Study on the Role of Robust Study Summaries in Hazard Assessment

Analysis of the accuracy of Robust Study Summaries

Work package 2 final report delivered under contract ECHA/2021/46[1]

21 October 2022, edited by ECHA for publication in March 2023

[1] https://echa.europa.eu/about-us/procurement/closed-calls

Study prepared by Yordas Group for the European Chemicals Agency
www.yordasgroup.com

Executive Summary

According to the EU REACH Regulation[2], article 3(28), 'a Robust Study Summary (RSS) is a detailed summary of the objectives, methods, results, and conclusions of a full study report, providing sufficient information to make an independent assessment of the study, minimising the need to consult the full study report'.

The OECD has developed Harmonised Templates[3] (OHT) to report the relevant information in the context of the risk assessment of chemicals, including Robust Study Summaries for several regulatory endpoints. IUCLID 6[4], a software tool developed by the European Chemicals Agency[5] (ECHA), in collaboration with the OECD, serves as the reference implementation for the OHTs, and provides data entry screens for users to provide the relevant information in an agreed format, within a regulatory context

As part of an OECD project, ECHA commissioned a study to Yordas Group, referred to as "the contractor" in this report, to evaluate the confidence in the RSS approach for hazard assessment and to identify potential improvements. Stakeholder engagement activities were conducted, including a survey and semi-structured interviews, to capture the comments and suggestions of RSS users in the first part of this project. The findings were published on the ECHA website in April 2022[6] in the report "*Study on the role of robust study summaries in hazard assessment"*.

The work of this study has been divided into three work packages (WP). This report is part of WP2 which has analysed the quality and the accuracy of a series of RSS. The results from WP1 and WP2 that will be used in the last WP of this project to identify areas of improvement that will increase the usefulness and trust in the RSS concept.

The objectives of WP2 were to evaluate:

1) The quality of the RSSs provided by companies (mainly EU REACH registrants)
2) Whether each company RSS can accurately summarise studies so that hazard can be properly assessed without accessing the full study report

A Standard Operating Procedure (SOP) was written to guide authors to write ideal RSS from the full study reports provided by ECHA for this study. The SOP-guided RSSs were qualitatively and quantitatively compared with their corresponding registrant's RSS along with the full study report and the findings of the comparison exercise were recorded in a spreadsheet-based database to allow statistical analysis.

[2] https://eur-lex.europa.eu/legal-content/en/TXT/HTML/?uri=CELEX:02006R1907-20220301
[3] https://www.oecd.org/ehs/templates/introduction.htm
[4] https://iuclid6.echa.europa.eu/project-iuclid-6
[5] https://echa.europa.eu
[6] https://echa.europa.eu/technical-scientific-reports

The data analysis assessed the following three parameters:

(a) RSS Score (assigned to each RSS based on their accuracy and completeness)
(b) Access to the full study report was needed
(c) Effect of deficiencies on the interpretation of the results/hazard conclusion

The SOP-guided RSSs showed higher accuracy and completeness and a lower number of deficiencies leading to a more accurate interpretation of the results and the hazard conclusions and reducing the number of instances when access to the full study report could be needed. Within registrant's RSSs the accuracy and completeness of the human health toxicology group of endpoints were significantly higher than ecotoxicology, environmental fate, and physicochemical endpoints. Particularly, among registrant's RSSs, a higher number of deficiencies that could have led to a different interpretation of results/hazard or required access to the full study report were identified for ecotoxicology and environmental fate. However, there were no statistically significant differences in the three parameters across different endpoints within a group of endpoints.

Additional analysis was conducted on registrant's RSSs to determine the effect of selected factors: full study report type (lab report and publication), Klimisch[7] score of the report (Klimisch score 1 and 2), and substance type (inorganic, organic, UVCB), on all three parameters of the RSSs. In this additional analysis, no statistically significant differences were observed between RSSs written from different full study report types, Klimisch score or substance type. Due to the lack of effect of these factors on RSSs, potential reasons for low scores across the three parameters (RSS score, access to the full report, and effect of deficiency on the interpretation of the results and on the hazard conclusion) for the registrant's RSSs could be one or several factors among author experience, poor use of existing guidance, deficiencies in existing guidance, insufficient review process by both authors and regulators or templates with a confusing structure. These factors will be examined in greater detail in Work Package 3 of this project.

[7] Klimisch, H.J.; Andreae, M.; Tillmann, U. (1997). "*A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data*"

Table of Contents

Glossary of Key Terms

Term	Definition
Author(s) / RSS author(s)	The terms author(s) and RSS author(s) in this report refers to the person(s) responsible for preparing the RSS. In this report, when this term is used in context to RSS authored from SOP (see below), it refers to person(s) from the contractor side who prepared a new RSS for this project from the full study report provided by ECHA; whereas, when this term is used in context of registrant's RSSs, it refers to person(s) from industry who prepared the RSS for REACH registration dossiers submitted to ECHA.
Full study report	A complete and comprehensive description of the activity performed to generate the information. This covers the complete scientific paper as published in the literature describing the study performed or the full report prepared by the test house describing the study performed Source: REACH Regulation: Retrieved: https://eur-lex.europa.eu/legal-content/en/TXT/HTML/?uri=CELEX:02006R1907-20210215
Lab report	A full study report prepared by a test house.
Peer reviewed scientific publication	A study report published in a scientific journal after peer review.
Robust Study Summary (RSS)	A detailed summary of the objectives, methods, results, and conclusions of a full study report providing sufficient information to make an independent assessment of the study minimising the need to consult the full study report Source: How to report robust study summaries, by ECHA. Retrieved at https://echa.europa.eu/practical-guides
Standard Operating Procedure (SOP)	SOPs are detailed written instructions to achieve uniformity of the performance of a specific process, which is repetitive and can be standardised.
Completeness Check	A check performed by ECHA on incoming REACH registration dossiers to ensure information as per Article 20 of the REACH Regulation has been provided. It also includes a manual verification by ECHA staff.

Glossary of Acronyms

Acronym	Terms
BPR	Biocidal Products Regulation
CRED	Criteria for Reporting and Evaluating Ecotoxicity Data
ECHA	European Chemicals Agency
IUCLID	International Uniform Chemical Information Database
KL	Klimisch Score
OECD	Organisation for Economic Co-operation and Development
OHTs	OECD Harmonised Templates
REACH	Registration, Evaluation, Authorisation, and restrictions of Chemicals
Registrant's RSS	RSS included in a dossier submitted to ECHA for REACH or BPR purposes.
RSS	Robust Study Summary
SciRap	Science in Risk Assessment and Policy
SOP	Standard Operating Procedure
SOP-guided RSS	RSS created for this project by the contractor using the full study reports included in the registrant's RSS.
UVCB	Unknown or variable composition, complex reaction products or of biological materials
WoE	Weight of Evidence
WP	Work Package

1. Introduction

REACH[8], which stands for *Registration, Evaluation, Authorisation and Restriction of Chemicals*, is a regulation of the European Union that invites companies to demonstrate to the *European Chemicals Agency* (ECHA) how the substance they manufacture, or import can be safely used along with the risk management measures to the users. *IUCLID* is a tool that can be used to prepare the registration dossiers that must be submitted to ECHA. Moreover, and following the *Biocidal Products Regulation* (BPR)[9], all biocidal products require an authorisation before they can be placed on the market, and the active substances contained in that biocidal product must be previously approved. IUCLID is also used for preparing these applications.

For each registration submitted to ECHA under REACH, and to ensure that the dossiers include all the information that is required, a completeness check[10] of the provided data is performed. Completeness checks are performed both on new registrations and updates of existing registrations. The completeness check can only be successful if all the information in the dossier is complete. The dossiers used for this study have passed this check and allow a reliable assessment of the quality of the data entered in the RSS. Note that the information submitted in the registration dossier may or may not be compliant with the legal requirements. This compliance is part of an additional process called compliance check[11] which evaluates the substance identity and the safety information in the dossier. Therefore, the samples used for the purpose of this project may or may not be compliant, but the scope of this project is to fully understand how RSS are used and to suggest improvements in areas that will increase the trust and reliability of the RSS concept.

The *Robust Study Summary* (RSS) is intended to summarise, in a standardised format, key details from a lengthier full study report. When the data and results are entered in this format it allows assessors to review study outcomes and relevant remarks on the quality of the data much more efficiently in comparison to reviewing the full study report. However, RSSs are only designed to capture a limited amount of data fields across a range of endpoints potentially leading to unclear and incomplete reporting that may confound the interpretation of the results.

To understand how RSSs are currently used by hazard assessors and what factors influence the assessor's confidence in the quality of RSSs, the contractor evaluated the confidence in the RSS approach in hazard assessment and identified potential improvements to the process. The work has been divided into three work packages (Table 1) allowing ECHA to review the outcomes from one before progressing further.

[8] REACH: https://echa.europa.eu/regulations/reach/understanding-reach
[9] BPR: https://echa.europa.eu/regulations/biocidal-products-regulation/understanding-bpr
[10] Technical completeness check: https://echa.europa.eu/technical-completeness-check
[11] Compliance check: https://echa.europa.eu/regulations/reach/evaluation/compliance-checks/5-percent-compliance-checks-2010-registration-dossiers/what-is-compliance-check

Table 1. Work packages (WP) of the study

Work Package	Description
WP1	Examine the role of robust study summaries in hazard assessment
WP2	Analysis of the accuracy of robust study summaries;
WP3	Improving the usefulness of and trust in robust study summaries

Before the work detailed in this report, a stakeholder engagement analysis and a literature review were undertaken in WP1 to understand stakeholders' views on the role of RSSs in hazard assessment. Stakeholders' comments and suggestions were collected and analysed to understand how RSSs are currently used by hazard assessors and the factors that influence their confidence in RSSs. The WP1 findings, which were published on the ECHA website[12], will be used alongside the WP2 results to suggest improvements to increase the usefulness and trustworthiness of RSSs. This will be done as part of WP3.

The overall objective of WP2 is to assess the quality of the RSS included in the registrants' dossiers submitted to ECHA under REACH and BPR. This was done by performing a qualitative and quantitative comparison of pairs of registrant's RSS and RSS prepared by the contractor following a predefined standard operating procedure (henceforth referred to as 'SOP-guided RSS'), both based on the same full study report across a range of regulatory endpoints as already introduced in this chapter. It was determined early in the project that a SOP and robust review process were crucial for the new RSSs that were to be written for this work package; therefore, this report also has dedicated chapters for describing the development and review process of the SOP and the implementation of an internal review of the RSSs. The report also provides details on the comparison among SOP-guided RSSs, registrant's RSSs, and full study reports in addition to findings of the data analysis of the RSS quality.

At the beginning of WP2, ECHA, the appointed OECD Steering Committee which supervised this project, and the contractor agreed on the selection criteria to be used to identify, from the ECHA IUCLID Database, the dossiers containing relevant RSS and the corresponding full study reports. These dossiers had been submitted by registrants and applicants following their obligations under the REACH and BPR regulations. For the relevant endpoints, ECHA extracted the full studies and gave them to the contractor so that they could generate their own RSS. Once the contractor-RSSs were created, ECHA provided the original RSS that the registrants and applicants had included as part of their dossiers so that the contractor could perform a critical comparison and assess their quality.

[12] https://echa.europa.eu/technical-scientific-reports

2. Development of RSS Authoring Standard Operating Procedure and its Testing in a Pilot Study

2.1 Introduction

It was determined in the planning of the project that the use of a standard operating procedure (SOP) for authoring RSSs and a rigorous review mechanism was of critical importance to ensure that the new RSSs being written for this work package were of the highest quality and that this quality could be maintained across multiple authors. Therefore, before commencing the RSS authoring, the contractor established an internal SOP that gave detailed instructions on how to write an RSS.

2.2 Methodology

2.2.1 SOP Development

The SOP was based on ECHA's Practical Guide 3[13] '*How to report robust study summaries'*, the OECD Harmonised Templates (OHTs) along with the experience gained by the contractor during the preparation of REACH registrations. In addition, the technical lead for creating the SOP also reviewed the final '*Stakeholder Engagement Report'* created in WP1. The Stakeholder Engagement Report identified, through surveys and interviews, some of the strengths and weaknesses in the RSS structure for all endpoints. Several questions in the survey focused on the factors that will increase confidence in the concept of RSS and these were taken into consideration when writing the SOP. Some examples of survey findings and how these were addressed in the SOP are:

- '*Any other information on results incl. Tables' should include tables with quantitative results and relevant raw data*.
  The SOP instructed authors to use this field extensively to provide as much data as possible
- '*It can be difficult to tell if information has been left out or misinterpreted*'.
  The SOP instructed authors to always write 'not specified' if any information was not available.
- '*Sometimes there are translation errors and misunderstandings of the text*'.
  The SOP instructed authors to always review copy and pasted content carefully after it was included in the RSS
- '*A lack of information on the quality and reliability of full study report data, particularly if it is not compliant with test guidelines*.
  The SOP instructed authors to use the remarks section below '*Rationale for reliability incl. Deficiencies'* to provide as much detail as possible on why authors think the chosen reliability score is appropriate.

The literature search in WP1 identified opportunities for a more comprehensive quality assessment strategy of both full studies and RSSs by using approaches beyond currently used Klimisch criteria. CRED (Criteria for Reporting and Evaluating Ecotoxicity Data) and

[13] *Practical Guide 3: How to report robust study summaries* is available at: https://echa.europa.eu/practical-guides

SciRap (Science in Risk Assessment and Policy)[14], were identified as examples of tools that could measure the reliability and the relevance of full study reports. The SOP instructed authors to cross-check the completed RSSs against these criteria to make sure that the critical information required for hazard assessment is included.

A draft SOP was tested in a pilot study (chapter 2.2.2) after it had been reviewed internally at the contractor side. During the internal review of the SOP, comments were gathered from internal RSS authors as well as other internal consultants knowledgeable in this area. In addition, the SOP was also reviewed externally by experts at ECHA and from the OECD Steering Committee.

2.2.2 Pilot Study

The RSS authors and reviewers familiarised themselves with the SOP before their participation in the pilot session. In the pilot session, four different full study reports from the contractor's internal resources were assigned to six authors of varied scientific backgrounds: human health toxicology, chemistry, and ecotoxicology. The purpose of including authors of varied backgrounds was to ensure that the SOP was equally accessible to the authors with and without domain expertise. The full study reports chosen for the pilot included ecotoxicology, environmental fate, and human health toxicology endpoints, and included both lab reports as well as one peer reviewed scientific publication to ensure the SOP could be applied across the range of full study reports used. The four full study reports that were included in the pilot session are described in Table 2.

Table 2. Full study reports used for creating RSSs in the pilot study

S. No	Endpoint	Full study report type	Number of RSSs (1 RSS per author)
1	Invertebrate reproduction (daphnia)	Lab report	6
2	Invertebrate reproduction (daphnia)	Peer reviewed scientific publication	6
3	Ready biodegradability	Lab report	6
4	13-week oral toxicity study	Lab report	6

The authors were instructed to disregard their prior writing practices and to solely rely on the SOP to draft the RSS. The pilot session lasted for two weeks and during this time, two Q&A sessions were organised: one at the beginning and one during the middle of the pilot session. The purpose of the Q&A sessions was to answer all queries related to the RSS and SOP. Twenty-four RSSs were written in total. These RSSs were reviewed by four experienced RSS authors. Each reviewer examined the six RSSs related to a single full study and used the SOP to support their reviews. All general and specific comments and feedback were consolidated, and the SOP was revised and approved accordingly.

[14] Science in Risk Assessment and Policy. Accessed on Sep 3, 2022. Available at: http://www.scirap.org/Page/Index/9ced3317-ab2b-4617-86f4-f2d3b86a419f/reporting-checklist

2.3 Results and Discussion

The SOP consists of four chapters and three appendices. The four chapters are:

A) Purpose and Scope
B) Description of the Contents of the SOP
C) General steps for authoring and reviewing RSS and
D) General aspects of the RSS (common for all endpoints)

Chapter C establishes the workflow of WP2 and chapter D describes the specific information that is common for all endpoints for the IUCLID sections on administrative data, data source, material and methods, and results.

The Appendices provide more granular support on specific aspects of RSS authoring.

Appendix 1: Endpoint specific information for RSS authoring
Appendix 2: SciRap quality assessment
Appendix 3: CRED criteria

2.3.1 External SOP Review

The draft SOP was sent to ECHA and the OECD Steering Committee for an external review. There were no major comments in chapters A, B, and C so no major changes in the general steps of authoring and reviewing were required. All suggestions for minor changes were incorporated in chapters A, B, and C. Numerous comments were received for chapter D (General aspects of the RSS (common for all endpoints)) and Appendix 1 (Endpoint specific information). Comments could be grouped into the following categories:

1. Order in which the information was arranged in the SOP:
   To make it easier for the authors to follow, it was suggested that the information presented in the SOP should be in the same order as it appears in IUCLID.

2. Relocating the information from one section to another:
   Some information should be described in a different section of the RSS. For example, ‘A summary on how any effects observed in the study are relevant for classification and labelling’ was initially specified by the SOP to be written in the results and discussion section which was suggested by the OECD to be moved to the ‘applicant’s summary and conclusion’ section.

3. Format of the executive summary:
   It was suggested that authors could also be directed to the OHT website where predefined Executive Summaries and Tables are available for most of the selected endpoints.

4. Technical information on the endpoints:
   Most of the comments provided were related to technical specifications for the endpoints; these comments either requested changes or required clarification on the existing information. These included suggestions related to changes in units of measurement in some endpoints, clarification in the language, and the addition of more information to improve the quality of the RSS for hazard assessment purposes.

The contractor reviewed all the comments and made all the necessary amendments to the SOP. A tracked changes version and a clean version of the revised SOP were shared with ECHA.

2.3.2 Pilot Study Findings

Each reviewer made several comments after assessing the six RSSs related to a single full study. During the assessment reviewers used the SOP to support their review. The general comments provided by the reviewers during the pilot study are provided in Table 3.

