EUROPEAN COMMISSION
DIRECTORATE-GENERAL JOINT RESEARCH CENTRE

Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Report on the Full Test Submission Assessment

TM2016-02

Bioelution - Bioaccessibility testing (Bioelution) of metals, inorganic metals compounds and metals-containing materials: simulated gastric fluid

January 10, 2019

Report on the Full Test Submission Assessment

Test Method: ***Bioelution – TM2016-02***

SUMMARY OF THE ASSESSMENT

The bioelution test method is an *in vitro* extraction method used to measure metal releases from various metal containing materials after 2 hours of exposure in artificial gastric fluid (pH 1.5). It is proposed to provide an *in vitro* estimate of oral bioaccessibility, which is defined as the fraction of a substance that is released in the gastrointestinal tract and becomes available for absorption. Bioaccessibility is proposed as a conservative measure of bioavailability.

The bioelution test method is not intended to predict *in vivo* bioavailability values but to allow comparisons of relative bioaccessibility of a given metal between test materials. The Submitter has proposed the use of relative bioaccessibility for two regulatory applications, called 1-Grouping and 2-Alloy classification. Under Grouping the *in vitro* relative bioaccessibility is used in a weight of evidence approach to support grouping of substances and read across to assess if target and reference substances can be predicted to have similar or different oral bioaccessibility and systemic effects. Under Alloy classification the calculated bioaccessible gastric concentration of the metal ion, rather than the bulk concentration, is used to assign CLP classification.

From the information provided in the Submission, the test method is applicable to metals and metalloids, inorganic metal compounds, and complex metal-containing materials (like alloys, UVCBs, pigments) in micron-size powder or massive form. It is not meant to be applied to nanomaterials, organometallic materials, no-metal or no-mineral materials or matrices for which no metal releases data has been generated to date (e.g. plastic, rubber, glass).

The Submitter requested the evaluation and validation of a protocol through the assessment of available data including the information generated in an inter-laboratory study (Henderson et al., 2014).

Full submission was received on 16 February 2018. Following evaluation, on 9 November 2018 EURL ECVAM received a revised full submission together with responses to the comments raised in June 2018.

Modules 1, 2, 4, 5 and 6 of the TST have been evaluated.

Following the evaluation of the Submission, EURL ECVAM concludes that the information provided is sufficient for an independent evaluation of the (mechanistic) relevance, reproducibility of the method and the quality of the inter-laboratory study conduct and the obtained results.

1. MODULE 1: TEST DEFINITION

1.1 Human health/environmental/biological effect addressed by the test method and biological/ mechanistic relevance

The bioelution test method does not address a biological effect. It provides *in vitro* estimates of oral bioaccessibility, which is the first step that leads to oral absorption. In this way oral bioaccessibility is related to systemic bioavailability of metal ions.

The claim made by the Submitter is that the systemic effects of metal containing materials are related to the bioavailability of the metal ion after systemic absorption. EURL ECVAM asked the Submitter to justify this claim and to flag eventual exceptions.

The Submitter provided detailed answers and confirmed that for massive and micron-size powders of metals, inorganic metal compounds or metal–containing complex materials, the bioavailable metal ion is considered to be responsible for the systemic toxicity. It was further clarified that nanoparticles of metals and metal compounds may be absorbed as such from the gastrointestinal tract; however, the proposed protocol is not applicable to nanomaterials. Other exceptions identified by the Submitter are some organometallic compounds such as methylmercury and nickel carbonyl, for which the gastric protocol proposed in the Submission is, nevertheless, not applicable. The Submitter pointed also to some reviews that mentioned cadmium oxide (CdO) as a compound that may have different solubility *in vitro* and *in vivo*. The Submitter explained that this is the case if *in vitro* water solubility is used for comparison. However, if the solubility in gastric juice (which is assumed to correspond to the *in vivo* situation) is taking into consideration, the differences in solubility between cadmium compounds would be less marked. In fact, as reported in the context of the EU Cadmium metal/cadmium oxide risk assessment, the solubility of CdO was estimated to be 94% in artificial gastric fluid.

The Submitter explained that the proposed protocol is relevant for systemic health effects of metals and metal compounds and not for local effects that may occur after exposure to inorganic metals and metal compounds.

With regard to the proposed gastric fluid as the worst case scenario for the oral route of exposure, the Submitter explained that the metal release in gastric fluid is higher than the metal release in any of the other artificial biological fluids considered for the oral route (i.e. saliva and intestinal fluid either separate or in sequence). This is the case for nickel, cobalt, lead, arsenic, cadmium, chromium, titanium, copper, iron, indium, and zinc. The Submitter has illustrated it for nickel-containing samples (see Table 1 of the attachment 2 included in Attachment 18). In brief, nickel release in synthetic intestinal fluid (24h) was much lower than that in synthetic gastric fluid (2h), even for the water-soluble substances. For each substance where nickel ion release in the intestinal fluid was >1%, the substance had already released 82-91% of its available nickel ion after the first 2 hours in gastric fluid, suggesting that the contribution to systemic nickel ion levels from intestinal dissolution would be minimal. For other metals, i.e. silver, molybdenum, silicon, antimony, tungsten, and mercury, there is not enough evidence to conclude that gastric fluid will return the highest metal release values for the oral route and, therefore, the Submitter recommends testing them in saliva and/or intestinal fluid. The Submitter also explained that simple metal speciation tools that are used in e.g. ecotoxicity testing can be used to predict, for most test materials, maximum solubility. The use of these tools ahead of the bioelution testing could be useful in order to flag elements that might pose problems *in vitro*.

