining acceptable variation between laboratories. For example, in one study using a saliva migration test for organic plasticizers, where 15 labs performed validation of the SOP, an RSD of 30% was found to be the best obtainable reproducibility (EUR 19826 EN, 2001). Similarly, in a study to validate the now OECD accepted method (Transformation Dissolution protocol, OECD 2001) for environmental assessment of metals, Skeaff et al. (2011) reported that the inter-laboratory variability ranged according to analysis by % Coefficient of Variance (%CV; similar to %RSD). In this study, 12/37 measurements had %CV values between 25–56% and 10/37 had values ≥57%.

The above discussion applies exclusively to estimates of absolute metal releases into the solution. However, for the two applications discussed in this document, only measurements of relative metal release from two or more forms of the same metal are needed (e.g. 2 compounds of the same metal for application 1; alloy and pure ingredient for application 2), diminishing the requirement for high inter-laboratory reproducibility in absolute metal releases. The high within-laboratory repeatability (RSD < 4%) and inter-laboratory reproducibility (RSD < 12%), supports the use of these methods for the assessment of relative metal release and allow use of these data for grouping and read-across of substances of the same metal (application 1-Grouping) or calculation of bioaccessible concentration of metals in alloys where a matrix effects can be present (application 2-Alloy classification).

While a proper transferability study has not been carried out, the round robin study provides indirect information on transferability. Some of the participating labs had experience with a "gastric fluid bioelution test" and some did not (they had experience with T/Dp testing). All the labs tested the same samples using the same gastric fluid SOP and the results showed good reproducibility. Yet, the inter-laboratory exercise suggested that the degrees of freedom within the SOP could be further minimized in order to achieve better concordance in absolute metal releases. As such, recommendations have since been implemented into the present SOP as discussed in Section 2.1.6 below.

2.1.6 Recommendations

Considering the study objectives and outcome, please provide recommendations, where applicable, regarding a) possible improvements of the test method (e.g. in relation to the SOP, prediction model etc.), b) future activities to be undertaken to better characterise the performance of the test method in view of its envisaged use (e.g. better characterisation of the PC by testing additional coded chemicals), c) the possible current use of the test

method (e.g. screening method, partial replacement method as part of a testing strategy, full replacement method), d) any other recommendation on future activities.

The inter-laboratory exercise suggested that the degrees of freedom within the SOP could be further minimized in order to achieve better concordance in absolute metal releases. Careful comparison of each of the laboratory reports revealed that the SOP might not have been precise enough for some parameters. In addition, a systematic comparison between the SOP and the reports from the five laboratories also identified a number of methodological issues where differences were noted. These differences, and any potential corresponding impact on the outcome of the study, were first evaluated by an independent third party reviewer (R. Henderson of ToxStrategies) during the drafting of the manuscript.

For example, in some labs the control of the pH presented challenges given the low pH required to maintain the gastric environment. Following this assessment, methodological differences warranting further consideration were discussed in detail with each of the laboratories and sponsors. The SOP was subsequently refined in 2015 and updated to address and improve upon these findings. For example, the revised SOP now instructs the initial pH of the gastric fluid to be set at 1.5. For each sample, the pH is monitored at the end of the 2-h period to assess any deviations that may have taken place over the duration of the test.[16]

With regards to testing in the simulated gastric fluid, the directions and specifications for the following steps were refined in the current SOP presented as part of this TST:
1. Sample loading precision, with additional emphasis on correcting all release values for the actual loading for each sample
2. Sample and fluid loading procedure defined in a step-wise approach, with additional step added to swirl vessels containing massive samples
3. Agitation to be limited to one speed, independent of instrument used
4. pH procedure defined in a step-wise approach, with additional requirement for separate test vessel for measuring pH after adding test medium to the test material (t=0)
5. Extraction and particle/solution separation procedure defined in a more specific step-wise approach
6. Analytical information, reporting, and record keeping are better specified and defined

The limits of detection (LODs) in the Henderson et al. (2014) round robin study varied depending upon the metal, fluid, loading, and analytical methodology used by each laboratory (e.g. AAS-flame, AAS-GF, ICP-MS, ICP-AES). The variable LODs precluded the possibility of using the measurements that were below the LOD, either by substituting them with the LOD or replacing them by a fraction of the LOD. In addition, because LODs/limits of quantification (LOQs) were 1) not both required or 2) not defined as part of the original SOP, the necessary information was not collected and/or clarified until after the study and could not be utilized in data analysis. The revised SOP now contains additional stipulations regarding all analytical information required, including reporting limits for each element (such as LOQ or at a minimum the LOD).