Table 3. Major comments received from the reviewers during the pilot study

Endpoints	Major comments	Root cause	Corrective action
Ecotoxicology (lab report)	1) Two of the six RSSs mentioned that the study had deviations, but the authors had not recorded those deviations.	1) No shortcoming in SOP	1) No change in SOP required. Authors were informed of the shortcoming and the RSS was corrected
	2) In two RSSs, the wrong water media type was reported.	2) No shortcoming in SOP	2) No change in SOP required. Authors were informed of the shortcoming and the RSS was corrected.
	3) One RSS wrongly identified the study as a limit test.	3) No shortcoming in SOP	3) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
	4) Same information was repeated multiple times in separate sections.	4) No shortcoming in SOP	4) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
Biodegradation (lab report)	1) One RSS did not report the correct degradation parameter.	1) No shortcoming in SOP	1) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
	2) Several RSSs reported information in the incorrect section of the RSS.	2) No shortcoming in SOP	2) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
	3) Study report did not have test material details; which was properly highlighted in all six RSSs; however, this led to some differences of opinion on the reliability score for the study report.	3) SOP required an additional citation	3) Authors were instructed to review the original Klimisch paper. Paper cited in SOP.

Endpoints	Major comments	Root cause	Corrective action
Human health toxicology (lab report)	1) Some RSSs showed insufficient clarity in the difference between adverse and non-adverse effects.	1) SOP required more clarity	1) SOP was modified to explicitly address this information.
	2) Some RSSs had insufficient clarity on the reversibility of observed effects.	2) SOP required more clarity	2) SOP was modified to explicitly address this.
	3) Indication and discussion of observed effect severity was not very clear in some RSSs.	3) SOP required more clarity	3) SOP was modified to explicitly address this.
	4) Insufficient clarity on where detailed histopathology data should go	4) No shortcoming in SOP	4) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
	5) Lack of full dose-response tables in several RSSs.	5) SOP required more clarity	5) SOP was modified to explicitly address this.
	6) In general, use of 'insert template' functionality is very useful and leads to harmonisation in the reporting.	6) SOP required more clarity	6) SOP was modified to address this
	7) When pasting blocks of text from the full report, be very careful not to bring over a reference to something that is not in the RSS.	7) SOP required more clarity	7) SOP was modified to explicitly address this.
	8) Insufficient test material information in some RSS and there was nothing mentioned in the RSS regarding the homogeneity data (should be written 'none' if not provided in the report instead of leaving it blank).	8) SOP required more clarity	8) SOP was modified to explicitly address this.

Endpoints	Major comments	Root cause	Corrective action
Ecotoxicology (peer reviewed scientific publication)	1) Variation among the six RSS, in the way information was presented such as misplaced information.	1) No shortcoming in SOP (this was probably due to authors bringing in their own prior experience in RSS instead of relying on SOP in some sections)	1) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
	2) Publication included many test substances but only one substance was relevant for RSS. This led to confusion for some authors.	2) No shortcoming in SOP	2) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.
	3) Formatting errors while copying/pasting text from the PDF into IUCLID	3) SOP required more clarity	3) SOP was modified to address it.
	4) Some RSSs assigned different Klimisch scores. This was due to the subjective element of the Klimisch assessment which led to a different interpretation by some authors.	4) SOP required more clarity	4) Authors were instructed to review the Klimisch paper. The paper is also cited in SOP.
	5) 'Sampling and Analysis' section was missing in one RSS and 'Test solutions' information was missing from another RSS.	5) No shortcoming in SOP	5) No change in SOP. Authors were informed of the shortcoming and the RSS was corrected.

All reviewers agreed that, in general, the RSSs compiled in the pilot study had sufficient information to understand the test methodology, results, and conclusion of the studies. RSS authors and reviewers noted a variation between the order in which information was presented in the SOP and how it was structured in IUCLID. They highlighted that replicating the order in which information is presented in the SOP would improve its effectiveness. Some comments, especially the ones provided for the human health toxicology RSS, required some changes to be made in the SOP. These required changes were made in the SOP and the updated version has been shared with ECHA.

Two team meetings were organised during the pilot study, where all major comments listed in Table 3 were discussed and an approach to addressing them was agreed upon. In addition to the major errors reported in Table 3, a few other minor errors were identified in some RSS but similar to most of the errors reported in Table 3, these were user errors and hence did not require any changes in the SOP, rather a recommendation was made that the author better understand the SOP before writing the RSS.

3. Extraction of full study reports from ECHA's database and review

3.1 Introduction

The tender specifications for this project[15] specified that a broad range of endpoints should be included in this research. The call also specified the minimum number of full study reports that are required for each critical endpoint. Using this information as a starting point, selection criteria for extracting full study reports and RSS pairs were designed by the contractor resulting in ECHA providing 103 full studies for the analysis.

3.2 Methodology

As part of the compliance check that is performed on the registration dossiers, ECHA focuses on the so-called "key endpoints"[16] that are used for the identification of substances of concern, so they were included in this analysis. These key endpoints include five key human health toxicology endpoints (genotoxicity, repeated-dose toxicity, prenatal developmental, reproduction toxicity, carcinogenicity) and a few ecotoxicology and environmental fate endpoints (long-term aquatic toxicity, biodegradation, and bioaccumulation). Overall, the tender required a minimum of four human health toxicology, four ecotoxicology, four physicochemical, and three environmental fate endpoints. All endpoints selected for this project are listed in Table 4.

In this report, the terms 'Long-term aquatic toxicity' or 'chronic aquatic' includes RSS for long-term fish toxicity and long-term invertebrate toxicity and the terms 'Short-term aquatic toxicity', or 'acute aquatic' includes RSS for short-term fish toxicity, short-term invertebrate toxicity, and algal toxicity. Similarly, the endpoints toxicity of soil microorganisms, toxicity of soil macro-organisms except for arthropods, and Toxicity to terrestrial plants are collectively referred to as 'soil toxicity' in this report. This grouping of endpoints was performed because the sample size of individual endpoints was too small to perform statistical analysis.

Table 4. Endpoints for which full study reports were requested and RSSs were written

Study No	Endpoint Group	Endpoints
1	Physicochemical	Vapour pressure
		Partition coefficient
		Water solubility
		Flammability (solid)

[15] Tender specs for ECHA/2021/46 at https://echa.europa.eu/about-us/procurement/closed-calls
[16] Compliance checks: https://echa.europa.eu/regulations/reach/evaluation/compliance-checks

Study No	Endpoint Group	Endpoints
2	Environmental fate	Hydrolysis
		Bioaccumulation: aquatic / sediment studies
		Biodegradation in water: screening
		Biodegradation in water: simulation
3	Ecotoxicology	Long-term aquatic toxicity (three trophic levels)
		Short-term aquatic toxicity (three trophic levels)
		Toxicity to aquatic microorganisms (sludge respiration)
		Toxicity of soil microorganisms and macroorganisms except for arthropods
		Toxicity to terrestrial plants (added later by OECD steering committee)
4	Toxicology	Genetic toxicology in vitro and in vivo
		Repeated dose toxicity oral and inhalation
		Developmental toxicity
		Toxicity to reproduction
		Carcinogenicity
		Skin sensitization in vivo and in vitro
		Skin irritation in vitro

Within each endpoint listed in Table 4, the main criteria to select the RSSs and full study report pairs was the Klimisch (KL) score. For this study, only KL1 and KL2 lab reports / peer reviewed scientific publications were selected because they are usually reported as ‘Key studies’ in the registration dossiers and are used to draw conclusions on the hazard classification and for possible use in risk assessment. Any endpoint in a registration dossier that only uses KL3 and KL4 studies needs to use a weight-of-evidence (WoE) approach and WoE is not within the remit of this project. As discussed with ECHA, supporting studies were not considered for the purpose of this study.

Three KL1 and three KL2 full study reports were requested per endpoint to provide equal weighting to each type of report. Within each endpoint, different types of full studies were examined with at least two peer-reviewed scientific publications and four lab reports included out of a total of six full study reports. Various other parameters depending on the endpoint were added in the selection criteria of full study reports such as organism/trophic level/route/type, substance type (e.g., mono-constituent, UVCB), and value range/hazard. The purpose of these additional criteria is to create diversity in the type of full study reports and avoid any bias in the analysis of selected RSS and full study report pairs.

The authors of the SOP-guided RSSs performed a preliminary assessment of the 103 full study reports provided by ECHA to ensure their suitability for drafting the RSS. The Klimisch score of 36 study reports were re-assigned following the assessment of the contractor. This is in line with the feedback received in WP1 (*see Figure 3.25, page 49*) where it was noted that 22% of evaluators dispute the reliability score assigned to a study in the RSS frequently. More details on this can be found in chapter 5.3.3.2 of this report.

The assessment included an evaluation of the content to make sure it is an actual full study report that is suitable to write a complete RSS and the language is English. Any other unusual findings were also noted, for example, whether the full study report included a single or multiple substances or whether the full study report is a full report or just a summary.

3.3 Results and Discussion

Out of 103 studies with a unique combination of criteria proposed by the contractor, 38 combinations did not find the exact match in the ECHA database for the purpose of the study. Therefore, the criteria for these studies needed some adjustments and, in most cases, changing the substance type among ‘UVCB/multiple constituents’, ‘inorganic’, and ‘organic mono constituent' resulted in the availability of a full study report matching the remaining criteria. In few cases, changes in the criteria titled ‘value range/hazard’ or ‘type of full study report (lab report / peer reviewed scientific publication)’ were also needed. Overall, the changes were expected not to cause any effect on the study because the main criteria i.e., KL score and endpoint tier were not altered.

4. RSS Authoring and Reviewing

4.1 Introduction

RSS authoring and reviewing are the principal tasks of WP2. The objective of WP2 is to assess the quality of RSS submitted to regulators by critically comparing the content of pairs of RSSs with the content of the corresponding full study report. The project achieved this by first writing new RSSs from the full study reports provided by ECHA following the new SOP (referred to as 'SOP-guided RSS') and then comparing the content of these SOP-guided RSSs with the content of the associated full study report and the corresponding RSSs provided by ECHA (referred to as 'registrant's RSS'). The new SOP-guided RSSs were subject to quality control review to ensure they met the standards set in the SOP, which in turn had been reviewed by ECHA and the OECD Steering Committee. The registrant's RSSs that were provided by ECHA were originally submitted by registrants for REACH or BPR. Pairs of registrant's RSS and its corresponding full study reports were identified from ECHA's existing database using the criteria outlined in chapter 3.

4.2 Methodology

The team of RSS authors that authored the SOP-guided RSSs consists of consultants in chemistry, toxicology, and ecotoxicology (>1-2 years of experience in drafting RSSs). Authors were assigned endpoints in line with their area of expertise. As with the pilot study, authors and reviewers used the SOP developed earlier in the work package to author their RSS.

5. Analysis and Comparison with registrants' RSSs

5.1 Introduction

As indicated in chapter 3.1, ECHA provided the contractor with the 103 full studies that industry had used to create the dossiers submitted to ECHA. The contractor then used the SOP that they developed in WP2 to write their own SOP-guided RSS for the selected endpoints. Only when ECHA had received these, ECHA then provided the registrants' RSS so that the contractor could perform a critical comparison of these pairs of RSSs. The objective of this comparative analysis was to evaluate: the quality of the RSS, the self-sufficiency of the registrant's RSS (no requirements for access to full study report), and the ability of RSSs to accurately inform on hazard conclusions.

5.2 Methodology

To address each of the objectives defined above, there were three main parameters assigned to each RSS and then statistical analysis was performed among RSSs for each of these parameters:

(a) RSS Score
(b) Access to full study report needed (yes/no); and
(c) Effect of deficiencies on the interpretation of results/hazard conclusion (yes/no)

The three parameters chosen for data analysis represent three important characteristics of the RSSs. The RSS Score represents the overall accuracy and completeness of the RSS, hence a representation of overall RSS quality. One of the key objectives of the RSS is to speed up hazard assessments by making the RSS self-sufficient and reducing reliance on the full study reports. Therefore, the parameter (b) 'Access to full study report needed (yes/no)' was included in this study to estimate the independence of RSS from the full study report. Finally, the parameter (c) 'Effect of deficiencies on the interpretation of results/hazard conclusion (yes/no)' represents the ability of RSS to accurately inform on hazard conclusions i.e., the presentation of an accurate results/hazard conclusion was well supported by the information present within the RSS.

5.2.1. Spreadsheet database for data collection

A spreadsheet-based database was developed to create an effective tool to collate all the relevant information from the RSSs (registrant and SOP-guided RSSs) so they could be compared quantitatively. Once collected, all important data was summarised, and statistical analysis was performed.

RSS score

The nature of any deficiency noted in a particular field of the RSS was described in the respective field and then each deficiency was classified as follows:

a) **Misplaced information**:
Important information present but at an inappropriate location within the RSS.

b) **Misplaced information without any major effect**:
When a misplaced piece of information was relatively easy to identify and was predictable, these instances were labelled as 'Misplaced information without any major effect'.

The distinction between categories (a) 'Misplaced information'; and (b) 'Misplaced information without major effect' was dependent on the context and reviewers were instructed to apply their best judgement. As a rule, if the misplaced information was readily identifiable without investing much time, then it should have been classified as 'Misplaced information without major effect'. For example, a method related piece of information that was presented within the method chapter but in a different field doesn't have a major effect and can be relatively easily located. On the other hand, if method information was presented in the results or summary section or vice versa, then it was treated as 'Misplaced information'.

c) **Missing information**:
Field in RSS left blank.

d) **Partial / Incomplete information**:
Several fields of IUCLID endpoint study record / OHT require large amounts of information. Some information was included but some important aspects required / expected to be reported in a specific field were missing.

e) Wrong Information:
Information present but was erroneous. Examples of wrong information included wrong units, wrong values, wrong species, wrong methodology, wrong identification of adverse effects or no adverse effects.

Some examples of these categories from the comparison exercise are provided in Figure 1.

Sections of RSS	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*
Summary and Conclusions	Conclusions	No conclusions. Something that resembles a conclusion was included in the any other information on results section.	Misplaced information	0.50

Sections of RSS	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*
	Details of test system and experimental condition	Left blank but all information is provided in 'any other detail on method'	Misplaced information without major effect	1.0

Sections of RSS	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*
	Study Period	Not provided	Missing information	0

Sections of RSS	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*
	Conclusions	Conclusion not adequately described. Only listed results with no interperetation	Partial/incomplete information	0.50

Sections of RSS	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*
Study design	Test type, Study type, Substrate type, Limit test, exposure duration, post exposure duration and any other remark	1) Wrong test type. Wrote: "seedling emergence toxicity test". Correct is "seedling emergence and seedling growth test". 2) Limit test listed as no when it is yes.	Wrong information	0.00

Figure 1. Examples of each deficiency category, taken from the comparison spreadsheet database.

When no deficiency was identified in a given field in the RSS, the deficiency description column was left blank and the score was set to ‘1’.

To quantify the importance of each deficiency when assessing the quality of the data entered, the different types of deficiencies were scored as described in Table 5:

Table 5. Deficiency class and the associated score used in RSS comparison

Deficiency Class	Score
No deficiency identified (correct information as per SOP entered in the appropriate field)	1
Wrong information entered in a field	0
Missing information in a field	0
Misplaced information	0.5
Misplaced information without major effect	1
Partial / incomplete information'	0.5 or 0*

* The 'partial / incomplete information' type of deficiency was subject to judgement. On a case-to-case basis, depending on the type and extent of missing information in a field as well as its impact on the overall quality of the RSS, a score of either 0.5 or 0 was assigned.

If there were instances where multiple deficiencies were recorded in a given cell, we took a conservative approach and the worst-case scenario was assumed, where a description of all deficiencies were provided but the worst category was selected for the field in question. The following order was used to decide the worst case: Wrong information > missing information > misplaced information > misplaced information without major effect.

Chapter 5.3.3.5 of this report provides more details about the differences between the terms "missing information" and "wrong information".

In addition to assigning scores to each deficiency, additional columns were also added to the spreadsheet to evaluate each deficiency for the following two parameters:

Access to the full study report is needed? (Yes / No)

Against each deficiency identified, a Yes/No type of response was recorded depending on the judgement of the RSS reviewer regarding whether access to the full study report was needed to resolve the deficiency in question (last column in the Figure 2).

Sections of RSS	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*	*Deficiency due to issues with study report (Yes/No)*	*Access to full report needed (yes/No)*
Sampling and analysis	Analytical monitoring. If yes, details on sampling method and analytical methods	No sampling information entered	Missing information	0.00	NO	Yes
	Nominal and measured concentrations	Listed in the "Any other information" section	Misplaced information without major effect	0.50	No	No

Figure 2. Examples of some deficiencies with response 'yes' or 'no' for the parameter 'Access to full report needed', taken from the comparison spreadsheet database

Any effect of deficiencies on the interpretation of results/hazard conclusion? (Yes / No)

Each deficiency identified in the spreadsheet was evaluated in terms of its effects on the interpretation of the results and on the hazard conclusion. The response was recorded as Yes/No and the rationale for selection was also provided (Figure 3).

Deficiencies in registrant's RSS	*Classification of the deficiency noted*	*Score*	*Deficiency due to issues with study report (Yes/No)*	*Access to full report needed (yes/No)*	*Any effect of deficiencies on the interpretation of results / Hazard conclusion (yes/no)*	*Reasoning for column K*
Only treatment group reported. Numbers in control group is missing (number of animals were different in control)	Wrong information	0.0	No	Yes	No	Although the number in control is different from treatment group, the number still meets the Guideline requirements, hence no effect on hazard conclusion.