The relevance of the composition and the pH of the gastric fluid in relation to the human situation was also explained by the Submitter. Considering the reported pH ranges for fasted and fed gastric fluid (1.2-1.7 and 3-6.4, respectively), the proposed gastric fluid at pH 1.5 can be considered representative of worst case fasted state.

The addition of pepsin, surfactants or glycine to the medium does not increase metal dissolution. Furthermore, the Submitter explained that complete and realistic sequential testing would introduce more technical problems and a lot of variability and does not seem to improve predictivity over the simpler (gastric fluid) approach. Therefore, the proposed simple fluid composition (HCl at pH 1.5) is considered as a predictive fluid for the oral route.

The Submitter justifies the two hours' time of exposure proposed based on an average half time for gastric emptying of 17.7 minutes and complete emptying of the stomach of 91 minutes in human volunteers. This is also consistent with existing standards (ASTM D5517 – see Submission files) and available research data.

Despite the fact that the proposed method is static and issues of saturation of dissolution and/or underestimation of bioaccessibility may occur *in vitro* in comparison to *in vivo*

where dissolution process is dynamic, the Submitter explained that the lower loading of 0.2 g/L, the short duration of the assay (2 hours) and the relative estimates will make saturation improbable. This argument was further supported by the results of two kinetics studies on powder and massive samples provided with the Submission: ECTX X02a-145 (stainless Steel 316L Powder) and ECTX X02a-146 (Z45 Massive). The results indicate that no saturation of metal release after 2 hours in gastric fluid is evident for the alloy examples tested. Other published data showed that metal release is not limited by the test system but rather it depends on the nature of the material tested.

In vivo processes like metal ion competition for oral absorption, uptake of intact particles, and speciation, have an impact on metal bioavailability. The Submitter has commented on all of them in relation to the *in vitro* bioelution test method. With regard to uptake of particles the conclusion from a review provided with the submission (see Annex 3 rev: gastrointestinal absorption of microparticles and nanoparticles after oral administration), is that it appears quite significant for nanoparticles (outside of the applicability domain of the proposed protocol) but it is expected to be below 1% for micron size particles. When the speciation of metals released from target and reference substances in the gastrointestinal tract are similar, the relative bioaccessibility estimates are expected to reflect relative *in vivo* metal bioavailabilities and toxicity. Metal competition cannot be excluded; however, the results from soil studies consistently indicate that differences in gastric bioaccessibility data for different forms of a metal can predict differences in relative oral bioavailability. The same holds true for the impact of intestinal bacteria on metal bioaccessibility, where absolute *in vivo* bioavailabilities of metals from soils are usually lower than bioaccessibilities (see the submitted file 180216PAREREAnnex2 on Absolute *in vivo* bioavailability of metal compounds).

Several studies and published data support the justifications and explanations provided by the Submitter in relation to the biological relevance of the test method and summarised above (see also file 181109Response to PAREREquestionsrev clean).

1.2 Purpose and need of the test method

The bioelution test method is proposed to generate relative bioaccessible data by comparing the *in vitro* metal ion releases from target and reference material for the oral route of exposure. The relative metal ion release is then use to make predictions on how the test material will behave compared to the reference material, as long as *in vitro* and *in vivo* speciation of the dissolved metal ions are the same.

The Submitter proposes to use the relative bioaccessibility measured *in vitro* for two regulatory applications, namely 1-Grouping and 2-Alloy classification.

Under application 1-Grouping the bioaccessibility results of a given metal released from different compounds of the same metal or from different materials containing this metal are compared. The resulting relative bioaccessibility is used in a subsequent step to support grouping of substances of a given metal and predict if they have similar or different oral bioaccessibility. Since relative oral bioaccessibility is considered to be predictive of relative oral bioavailability, bioaccessibility data can contribute in a weight of evidence (WoE) approach to bridging and read across to fulfil human health data requirements under REACH and/or classification under chemical management programmes. In other words, the information generated is used in a WoE approach to determine if target and reference substances can be predicted to have similar or different oral systemic effects.

Under application 2-Alloy classification, the data can be used to assign CLP classifications for alloys. First, the bioaccessible gastric concentration of the metal ion is calculated and compared to the bulk concentration of the metal ingredient in order to assess the presence of a matrix effect. This would then warrant the use of the

bioaccessible concentration for the purpose of alloy classification for systemic health effects.

Examples illustrating both applications are included in the Submission.

The Submitter claims that, for both applications, the information generated 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 and its speciation. With regard to other routes of exposure, the Submitter explained that the results obtained with the proposed test method are also relevant for the fraction of any inhaled material that is swallowed and absorbed through the gastrointestinal tract. It is also clarified in the Submission that dermal exposure does not contribute significantly to the systemic bioavailability of most inorganic metals.

The Submitter has provided with the Submission a review of the available information to understand the relationship between *in vivo* bioavailability and *in vitro* bioaccessibility for metals (see Attachment 18 of TST). This review provides a summary of 39 studies from the published literature regarding the bioavailability and bioaccessibility of metals from soil. The evaluation focused on specific target metals: arsenic, cadmium, cobalt, lead, mercury, nickel, selenium, beryllium, and antimony. For all cases evaluated, a relationship between *in vitro* bioaccessibility and *in vivo* relative bioavailability was established. According to the Submitter the data are considered robust enough to support applications 1 (Grouping) and 2 (Alloy classification).

