

## Appendix C – Approach for assessing toxicokinetic and toxicity studies retrieved in the literature search

### 1. Background

The validity of a study with respect to risk assessment may be considered to have two dimensions. The first dimension is relevance of the endpoint(s) used for the risk assessment. This is often described as 'external validity'.

The second dimension is reliability, which considers systematic error (accuracy, risk of bias) and random error (imprecision). Risk of bias (or internal validity) is defined as 'the extent to which the design and conduct of a study are likely to have prevented bias'. Risk of bias relates to the propensity of a study to be affected by systematic error. Biases can operate in either direction and can lead to underestimation or overestimation of the true intervention effect.

Bias should not be confused with imprecision. Bias refers to systematic error, meaning that multiple replications of the same study would reach the wrong answer on average.

Imprecision refers to random error, meaning that multiple replications of the same study will produce different effect estimates because of sampling variation even if they would give the right answer on average. The results of studies with small sample numbers are subject to greater sampling variation and hence are less precise. Imprecision is reflected in the confidence interval around the intervention effect estimate from each study and in the weight given to the results of each study in a meta-analysis. More precise results are given more weight (Higgins and Green, 2011).

The Panel considered and agreed on a set of criteria to be followed for the evaluation of the evidence that will be included in the scientific opinion, compatible with the timeline and the scope of the mandate. The criteria cover the two dimensions of:

- Relevance
- Reliability

### 2. Appraising relevance (external validity) of the studies

In the current assessment for E 171, two main research questions were considered:

**1Q.** Does the food additive E 171 raise a concern for genotoxicity?

As genotoxicity is considered a standalone endpoint, the EFSA cross-cutting (cc)WG Genotoxicity developed a procedure for the appraisal of the scientific evidence addressing this question (see Appendix D).

**2Q.** Does exposure to the food additive E 171 **through the diet** raise a safety concern?

The latter question defines the overall boundaries of this risk assessment. However, to facilitate an efficient assessment of the literature database, evidence was included if it was considered relevant to the following key sub-questions:

- What is the extent of absorption of the food additive E 171 including any TiO<sub>2</sub> NPs that may be present in E 171, following oral intake?
- What is its distribution?
- Is bioaccumulation possible?
- Does E 171 cause any adverse effect on the GIT, including tumour promotion?
- Does E 171 cause any adverse effect on the liver?
- Does E 171 cause any adverse effect on the spleen and immune system?
- Does E 171 cause any adverse effect on the reproductive system or does it affect development?
- Does E 171 cause any adverse effect on the (developing) nervous system?

- Does E 171 cause any adverse effect on other organs/systems?
- Does E 171 cause GIT microbiota dysbiosis?

### 2.1.1. Initial screening for relevance

An initial exercise of screening for relevance was conducted by EFSA staff on the publications retrieved from the literature search covering the last 5 years (Appendix A), in consultation with the EFSA ccWG Nanotechnology (Appendix B). The aim of this initial step was to exclude from further screening those studies conducted with test material not relevant for E 171 and with routes of exposure or test systems deemed as clearly non-relevant.

Taking into account the advice from the EFSA ccWG Nanotechnology the criteria for inclusion and the exclusion were defined (Appendix B), the screening included toxicological studies performed with (i) the food additive titanium dioxide (E 171), (ii) titanium dioxide - other than E 171 – containing a fraction of particles <100 nm (TiO<sub>2</sub> (X% nano))<sup>1</sup> or (iii) nano titanium dioxide (TiO<sub>2</sub> NPs).

Upon advice from the EFSA ccWG Nanotechnology, an exercise of prioritisation was also made at this stage, giving precedence to the assessment of the available *in vivo* toxicity studies over the evidence obtained from *in vitro* tests and giving more weight to the studies with E 171.

### 2.1.2. Second screening for confirmation of relevance

A second screening for relevance was carried out by Expert members of the FAF WG Titanium dioxide (E 171) on the publications identified in the step above.