Deficiencies in registrant's RSS	*Classification of the deficiency noted*	*Score*	*Deficiency due to issues with study report (Yes/No)*	*Access to full report needed (yes/No)*	*Any effect of deficiencies on the interpretation of results / Hazard conclusion (yes/no)*	*Reasoning for column K*
1) 'Sampling and Analysis' subsection: Details on sampling are missing completely which would have greatly increased the confidence in measured concentrations 2) 'Details on analytical methods': several key information such as centrifugation, solvent extraction, and linerity range are missing	Missing information	0.0	No	Yes	Yes	would have greatly increased the confidence in measured concentrations
A more transparent description of the test solution preparation is possible based on the information available in the report. This include the amount of substance in amount and conentration of solvent solution and more information on the findings of the range finding test which prompted the use of organic solvent.	Missing information	0.0	No	Yes	Yes	critical information on exposure concentrations and used solvent

Figure 3. Examples of some deficiencies with response 'yes' or 'no' for the parameter 'Any effect of deficiencies on the interpretation of results / hazard conclusion', taken from the comparison spreadsheet database

5.2.2. Data analysis

As discussed in chapter 5.1, there were three parameters of relevance assigned to each RSS and then statistical analysis was performed among RSSs for each of these parameters:

(a) RSS score
(b) Access to full study report needed (yes/no) and
(c) Effect of deficiencies on the interpretation of results/hazard conclusion (yes/no)

An example with the endpoint vapour pressure for one of the registrant's RSSs is shown in Figure 4 below demonstrating how the three parameters were calculated.

(a) RSS Score

The scores in all fields were added and a percentage score was calculated using the following formula:

$$RSS\ Score\ =\ (sum\ of\ scores\ from\ each\ field\ in\ RSS\ /\ total\ number\ of\ fields\ in\ RSS)\ *\ 100$$

The RSS score was presented as a percentage instead of an absolute score because different endpoints have different numbers of fields in their templates depending on the complexity of the endpoint; therefore, scores as absolute values would not have led to accurate comparison across different endpoints (Figure 4).

(b) Access to full study report needed (yes/no)

This parameter was presented as absolute numbers. Essentially, the number of instances in each RSS where a deficiency resulted in the selection of 'Yes' were summed and used as a dependent variable for the analysis (Figure 4). Percentage scores for this parameter were deemed to be not suitable because, in some scenarios, the percent score will not reflect the results as intended. For example, in a hypothetical scenario, where there were in total four deficiencies in the RSS, out of which two needed access to the full study report, this would result in 50% deficiencies needing access to full study reports; whereas if there was overall only one deficiency in RSS and if that required access to the full study report, it would result in 100% deficiencies requiring access to the full study report.

(c) Effect of deficiencies on the interpretation of results/hazard conclusion (yes/no)

This parameter too was presented as absolute numbers. Briefly, the number of instances in each RSS where a deficiency resulted in the selection of 'Yes' were summed and used as a dependent variable for the analysis (Figure 4). Percentage scores for this parameter was deemed to be not suitable for similar reasons as discussed in parameter (b).

Figure 4. An example of the calculation of the parameters of interest: RSS score; Access to full study report needed; and Effect of deficiencies on the interpretation of results/hazard conclusion, using the vapour pressure endpoint as an example from registrant's RSS (Note: some rows are hidden in this example figure to accommodate the entire table in one figure).

					Registrant's RSS					
Endpoint	*Study number*	*Study type*	*Sections of RSS*	*Sub-sections*	*Deficiencies in registrant's RSS*	*Classification of the deficiency noted*	*Score*	*Deficiency due to issues with study report (Yes/No)*	*Access to full report needed (yes/No)*	*Any effect of deficiencies on the interpretation of results / Hazard conclusion (yes/no)*
Vapour pressure	74	Publication	Administrative	Endpoint (picklist)			1			
				Adequacy (pick-list)	Should be supporting rather than key	Wrong information	0	No	No	Yes
				Study Period	Only year given. If no month available this should be specified.	Partial/incomplete information	0.5	No	Yes	No
				Reliability (pick-list)	Should be KL3 rather than 2 in Yordas' opinion	Wrong information	0	No	Yes	Yes
				Rationale for reliability incl. Deficiencies	As above	Wrong information	0	No	Yes	No
			Data source	Reference			1			
				Data access and Data protection claims			1			
			Materials and methods	Guideline followed			1			
				Version			1			
				Deviation			1			
				Principles of method if other than guideline	Not relevant		--			
				GLP compliance (pick-list)			1			
			Results and Discussion	Vapour pressure table / transition-decomposition table whichever is relevant			1			
			Summary and Conclusions	Conclusions			1			
				Executive summary	Could have been summarised rather than just stating the result	Partial/incomplete information	0.5	No	No	No
Total score							14	0	3	2
Percentage							77.8	0	60	40

Once the three parameters were calculated for each registrant's RSS as well as for each SOP-guided RSS, statistical analysis was performed among RSS, on the three parameters, at two levels, to gain insight into how groups of endpoints differ from each other as well as how different endpoints within a group differ from each other.

Level 1 analysis: Endpoints were divided into four groups (ecotoxicology, human health toxicology, environmental fate, and physicochemical) and statistical analysis was performed to evaluate differences in RSS quality between these groups of endpoints. Comparisons were made across registrant's RSSs, across SOP-guided RSSs, and between registrant's RSSs and SOP-guided RSSs.

Level 2 analysis: Each group of endpoints was further analysed at a deeper level to understand the differences among the endpoints (e.g., how an acute aquatic toxicity RSS differs from chronic aquatic toxicity). These comparisons were made among registrant's RSSs, among SOP-guided RSSs, and between registrant's RSSs and SOP-guided RSSs.

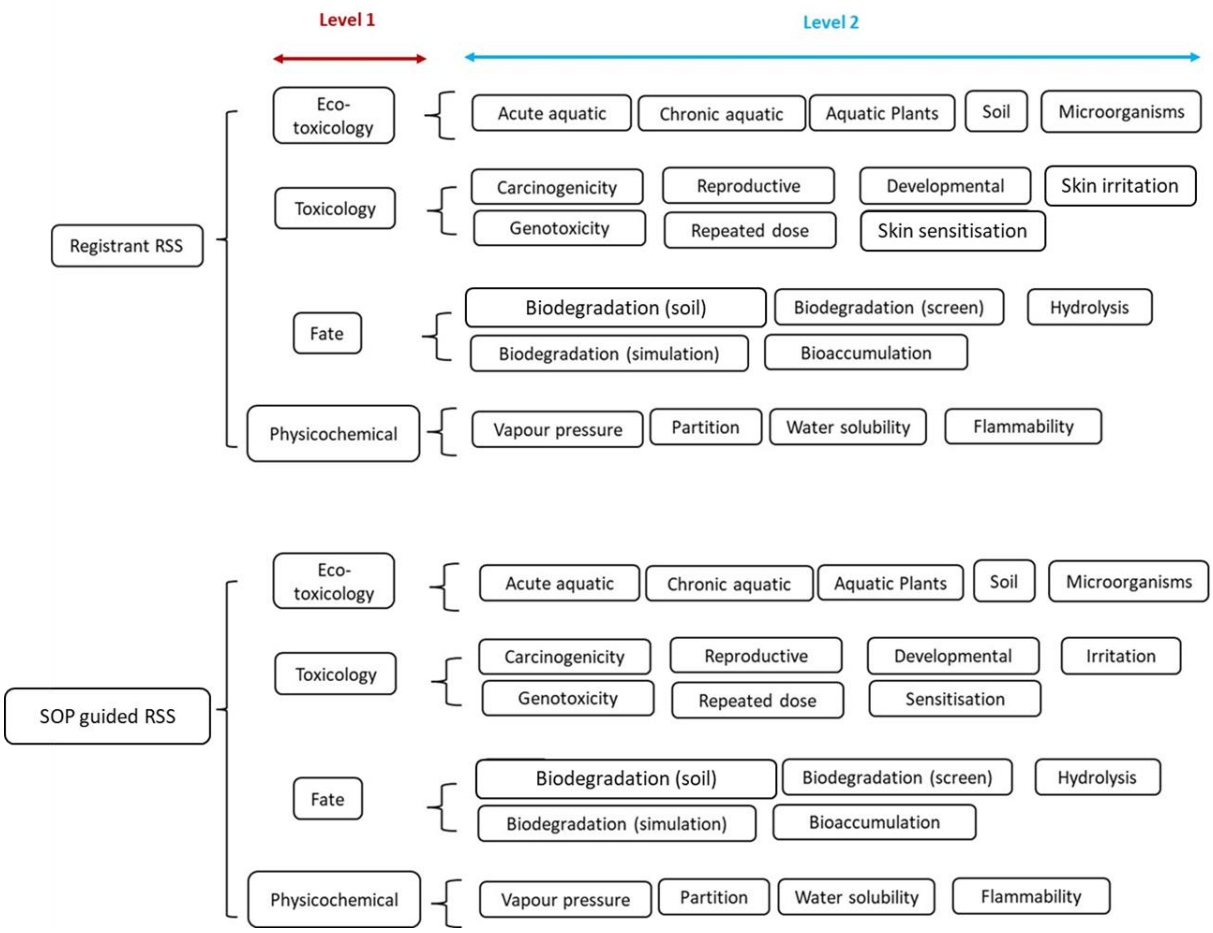


Figure 5. Schematic representation of the analysis performed in WP2

Additional analysis:

In addition to the Level 1 and 2 analysis (Figure 5), four additional quantitative and qualitative analyses were performed for registrant's RSSs:

a) **Report type (lab report versus peer reviewed scientific publications)**
   Firstly, within each endpoint group of the registrant's RSSs, statistical analysis was performed to determine the effect of full study report type (lab report versus peer reviewed scientific publications) on all the three parameters of the RSS. The purpose of this analysis was to evaluate the effect of study report type on the quality of the RSS.

b) **Klimisch score of the full study report**
   Within each endpoint group of the registrant's RSSs, statistical analysis was performed to determine the effect of the Klimisch score of the full study report from which the RSSs were written on all the three parameters of RSS. The purpose of this analysis was to evaluate the effect of full study reports with different Klimisch scores on the quality of the RSS.

c) **Substance type (inorganic, organic mono-constituent, and UVCB / multi-constituent)**
   Within each endpoint group of the registrant's RSSs, statistical analysis was performed to determine the effect of substance type on all the three parameters of RSS. The purpose of this analysis was to evaluate the effect of full study reports of different substance types on the quality of the RSS.

d) **Determining the registrant's RSS with no deficiency**
   Within both Level 1 and Level 2 analysis for registrant's RSSs, the number of RSSs within each endpoint that could be regarded as an ideal RSS was evaluated (i.e., no such deficiency identified that affected the interpretation of results/hazard conclusion). The objective was to gain qualitative insight on which endpoints and endpoints groups have higher numbers of 'ideal' RSSs compared to other endpoints and endpoint groups.

In all analyses, the parameters of interest (RSS score, access to full study report, effect on interpretation of results/hazard conclusion) were analysed by an analysis of variance (ANOVA) using frequentist statistical methods, where the p-values establish the statistical significance of the results. In level 1 analysis, the two independent variables were endpoint groups (ecotoxicology, human health toxicology, environmental fate, and physicochemical) and RSS author (SOP-guided, registrant), while the dependent variables were RSS Score, the number of instances where access to the full study report was needed for interpretation and number of deficiencies per RSS affecting the interpretation of results for the hazard conclusion. In level 2 analysis, the two independent variables were the endpoints and RSS author (SOP-guided, registrant), and the dependent variables were aligned with the above. Whenever a statistically significant result was observed in ANOVA, a Tukey post-hoc test was completed to determine where the differences are found. A p-value of <0.05 was statistically significant while comparing different treatments.

5.3 Results and Discussion

5.3.1. Level 1 analysis: Endpoint groups

Endpoints were grouped into four groups (ecotoxicology, human health toxicology, environmental fate, and physicochemical) and statistical analysis was performed to evaluate the differences among each group of endpoints using the three parameters introduced in Chapter 5.2.

5.3.1.1. RSS Score

The ‘RSS Score’ represents the overall accuracy and completeness of the RSS and hence is a representation of overall RSS quality. The objective of this analysis was twofold: to do a comparative analysis of overall RSS quality across groups of endpoints among registrant’s RSSs and SOP-guided RSSs, and to compare registrant’s RSS and SOP-guided RSSs for all groups of endpoints.

Among the registrant’s RSSs, the mean percentage score of the human health toxicology group of endpoints was significantly higher than ecotoxicology, environmental fate, and physicochemical endpoints, whereas there was no significant difference between ecotoxicology, environmental fate, and physicochemical endpoints. This indicates that registrants filled the human health toxicology endpoints RSS with more accuracy and completeness compared to the other endpoint groups. There was no statistically significant difference among any groups of the SOP-guided RSSs indicating all types of RSSs authored using the SOP were of similar accuracy and completeness (Figure 6).

Within each endpoint group, the SOP-guided RSSs scored higher than the corresponding RSSs submitted by registrants, with the biggest difference noted in ecotoxicology and environmental fate RSSs (mean difference of 28.5 and 22.9 percentage points, respectively). The difference between SOP-guided and registrant’s RSSs for human health toxicology and physicochemical endpoints was also statistically significant although the difference was slightly smaller (mean difference of 11.6 and 18.5 percentage points, respectively) (Figure 6).

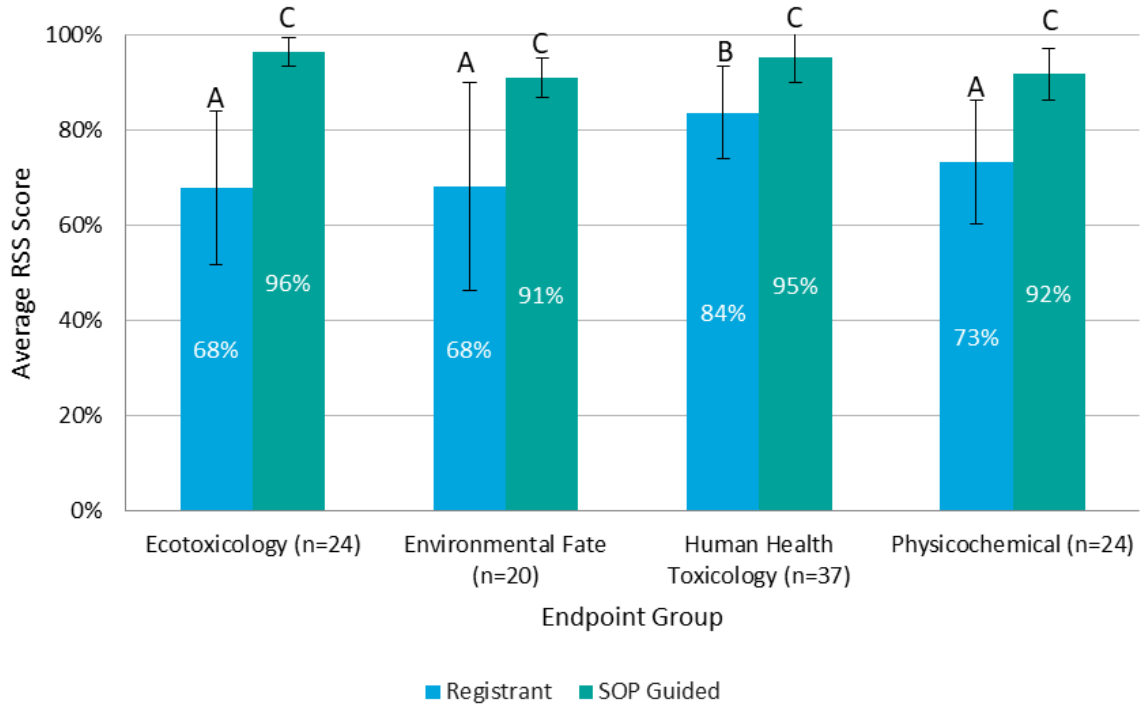


Figure 6. The average RSS score and standard deviation for registrant and SOP-guided RSSs per endpoint group. Higher score in this figure demonstrates higher accuracy and completeness of an RSS. Letters that are different show statistically significant differences ($p < 0.05$). Same letters on two bars demonstrates no statistically significant difference.

Overall, the results indicated that the SOP-guided RSSs were written with greater accuracy and completeness than RSSs authored by registrants. Moreover, the accuracy and completeness of SOP-guided RSSs were similar across all groups of endpoints whereas, for registrant's RSSs, the accuracy and completeness of the human health toxicology group of endpoints was significantly higher than ecotoxicology, environmental fate, and physicochemical endpoints.

The results clearly demonstrate that implementing a SOP and a robust review process increases the accuracy and completeness of RSSs across all groups of endpoints. This reduces variability in the SOP-guided RSSs across all groups and also demonstrate greater consistency in the quality of RSS in all the groups. Moreover, among registrant's RSSs, human health toxicology RSSs scored higher than ecotoxicology, fate, and physicochemical groups of endpoints. The reasons for lower quality/completeness scores are unknown. As indicated in later parts of this report, the 2-way ANOVA analysis of factors such as substance type, Klimisch score of full study reports, and the type of full study report did not reveal any statistically significant difference among endpoints groups, suggesting some other reasons account for the lower scores observed in RSSs for the ecotoxicology, fate, and physicochemical endpoint groups.

5.3.1.2. Access to full study report needed

As discussed in chapter 2.2, one of the key objectives of RSSs is to speed up hazard assessments by reducing the need to return to/review the full study reports. Therefore, the parameter '*Access to full study report needed (yes/no)'* was included in this study to estimate

whether a hazard assessment can be made using RSSs only without reference to a full study report. An RSS with a higher number of instances where access to the full report is needed is more dependent on the availability of the full study report, compared to an RSS with a lesser number or no instances. The objective of this analysis was twofold: to do a comparative analysis of this parameter across groups of endpoints among registrant's RSS and among SOP-guided RSSs, and to do a comparison between registrant's RSS and SOP-guided RSSs for all groups of endpoints.

For the purpose of this project, deficiencies are defined as an unexpected value in a field in IUCLID. Deficiencies can potentially increase the number of times an assessor would be inclined to access the full study report.

Having zero deficiencies can be seen as ideal and when writing this report, the contractor aimed at reaching 'perfect' RSS. This means that amongst all identified deficiencies, minor ones are also included. Some examples of these minor deficiencies are the absence of the month in the study period, the absence of the formal indication of no deviations from the guidelines even when no deviations were indicated in the study, or when the batch number was missing in a test material.