The Submitter claims that there is enough evidence for the need of the test method for both regulatory applications. With regard to application 1-Grouping, the use of bioaccessibility data can be considered as a refinement over the simple water solubility often used as surrogate of *in vivo* oral bioavailability and starting point for read across for a number of metal compounds in the context of grouping and read-across. In the case of application 2-Alloy classification, it is a refinement of the bulk concentration used today under CLP for classifying alloys. As explained by the Submitter, the use of the current mixture approach using the bulk concentration of the classified metal ingredients may lead to an overestimation of the risk and then the promotion of the deselection of safe materials. In some cases it can also lead to an underestimation of the risk.

With regard to the impact on the 3Rs, the Submitter claims that the use of *in vitro* bioaccessibility data in the WoE approach for grouping and read across would significantly decrease the number of animals used to fulfil data requirements under REACH. Using as example the number of metal compounds registered in 2010 (around 324) and the number of human health endpoints for which data are required (about 10), the Submitter estimated that 3.200 animal tests would have been needed. However, the Submitter explains that Industry reduced that number considerably by including bioelution data in the dossiers submitted to ECHA.

1.3 Protocol(s) of the test method

A detailed protocol (SOP) has been provided as Attachment 1a to the TST (Version submitted in February 16 2018) and revised following requests for clarification by EURL ECVAM (Version submitted in November 9 2018).

Three different forms of a DB-ALM protocol (long and short version) or Method summary were submitted as Attachment 1b (version submitted February 16 2018). Upon request by EURL ECVAM only the DB-ALM Protocol (Attachment 1b long) was revised and submitted in November 9 2018. This version was then harmonised with the revised protocol (SOP).

Following the comments raised by EURL ECVAM, the revised protocol submitted clearly describes the test system used, the endpoint measured as well as the quality criteria applied to test system, the acceptance criteria applied to the results and the positive and

the negative controls used and the proficiency materials. Certified standard metal solutions are also included and described.

The test system consists of a minimum of 50 mL of simulated gastric fluid (0.07 N HCl), pH 1.5, at 37°C in a test vessel. The pH of the test medium at the beginning of the test and the negative control at the end of the test are used as quality criteria of the test system and should fall within the test boundaries conditions: pH 1.5 ± 0.1 and temperature 37°C ± 1 °C.

With regard to proficiency materials, a list is provided as Attachment 19rev and the Submitter proposes to set up a physical repository and publically available database that will contain repeated measurements obtained with these materials. The ultimate aim is to allow the establishment of the upper and lower confidence interval for each proficiency material included in that list.

The Submitter has provided the rationale behind the selection of the two loadings used in the protocol (0.2 g/L and 2 g/L) claiming that they are a good compromise between having sufficient amounts of sample to accurately detect metal releases. With regard to the analytical methods, the protocol mentions ICP-MS or ICP-AES. They are provided as examples of the most commonly used methods for measuring dissolved element concentrations and, therefore, they are recommended. The protocol highlights, nevertheless, the importance of carrying the analysis according to an acceptance quality system that assures validated analytical instrumentation and qualified personnel.

Issues related to the nature of the test items used with the bioelution test method and handling are well described in the revised protocol. In particular, the preparation of epoxy embedded samples is well detailed in the version submitted in November 9 2018.

There are not elements in the protocol covered by proprietary issues.

The information provided in the TST with regard to the protocol is consistent with Attachments 1a and b.

Furthermore, the test submitter has clarified the issues raised by EURL ECVAM with regard to the protocol(s) (see file 181109RCOM ECVAM).

1.4 Data analysis and prediction model

The bioelution test method is proposed to generate relative bioaccessible data for the oral route of exposure by comparing the *in vitro* metal ion releases from target and reference material. The test submitter has clarified the rationale behind the relative measurement of bioaccessibility between reference and target samples and provided examples of how data are used in the context of the two regulatory applications proposed (i.e. Grouping and Alloy classification).

The information collected and analysed in Attachment 18 provides the basis to understand how well the *in vitro* bioelution test data correlates with *in vivo* relative bioavailability. The data presented come from studies of soil and focus on specific target metals (arsenic, cadmium, cobalt, lead, mercury, nickel, selenium, beryllium, and antimony). The studies included in the review evaluated metal compounds. The claim made is that once the relationship between gastric bioaccessibility and bioavailability 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 the same metal (e.g. alloys).

In the case of application 1-Grouping, relative bioaccessibility data is considered in a WoE approach to assess if target and reference substances are predicted to have similar or different oral systemic toxicities. The Submitter has included in the TST data from Henderson et al, 2012 to illustrate the use of bioelution data for grouping and read across.

For application 2-Alloy classification, the bioaccessible concentration is calculated and compared to the bulk concentration in order to assess the presence of a matrix effect, which will then warrant the use of bioaccessible concentration of the classified metal ingredient to assign the CLP classification. The Submitter has summarised in the TST data from Hillwalker and Anderson, 2014 to illustrate calculation of bioaccessible concentration and alloy matrix effects. Also additional experiments carried out by ECTX [i.e. ECTX X02a-144 (SS316L Powder), ECTX X02a-130 (Fe-Cr-Ni Powder Mixture)] and provided with the Submission demonstrate the presence of matrix effect in alloys and the calculation of bioaccessible concentration (see also summary of ECTX experiments in Attachment 20).