[16] At the end of the test, the pH of the triplicate test item vessels may deviate significantly from 1.5, depending on the nature of the test material (e.g. tin) and should be recorded. In such a situation, the pH of the triplicate samples test material deviating significantly, it is recommended to compare the changes in pH between the high and low loadings. If the pH is above 4.5 and there is a difference between the low and high loadings, it shall be considered to use the results from the loading (expected to be the lower one) at which the pH is closer to 1.5. If the pH among the 3 test material replicates at the end of the test differ by more than 10%, contamination could be suspected and the test should be repeated.

Finally, other recommendations considering the study objectives and outcome include ensuring that substances that are being compared (e.g. Cu metal and Cu alloy) should always be tested side-by-side or at least in the same lab. In addition, two particle loadings are recommended (0.2 g/L and 2 g/L). The two particle loadings aim to minimize effects such as agglomeration and abrasion on one hand and to overcome the variability associated with low metal releases close to the LOD on the other hand. In all cases, realistic conditions need to be considered. The main differences between round robin and current SOP are noted in Table 11 (section 1.6.2).

2.2 Essential information for a Retrospective Validation Study

Retrospective validation studies of test methods are performed on the basis of available information/data collected from different sources (such as validation studies, in-house data, unpublished company data, publications, communications, etc).

2.2.1 Information/Data Collection

2.2.1.1 Information/data sources

Please detail the sources used for collecting the information/data.

Not completed as information is under 2.1.

2.2.1.2 Selection criteria applied to the collected information/data

In a retrospective validation study not all the retrieved information/data might further be considered to assess the test method. Collected information/data are usually included/excluded based on pre-defined criteria, such as e.g. quality of study/data. Please list the inclusion/exclusion criteria applied to the collected information/data.

[…]

2.2.1.3 Information/data collected

Please summarise and number in a table the information/data sets collected to compile Section 1 of this TST, and identify the source for each information/data set, its inclusion or exclusion and the criteria used for inclusion/exclusion.
Please append a Table in the following format as Attachment 15.

Continuous number of information/data set	*Identification of information/data set retrieved*	*Proprietary data included (YES/NO)*	*Source*	*Included (YES/NO)*	*Inclusion/Exclusion criteria used*

2.2.2 Protocols identified

Information/data sets considered to assess the test method in the retrospective validation study might have been generated using different protocols. Please provide a summary (e.g. in the form of a table) of the main protocols identified and highlight important differences between them. Please link each identified protocol to the appropriate numbered information/data set(s) listed in the table under section 2.2.1.3 (Attachment 15). Propose critical components of the protocol which may impact on the reproducibility and/or predictive capacity of the test method. Indicate the preferred or recommended protocol(s), if applicable.

[…]

2.2.3 Data analysis and Prediction Models

Please describe the various procedures that have been used in identified information/data sources to translate the in vitro data into predictions of the effect in the target organism, system or organ.(e.g. Prediction Model), and how they were derived. Please state if any of those were/shall be used in the retrospective validation study or if a new procedure has been developed. In the latter case, please describe the new procedure.

[…]

2.3 Essential information for a Validation Study based on Performance Standards (for similar or updated test methods)

2.3.1 Performance standards based on the validated reference test method(s)

Please identify the performance standards to be considered for assessing similarity, reliability and accuracy of the submitted test method when compared to the validated reference test method(s). Please append the performance standards as Attachment 16.

Not applicable

2.3.2 Essential test method components (used for similarity assessment)

Please describe how the submitted test method adheres to/fulfils the essential test method components as specified in the performance standards.

[…]

3. ADDITIONAL INFORMATION

3.1 Additional information

Please add any additional information that was not covered in the TST.