For registrant's RSSs, the mean number of the deficiencies in RSSs that required access to the full study report were 10.7 and 10.6 in ecotoxicology and environmental fate RSSs compared to human health toxicology and physicochemical RSSs (3.7 and 2.5, respectively) (Figure 7). Thus, ecotoxicology and environmental fate RSSs were statistically similar to each other ($p = 1.0$). However, both ecotoxicology and environmental fate RSSs were significantly different from human health toxicology and physicochemical groups of RSSs ($p < 0.001$). There was no statistically significant difference between human health toxicology and the physicochemical group of endpoints ($p = 0.9$) (Figure 7). The SOP-guided RSSs showed no significant differences among any group of endpoints indicating that group of endpoints had no effect on the dependence of RSS on the full study report (Figure 7).

Within each endpoint group, the SOP-guided RSSs for ecotoxicology, human health toxicology and environmental fate required access to full study reports at a significantly lower number of instances than the corresponding RSSs submitted by registrants, with the biggest difference noted in ecotoxicology and environmental fate RSSs (mean difference in the number of times access to full study report needed: 9.4 and 7.5 points, respectively). The difference between SOP-guided and registrant's RSSs for human health toxicology endpoints was also statistically significant although the difference was slightly smaller (mean difference in the number of times access to full study report needed: 3 points). There was no statistically significant difference in the mean number of times access to the full study report needed for physicochemical endpoints RSSs (mean difference 1.3 points, $p = 0.9$) (Figure 7).

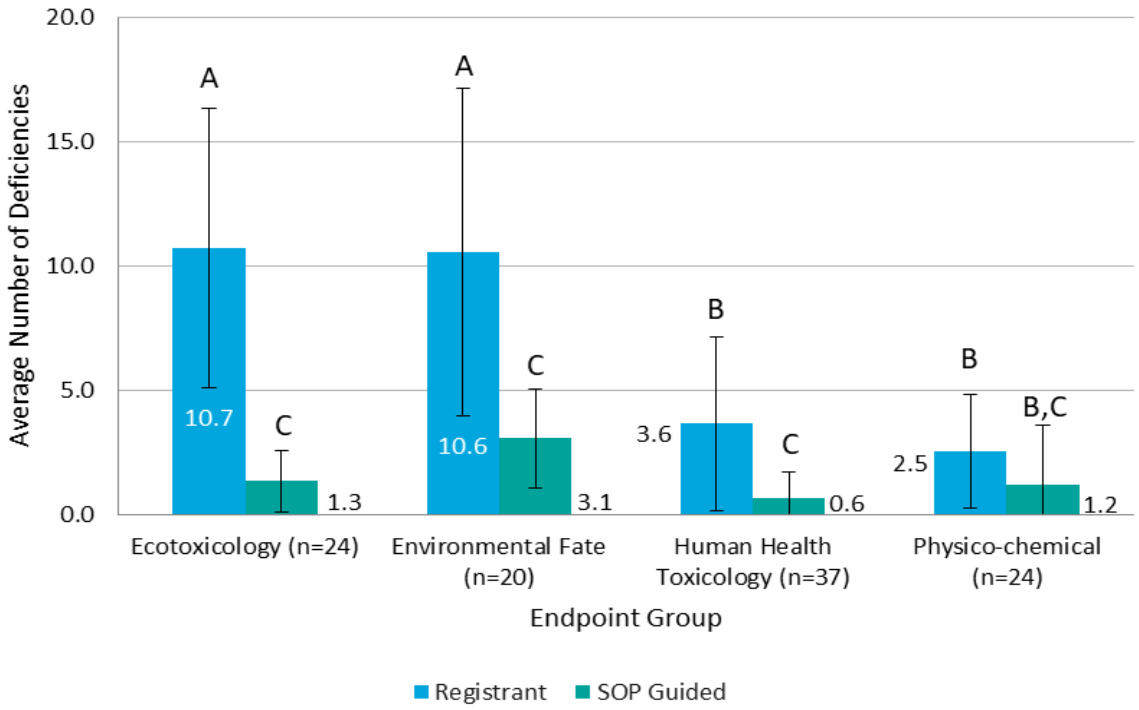


Figure 7. The average number of instances per RSS where access to the full study report was needed and standard deviation for registrant and SOP-guided RSSs for each per endpoint group. Lower number for an RSS in this figure demonstrates reduced dependency on the full study report for that RSS. Letters that are different show statistically significant differences ($p < 0.05$). Same letters on two bars demonstrates no statistically significant difference.

Overall, the results indicate that for ecotoxicology, environmental fate and human health toxicology groups of endpoints, the discrepancies between the full study report and the RSS were lower for SOP-guided RSSs compared to registrant's RSSs, meaning the dependence of the RSSs on full study reports to make a hazard assessment was lower for SOP-guided RSSs compared to registrant's RSSs. There was little difference in the number of times access to the full study report was needed during the review of RSS for either type of RSS for physicochemical endpoints.

Overall, these results demonstrated that implementing a SOP and a robust review process resulted in fewer number of deficiencies across all groups of endpoints. For physicochemical endpoints there was no difference between SOP-guided and registrant's RSS because the registrant's RSSs already had a very low number of deficiencies that required to access to the full study report. SOP-guided RSSs were statistically similar across the endpoint groups, whereas the registrant's RSSs for ecotoxicology and environmental fate RSSs were similar to each other but both two groups of RSSs were significantly different (higher dependence on the full study report) from human health toxicology as well as the physicochemical group of RSSs.

The reasons for higher dependence on full study reports in the registrant ecotoxicology and fate RSSs are unknown. In later parts of this report, the 2-way ANOVA analysis of factors such as substance type, Klimisch score of full study reports, and the type of full study report did not reveal any statistically significant difference among endpoints groups.

5.3.1.3. Effect of deficiencies on the interpretation of results/hazard conclusion

One of the key goals of the RSS is to present an accurate results/hazard conclusion well supported by the information present in it. Therefore, the parameter '*Effect of deficiencies on the interpretation of results/hazard conclusion (yes/no)*' was included in this study to evaluate RSSs in terms of their ability to achieve that goal. The objective of this analysis was twofold: to do a comparative analysis of this parameter across groups of endpoints among registrant's RSS and among SOP-guided RSSs, and to do a comparison between registrant's RSS and SOP-guided RSSs for all groups of endpoints.

RSSs are meant to capture key details from a full study report. For the purpose of this project, the contractor considered that, although an RSS already included all the required key information, additional information could still have been provided by the author (e.g. by including more details in tables). Also, not indicating details about the principles of the study or methods used, not including an executive summary as part of the conclusions are aspects that were considered to, potentially, impact the interpretation of the results.

For registrant's RSSs, the mean number of times errors that could affect the interpretation of results were 5.7 and 6.8 in ecotoxicology and environmental fate RSSs, respectively compared to human health toxicology and physicochemical RSSs (1.6 and 1.4, respectively) and the differences were statistically significant. There was no statistically significant difference between human health toxicology and the physicochemical group of endpoints ($p = 1$) or ecotoxicology and environmental fate ($p = 0.8$). This analysis indicates that among the registrant's RSSs, the average number of errors that could affect the interpretation of results for ecotoxicology and environmental fate were significantly higher than that for human health toxicology and physicochemical RSSs (Figure 8). The SOP-guided RSSs showed no significant differences among any group of endpoints in terms of the average number of errors that affected the interpretation of results, indicating that type of endpoints had no effect on how the deficiencies affected the interpretation of results (Figure 8).

Within each endpoint group, the mean number of deficiencies that could affect the interpretation of results for the SOP-guided RSSs were significantly lower for ecotoxicology and environmental fate than the corresponding RSSs submitted by registrants (difference in the mean number of deficiencies that affected interpretation: 4.9 and 5 points, respectively). However, between SOP-guided and registrant's RSS, there was no statistically significant difference in the human health toxicology and physicochemical endpoints RSSs (mean difference 1.2 points ($p = 0.5$) and 0.5 points ($p = 0.9$), respectively (Figure 8).

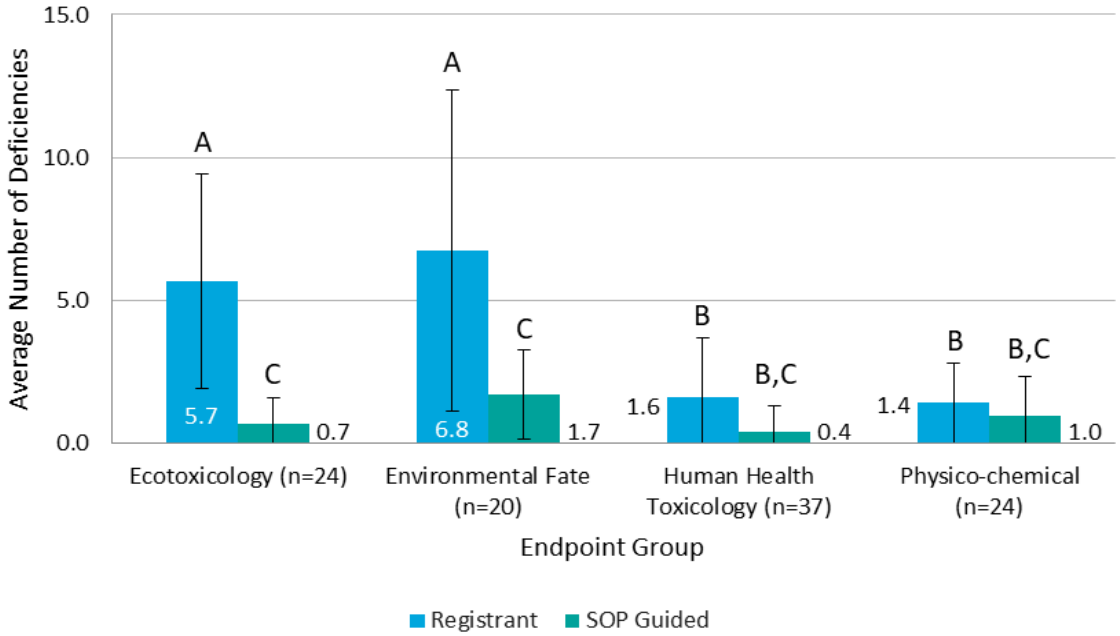


Figure 8. The average number of deficiencies per RSS that could affect the interpretation of results or the hazard conclusion and standard deviation for registrant and SOP-guided RSSs per endpoint group. Lower number for an RSS in this figure demonstrates that there are fewer deficiencies in that RSS that affected the results / hazard conclusion. Letters that are different show statistically significant differences ($p < 0.05$). Same letters on two bars demonstrates no statistically significant difference.

Overall, the results indicate that for ecotoxicology and environmental fate groups of endpoints, the number of deficiencies that could affect the interpretation of results was higher for RSS's authored by registrants compared to SOP-guided RSSs. However, for the human health toxicology and physicochemical endpoints, the registrant and SOP-guided RSSs were statistically similar in terms of the number of deficiencies that could affect the interpretation of results.

These results demonstrate that implementing a SOP and a robust review process improved the ecotoxicology and environmental fate RSSs in terms of achieving more accurate results and supporting the hazard conclusion on its own thanks to the information contained in it. For physicochemical and human health toxicology RSSs the registrant's RSSs already had a very low number of deficiencies that could affect the interpretation of the results and no difference was observed between the registrant's RSS and SOP-guided RSS in this regard. SOP-guided RSSs were statistically similar across the groups of endpoints, whereas the registrant's RSSs for ecotoxicology and environmental fate were similar to each other but different from the human health toxicology and physicochemical properties (which were similar to each other). However, both two groups of endpoints demonstrated a significantly higher number of deficiencies that could affect the interpretation of the results compared to human health toxicology as well as the physicochemical group of RSSs.

The reasons for a higher number of deficiencies affecting the interpretation of the results and the hazard conclusion in the registrant ecotoxicology and fate RSSs are unknown.
In later parts of this report, the 2-way ANOVA analysis of factors such as substance type, Klimisch score of full study reports, and the type of full study report did not reveal any

statistically significant difference between endpoints groups, suggesting some other reasons for the cause of the variability in the quality observed.

5.3.2. Level 2 analysis: Endpoint types within endpoint groups

Level 2 analysis takes a more granular approach. Each group of endpoints (ecotoxicology, human health h toxicology, fate, physicochemical) were further analysed at a deeper level to understand the differences between the endpoints within each group. For example, within the ecotoxicology group, how the acute aquatic toxicity RSSs differ from chronic aquatic toxicity. The three parameters of interest were the same as in the level 1 analysis. The analysis was performed within registrant's RSSs, within SOP-guided RSS, and between registrant and SOP-guided RSS using a 2-way ANOVA method followed by post-hoc testing.

5.3.2.1. Conclusions on the RSS score level 2 analysis

Overall, based on the level 2 analysis of RSS scores for each endpoint group, it can be concluded that the accuracy and completeness of the registrant's RSSs did not depend on the endpoint within a group. Moreover, the accuracy and completeness of the RSSs did not differ between the SOP-guided and registrant's RSSs for most endpoints except for four ecotoxicology endpoints (chronic aquatic, aquatic microorganisms, aquatic plant, and soil toxicity). These differences in quality between SOP-guided and registrant's RSSs for various ecotoxicology endpoints suggest a lack of proper guidance, lack of a robust review process and author experience could be potential reasons for the differences.

For the details of this analysis, see *Annex I Conclusions for the RSS score level 2 analysis.*

5.3.2.2. Conclusions on the access to full study report needed level 2 analysis

Overall, the general trend from the level 2 analysis indicates that within each endpoint group, there is no difference among endpoints for registrant's RSSs as well as for SOP-guided RSSs, in terms of the average number of times access to the full study report was needed. The only exception was aquatic microorganism toxicity in registrant's RSSs, where the number was significantly higher for aquatic microorganism toxicity when compared to acute aquatic as well as soil toxicity.

When corresponding endpoints between SOP-guided and registrant's RSSs were compared, four out of five endpoint types in the group ecotoxicology (long-term aquatic, aquatic microorganisms, aquatic plant, and soil toxicity) showed significant differences, where the number of times access to full study report needed was significantly higher for registrant's RSSs compared to SOP-guided RSSs. For all other endpoint groups i.e., human health toxicology, environmental fate, and physicochemical, there were no significant differences between SOP-guided and registrant's RSSs.
As discussed in the level 1 analysis (chapter 5.3.1), the parameter 'Access to full study report needed (yes/no)' was included in this study to estimate the independence of an RSS from the full study report. An RSS with a higher number of instances where access to the full study report is needed is more dependent on the full study report compared to an RSS with a lesser number of instances.

Therefore, the results suggest that the implementation of a SOP and a robust review process decreased the dependence of the SOP-guided RSSs on the full study report, compared to the registrant's RSS, for most of the ecotoxicology endpoints. Whereas for endpoints in other groups, no effect of SOP implementation was apparent, which is likely due to the relatively low sample size compared to level 1 analysis.

For the details of this analysis, see *Annex II Conclusions for the access to full study report needed level 2 analysis.*

5.3.2.3. Conclusions on the effect of deficiencies on the interpretation of the results/hazard conclusion level 2 analysis

Overall, the general trend of the level 2 analysis indicates that within each endpoint group, there is no difference among endpoints for registrant's RSSs as well as for SOP-guided RSSs, in terms of the average number of deficiencies that affected the hazard conclusion. In the level 1 analysis (chapter 5.3.1), it was shown that the registrant ecotoxicology endpoint group had a significantly higher number of deficiencies affecting the result interpretation/hazard conclusion than human health toxicology and physicochemical endpoints. Hence, together with level 2 results, the deficiencies identified in the ecotoxicology group are equally distributed among all endpoints of ecotoxicology and there is no effect of endpoint tier (such as acute versus chronic) on the number of deficiencies affecting results of registrant's RSSs.

When we compared corresponding endpoints between SOP-guided and registrant's RSSs, the only endpoint that showed a significant difference in terms of the average number of deficiencies with an effect on hazard conclusion was chronic aquatic, where the average number was higher for registrant's RSSs compared to SOP-guided RSSs. For all other endpoint groups, there were no significant differences between SOP-guided and registrant's RSSs. The level 1 analysis demonstrated that implementation of SOP and a robust review process improved the ecotoxicology and environmental fate RSSs in terms of reaching one of their key goals i.e., an accurate results/hazard conclusion well supported by the information present within the RSS; therefore, lack of similar results in level 2 analysis could be due to the relatively low sample size compared to level 1 analysis.

For the details of this analysis, see *Annex III Conclusions for the Effect of deficiencies on the interpretation of the results/hazard conclusion level 2 analysis.*

5.3.3. Additional analyses

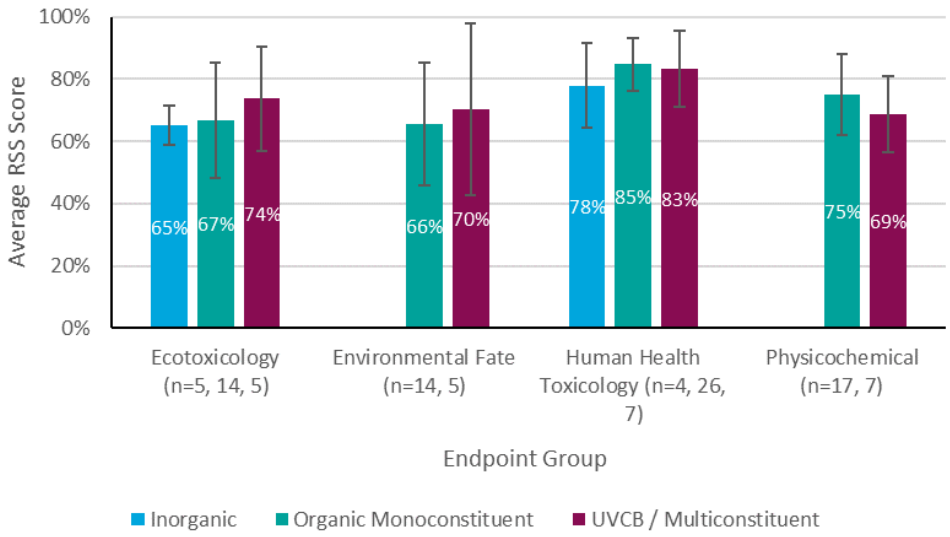
As discussed in the methodology (chapter 5.2.2), a few additional analyses were performed on registrant's RSSs with the purpose to further investigate the lower scores and higher number of deficiencies in the registrant's RSSs. This was achieved by evaluating the effects of various additional factors: different types of substances; full study reports of different Klimisch scores; and types of full study reports.
Additionally, an analysis of the registrant's RSSs with no deficiencies affecting the interpretation of the results and the hazard conclusion was carried out to evaluate the number of registrant's RSSs within each endpoint and endpoint groups that could be regarded as an ideal RSS.