1.5 Test items used for developing and optimising the test method (protocol and prediction model)

The Submitter has provided, as Attachment 2rev, a long list of test items used during the development and optimisation phases of the test method protocol. Several papers are cited that served as the basis for the protocol used in the inter-laboratory study reported by Henderson et al., 2014. They include bioaccessibility of various cobalt and nickel compounds, as well as measurement of Cu, Cr, Mn, Mo, Si, Ti, Zn, and Zr releases from compounds and alloys. In addition, other metals and metal containing samples (Rh, Ru, Pd, Ag, Au, Re, Cu, In, Ge, Cd, Pb, Si, Al, Zr, Sb, V, Sn) from unpublished datasets are included in the list.

The list provided contains metals, inorganic metal compounds and complex metal-containing materials, which are the type of materials to be tested with the proposed bioelution test protocol. In addition, powders and massive forms are represented.

In addition to citing the relevant study or publication in which the test item has been used, the Submitter has included in the table of Attachment 2rev information regarding the gastric fluid composition together with other relevant details of the protocol, such us loading, incubation time and shaking steps.

1.6 Occurrence of non-qualified test

Test that do not meet the proposed acceptance criteria were considered non-qualified tests by the Submitter.

• Availability of a LOQ 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.

The Submitter has applied retrospectively these acceptance criteria to the data from Hillwalker and Anderson, 2014 and Henderson et al. 2014 to illustrate the occurrence of non-qualified tests. Henderson et al. 2014 reports the outcome of the inter-laboratory study used as basis for Module 2 and Module 4 described in the Submission. The data available in these two papers is not very extensive but the Submitter calculated, based on an analysis of CVs and negative control values, the percentages of disqualified tests. Based on that analysis, the Submitter concludes that the proposed acceptance criteria are reasonable.

- In Hillwalker and Anderson (2014), out of 63 measurements 27 (43%) were above LOD and among those 6 (22%) had CVs that did not meet the acceptance criteria.

- In Henderson et al. 2014 LOQs were not reported but calculated from LODs as 2xLOD. Out of 43 measurements 5 (12%) had CVs that did not meet the acceptance criteria.

- Data on positive control were only available from Hillwalker and Anderson, 2014 that reported recoveries of fortification samples and [instrument] check standards from 85 to 110% and 93 to 109%, respectively. According to test submitter, the data acceptance criterion for the positive control (±10% or between 90-110% of value) was likely met.

- Negative control values were not reported in any of the two papers. Some values were reported from the laboratories participating in the round robin exercise that were reported by the test submitter: out of 58 measurements, 8 (14%) were above 20xSD of negative control, suggesting contamination and only 2 (3.5%) would have warranted a repeat of the test.

1.7 Issues and drawbacks

According to the Submitter one important limitation is that nanomaterials are not covered with the proposed bioelution test method.

Furthermore, it is not meant to be applied to organometallic materials, no-metal or no-mineral materials or matrices for which no metal releases data has been generated to date (e.g. plastic, rubber, glass).

Other than that, only minor drawbacks are mentioned in relation to the need of specialised equipment to prepare epoxy embedded and polished massive samples. In the case of massive samples up to 1 L vessels may be needed.

1.8 Essential information for a Validation Study based on Performance Standards [for similar or updated test method(s)]

Not applicable

1.9 Overall conclusions on Module 1

The information provided appears to have an overall satisfactory completeness and quality for the purpose of assessing Module 1. The comments made by EURL ECVAM on the original full submission have been addressed by the test Submitter.

2. MODULE 2: WITHIN-LABORATORY REPRODUCIBILITY

2.1 Test items used for assessing within-laboratory reproducibility

Six inorganic metal-containing materials were tested in the inter-laboratory study (Henderson et al., 2014) and used by the Submitter to demonstrate within-laboratory reproducibility. They are listed in Attachment 3.

These test items represent water soluble metal compounds, insoluble metal compounds, pure metals, and complex materials (alloys and ore concentrates). They cover the various types of metals and materials the proposed test method is intended to be applied to. They also represent a variety of physicochemical properties that impact the metal releases.

All materials were powders with median particle size <60 µm in diameter (relevant for oral exposure) and were tested as received.

2.2 Assessment of within-laboratory reproducibility

The assessment of within-laboratory reproducibility was done with 6 inorganic metal-containing materials, tested in five laboratories, as reported in Henderson et al., 2014. Results from the laboratories were randomly coded before statistical analysis.

Measurements below LOD were excluded from the analysis of within-laboratory reproducibility.

Two independent statistical analyses were carried out. One was based on the international standard ISO 5725-2 [*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, 1994]* and used Mandel's k statistics and Cochran test to evaluate within-laboratory consistency and to identify outliers, respectively. The within laboratory precision was evaluated calculating the repeatability standard deviation (Sr). Sr values for the six compounds are shown in Attachment 4. The lowest value was 0.009 (nickel compound) and the highest 0.083 (Fe-Inconel alloy 718).

The second approach used to assess within-laboratory reproducibility was based on the relative standard deviation (RSD) analysis of the log concentration and was expressed in percentages. RSD < 10% was the criteria used for acceptable repeatability. All RSD values were below 4%. The highest and lowest variability was obtained with Fe-Inconel alloy 718 (3.1%) and nickel compound (0.2%). This outcome is consistent with the above repeatability standard deviation approach.