[…]

3.2 List of references

Please provide a list of scientific or other appropriate references, briefly describing their relevance with regard to the submitted test method. Please include all publications that provide background information on the test method's biological and mechanistic relevance. If available, please include publications that provide direct information of the performance of the proposed test method. In cases of submissions of putative similar or updated test methods, please include publications relating to the validated reference test method(s) on which – if available – the performance standards (PS) are based. Please append publications as electronic files (pdf or scans of paper copies) as Attachment 17.

This is the complete list of references used in the TST. All these papers are referred to in the TST text as supporting either the test protocol, its relevance, or the predictive capacity. Copies of key studies are provided in Attachment 17 (zip file).

- Alloy LLC. 2017. Bioavailability of Metals from Soil: Review of In Vitro Methods to Predict RBA (unpublished).
- ASTM (American Society for Testing and Materials). 2007. Standard test method for determining extractability of metals from art materials. In: ASTM, Annual Book of ASTM Standards, vol. 06. 02, D5517–03. ASTM, Philadelphia, US.
- Axson JL, Stark DI, Bondy AL, Capracotta SS, Maynard AD, Philbert MA, Bergin IL, Ault AP. 2015. Rapid Kinetics of Size and pH-Dependent Dissolution and Aggregation of Silver Nanoparticles in Simulated Gastric Fluid. J Phys Chem C Nanomater Interfaces, 119(35):20632-20641.
- BARGE. 2016. The BARGE unified bioaccessibility method. The Bioaccessibility Research Group of Europe. https://www.bgs.ac.uk/barge/ubm.html.
- BS (British Standard) EN 71–3. 2013. Safety of toys. Migration of certain elements. Cordeiro F, Baer I, Robouch P, Emteborg H, Got JC, Kortsen B, de la Calle B. 2012. IMEP-34: Heavy metals in toys according to EN 71–3:1994; Interlaboratory Comparison Report. JRC Scientific and Policy Reports. EUR 25380 EN.
- BS (British Standard) EN 1811. 2011. Reference test method for release of nickel from all post assemblies which are inserted into pierced parts of the human body and articles intended to come into direct and prolonged contact with the skin.
- Brandon EFA, Oomen AG, Rompelberg CJM, Versantvoort CHM,Van Engelen GM, Sips AJAM. 2006. Consumer product in vitro digestion model: bioaccessibility of contaminants and its application in risk assessment. Regulatory Toxicol Pharmacol, 44: 161-171.
- Brattin W, Drexler J, Lowney Y, Griffin S, Diamond G, Woodbury L. 2013. An in vitro methodfor estimation of arsenic relative bioavailability in soil. J Toxicol Environ Health, Part A, 76: 458-478.
- Cage. 2003. Copper compounds: In vitro dermal penetration study through human skin. Huntingdon Life Sciences, Report no. CSV 004/023929 (unpublished).