A decision was taken to classify relevant papers in the following categories, according to a limited set of criteria:

- **Relevant for hazard characterisation:** the study should be designed in such a way that would allow its use for identifying a reference point (RP) that could be used for deriving a health-based guidance value (HBGV):
  - o Different dose levels tested with repeated administration (for toxicity studies)
  - o A minimum duration of 28 days for repeated dose toxicity studies (except for window dosing in developmental studies)
- **Relevant for providing supporting evidence:** this category covers those studies that could provide supporting information to the risk assessment, but that would not be appropriate to use for identification of an RP. For example:
  - o Studies with a single dose level (including repeated administration), including studies in which an adverse effect is purportedly triggered by TiO<sub>2</sub>.
  - o Studies with a repeated dose range duration shorter than 28 days
- **Not relevant:**
  - o Single administration (except for toxicokinetic studies)
  - o Route of administration different from oral (except for toxicokinetic studies)

This second screening step for relevance was conducted in parallel independently by two Experts and when a conflict in classification of the study was identified, ad hoc discussion between the reviewers took place prior to a final agreed classification.

## 3. Appraising reliability of evidence

*In vivo* studies that were confirmed as relevant were individually appraised for their reliability.

<sup>1</sup> Where X refers to the percentage of particles <100 nm as described in the publication.

In other assessments, ongoing and completed by EFSA Scientific Panels supported by the Food Ingredients and Packaging Unit (FIP Unit),<sup>2,3,4</sup> the Office of Health Assessment and Translation (OHAT) Risk of Bias Rating Tool for Human and Animal Studies<sup>5</sup> has been modified and used for the assessment of experimental study (animal and human) risk of bias.

The original OHAT tool is based on sets of questions grouped under the following bias domains:

- Selection
- Performance
- Attrition/exclusion
- Detection
- Selective reporting
- Other threats to internal validity

Moreover, in the case of the current assessment of E 171 and considering that the particle size distribution of the food additive allows for the presence of a fraction of particles falling within the nanoscale (less than 50% for the percentage of constituent particles by number with a median minimal external dimension below 100 nm (FAF Panel, 2019)), two additional customised elements were identified as specific for the appraisal of the overall reliability of the evidence from *in vivo* studies and were evaluated separately by experts for each study:

- Nanoscale considerations related to specific aspects of study design (see Section 3.1.2)
- Examination of internal exposure (see Section 3.1.3)

### 3.1.1. Reliability regarding the specific 'toxicity' endpoint

Publications were allocated to two Experts, members of the FAF WG Titanium dioxide (E 171) with expertise relevant to the endpoint reported. The screening step for reliability was also conducted in parallel independently by two Experts and when a conflict in classification of the study was identified, *ad hoc* discussion between the reviewers took place prior to a final agreed classification and a final rating was assigned to the endpoint(s) investigated through consensus.

The following rating was used to indicate whether the design and conduct of the study supported the association (or the lack of it) between the exposure to the test material and the reported outcome:

1. Yes, only minor or no limitations (Reliability 1)
2. Overall yes, with some limitations (Reliability 2)
3. There are important flaws that may hamper evaluation of the association (Reliability 3)
4. No, there are major limitations (Reliability 4)

The rating assigned by the Experts took into account the following aspects of the design and performance of the studies, borrowed from the OHAT Risk of Bias tool:

Bias domain	Questions	Elements considered for the rating of reliability
-------------	-----------	---

<sup>2</sup> Protocol for risk assessment for peri- and post-menopausal women taking food supplements containing isolated isoflavones. Available at: <https://efsa.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.2903%2Fj.efsa.2015.4246&file=4246ax1-sup-0001.pdf>

<sup>3</sup> Bisphenol A (BPA) hazard assessment protocol. Available at: <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2017.EN-1354>

<sup>4</sup> Draft protocol for the assessment of hazard identification and characterisation of sweeteners. Available at: <https://efsa.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.2903%2Fsp.efsa.2020.EN-1803&file=efs31803e-sup-0001-annex.pdf>

<sup>5</sup> [https://ntp.niehs.nih.gov/ntp/ohat/pubs/riskofbiastool\\_508.pdf](https://ntp.niehs.nih.gov/ntp/ohat/pubs/riskofbiastool_508.pdf)