The Submitter mentions issues related to the handling of particles in solution (e.g. aggregation, sedimentation and separation) and differences related to metastable corrosion processes inherent to metal and alloy samples, as possible sources of variability.

Besides the data of the inter-laboratory study, the Submitter has reported the variability from triplicate measurements of two different alloy embedded massive samples (Z45 Massive and SS316L Massive) and different metals (Cu, Ni and Co). These data were obtained in two additional studies provided with the Submission and performed by ECTX [X02a-146 (4 hours kinetics study) and X02a-081 (2 hours extraction)]. The Submitter reported these data in section 1.1.8 of the TST even though they do not seem to be linked to the occurrence of non-qualified tests. EURL ECVAM considered these extra data for the analysis of reproducibility. Good repeatability (< 10% criterion) was shown for all except one case (CV 14% for Z45 samples collected at 1h extraction) (see also Table 8A of TST).

2.3 Identification and discussion of outlying values

Outlying values were identified from the results obtained in the inter-laboratory study and properly discussed. Following the LOD criteria, Cr measurements from Inconel alloy were excluded from the analysis. According to the Cochran test, measurements for iron from the Inconel alloy in one laboratory were discarded (the result of the outlier test was significant at the 1% level).

According to the statistical approach used, stragglers were identified if the outlier test result was significant at 5% level. Nickel from Ni compound was identified as a straggler but retained.

2.4 Overall conclusions on Module 2

The information provided appears to have an overall satisfactory completeness and quality for the purpose of assessing Module 2. The statistical approaches used to evaluate within-laboratory reproducibility, although different from the traditional ones used in validation studies, follow international standards for the determination of repeatability and reproducibility.

The comments made by EURL ECVAM on the original full submission have been addressed by the test Submitter.

3. MODULE 3: TRANSFERABILITY

3.1 Test items used for assessing the transferability

A single transfer phase was not carried out. The Submitter, in the replies to the comments raised by EURL ECVAM explained that, according to them, information on transferability could be obtained indirectly from the inter-laboratory study (Henderson et al 2014). Therefore, the same 6 inorganic chemicals described under Module 2 were used to assess transferability.

Furthermore, the Submitter has proposed the use of one or more proficiency materials whenever the bioelution method needs to be transferred to another laboratory. A list of potential proficiency materials is provided as Attachment 19 to the TST and in an annex to the protocol (Attachment 1a).

3.2 Assessment of the test method transferability to other laboratories

From the information provided by the Submitter, it appears that the method can be easily transferred. The Submitter claims that no special training is needed and any laboratory with analytical capability, working according to a standard quality system and having knowledge of metal analysis should be able to implement the test method by following the test protocol.

The Submitter based this claim on the experience gained in the inter-laboratory study (Henderson et al., 2014), in which some of the participating laboratories had experience with a “gastric fluid bioelution test” and some did not. Nevertheless, all the laboratories tested the same samples of the test material using the same SOP and the results showed good reproducibility (see Module 4).

3.3 Overall conclusions on Module 3

Even if module 3 has not been carried out, the information provided by the Submitter with regard to training needs appears to be sufficient.

4. MODULE 4: BETWEEN-LABORATORY REPRODUCIBILITY

4.1 Test items used for assessing between-laboratory reproducibility

The chemicals used for the evaluation of the between-laboratory reproducibility are the same as those used for the evaluation of the within-laboratory reproducibility (see section 2.1 of the assessment report).

4.2 Assessment of between-laboratory reproducibility of experimental data

Results from the five laboratories that participated in the inter-laboratory study (Henderson et al., 2014) were randomly coded before statistical analysis.

Measurements below LOD were excluded from the analysis of between-laboratory reproducibility.

Two independent statistical analyses were carried out. One was based on the international standard, ISO 5725-2 and used Mandel's h statistics and Grubbs test to

evaluate between-laboratory consistency and to identify outliers, respectively. The between laboratory precision was evaluated calculating the reproducibility standard deviation (sR). Sr values for the six compounds are shown in Attachment 4. The lowest value was 0.024 (nickel compound) and the highest 0.272 (cobalt powder) followed by iron from the Inconel alloy (which also demonstrated high within-laboratory reproducibility).

The ratio of repeatability standard deviation and reproducibility standard deviation (sR:Sr) of the log-concentration was used to compare the mean results of the laboratories. Good, fair or poor agreement between laboratories was indicated by ratios up to 3 (Co compound, Ni from Inconel alloy 718 and nickel compound), between 3 and 6 (copper concentrate, Cu from leaded brass alloy, Fe from Inconel alloy 718, Pb and Zn from leaded brass alloy), and above 6 (Co from Co powder), respectively.

The second approach used to assess between-laboratory reproducibility was based on the relative standard deviation (RSD) analysis of the log concentration and was expressed in percentages. RSD_R < 20% was the criteria used for acceptable laboratory-to-laboratory result variability. All RSD_R values were below 12%. The highest and lowest variability was obtained with Fe-Inconel alloy 718 (11.6%) and Nickel compound (0.4%).