- Casarett and Doull's Toxicology: The Basic Science of Poisons, 8th edition, Editor C. D. Klaassen. 2013. ISBN 978-0071769235, McGraw-Hill Education.
- Denys S, Caboche J, Tack K, Rychen G, Wragg J, Cave M, Jondreville C, Feidt C. 2012. InVivo Validation of the Unified BARGE Method to Assess the Bioaccessibility of Arsenic, Antimony,Cadmium, and Lead in Soils. Environ Sci Technol, 46: 6252-6260.
- Diamond GL Bradham KD, Brattin WJ, Burgess M, Griffin S, Hawkins CA, Juhasz AL, Klotzbach JM, Nelson C, Lowney YW, Scheckel KG, Thomas DJ. 2016. Predicting oral relative bioavailability of arsenic in soil from in vitro bioaccessibility. J Toxicol Environ Health, 79(4):165–173.
- Drexler JW, Brattin WJ. 2007. An in vitro procedure for estimation of lead relative bioavailability: with validation. Hum Ecol Risk Assess 13, 383–401.
- ECTX. 2017. X02a-130: A 2 hours Bioelution Study on Fe-Cr-Ni Powder Mixture (68.6/16.4/10.7 mass fraction) at a 0.2, 1 and 2 g/L loading in a simulated gastric fluid.
- ECTX. 2016: X02a-085: 2 hours bioelution study on nickel metal powder at a 0.2 g/L loading in simulated gastric fluid
- ECTX. 2016: X02a-084: 2 hours bioelution study on stainless steel 316L powder at a 2 g/L loading in simulated gastric fluid
- ECTX. 2018a. X02a-144: 2 hours bioelution study on stainless steel 316L powder at a 2 g/L loading in simulated gastric fluid- effect of mass sample.
- ECTX. 2018b. X02a-145: 4 hours bioelution kinetics study on stainless steel 316L powder at a 2 g/L loading in simulated gastric fluid.
- ECTX. 2018c. X02a-146: 4 hours bioelution kinetics study on Z45 (massive) at a 2 g/L loading (surface equivalent) in a simulated gastric fluid
- ECTX. 2017. X02a-081: 2 hours bioelution study on stainless steel 316L Massive (epoxy embedded) at a 2 g/L loading (surface equivalent) in a simulated gastric fluid.
- ESTCP. 2012. Validation of an In Vitro Bioaccessibility Test Method for the Estimation of the Bioavailability of Arsenic from Soil and Sediment Cost and Performance Report Environmental Security Technology Certification Program U.S. Department of Defense (ER-200916).
- EU definition of nanomaterials: Commission Recommendation of 18 October 2011 on the definition of nanomaterial (2011/696/EU). https://ec.europa.eu/research/industrial_technologies/pdf/policy/commission-recommendation-on-the-definition-of-nanomater-18102011_en.pdf
- EU CLP (Classification and Labelling Packaging Regulation. 2008. (EC) No 1272/2008: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02008R1272-20170101
- EU REACH. 2007. Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02006R1907-20161011&from=EN
- EUR 19826 EN. 2001. Validation of methodologies for the release of diisonoylphthalate (DINP) in saliva stimulant from toys
- EU (RAR) Risk Assessment Report Cd metal and Cd oxide. 2004. https://echa.europa.eu/documents/10162/4ea8883d-bd43-45fb-86a3-14fa6fa9e6f3
- Filon FL. 2004. In vitro percutaneous absorption of cobalt. Int Arch Occup Environ Health 77: 85-89.
- Gamer AO, Leibold E, van Ravenzwaay B. 2006. The in vitro absorption of microfine zinc oxide and titanium dioxide through porcine skin. Toxicology in vitro, Vol. 20, Issue 3, 301-307.

- Grötsch. Cutaneous Permeation of Zinc Oxide and Zinc Sulphate through Pig Skin In Vitro. Final Report. Study Nrs. 02073979/02073989. Labor L+S AG, Bad Bocklet, Germany, 1999.
- Hedberg YS, Herting G, Latvala S, Elihn K, Karlsson HL, Odnevall Wallinder I. 2016. Surface passivity largely governs the bioaccessibility of nickel-based powder particles at human exposure conditions. Regulatory Toxicol Pharmacol, 81:162-170.
- Hedberg YS, Odnevall Wallinder I. 2015. Metal release from stainless steel in biological environments – a review. Biointerphases – special issue on Ions and Solvation at Biointerphases, 11(1).
- Hedberg Y, Midander K. 2014. Size matters: Mechanism of metal release from 316L stainless steel particles is governed by size-dependent properties of the surface oxide. Mater Lett 122, 223-226.
- Hedberg Y, Hedberg J, Odnevall Wallinder I. 2012. Particle characteristics and metal release from natural rutile (TiO2) and zircon particles in synthetic body fluids. Journal of Biomaterials and Nanobiotechnology, 3: 37-49.
- Hedberg Y, Gustafsson J, Karlsson HL, Möller L, Odnevall Wallinder I. 2010a. Bioaccessibility, bioavailability and toxicity of commercially relevant iron- and chromium-based particles: in vitro studies with an inhalation perspective, Particle and Fibre Toxicology, 7:23.
- Hedberg Y, Jiang T, Odnevall-Wallinder I. 2010b. Bioaccessibility of antimony released from four different antimony compounds in synthetic biological media. Commissioned by International Antimony Association (i2a).
- Henderson RG, Cappellini D, Seilkop SK, Bates HK, Oller AR. 2012. Oral bioaccessibility testing and read-across hazard assessment of nickel compounds. Regul Toxicol Pharmacol, 63 (1): 20–28.
- Henderson R, Verougstraete V, Anderson K, Arbildua JJ, Brock TO, Brouwers T, Cappellini D, Delbeke K, Herting G, Hixon G, Wallinder IO, Rodriguez PH, Assche FV, Wilrich P, Oller AR. 2014. Interlaboratory validation of bioaccessibility testing for metals. Regul Toxicol Pharmacol, 70(1):170-181.
- Herting G, Jiang T, Sjöstedt C, Odnevall Wallinder. 2014. Release of Si from silicon metalloids, a FeSi alloy and a silicate mineral in simulated biological media. PLoS ONE, 9(9).
- Herting G, Odnevall Wallinder I, Leygraf C. 2006. Factors that influence the release of metals from stainless steels exposed to physiological media, Corrosion Science, 48: 2120-2132.
- HERAG (Health Risk Assessment Guidance for metals). 2008. Assessment of occupational dermal exposure and dermal absorption for metals and inorganic metal compounds. https://www.ebrc.de/downloads/HERAG_FS_01_August_07.pdf
- Hillwalker WE, Anderson KA. 2014 Bioaccessibility of metals in alloys: Evaluation of three surrogate biofluids. Environmental Pollution, 185: 52-58.
- Hostýnek JJ, Dreher F, Pelosi A, Anigbogu A, Maibach HI. 2001. Human stratum corneum penetration by nickel. In vivo study of depth distribution after occlusive application of the metal as powder. Acta Derm Venereol Suppl, 212:5-10.
- Hostýnek JJ1, Dreher F, Nakada T, Schwindt D, Anigbogu A, Maibach HI. 2001. Human stratum corneum adsorption of nickel salts. Investigation of depth profiles by tape stripping in vivo. Acta Derm Venereol Suppl, 212:11-8.
- ISO/IEC:2017. 2017. General requirements for the competence of testing and calibration laboratories. https://www.iso.org/standard/66912.html