5.3.3.1. Comparison of registrant's RSSs created for different types of substances

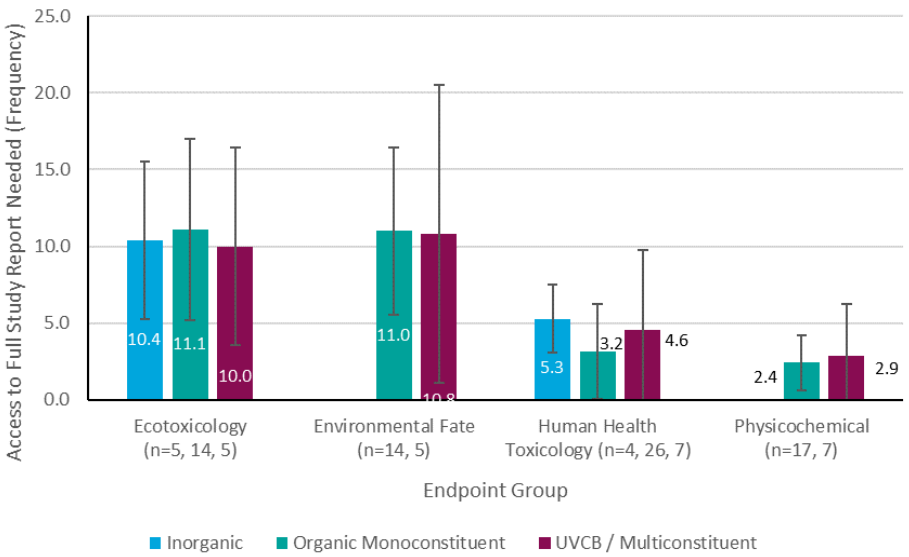
Among the endpoint groups of the registrant's RSSs, statistical analysis was performed to determine the effect of substance type (organic mono-constituent, UVCB / multi-constituent, and inorganic) on various parameters. This analysis revealed that there was no statistically significant difference in the registrant's RSSs created for different types of substances.

These results demonstrated that there was no effect on the registrant's RSSs irrespective of whether the RSSs were created for organic mono-constituent, UVCB / multi-constituent, or inorganic substances and the same trend was observed across all groups of endpoints. One caveat in this result is that for inorganic substances, there were no environmental fate and physicochemical RSSs available (Figure 21).

a)



b)



c)

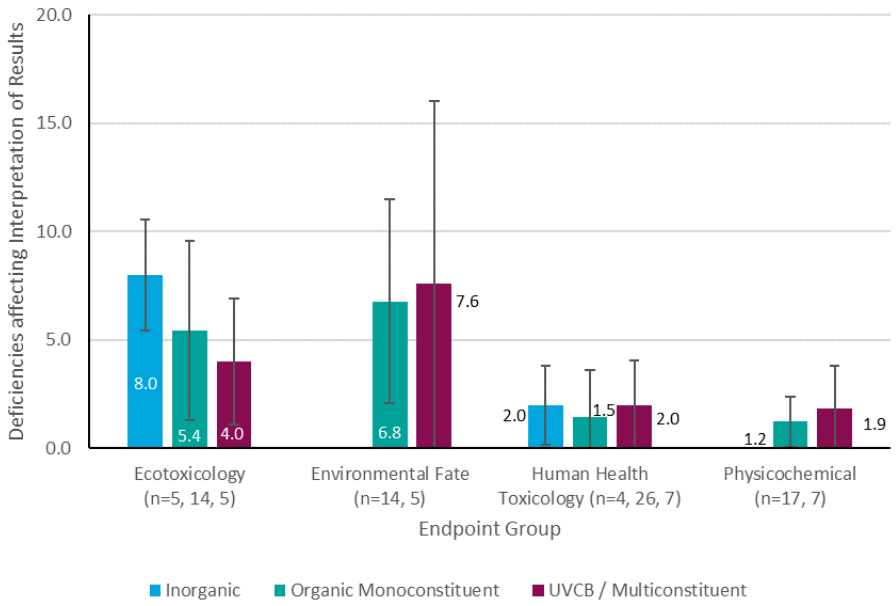


Figure 21. Comparison of registrant's RSSs created for different types of substances. The X-axis represents the endpoint groups and different coloured bars represent different types of substances. No statistically significant differences were observed in any groups. a) The average RSS score and standard deviation for registrant's RSSs created for different types of substances. (b) The average number of instances per registrant's RSS where access to the full study report was needed and standard deviation. (c) The average number of deficiencies per registrant's RSS affecting the interpretation of results for the hazard conclusion and standard deviation.

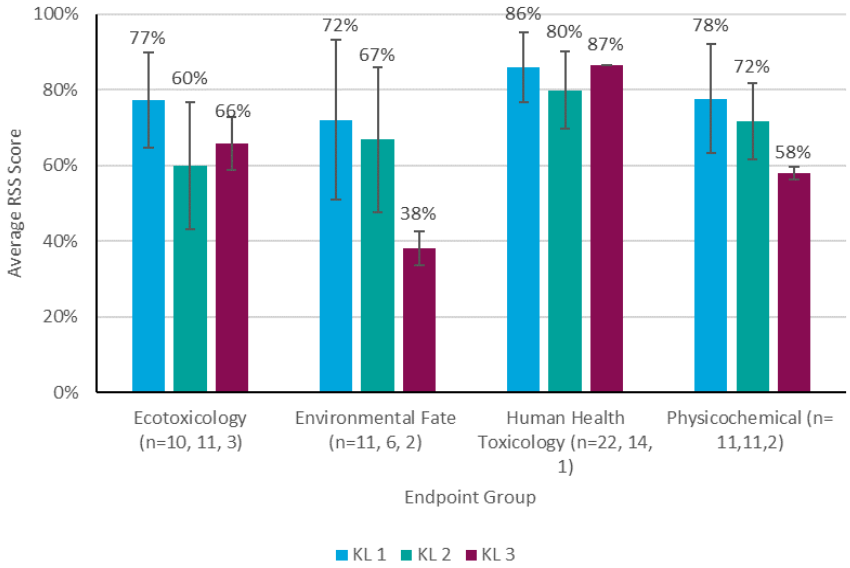
5.3.3.2. Comparison of registrant's RSSs created with full study reports of different Klimisch scores

Among endpoint groups (ecotoxicology, human health toxicology, environmental fate, and physicochemical) of the registrant's RSSs, statistical analysis was performed to determine the effect of full study reports of different Klimisch scores on various parameters.

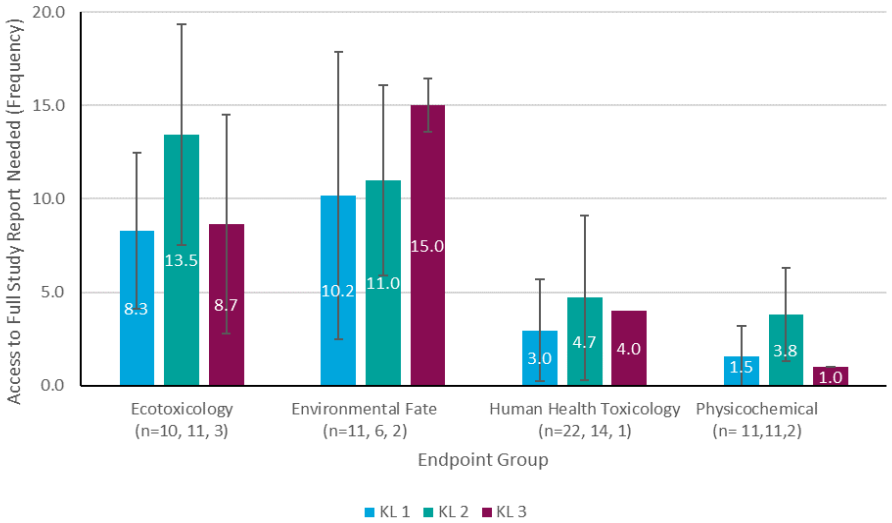
It is noteworthy that before this analysis, during the comparison between the registrant's RSS and SOP-guided RSS, Klimisch scores were also compared and any discrepancy between the Klimisch score assigned in registrant's RSSs and SOP-guided RSSs was noted. For the current analysis, which only focuses on registrant's RSSs, Klimisch scores assigned in the corresponding SOP-guided RSSs were considered as the correct Klimisch scores of the full study report, and hence used in this analysis.

This analysis revealed that, although there was no statistically significant difference in the registrant's RSSs created from full study reports with Klimisch scores 1 or 2, the results demonstrate that registrants' RSSs based on full studies with Klimisch score 3 had consistently a higher number of deficiencies affecting the interpretation of the results (Figure 22).

a)


100%
80%
60%
40%
20%
0%
Average RSS Score
77%
60%
66%
72%
67%
38%
86%
80%
87%
78%
72%
58%
Ecotoxicology (n=10, 11, 3)
Environmental Fate (n=11, 6, 2)
Human Health Toxicology (n=22, 14, 1)
Physicochemical (n= 11,11,2)
Endpoint Group
KL 1
KL 2
KL 3

b)


20.0
15.0
10.0
5.0
0.0
Access to Full Study Report Needed (Frequency)
8.3
13.5
8.7
10.2
11.0
15.0
3.0
4.7
4.0
1.5
3.8
1.0
Ecotoxicology (n=10, 11, 3)
Environmental Fate (n=11, 6, 2)
Human Health Toxicology (n=22, 14, 1)
Physicochemical (n= 11,11,2)
Endpoint Group
KL 1
KL 2
KL 3

c)

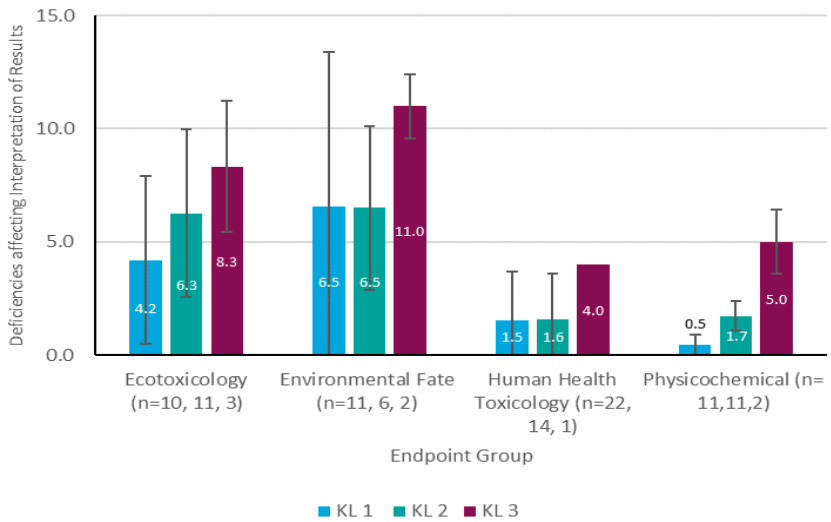

15.0
10.0
5.0
0.0
Deficiencies affecting Interpretation of Results
4.2
6.3
8.3
6.5
6.5
11.0
1.5
1.6
4.0
0.5
1.7
5.0
Ecotoxicology (n=10, 11, 3)
Environmental Fate (n=11, 6, 2)
Human Health Toxicology (n=22, 14, 1)
Physicochemical (n= 11,11,2)
Endpoint Group
KL 1
KL 2
KL 3

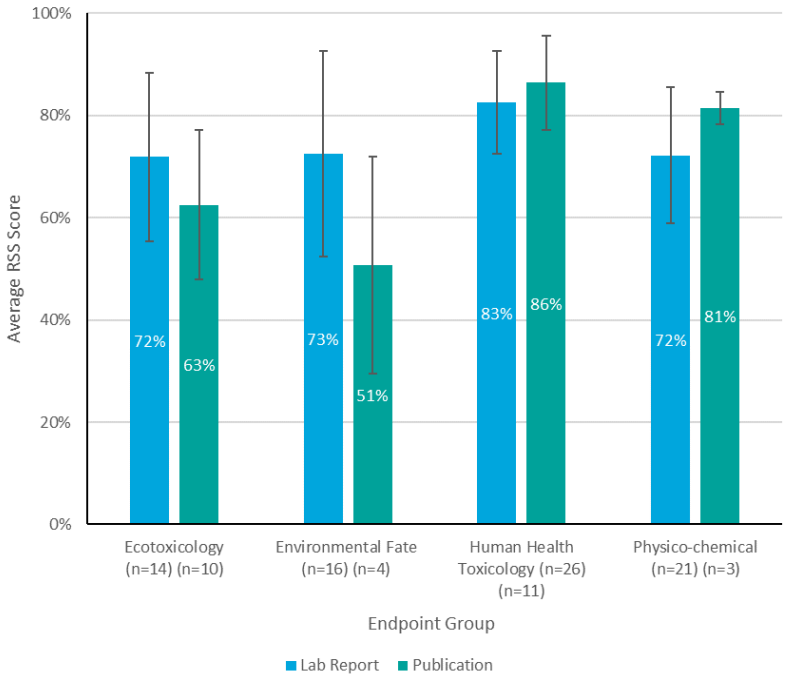
Figure 22. Comparison of registrant's RSSs created with full study reports of different Klimisch scores. The X-axis represents the endpoint groups and different coloured bars represent RSS created from full study reports with different Klimisch scores. No statistically significant differences were observed in any groups except for registrants' RSSs based on full studies with Klimisch score 3 which consistently have a higher number of deficiencies affecting the interpretation of the results in all groups. (a) The average RSS score and standard deviation for registrant's RSSs created for different types of substances. (b) The average number of instances per registrant's RSS where access to the full study report was needed and standard deviation. (c) The average number of deficiencies per registrant's RSS affecting the interpretation of results for the hazard conclusion and standard deviation.

5.3.3.3. Comparison of registrant's RSSs created from lab reports and peer reviewed scientific publications

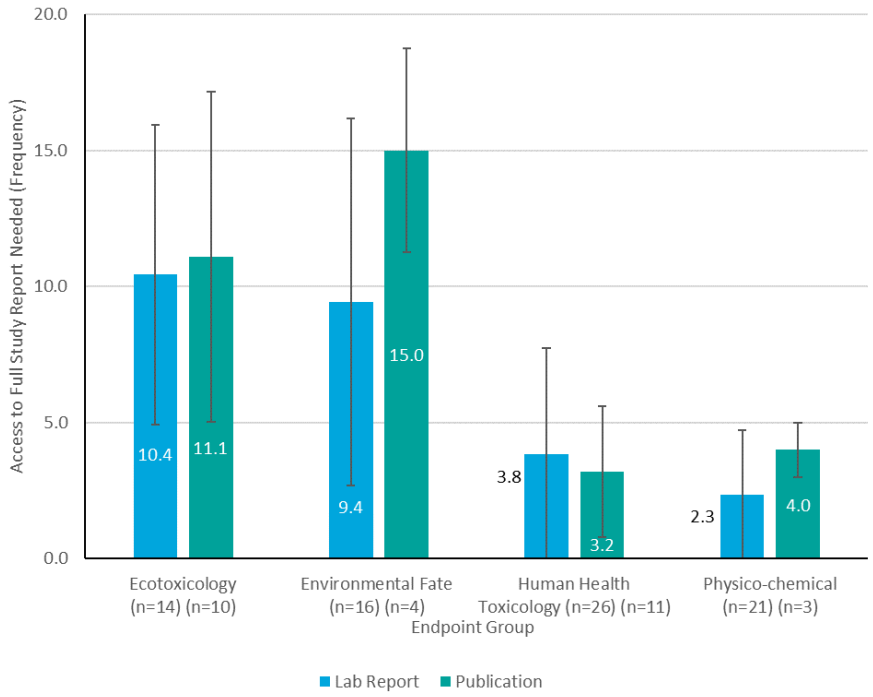
Among endpoint groups of the registrant's RSSs, statistical analysis was performed to determine the effect of full study report type (lab report versus peer reviewed scientific publications) on various parameters. This analysis revealed that there was no statistically significant difference in the registrant's RSSs created from lab reports and peer reviewed scientific publications for any of the endpoint groups.