Prior to the conduct of the inter-laboratory study, the analytical procedures among the laboratories were evaluated by asking the laboratories to analyse samples of interstitial fluid spiked with known metal concentrations. In all cases the ratios sR:Sr were about 6, indicating lack of harmonisation between the laboratories. As reported in Henderson et al., 2014, following this exercise several recommendations for improving reproducibility were subsequently implemented in the protocol used in the inter-laboratory study.

Overall, the Submitter has identified as possible sources of variability affecting the between-laboratory reproducibility, differences in sensitivity of the analytical instrumentation and metal analysis techniques. Furthermore, the Submitter has indicated that methods of pH control, sample weighing and loading, agitation, and sample filtering may have varied between laboratories during the inter-laboratory study.

Besides the outcome of the inter-laboratory study, the Submitter has also included information with regard to embedded massive samples. The Submitter reported these data in section 1.1.8 of the TST even though they do not seem to be linked to the occurrence of non-qualified tests. EURL ECVAM considered these extra data for the analysis of reproducibility. Data obtained with one alloy (Z45) showed good reproducibility between two independent laboratories (see also Table 8B of TST).

4.3 Identification and discussion of outlying values

Outlying values were identified from the results obtained in the inter-laboratory study and properly discussed. Following the LOD criteria, Cr measurements from Inconel alloy were excluded from analysis. According to the Double high Grubbs test, only values for Ni in the Inconel alloy in two laboratories were discarded.

Cobalt in Co compound in one laboratory was identified as a straggler in the Single High Grubbs test but retained.

4.4 Overall conclusions on Module 4

The information provided by the Submitter appears to have an overall satisfactory completeness and quality for the purpose of assessing Module 4. The statistical approaches used to evaluate between-laboratory reproducibility, although different from the traditional ones used in validation studies, follow international standards for the determination of reproducibility.

The comments made by EURL ECVAM on the original full submission have been addressed by the test Submitter.

5. MODULE 5: PREDICTIVE CAPACITY

5.1 Test items used for assessing predictive capacity

The Submitter, in reply to EURL ECVAM's comments, has provided with the Submission a review of available information to understand the relationship between *in vivo* bioavailability and *in vitro* bioaccessibility for metals (see Attachment 18).

As reported in the review, the evaluation focused on specific target metals for which exposure is associated with systemic effects via the oral route: arsenic, cadmium, cobalt, lead, mercury, nickel, selenium, beryllium, and antimony. The majority of studies included in the review have evaluated metal compounds (as opposed to pure metals), either as the compounds are present in contaminated soils, or as pure compounds.

None of the studies dealt with characterisation of metal alloys. However, the review claims that it might be expected that *in vitro* bioaccessibility (IVBA) would be as predictive of the *in vivo* relative bioavailability (RBA) of metal from alloys as from soils, since oral absorption of metals from alloys or soil depends on the bioaccessible fraction.

5.2 Assessment of the predictive capacity of the test method

Twenty two out of the 39 studies reviewed by Alloy LLC (see Attachment 18), included information on both RBA measured in mouse, rat, and/or swine and IVBA measured in laboratory extraction tests. RBA corresponds to the bioavailability of metals from soils relative to that from fully water soluble metal compounds. The review explained that there are very few studies that have evaluated the ability of IVBA to predict absolute bioavailability of metals. The Submitter claims that 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 (8 studies) and arsenic (8 studies), with fewer available data for Sb (2 studies), Cd (3 studies), Co (1 study), Ni (6 studies), and Hg (1 study), and no data for Be and Se.

Antimony: very low RBA of antimony for all soils were tested. Bioaccessibility values were similarly low, but no strong predictive relation was identified (as assessed by $R2 < 0.3$).

Arsenic: On average, across soils tested, IVBA tends to over-predict RBA (as indicated by a regression line with slope <1), but for individual soils, IVBA may not provide a conservative estimate of RBA.

Cadmium: For assessing whether IVBA is a conservative assessment of RBA, the study results disagree. The weight of evidence would suggest that IVBA is a reasonably conservative predictor of RBA (i.e., based on the 26 soils studied by Denys et al. [2012] and Schroder et al. [2003] vs. the results for one soil reported by Kang et al. [2016]).

Cobalt: From unpublished data provided in Attachment 18, the Submitter shows bioaccessibility data for four cobalt compounds (cobalt, $CoCl_2$, Co_3O_4, and CoS), and evaluation of oral RBA for $CoCl_2$ and Co_3O_4. The Submitter explained that the data indicate that the bioaccessibility of cobalt compounds can vary considerably (from 0.2% to 86%) and that this solubility appears to be reflected in bioavailability. For instance, the compound Co_3O_4 had 0.2% bioaccessibility relative to $CoCl_2$, and 1% RBA relative to $CoCl_2$.

Lead: On average, across soils tested, IVBA tends to over-predict RBA (as indicated by a regression line with slope <1), but for individual soils, IVBA may not provide a conservative estimate of RBA.

Mercury: the available literature is too limited to draw firm conclusions regarding the ability of IVBA to predict RBA for mercury in soil.

Nickel: Information on whether IVBA provides a conservative estimate of RBA is mixed. A study by Kang et al. (2016) suggests that IVBA under predicts RBA, however as indicated in the review, this study is limited to a single soil and appears to suffer from some issues in design or interpretation. Vasiluk et al. (2011) indicates that IVBA is a conservative estimate of RBA for about half of the soils investigated. The Submitter has also included in the submission additional information on nickel showing a reasonable correlation between bioaccessibility in gastric fluid and *in vivo* oral bioavailability (compiled by NiPERA as Attachment 2 to the reviewed by Alloy LLC). Then, based on all the information provided, the review concluded that the exact relationship between RBA and IVBA cannot be fully deciphered. However, taken together the available information on nickel indicates that the solubility of nickel is predictive of an absorbed dose, including for saline and simulated gastric fluid.