- ISO (International Organization for Standardization) 5725–2, 1994. Accuracy (Trueness and Precision) of Measurement Methods and Results, Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method. International Organization for Standardization, Geneva.
- Jantratid E, De Maio V, Ronda E, Mattavelli V, Vertzoni M, Dressman JB. 2009. Application of biorelevant dissolution tests to the prediction of in vivo performance of diclofenac sodium from an oral modified-release pellet dosage form. Eur J Pharm Sci, 37(3-4):434-41.
- Jiang T, Odnevall Wallinder I, Herting G. 2012. Chemical stability of chromium carbide and chromium nitride powders compared with chromium metal in synthetic biological solutions. ISRN Corrosion.
- Juhasz AL, Scheckel KG, Betts AR, Smith E. 2016. Predictive capabilities of in vitro assays for estimating Pb relative bioavailability in phosphate amended soils. Environ Sci Technol, 50(23):13086-13094.
- Juhasz AL, Herde P, Herde C, Boland J, Smith E. 2014. Validation of the predictive capabilities of the Sbrc-G in vitro assay for estimating arsenic relative bioavailability in contaminated soils. Environmental Science & Technology, 48: 12962-12969.
- Juhasz AL, Weber J, Smith E, Naidu R, Marschner B, Rees M, Rofe A, Kuchel T, Sansom L. 2009. Evaluation of SBRC-Gastric and SBRC-intestinal methods for the prediction of in vivo relative lead bioavailability in contaminated soils. Environmental Science & Technology, 43:4503- 4509.
- Kalantzi L, Goumas K, Kalioras V, Abrahamsson B, Dressman JB, Reppas C. 2006. Characterization of the human upper gastrointestinal contents under conditions simulating bioavailability/bioequivalence studies. Pharm Res, 23: 165−176.
- Koch I, Reimer KJ, Bakker MI, Basta NT, Cave MR, Denys S, Dodd M, Hale BA, Irwin R, Lowney YW, Moore MM, Paquin V, Rasmussen PE, Repaso-Subang T, Stephenson GL, Siciliano SD, Wragg J, Zagury GJ. 2013. Variability of bioaccessibility results using seventeen different methods on a standard reference material, NIST 2710. Journal of Environmental Science and Health Part a-Toxic/Hazardous Substances & Environmental Engineering, 48(6):641–655.
- Latvala S, Hedberg J, Di Bucchianico S, Möller L, Odnevall Wallinder I, Elihn K, Karlsson HL. 2016. Nickel Release, ROS Generation and Toxicity of Ni and NiO Micro- and Nanoparticles. PLoS One. 19, 11(7).
- Maltby JR, Lewis P, Martin A. 1991. Gastric fluid volume and pH in elective patients following unrestricted oral fluid until three hours before surgery. Canadian Journal of Anaesthesia, 38 (4): 425.
- Midander K, de Frutos A, Hedberg Y, Darrie G, Odnevall Wallinder I. 2010. Bioaccessibility studies of ferro-chromium alloy particles for a simulated inhalation scenario. A comparative study with the pure metals and stainless steel. Integrated Environmental Assessment and Management, 6(3): 441-455.
- Midander K, Pan J, Wallinder IO, Leygraf C. 2007a. Metal release from stainless steel particles in vitro-influence of particle size. Environ Pollut. 145(1):51-9.
- Midander K, Wallinder IO, Leygraf C. 2007 b. In vitro studies of copper release from powder particles in synthetic biological media. J Environ Monit, 9(1):74-81.
- Midander K. 2006. Licentiate Thesis (co- supervisor), “Metal release from powder particles in synthetic biological media”.