These straightforward results demonstrated that there was no effect on the registrant's RSSs irrespective of whether the RSSs were created from lab reports or peer reviewed scientific publications and the same trend was observed across all groups of endpoints (Figure 23).

a)



b)



c)

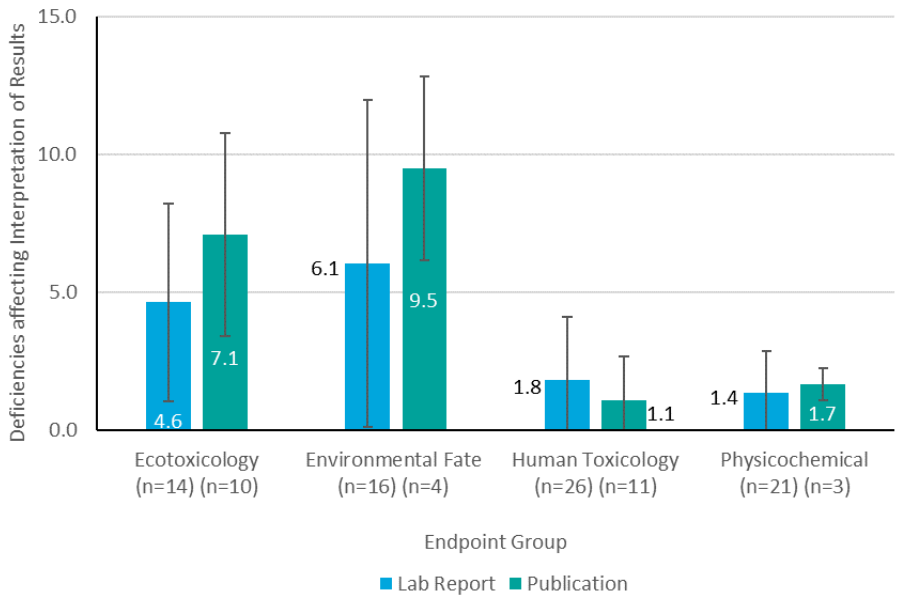


Figure 23. Comparison of registrant's RSSs created with different full study report types (Lab report or peer reviewed scientific publication). The X-axis represents the endpoint groups and different coloured bars represent different full study report types. No statistically significant differences were observed in any groups.
(a) The average RSS score and standard deviation for registrant's RSSs created for different types of substances. (b) The average number of instances per registrant's RSS where access to the full study report was needed and standard deviation. (c) The average number of deficiencies per registrant's RSS affecting the interpretation of results for the hazard conclusion and standard deviation.

5.3.3.4. Registrant's RSSs with no deficiencies that could affect the interpretation of results/hazard conclusion

The purpose of this exercise was to evaluate the number of registrant's RSSs within each endpoint and endpoint groups that could be regarded as an ideal RSS (i.e., no such deficiency identified that could affect the interpretation of results/hazard conclusion). As already introduced in chapter 5.3.1.3, the contractor took a conservative approach when assessing the RSSs.

The analysis revealed that within the ecotoxicology group of endpoints, up to 8.3% RSS (2/24) reached an ideal status that the contractor aimed at achieving and had no deficiencies. Similar numbers were noted for environmental fate endpoints (5%; 1/20); however, physicochemical endpoints demonstrated a several times higher number of RSSs that could be considered ideal with no deficiency affecting the results/hazard conclusion (25%; 6/24). The number of RSSs that showed no impact of deficiencies on results/hazard conclusion and thus can be termed 'ideal' was significantly higher for the human health toxicology group compared to other groups (48.7%; 18/37). Interestingly, further breaking the human health toxicology group into individual endpoints revealed that the number of ideal RSSs were higher in high tier endpoints such as carcinogenicity, and developmental toxicity compared to low tier endpoints such as irritation and sensitization. Similarly, high tier human health toxicology endpoints had a higher number of ideal RSSs compared to low tier endpoints in other endpoint groups such as partition coefficient, water solubility, vapour pressure, flammability, and acute aquatic toxicity.

These results demonstrate that low tier endpoints, whose RSS templates are generally considered relatively easier, do not necessarily lead to high quality RSSs. The full details are provided in Table 6.

The findings presented in Table 6 support the results of the level 1 analysis conducted in this work package and show that registrant' RSSs for ecotoxicology and environmental fate were similar to each other in terms of the number of deficiencies that affected the interpretation of results or hazard conclusion and both of these two groups of RSSs demonstrated a significantly higher number of deficiencies that affected the interpretation of the results compared to human health toxicology as well as the physicochemical group of RSSs. Therefore, these results demonstrate that registrants struggle more with ecotoxicology and environmental fate endpoints compared to human health toxicology and physicochemical endpoints, in terms of producing RSSs where results/hazard conclusions are well supported by the information contained within the RSS.

Table 6. Detailed analysis of registrant's RSSs with no deficiencies affecting interpretation of results/hazard conclusion

Endpoint (group)	Number of RSSs available (n)	RSSs with no deficiency affecting hazard conclusion	Percentage value (RSSs with no deficiency affecting hazard conclusion)
Ecotoxicology	24	2	8.3%

Endpoint (group)	Number of RSSs available (n)	RSSs with no deficiency affecting hazard conclusion	Percentage value (RSSs with no deficiency affecting hazard conclusion)
- Acute aquatic	6	1	16.7%
- Chronic aquatic	7	1	14.3%
- Soil	6	0	0
- Aquatic microorganism	3	0	0
- Aquatic plant	2	0	0
Human health toxicology	37	18	48.7%
- Carcinogenicity	6	4	66.7%
- Developmental	6	5	83.3%
- Genetic toxicity	6	2	33.3%
- Irritation	3	0	0
- Repeated dose	6	3	50%
- Reproductive	6	2	33.3%
- Sensitisation	4	2	50%
Environmental fate	20	1	5%
- Biodegradation screening	5	1	20%
- bioaccumulation	6	0	0
- biodegradation (soil)	1	0	0
- Biodegradation (simulation)	2	0	0
- Hydrolysis	6	0	0
Physicochemical	24	6	25%
- Vapour pressure	6	3	50%
- Partition coefficient	6	1	16.7%
- Water solubility	6	1	16.7%
- Flammability	6	1	16.7%

5.3.3.5. Analysis of the “missing information” and “wrong information” deficiencies

As defined in chapter 5.2.1, the term “wrong information” refers to information that was erroneous (i.e., wrong units, wrong values, etc.). An example of this is given in table 3 as the contractor identified the wrong water media type in a couple of RSSs or when a RSS incorrectly identified a study as a limit test when it was not. The term “missing information” is simply a field in an RSS left blank whether or not it is relevant for the assessment.

In the calculation for RSS score in this report, the deficiency classes labelled as ‘missing information’ and ‘wrong information’ had similar effects i.e., both lead to a score of zero because both classes of deficiency seem to indicate similar effects on the accuracy and completeness of the RSS. However, it is still valuable to recognise the difference between the two classes because they could indicate different causes of deficiencies. For example, ‘wrong information’ in an RSS field is more likely to indicate a problem in the author's experience and understanding for the endpoint, whereas ‘missing information’ is more likely to indicate issues with the lack of clarity in OHT structure or opportunities forimprovement of the technical completeness check tools, such as validation assistant for REACH submissions.

However, there could also be overlap between the two classes of deficiencies in certain situations. For example, for some deficiencies, missing information as well as wrong information could both point towards a need for improvement in the existing guidance.
In all endpoint groups, there was a clear trend of lower average number of wrong information per RSS compared to the missing information as shown in Figure 24. The same trend persisted across various endpoints within each group as shown in Figures 25-28. Therefore, it is clear that across all endpoints and endpoint groups, the number of deficiencies that were labelled as ‘missing information’ were higher than the number of deficiencies that were labelled as ‘wrong information’.

Looking at the data, “missing information” accounted for nearly 60% of all the deficiencies identified (“wrong information” and “partial information” accounted for 20% each). Moreover, “missing information” represented 63% of the cases where deficiencies had an effect in the interpretation of the results while “wrong information” followed with 21% and “partial information” with 16%.

Therefore, minimising the number of fields that are left blank (missing information) would have a greater impact on the quality of the RSSs. This would apply to all categories of endpoints, but especially to the genetic toxicity, developmental and sensitisation endpoints.

The following examples describe some of the most common causes for a deficiency to have an impact in the interpretation of the results:

- wrong information about the adequacy of the studies used (key study)
- questionable reliability of the study (Klimisch score)
- no indication whether the validity criteria were met

The individual deficiencies identified in WP2 will be evaluated more thoroughly during the last phase of this project, in WP3.

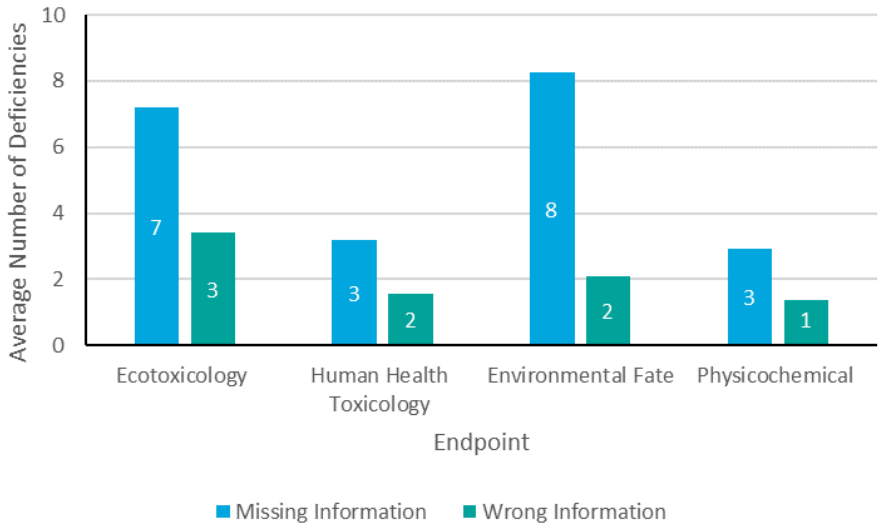


Figure 24. Graph demonstrating the average number of deficiencies that were labelled as 'missing information' as well as 'wrong information' in all endpoint groups.

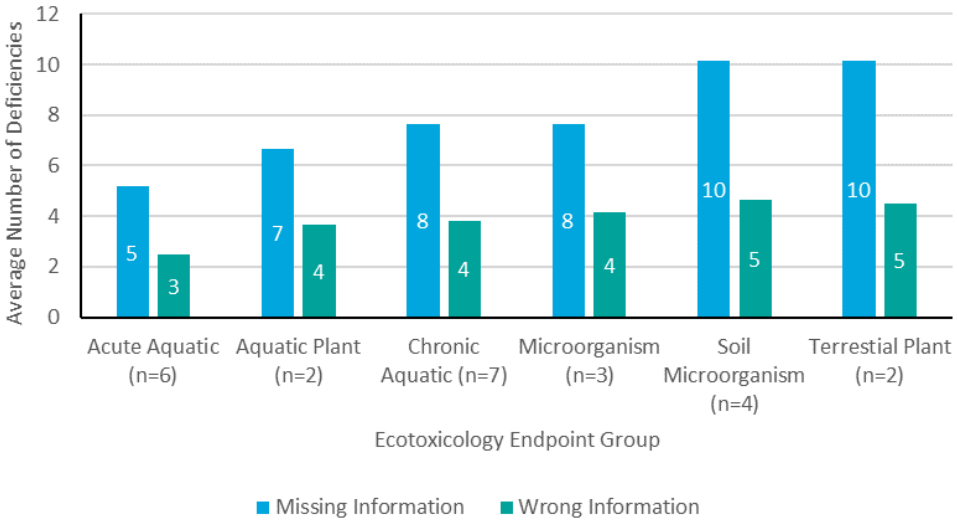


Figure 25. Graph demonstrating the average number of deficiencies that were labelled as 'missing information' as well as 'wrong information' in ecotoxicology endpoints.

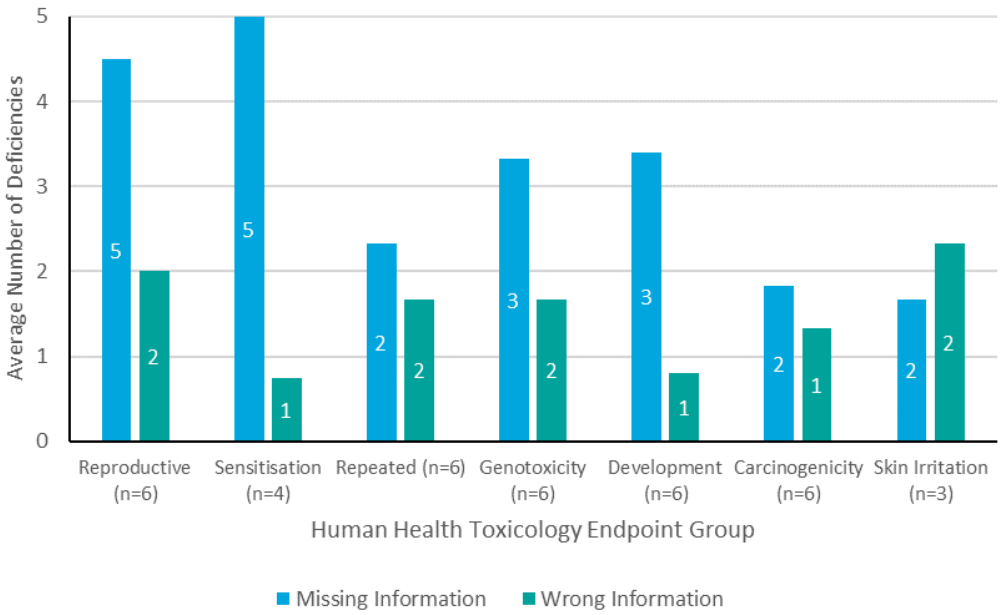


Figure 26. Graph demonstrating the average number of deficiencies that were labelled as 'missing information' as well as 'wrong information' in human health toxicology endpoints.

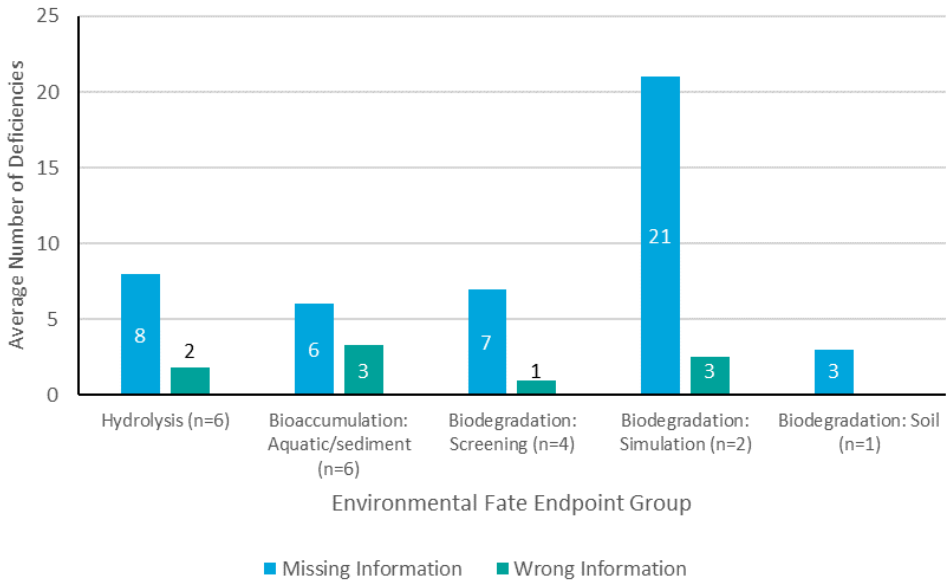


Figure 27. Graph demonstrating the average number of deficiencies that were labelled as 'missing information' as well as 'wrong information' in environmental fate endpoints.

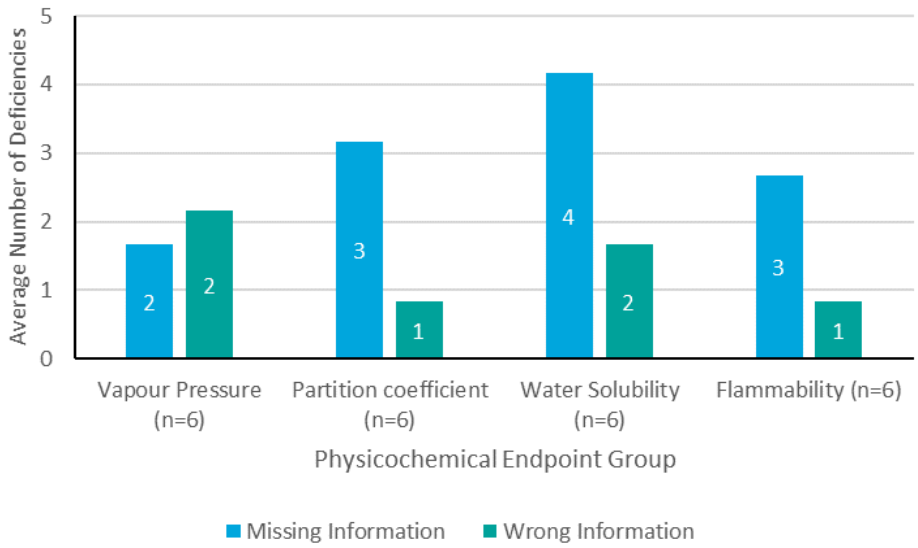


Figure 28. Graph demonstrating the average number of deficiencies that were labelled as 'missing information' as well as 'wrong information' in physicochemical endpoints.

5.4. Limitations of the study

The purpose of the level 2 analysis was to look deeper within each group of endpoints. To achieve this objective, four separate 2-way ANOVA analyses were performed, one within each group of endpoints i.e., ecotoxicology, human health toxicology, environmental fate, and physicochemical. One shortcoming of level 2 analysis was the reduction in sample size, which ranged between n = 1 to n = 7. Such low sample size is likely to be one of the reasons for the lack of statistical significance among most endpoints within the group of endpoints in level 2 analysis (chapter 5.3.2.1).

For example, within the human health toxicology group of endpoints, there was no significant difference in the RSS score between registrant and SOP-guided RSSs for carcinogenicity, developmental, genetic toxicity, irritation, repeated, reproductive, and skin sensitisation (chapter 5.3.2.1), where the sample size was 3-6, whereas, in level 1 analysis there was a significant difference in the RSS score between registrant and SOP-guided RSSs in the human health toxicology group (sample size 20-37) (chapter 5.3.1.1).

5.5. Overall Conclusions

The analysis revealed that SOP-guided RSSs showed higher accuracy and completeness as well as a lower number of deficiencies that could affect the resulting interpretation/hazard conclusion. This reduced the number of cases in which access to the full study report was required.

The conservative approach taken by the contractor towards highlighting all deficiencies to obtain ideal RSSs has, as expected, resulted in a low number of RSSs that could be qualified as 'perfect'. This is an accepted consequence as the purpose of this study was to identify areas of improvement.