Results from the review by Alloy LLC and other existing studies (e.g. Molina et al. 2013) seem to indicate that for several metals (As, Cd, Co, Cu, Mn, Ni, Pb, Sb, Zn) there is good statistical evidence to show that there is 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. The test submitters claim that in all instances a defined relationship between gastric bioaccessibility and *in vivo* bioavailability was established (see Attachment 18).

The Submitter concludes, based on the review by Alloy LLC that, in general, IVBA methods provide an efficient and effective method for predicting the relative bioavailability of metals.

The Submitter demonstrates the applicability of the bioaccessibility data to the two applications proposed in the Submission, with three examples.

The data showed by Henderson et al., 2012 illustrate how gastric bioaccessible data from metal compounds can be used for grouping and read-across in a weight-of-evidence approach (Application 1-Grouping).

With regard to application 2-Alloy classification, the test submitter refers to the experiments carried out by ECTX in 2017 and provided with the Submission as ECTX X02a-144 (Stainless Steel 316L Powder) and ECTX X02a-130 (Fe-Cr-Ni powder Mixture). The results show how the presence of a matrix effect in alloys can affect metal releases compared to those from simple mixtures. A 300-fold difference between the Ni releases from a simple mixture compared to release from the SS316L alloy was found (even if the concentration of Ni in both samples was the same). Furthermore, by comparing the bioaccessible concentrations of Ni in the simple mixture (12%) and SS316 L alloy (0.04%) is evident that in the simple mixture bioaccessible concentration coincides with its bulk concentration (11%), while in SS316L it is ~300-fold lower than its bulk concentration (11%) and 300-fold lower than the bioaccessible concentration of the simple mixture (12%). With these experiments the Submitter illustrates how the alloy matrix effect can affect metal bioaccessibility.

The Submitter also refer to other examples that illustrate how the bioaccessible concentration of a metal can differ from its bulk concentration due to alloy matrix effects [e.g. from Hillwalker and Anderson, 2014 and Mörsdorf et al, 2015 (both included with the Submission)]. Other references are cited but they report bioaccessibility data in lysosomal lung fluid rather than gastric fluid. Nevertheless, they all serve to illustrate the concept of alloy matrix effect.

5.3 Identification and discussion of false predictions

The soil studies included in the review provided as Attachment 18, have compared the gastric *in vitro* bioaccessibility (IVBA) of metals to their *in vivo* relative bioavailability (RBA). In some instances, the gastric IVBA values under-predict the RBA values. As explained in the report, this does not necessarily implies that IVBA under-predicts absolute bioavailability. This claim is justified in the report because RBA does not provide information on absolute bioavailability, which in any case depends on the extend of absorption of soluble forms of each metal. Furthermore, the Submitter has also provided with the submission (see file 180216PAREREAnnex2) information on *in vivo* absorption for the most water soluble of the metal compounds considered in the Alloy LLC review and demonstrated that the absolute bioavailabilities of the most water soluble metal compounds are always lower than 100%. The Attachment 2 to the Alloy LLC review includes also data on bioaccessibility and bioavailability for nickel compounds. These data showed that, although the relative bioavailability values can be higher than the relative bioaccessibility ones, the absolute values are generally lower than the bioaccessible values. With all pieces of evidence, the Submitter confirms that even if an IVBA is lower than an RBA, the IVBA will never be lower than the IVBA and it will not under-predict metal bioavailability.

Kang et al. 2016 evaluated the bioaccessibility of several heavy metals using *in vitro* assays and compared them with the relative bioavailability of heavy metals using *in vivo* mouse model. The authors concluded that there is "poor correlation between IVBA and RBA" and that "RBA provides more accurate information...". This paper is listed in the tables showed in Attachment 18 and a clarification note has been added suggesting interpretation issues and errors in the calculations of RBA: *Investigation of 1 soil: RBA of metals in soil with mouse model and comparison to IVBA. Authors conclude IVBA <RBA for all metals, by significant fraction – from Figure 4 in the paper, RBA appears to be higher than IVBA by more than 2X for all metals. "poor correlation between IVBA and RBA" and that "RBA provides more accurate information..." Note: averaged RBA values reported were >100% for 2 metals, and >100% for individual animals for 4 of 5 metals evaluated.*

5.4 Quality/variability/uncertainty of the reference data used to calculate predictive capacity

There is an extensive base of research to understand 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. As claimed in the review by Alloy LLC (see Attachment 18), findings from these studies of soil, including evaluation of *in vitro* extraction testing to predict bioavailability, could significantly advance the understanding of how bioelution can predict (or provide a conservative estimate of) the bioavailability of metals from alloys. In this regard the data showed in the report are considered adequate.