- Molina RM, Schaider LA, Donaghey TC, Shine JP, Brain JD. 2013. Mineralogy affects geoavailability, bioaccessibility and bioavailability of zinc. Environmental Pollution, 182: 217-224.
- Mörsdorf A, Odnevall Wallinder I, Hedberg Y. 2015 Bioaccessibility of micron- sized powder particles of molybdenum metal, iron metal, molybdenum oxides and ferromolybdenum. Importance of surface oxides. Regulatory Toxicology and Pharmacology, 72(3): 447-457.
- Mudie D M, Amidon G L, Amidon GE 2010. Physiological parameters for oral delivery and in vitro testing. Mol Pharmaceutics, 7: 1388−1405.
- OECD. 2014. OECD Series on Testing and Assessment. Guidance on Grouping of Chemicals, Second Edition ISBN: 9789264274679.
- OECD. 2001. OECD Series on Testing and Assessment Number 29. Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media. ENV/JM/MONO(2001)9.
- OSWER (Office of Solid Waste and Emergency Response). (2007) Estimation of relative bioavailability of lead in soil and soil- like materials using in vivo and in vitro methods. OSWER 9285.7-77.
- Priest ND. 2004. The biological behaviour and bioavailability of aluminium in man, with special reference to studies employing aluminium-26 as a tracer: review and study update. J Environ Monit, 6: 375-403.
- RAAF (Read-across assessment framework). 2017. ECHA, https://echa.europa.eu/documents/10162/13628/raaf_en.pdf
- Roper. 2003. The in vitro percutaneous absorption of Copper from various formulations through human skin. Inveresk Research, Report no. 22829 (unpublished).
- Roper and Stupart. 2006. The in vitro percutaneous absorption of diantimony trioxide through human skin, unpublished report on behalf of International Antimony Oxide Industry Association, Study Number 775440, report date 20.04.2006, Charles River Laboratories, Edinburgh, UK.
- Ruby MV, Schoof R, Brattin W, Goldade M, Post G, Harnois M, Mosby DE, Casteel SW, Berti W, Carpenter M, Edwards D, Cragin D, Chappell W. 1999. Advances in Evaluating the Oral Bioavailability of Inorganics in Soil for Use in Human Health Risk Assessment. Env Sci & Technol, 33 (21): 3697–3705.
- Schroder JL, Basta NT, Casteel SW, Evans TJ, Payton ME, Si J. 2004. Validation of the In Vitro Gastrointestinal (IVG) Method to Estimate Relative Bioavailable Lead in Contaminated Soils. J Environ Qual, 33: 513–521.
- Skeaff J, Adams WJ, Rodriguez P, Brouwers T, Waeterschoot H. 2011. Advances in metals classification under the United Nations globally harmonized system of classification and labeling. Integr Environ Assess Manag, 7 (4): 559–576.
- Stockmann-Juvala H, Hedberg Y, Dhinsa NK, Griffiths DR, Brooks PN, Zitting A, Odnevall Wallinder I, Santonen T. 2013. Inhalation toxicity of 316L stainless steel powder in relation to bioaccessibility. Hum Exp Toxicol, 32: 1137.
- Stopford W, Turner J, Cappellini D, Brock T. 2003. Bioaccessibility testing of cobalt compounds. J Environ Monit, 5: 675–680.
- Tanojo H, Hostýnek JJ, Mountford HS, Maibach HI. 2001. In Vitro Permeation of Nickel Salts Through Human Stratum Corneum. Acta Derm Venerol Suppl Vol. 212: 19-23.
- Toner and Roper. 2004. The In Vitro Percutaneous Absorption of Lead in Three Test Preparations Through Human Skin, Report Number 24569, Inveresk Research, Tranent, Scotland, UK.