With this in mind, overall, the accuracy and completeness of the human health toxicology group of endpoints in the registrant's RSSs was considered significantly higher than ecotoxicology, environmental fate, and physicochemical endpoints. Moreover, a significantly higher number of deficiencies that could affect the results as well as access to a full study report was observed for ecotoxicology and environmental fate compared to human health toxicology and physicochemical endpoints.

Therefore, considering this approach, when looking at the RSS submitted in the registrants' dossiers, human health toxicology and physicochemical RSSs showed better fidelity to the full study reports than the ecotoxicology and environmental fate RSSs. Furthermore, 8.3% of registrant ecotoxicology RSSs and 5% of registrant environmental fate RSS demonstrated an ideal status, compared to human health toxicology (48.7%) and physicochemical RSSs (25%). Deeper analysis revealed that the quality of registrant's RSSs, as measured using the three parameters, did not depend on the individual endpoints within a group as there were no statistically significant differences across different endpoints within a group.

This showed that the RSSs for high tier endpoints do not differ from low tier endpoints, although it should be noted that the sample size for this analysis was very small. Moreover, the additional analysis also revealed that there were no significant differences among RSSs created from different types of full study reports, full study reports with different Klimisch score (KL score 1 or 2), and substance type (mono-constituent organic, inorganic, or multi-constituent/UVCB). The low RSS scores and high number of deficiencies that affected the interpretation of the results and the hazard conclusions or that required access to full study reports for the registrant's RSSs (and even poorer performance of ecotoxicology and environmental fate endpoint groups) could be due to poor use of available guidance, lack of author's experience, lack of a proper review process, inadequacies in templates or guidance for these endpoint groups.

As discussed above, our analysis showed that SOP-guided RSSs consistently had improved scores for all three parameters across all endpoint groups; therefore, suggesting that the overall RSS quality can be improved significantly by implementing proper training, guidance and a RSS review process.

In WP3 the results from this analysis will be reviewed alongside the findings from WP1 to understand the causation of the results. Recommendations to improve the usefulness of RSS for the purpose of hazard assessment will then be made.

Annexes

Annex I Conclusions on the RSS score level 2 analysis

Ecotoxicology

Within both the registrant's RSSs and the SOP-guided RSSs, there was no significant difference in the scores among different endpoints. When the two types of RSS were compared, the scores for each endpoint were significantly higher for the SOP-guided RSSs compared to registrant's RSSs. The SOP-guided RSS scores ranged between 95.4 - 97.2% across the five endpoints compared to the registrant's RSSs which scored between 52.7 - 72.6% (Figure 9).

In the level 1 analysis, it was shown that the ecotoxicology endpoint group had a significantly lower score than the human health toxicology score. Hence, together with the level 2 results, the deficiencies identified in the ecotoxicology group are equally distributed among all endpoints of ecotoxicology and there is no effect of endpoint type (such as acute versus chronic) on the score of the RSSs.

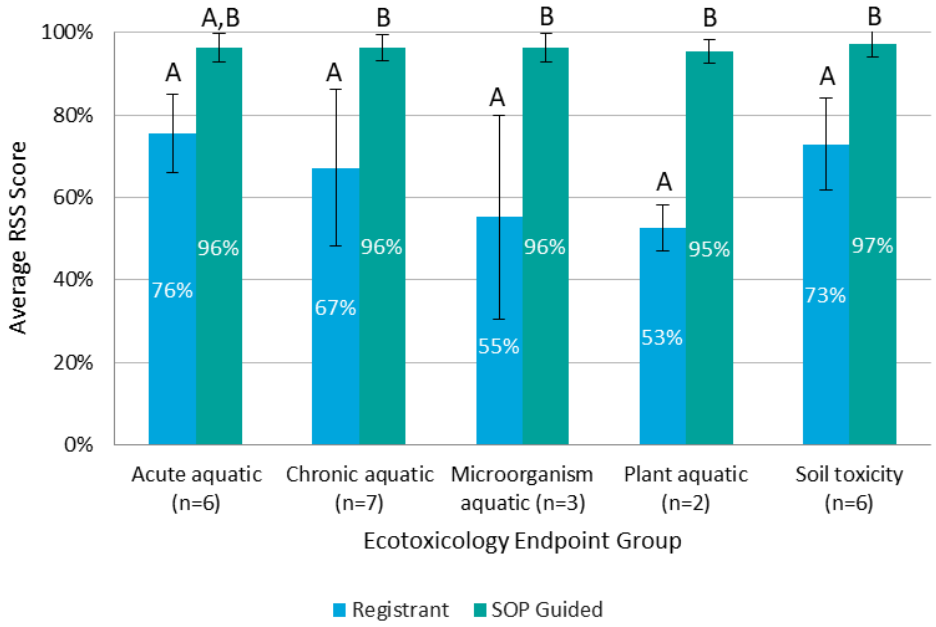


Figure 9. The average RSS score and standard deviation for registrant and SOP-guided RSSs per ecotoxicology endpoint group. Higher score in this figure demonstrates higher accuracy and completeness of an RSS. Letters that are different show statistically significant differences ($p < 0.05$). Same letters on two bars demonstrates no statistically significant difference.

Human Health Toxicology

There was no significant difference in the RSS scores among different endpoints within human health toxicology endpoint groups for both registrant as well as SOP-guided RSSs. In addition, there were no significant differences between the SOP-guided and registrant's RSS for each endpoint (Figure 10).

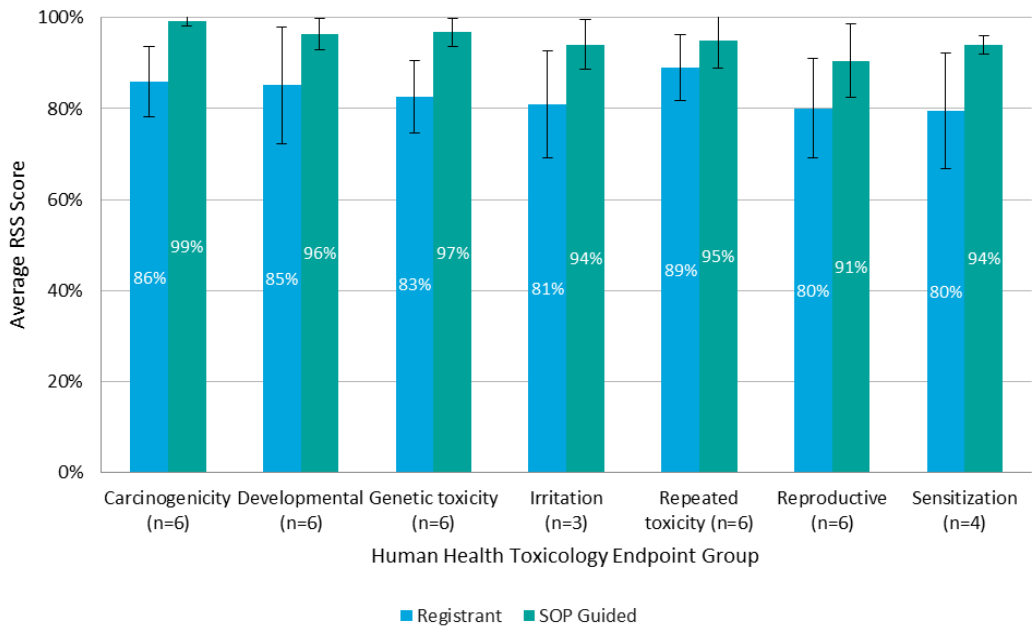


Figure 10. The average RSS score and standard deviation for registrant and SOP-guided RSSs per human health toxicology endpoint group. No statistically significant differences were observed in any of the groups.

Environmental fate

Within both registrant as well as SOP-guided RSSs, there was no significant difference in the scores among different endpoints of the environmental fate group. Figure 11 indicates that there was more than a 2-fold difference between the biodegradation in soil and biodegradation simulation; however, the results were still not statistically significant (p = 0.12), which was due to low sample sizes in these two endpoints and high standard deviation in biodegradation simulation endpoint. Between SOP-guided and registrant's RSSs, there were no significant differences in the RSSs for the corresponding endpoints (Figure 11).

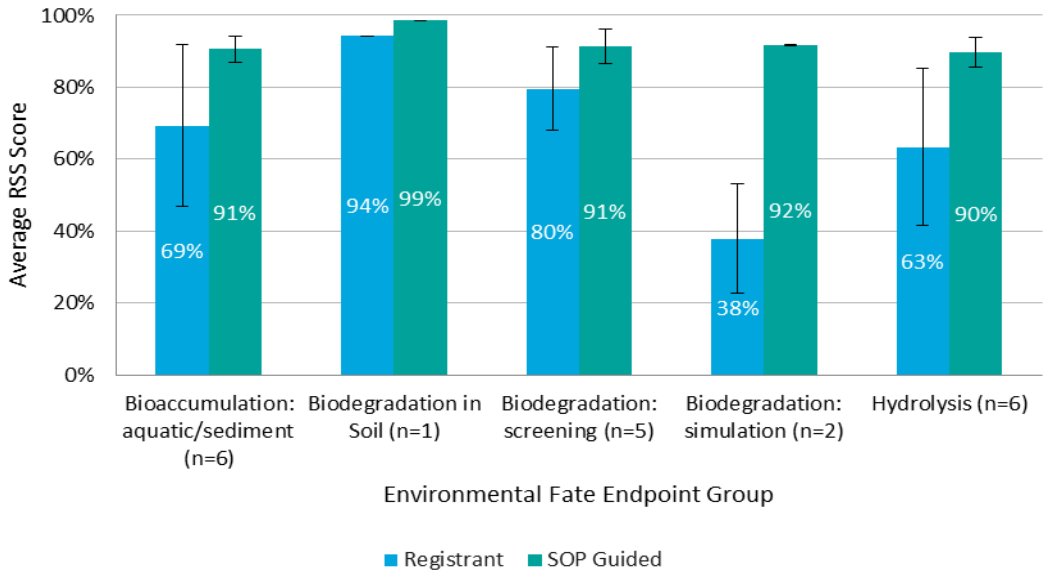


Figure 11. The average RSS score and standard deviation for registrant and SOP-guided RSSs per environmental fate endpoint group. No statistically significant differences were observed in any of the groups.

Physicochemical

Within both registrant and SOP-guided RSSs, there was no significant difference in the scores among different endpoints of the physicochemical group. Between the SOP-guided and registrant's RSSs, there were no significant differences in the RSSs for the corresponding endpoints (p values in all cases >0.5) (Figure 12).

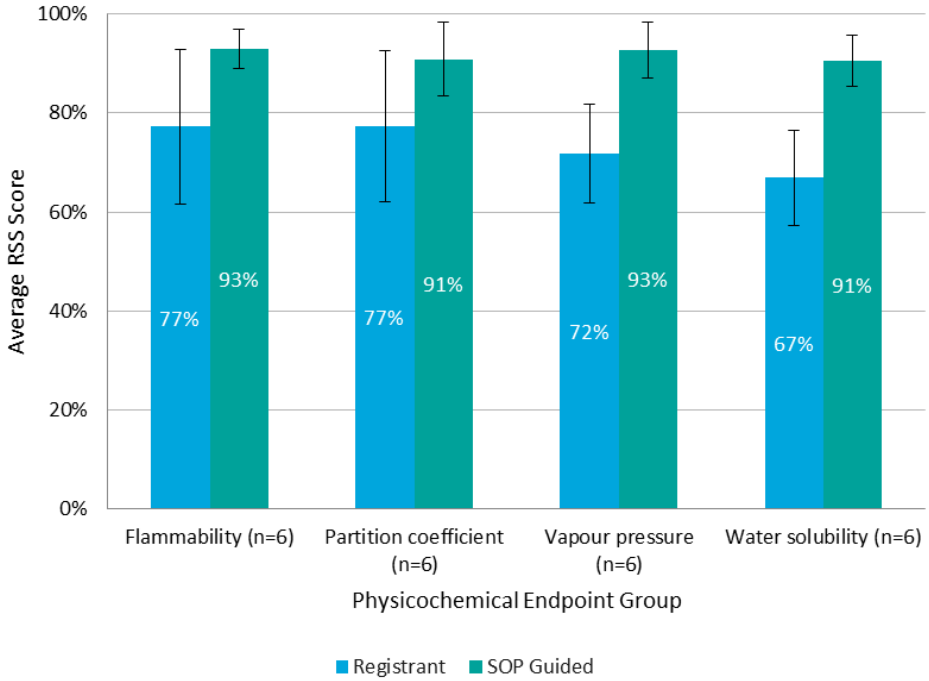


Figure 12. The average RSS score and standard deviation for registrant and SOP-guided RSSs per physicochemical endpoint group. No statistically significant differences were observed in any of the groups.

Annex II Conclusions on the access to full study report needed level 2 analysis

Ecotoxicology

Within registrant's RSSs, there was a significant difference between acute aquatic and aquatic microorganism, and between soil toxicity and aquatic microorganisms in terms of the number of times access to a full study report was needed. In both instances, the number of times access to the full study report was needed was higher for aquatic microorganism (mean value 17.3) than acute aquatic (mean value 7.8) and soil toxicity (mean value 8.3). Within the SOP-guided RSSs, there were no significant differences among the endpoints (Figure 13).

Between SOP-guided and registrant's RSSs, the number of times access to a full study report was needed for chronic aquatic, aquatic microorganisms, aquatic plant, and soil toxicity was significantly higher for registrant's RSSs compared to SOP-guided RSSs. The average number of times access to the full study report was needed was in the range of 7.8 - 17.3 for registrant's RSSs compared to 2.8 - 6.0 for SOP-guided RSSs (Figure 13).

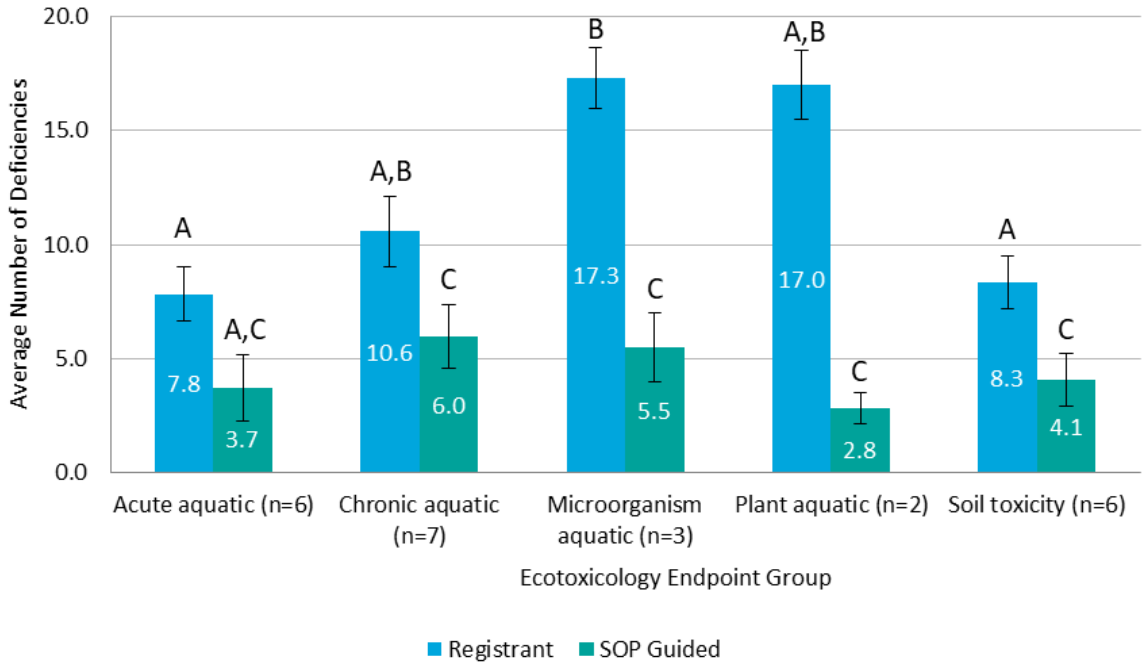


Figure 13. The average number of instances per RSS where access to the full study report was needed and standard deviation for registrant and SOP-guided RSSs for each endpoint within ecotoxicology. Lower number for an RSS in this figure demonstrates reduced dependency on the full study report for that RSS. Letters that are different show statistically significant differences ($p < 0.05$). Same letters on two bars demonstrates no statistically significant difference.

Human Health Toxicology

Within both registrant and SOP-guided RSSs, there was no significant difference among the different endpoints of the human health toxicology group. Between SOP-guided and registrant's RSSs, there were no statistically significant differences in the RSSs for the corresponding endpoints (p values in all cases >0.5) (Figure 14). The lack of statistical significance was probably due to low sample size and high variability in the dataset.

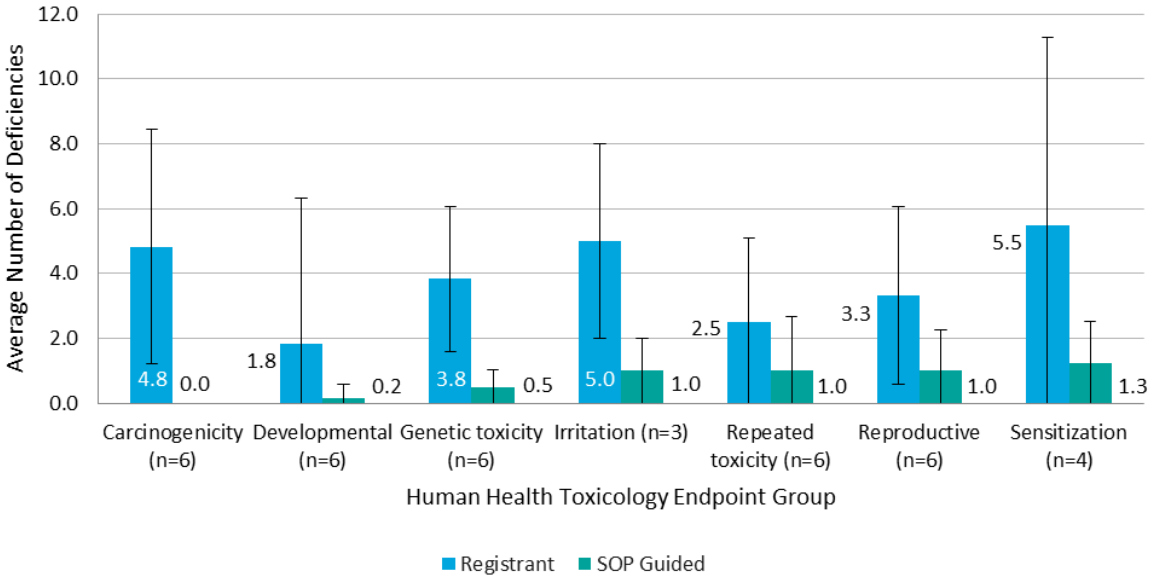


Figure 14. The average number of instances per RSS where access to the full study report was needed and standard deviation for registrant and SOP-guided RSSs for each endpoint within human health toxicology. No statistically significant differences were observed in any of the groups.