<p><b>Selection</b></p>	<p>Were there systemic differences between baseline characteristics of the groups that are compared?</p> <p>Did each animal have an equal chance of being assigned to study groups (randomisation of exposure)?</p> <p>Could the research personnel allocating animals to treatment groups (including the control group) foresee which administered dose or exposure level is going to be assigned at the start of a study?</p> <p>Did the research personnel know what groups the animals were allocated to?</p>	<p>Randomisation and allocation concealment methods are unlikely to be explicitly reported in the publications. Therefore, it will be assumed that, unless explicitly stated otherwise, there were no systemic differences between baseline characteristics of the group.</p>
<p><b>Performance</b></p>	<p>Were there systematic differences in the care provided to animals by study groups?</p>	<p>It is assumed that, unless stated otherwise, experimental conditions (other than use of appropriate vehicle for control animals) were identical across groups.</p>
<p><b>Attrition/Exclusion</b></p>	<p>Were there systematic differences in the loss or exclusion from analyses of animals from the study?</p> <p>How were they accounted for in the results?</p> <p>Were there very little missing outcome data?</p> <p>Were the reasons for missing animals unlikely to be related to outcome?</p> <p>Were the missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups?</p> <p>Did the missing outcomes impact the effect estimate?</p>	<p>Exclusion of individuals or animals from analyses should be clearly reported and outliers identified with appropriate statistical procedures.</p>
<p><b>Detection</b></p>	<p>Were there systematic differences between experimental and control groups with regard to how outcomes and exposures are assessed?</p> <p>Were the methods used to assess outcomes and exposures valid and reliable?</p> <p>Can we be confident in the exposure characterization?</p> <p>Can we be confident in the outcome characterization?</p>	<p>Confidence in the exposure characterisation will be largely addressed when appraising the evidence for the dispersion protocol applied to test material and when verifying whether internal exposure was measured in the study.</p> <p>When looking at each outcome relevant for this risk assessment, the WG experts will focus/focussed on the appraisal of the validity, reliability and sensitivity of the methods applied to assess the outcome consistently across the groups (i.e. using the same method and in the same time-frame).</p> <p>According to OHAT, there are three important factors for assessing bias in the outcome assessment:</p> <ol style="list-style-type: none"> <li>1) the objectivity of the outcome assessment</li> <li>2) consistency in measurement of outcomes, and</li> <li>3) blinding of the outcome assessors (for knowledge of the exposure).</li> </ol> <p>The experts will look/looked at the acceptability of the method used (e.g. gold standard, valid and reliable, insensitive), at</p>

		<p>the length of time at which the outcome was assessed (same across study groups).</p> <p>With respect to the blinding it is acknowledged that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures (e.g., automated red blood cell counts) than to subjective outcome measures (e.g. a behavioural outcome rated by a researcher).</p> <p>For some outcomes, particularly histopathology assessment, outcome assessors are not blind to study group as they require comparison to the control to appropriately judge the outcome, but additional measures such as multiple levels of independent review by trained pathologists can minimise this potential bias.</p>
<b>Selective reporting</b>	Were all measured outcomes reported?	It is acknowledged that this element is difficult to assess with confidence for most studies unless the study protocol is available. In the majority of the cases this element can only be assessed by comparing the 'Materials and methods' and the 'Results' sections of the publications.
<b>Other threats to internal validity?</b>	<p>Were statistical methods appropriate?</p> <p>Was the research question clear?</p>	One of the common statistical issues identified has been reporting of statistical tests that require normally distributed data (e.g., <i>t</i> -test or ANOVA) without reporting that the homogeneity of variance was tested or confirmed.

The rating was based on experts' judgement, with a narrative description of the elements considered for the overall rating. This information was managed and recorded using an online tool.

After piloting the criteria above with a set of publications reporting on toxicokinetic parameters, alone or in addition to other toxicity endpoints, it was noted that the main element for the appraisal (confidence in the analytical method used to determine TiO<sub>2</sub> or Ti in blood and/or tissues) of studies to be considered for the evaluation of ADME was already covered by the specific element described as internal exposure (See Section 3.1.3 Examination of internal exposure) and therefore this step of the reliability appraisal was omitted for those publications exclusively reporting information on the toxicokinetics of TiO<sub>2</sub> (absorption, distribution in tissues and organs, accumulation, excretion).