- Walsh PL, Bothe JR, Bhardwaj S, Hu M, Nofsinger R, Xia B, Persak S, Pennington J, Bak A. 2015. A canine biorelevant dissolution method for predicting in vivo performance of orally administered sustained release matrix tablets. Drug Dev Ind Pharm, 4:1- 9.
- Wragg J, Cave M, Basta N, Brandon E, Casteel S, Denys S, Gron C, Oomen A, Reimer K, Tack K, Van de Wiele T. 2011. An inter-laboratory trial of the unified BARGE bioaccessibility method for arsenic, cadmium and lead in soil. Sci Total Environ. 409: 4016–4030.

4. ATTACHMENTS

Attachment Number	Description	Tick if attached 181109	File name 16 February 2018	File name Revised Submission November 2018
Attachment 1a	Protocol(s) of the test method	X X	180216TSTAttachment1a	181109TSTAttachment 1a rev changes 181109TSTAttachment 1a rev clean
Attachment 1b	DB-ALM protocol (stand-alone-, short version, method summary)	X X - -	180216TSTAttachment1blong 180216TSTAttachment1bshort 180216TSTAttachment1bsummary	181109TSTAttachment 1blongrev changes 181109TSTAttachment 1blongrev clean - -
Attachment 2	List of test items used to develop and optimise the test method, including their CAS number	X	180216TSTAttachment2	181109TSTAttachment 2 rev
Attachment 3	List of test items used to assess WLR, including their CAS number and relevant properties	-	180216TSTAttachment3	-
Attachment 4	Data used for WLR assessment (Tables and/or Figures)	X	180216TSTAttachment4	181109TSTAttachment 4rev
Attachment 5	List of test items used to assess transferability, including their CAS number and relevant properties	NA	NA	NA
Attachment 6	Training protocol	NA	NA	NA
Attachment 7	Transfer protocol	NA	NA	NA
Attachment 8	Transfer report	NA	NA	NA
Attachment 9	List of test items used to assess BLR, including their CAS number and relevant properties	-	180216TSTAttachment3	-
Attachment 10	Data used for BLR assessment (Tables and/or Figures)	See attachment 4rev	180216TSTAttachment4	181109TSTAttachment 4rev

Attachment 11	List of test items used to assess PC, including their CAS number and relevant properties	-	180216TSTAttachment18	-
Attachment 12	Data used for PC assessment (Tables and/or Figures)	-	180216TSTAttachment18	-
Attachment 13	Project plan(s)	X	180216TSTAttachment13	181109TSTAttachment 13rev changes 181109TSTAttachment 13rev clean
Attachment 14	List of test items used in the validation study	-	180216TSTAttachment3	-
Attachment 15	Information/data collected for a retrospective validation study	NA	NA	NA
Attachment 16	Performance standards for validation of a similar or updated test method.	NA	NA	NA
Attachment 17	Key References	X	180216TSTAttachment17	181109Attachment17rev
Attachment 18	Bioavailability of Metals from Soil: Review of In Vitro Methods to Predict RBA (Alloy LLC)	-	180216TSTAttachment18	-
Attachment 19	List of possible benchmark materials	X	180216TSTAttachment19	181109TSTAttachment 19rev
Attachment 20	Summary of studies conducted by ECTX for Eurometaux at ECVAM's request	-	180216TSTAttachment20	-

NOTE: *Please label appended files by indicating the relevant Attachment Number at the beginning of the file name. If more than one file needs to be attached for each description, please use ZIP compression.*