Environmental Fate

Within both registrant as well as SOP-guided RSSs, there was no significant difference among different endpoints of the environmental fate group. For each endpoint, there were no significant differences between SOP-guided and registrant's RSSs. In biodegradation simulation and bioaccumulation, the differences between SOP-guided and registrant's RSSs were large (~5 and ~3 fold, respectively); however, the results were still not significant due to low sample size (n = 2 and 6, respectively) and high standard deviation in results (Figure 15). Similarly, it is noteworthy that the biodegradation in soil only had a sample size of 1, hence, it cannot be interpreted accurately from statistical perspective; nonetheless, the number of deficiencies that needed access to the full study report were 3-fold lower for the SOP-guided RSS compared to the registrant's RSSs.

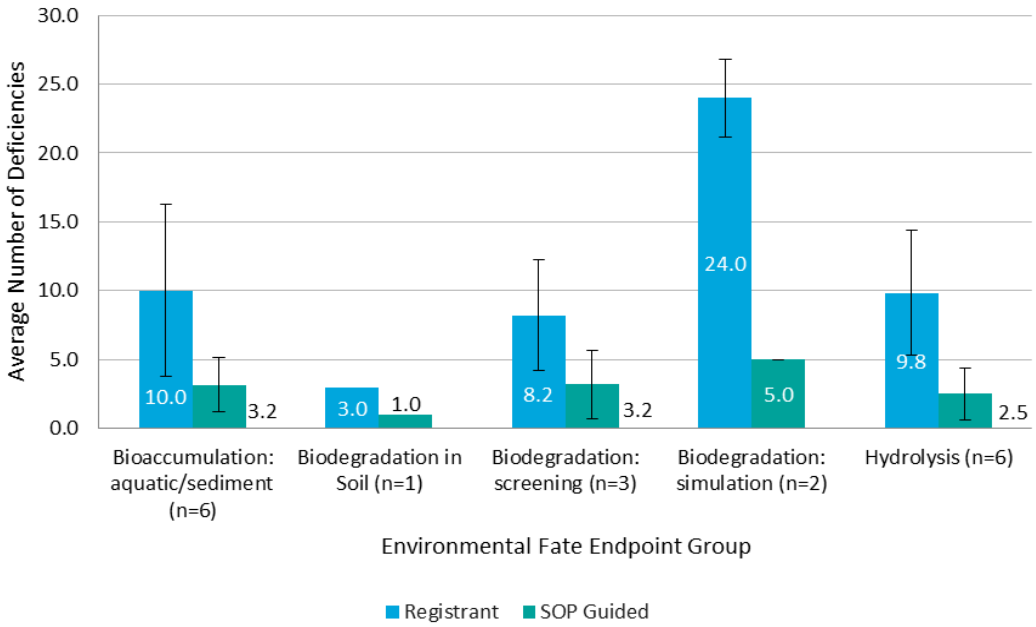


Figure 15. The average number of instances per RSS where access to the full study report was needed and standard deviation for registrant and SOP-guided RSSs for each endpoint within environmental fate. No statistically significant differences were observed in any of the groups.

Physicochemical

Within both registrant as well as SOP-guided RSSs, there was no significant difference among different endpoints of the physicochemical group. Between SOP-guided and registrant's RSSs, there were no significant differences in the RSSs for the corresponding endpoints (Figure 16).

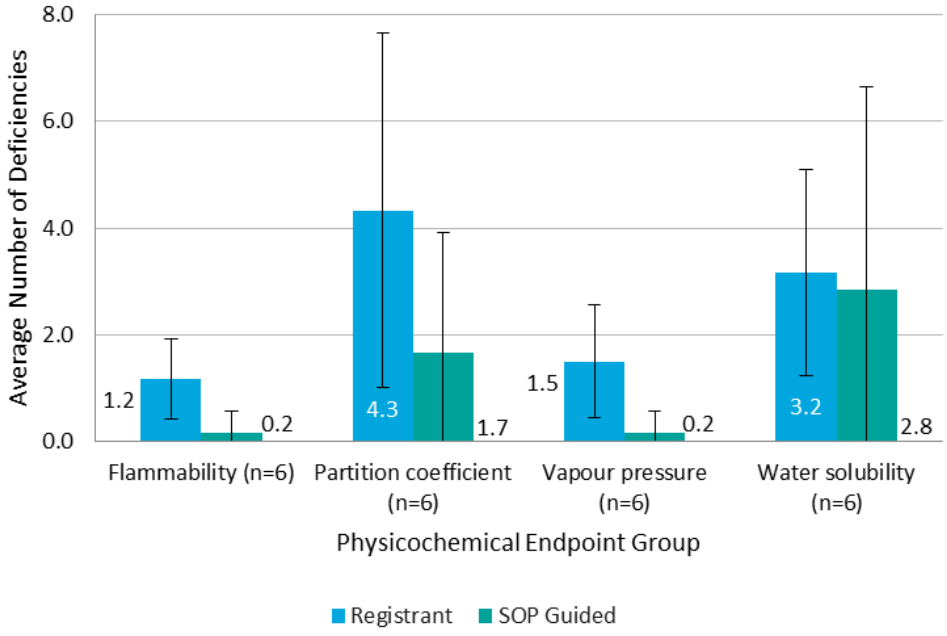


Figure 16. The average number of instances per RSS where access to the full study report was needed and standard deviation for registrant and SOP-guided RSSs for each endpoint within physicochemical. No statistically significant differences were observed in any of the groups.

Annex III Conclusions on the effect of deficiencies on the interpretation of the results/hazard conclusion level 2 analysis

Ecotoxicology

Within both registrants as well as SOP-guided RSSs, there was no significant difference among different endpoints of the ecotoxicology group in terms of the effect of deficiencies on the results/hazard conclusion (Figure 17).

Between SOP-guided and registrant's RSSs, the only endpoint within ecotoxicology that showed significant difference in terms of the effect on hazard conclusion was chronic aquatic, where the average number of deficiencies that affected the interpretation of results/hazard conclusion was 6.7 for registrant's RSSs compared to 0.9 for SOP-guided RSSs (Figure 17).

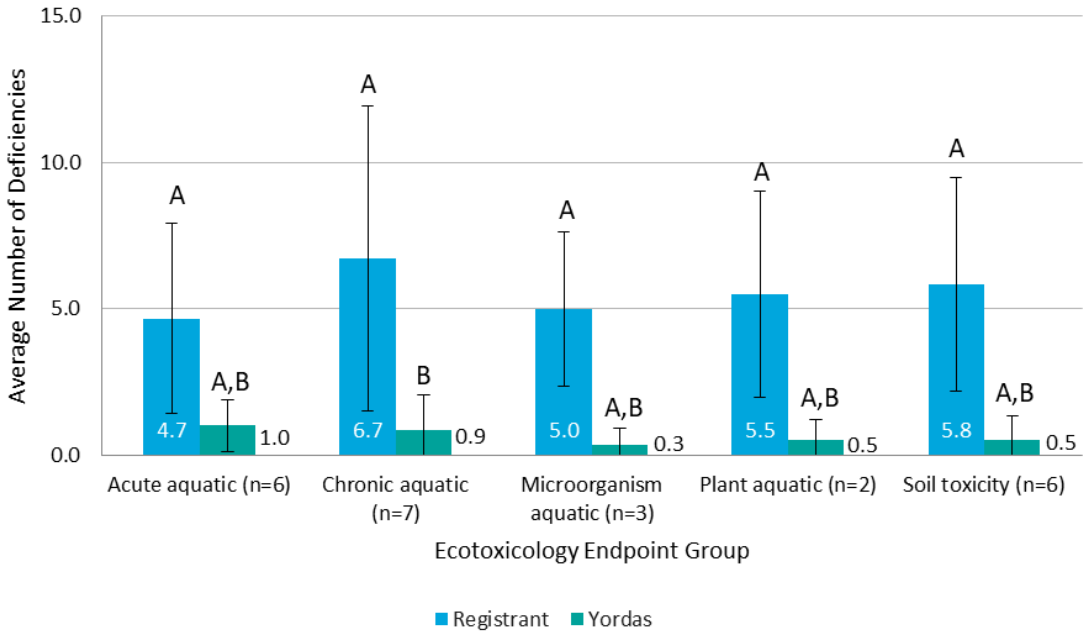


Figure 17. The average number of deficiencies per RSS that affect the interpretation of results for the hazard conclusion and standard deviation for registrant and SOP-guided RSSs for each endpoint within the ecotoxicology group. letters that are different show statistically significant differences ($p < 0.05$). Same letters on two bars demonstrates no statistically significant difference.

Human Health Toxicology

Within both registrants as well as SOP-guided RSSs, there was no significant difference in the scores among different endpoints of the human health toxicology group. Between SOP-guided and registrant's RSSs, there were no significant differences in the RSSs for the corresponding endpoints (Figure 18).

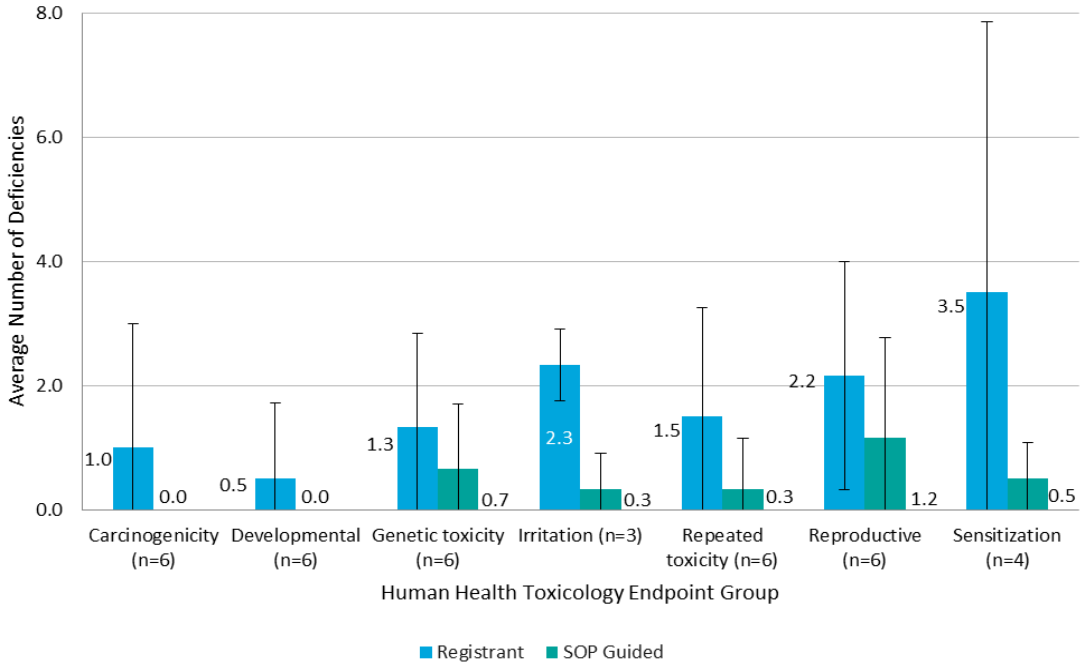


Figure 18. The average number of deficiencies per RSS that affect the interpretation of results for the hazard conclusion and standard deviation for registrant and SOP-guided RSSs for each endpoint within the human health toxicology group. No statistically significant differences were observed in any of the groups.

Environmental fate

Within both registrants as well as SOP-guided RSSs, there was no significant difference in the scores among different endpoints of the environmental fate group. Between SOP-guided and registrant's RSSs, there were no significant differences in the RSSs for the corresponding endpoints (Figure 19).

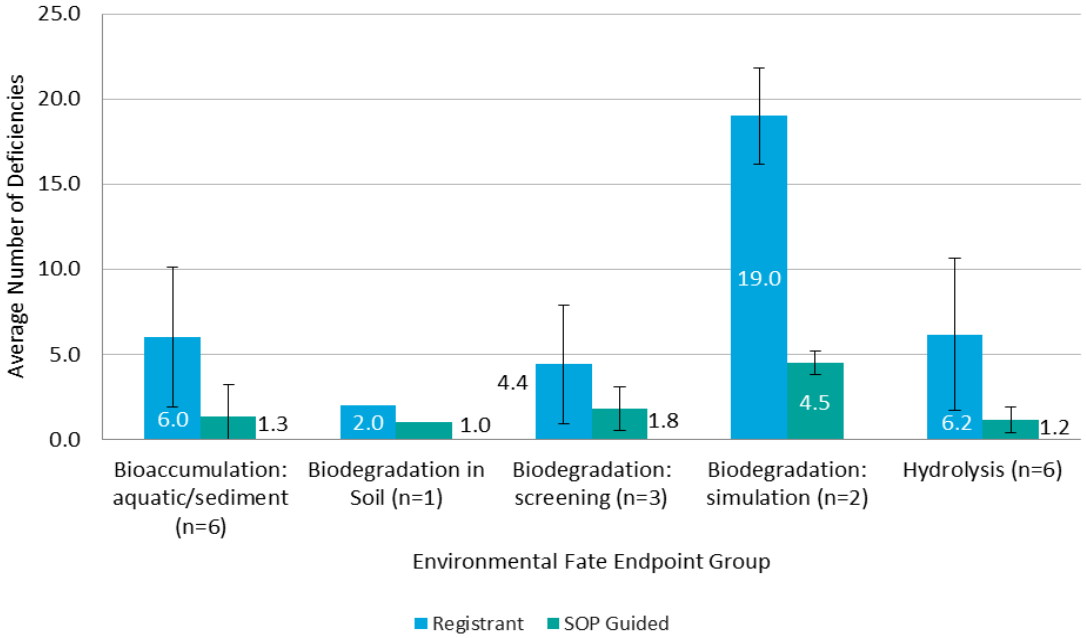


Figure 19. The average number of deficiencies per RSS that affect the interpretation of results for the hazard conclusion and standard deviation for registrant and SOP-guided RSSs for each endpoint within the environmental fate group. No statistically significant differences were observed in any of the groups.

Physicochemical

Within both registrants as well as SOP-guided RSSs, there was no significant difference in the scores among different endpoints of the physicochemical group. Between SOP-guided and registrant's RSSs, there were no significant differences in the RSSs for the corresponding endpoints (Figure 20).

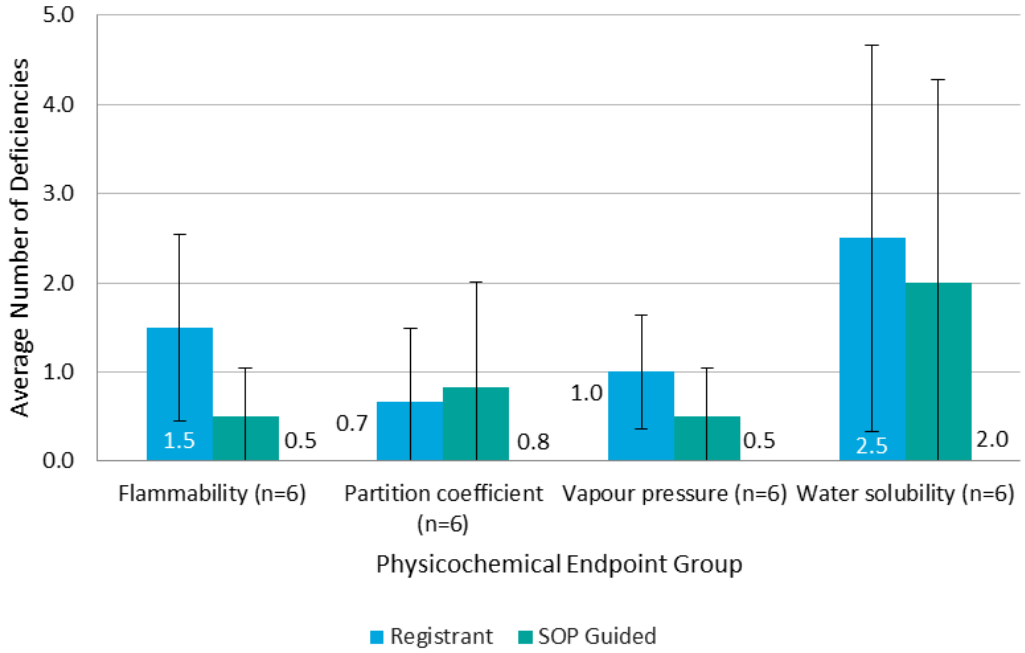


Figure 20. The average number of deficiencies per RSS that affect the interpretation of results for the hazard conclusion and standard deviation for registrant and SOP-guided RSSs for each endpoint within the physicochemical group. No statistically significant differences were observed in any of the groups.

Disclaimer

Reasonable efforts have been made throughout the review process to reach the conclusions and recommendations provided. The conclusions and recommendations given in this report are based upon and therefore limited to the information available and provided by the client at the time of writing. As such, Yordas Group accepts no liability if any regulating or enforcement bodies do not reach the same conclusions or recommendations.