Only studies, considered sufficiently reliable with respect to their internal validity (Reliability 1 and 2), were included in the assessment.

### 3.1.2. Nanoscale considerations related to specific aspects of study design

The appraisal of this specific element was done for toxicological studies considered sufficiently reliable with respect to their internal validity (Reliability 1 and 2). The nanoscale considerations were scored in a range from 1 (high) to 4 (low) based on information available in the publications/study reports applying the criteria described in Appendix E.

### 3.1.3. Examination of internal exposure

The Panel noted that the systemic availability of TiO<sub>2</sub> following oral administration to animals is low. In the re-evaluation of the food additive (E 171), an oral systemic availability within the range 0.02%–0.1% was estimated (EFSA ANS Panel, 2016). Furthermore, the gastrointestinal uptake is expected to depend on the degree of agglomeration of the particles. As toxicity may depend on the amount of TiO<sub>2</sub> that becomes systemically available, internal exposure was examined. In interpreting the results from toxicity studies, information on the levels of Ti in blood and tissues and/or presence of particles were considered as supportive in appraising the reliability of all reported systemic findings.

For all the studies retrieved from the literature that were considered either relevant for the assessment of toxicokinetics or sufficiently reliable with respect to the toxicity endpoint investigated, i.e. rated by the experts as '1, Yes, only minor or no limitations' (reliability 1) or '2. Overall reliable, with some limitations' (reliability 2) in accordance with the criteria described in Section 1.2.1, proof of internal exposure in the studies was examined and reported as:

- Yes, internal exposure was examined
  - Quantitative measurement in blood and/or tissues (total Ti, i.e. Ti as element, or TiO<sub>2</sub>)
  - Presence of TiO<sub>2</sub>-particles in blood or tissues detected by, e.g. EM
- No, internal exposure was not examined
- Not relevant (e.g. in the case of studies on GIT microbiota).

For those studies in which internal exposure was quantitatively examined, according to the information reported in the publication, a further appraisal of the reliability of the analytical methods applied was undertaken. For studies based on qualitative detection of TiO<sub>2</sub> particles by EM, the use of a methodology (e.g. EDX) enabling the elemental composition to be checked, with confirmation of the chemical nature of the particles as titanium dioxide, was considered essential when assessing the reliability. Accordingly, the following rating was given to the methodology used to examine the internal exposure:

1. Highly reliable
2. Reliable with some limitations
3. There are important flaws to the reliability of the results
4. Not reliable

The following guiding principles were used by the experts in assigning the reliability of the methodology:

Score	Analytical equipment	Sample treatment	Analytical quality control	LODs/LoQs	Treated vs controls
<b>1</b>	Instrument described and (i) able to overcome interferences (e.g. HR-ICP-MS or TripleQuad ICP-MS) or (ii) awareness of interferences and means to deal with them demonstrated	Mentioned and sound	Use of CRMs, RMs, quality control samples (or at least analytical recovery assessment, e.g. by spikes before sample treatment) OR participation to proficiency tests or other external QC schemes	Credible (based on the instrument used) and enabling quantification in at least part of the samples	Full quantitative data for treated vs controls for all the tissues/organs analysed
<b>2</b>	Instrument described	Mentioned and sound	Either reported (as described in #1) or not reported	Credible (based on the instrument used) and enabling quantification in at least part of the samples	Full quantitative data for treated vs controls for all the tissues/organs analysed
<b>3</b>	Instrument described	Mentioned and sound	Either reported (as described in #1) or not reported	Credible (based on the instrument used) and enabling quantification in at least part of the samples	Full quantitative data for treated vs controls for all the tissues/organs analysed not available

			Not enabling quantification in any sample	Quantitative data not provided
4	Instrument described	Not mentioned or unsound	Either reported (as described in #1) or not reported	As for #3
	Instrument not described or wrongly described	Mentioned and sound	Either not credible or as for #3	

This assessment was performed by one Expert member of the FAF WG Titanium dioxide (E 171); the opinion of a second Expert was sought in specific cases, e.g. related to higher complexity/uncertainty.