With regard to the uncertainty associated with the observed relationships between IVBA and RBA and in reply to one of the comments raised by EURL ECVAM on the original full submission, the Submitter pointed to a meta-analysis carried out by Diamond et al., 2016 (paper included in the Submission) that allowed to incorporate uncertainty in individual IVBA and RBA of arsenic from soil samples into the regression model. The authors concluded that their regression model is more robust than previously reported models because it includes a larger number of soil samples, and also accounts for variability in RBA and IVBA measurements made on samples collected from sites contaminated with different arsenic sources and conducted in different laboratories that have used different experimental models for estimating RBA. The Submitter concludes from this example

that, even taking into account the variability in the estimates of IVBA and RBA, the relationship between these parameters is robust.

5.5 Overall conclusions on Module 5

The information provided by the Submitter appears to have an overall satisfactory completeness and quality for the purpose of assessing Module 5. The nature of the relationship between bioaccessibility and relative bioavailability has been shown from available data from soil samples.

The comments made by EURL ECVAM on the original full submission have been addressed by the test Submitter.

6. MODULE 6: APPLICABILITY DOMAIN

6.1 Limitations of the test method

Based on the information provided by the Submitter, the test bioelution method is applicable to metals, inorganic metal compounds, or complex metal (metalloid)-containing materials (like alloys, UVCBs, pigments) in powder or massive form. When testing powders, it is recommended to test a sub-fraction of particles < 100 µm in diameter, which is relevant for the fraction of the inhalation exposure that is swallowed and, therefore, becomes available via the oral route.

For Ag Mo, Sb, Si, W and Hg gastric fluid alone should not be used for the oral route of exposure. In the case of Ag, the problem is the formation of precipitates of AgCl. For the other metals there is not enough evidence to conclude that gastric fluid will result in the highest metal release.

The bioelution test method is not meant to be applicable to nanomaterials, organometallic materials, no-metal or no-mineral materials or matrices for which no metal releases data has been generated to date (e.g. plastic, rubber, glass).

6.2 Overall conclusions on Module 6

The information provided on Module 6 appears to be consistent and sufficient.

7. MODULE 7: PERFORMANCE STANDARDS

7.1 Suggestions for essential test method components, possible reference items, accuracy and reliability values

Not applicable

7.2 Overall conclusions on Module 7

Not applicable

8. ADDITIONAL INFORMATION

8.1 Study organisation and management

The inter-laboratory study reported in the Submission was initiated in 2010. Five laboratories have participated and evaluated the within-laboratory repeatability and between-laboratory reproducibility of bioaccessibility testing in gastric fluid. The outcome of the study was published in 2014 by Henderson et al. The study organisation and management are properly described in the TST and the Attachment 13rev. Information is provided about study coordinator, study funding sponsors and steering team representatives, independent laboratories and designated Study Directors, and independent biostatisticians.

The objective and goals of the inter-laboratory study are clearly defined in the Submission.

8.2 Collection of existing (historical) information

Available information on metals *in vitro* bioaccessibility and *in vivo* bioavailability data were collected and analysed and provided as Attachment 18. The sources and criteria used are included in the review by Alloy, LLC.

8.3 Study conclusions

The conclusions provided by the Submitter are supported by the data generated in the inter-laboratory study and refer to the repeatability and reproducibility of the bioaccessibility of metals in the synthetic gastric fluid. There are not conclusions regarding whether the method is sufficiently accurate compared to the reference data because the bioelution is not meant to predict actual *in vivo* bioavailability values but to allow comparisons of relative bioaccessibility of a given metal between test materials.

The Submitter made comparisons between the outcome of the study and similar inter-laboratory validation studies of specific bioelution methods with regard to the reproducibility of the bioaccessibility data (absolute values).

Thinking on the proposed applications described in the Submission, the Submitter concludes that the high within-laboratory repeatability (RSD < 4%) and between-laboratory reproducibility (RSD < 12%), supports the use of the bioelution test method 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).

8.4 Study recommendations

The Submitter clearly explained and justified several recommendations that were made at the end of the inter-laboratory study, and that were meant to improve the protocol. They are all well reflected in the SOP submitted with the Submission (Attachment 1b). Table 11 of the TST shows the main differences between the protocol used in the inter-laboratory study and the submitted SOP.

9. OTHER

Appropriate information is provided in the Submission (section 1.2.4) with regard to the quality systems of the five participant laboratories in the inter-laboratory study.

10. OVERALL CONCLUSIONS ON THE OUTCOME OF THE ASSESSMENT

Following the evaluation of the revised Submission and the replies to the comments raised by EURL ECVAM in February and June 2018, EURL ECVAM concludes that the

modules 1, 2, 3, 4, 5 and 6 of the EURL ECVAM modular approach appear sufficiently fulfilled.

Therefore, EURL ECVAM proposes to progress the test method into the ESAC peer review process.

11. CLARIFICATIONS/ADDITIONAL INFORMATION TO BE REQUESTED TO THE TEST SUBMITTER

None

ANNEX: FOR INTERNAL USE ONLY (TO BE DELETED IN THE FINAL ASSESSMENT REPORT)

Please complete the following table by ticking for each module (M) the appropriate box (where feasible) with regard to the completeness and quality of the information provided in the Test Submission.

Completeness and quality of the information provided for each module	**Insufficient**	**Requires substantial additional work**	**Requires minor additional work**	**Complete**
M1 - Test Definition				**x**
M2 - Within-Laboratory Reproducibility				x
M3 - Transferability				**x**
M4 - Between-Laboratory Reproducibility				x
M5 - Predictive Capacity				**x**
M6 - Applicability Domain				**x**
M7 - Performance Standards				

Any additional comments from the assessment team: