Bayesian based similarity assessment of nanomaterials to inform grouping

Tsiliki, G.; Seleci, D.A.; Zabeo, A.; Basei, G.; Hristozov, D.; Jeliazkova, N.; ... ; Stone, V.

Citation
Tsiliki, G., Seleci, D. A., Zabeo, A., Basei, G., Hristozov, D., Jeliazkova, N., ... Stone, V. (2022). Bayesian based similarity assessment of nanomaterials to inform grouping. *Nanoimpact, 25*. doi:10.1016/j.impact.2022.100389

Version:	Publisher's Version
License:	Licensed under Article 25fa Copyright Act/Law (Amendment Taverne)
Downloaded from:	https://hdl.handle.net/1887/3304425

Note: To cite this publication please use the final published version (if applicable).

NanoImpact 25 (2022) 100389

Contents lists available at ScienceDirect

NanoImpact

journal homepage: www.elsevier.com/locate/nanoimpact

Bayesian based similarity assessment of nanomaterials to inform grouping

Georgia Tsiliki [a,*], Didem Ag Seleci [b], Alex Zabeo [c], Gianpietro Basei [c], Danail Hristozov [c], Nina Jeliazkova [d], Matthew Boyles [e], Fiona Murphy [f], Willie Peijnenburg [g,h], Wendel Wohlleben [b], Vicki Stone [f]

[a] Institute for the Management of Information Systems, Athena Research Center, Marousi, Greece
[b] Advanced Materials Research, Dept. of Material Physics and Analytics and Dept. of Experimental Toxicology and Ecology, BASF SE, Ludwigshafen, Germany
[c] GreenDecision Srl., Venezia, Italy
[d] Ideaconsult Ltd, Sofia, Bulgaria
[e] Institute of Occupational Medicine, Edinburgh, United Kingdom
[f] NanoSafety Group, Heriot-Watt University, Edinburgh, United Kingdom
[g] National Institute of Public Health and the Environment (RIVM), Center for Safety of Substances and Products, Bilthoven, the Netherlands
[h] Leiden University, Institute of Environmental Sciences (CML), Leiden, the Netherlands

ARTICLE INFO

Editor: Dr. Bernd Nowack

Keywords:
Similarity
Grouping
Dose-response data
Pairwise comparisons
Biological relevance

ABSTRACT

Nanoforms can be manufactured in plenty of variants by differing their physicochemical properties and toxicokinetic behaviour which can affect their hazard potential. To avoid testing of each single nanomaterial and nanoform variation and subsequently save resources, grouping and read-across strategies are used to estimate groups of substances, based on carefully selected evidence, that could potentially have similar human health and environmental hazard impact. A novel computational similarity method is presented aiming to compare dose-response curves and identify sets of similar nanoforms. The suggested method estimates the statistical model that best fits the data by leveraging pairwise Bayes Factor analysis to compare pairs of curves and evaluate whether each of the nanoforms is sufficiently similar to all other nanoforms. Pairwise comparisons to benchmark materials are used to define threshold similarity values and set the criteria for identifying groups of nanoforms with comparatively similar toxicity. Applications to use case data are shown to demonstrate that the method can support grouping hypotheses linked to a certain hazard endpoint and route of exposure.

1. Introduction

Grouping and read-across are powerful tools to reduce the amount of necessary experimental testing and time (OECD, 2014; ECHA, 2008) for newly produced nanomaterials (NMs) and nanoforms (NFs). REACH established the requirements to identify and characterize sets of similar NFs by providing clear evidence of certain physicochemical (PC) properties, i.e. the size, shape, surface chemistry and surface area of particles of the NFs (ECHA, 2019a; ECHA, 2019b). The identification of PC properties affecting the hazard potential of NFs is crucial for similarity assessment, even so as NFs and NMs are characterized not only by many material-specific intrinsic properties, but also by extrinsic properties that vary in dependence of the surrounding medium and can be expressed using dose-response data. Recent literature suggests that relevant nanospecific PC properties to support grouping according to their (eco)toxicological effects, are factors such as the shape and the surface (surface chemistry or reactivity), whereas hazard classes are identified on the basis of bio-persistence, morphology, reactivity, and solubility (Jeliazkova et al., 2021- this issue; Hund-Rinke et al., 2018). Moreover, NFs properties may change during their lifetime, for example due to aging, agglomeration or aggregation, corona formation, reactivity or dissolution (EU US Roadmap Nanoinformatics 2030, 2018), potentially influencing NFs toxicity, uptake or fate.

Several frameworks for grouping NFs are already available based on the above mentioned assumption that NFs with similar PC and toxicological profiles can be considered as a group (Knudsen et al., 2019; Oomen et al., 2018; Arts et al., 2015). Some of them employ statistical approaches, such as clustering and regression analyses (Knudsen et al., 2019; Drew et al., 2017), whereas others developed specific grouping rules based on fixed boundaries and pre-defined categories (Arts et al., 2015). Unsupervised learning algorithms, such as principle component analysis, were suggested for grouping NFs (Aschberger et al., 2019), but

* Corresponding author.
E-mail address: gtsiliki@athenarc.gr (G. Tsiliki).

https://doi.org/10.1016/j.impact.2022.100389
Received 29 September 2021; Received in revised form 29 December 2021; Accepted 2 February 2022
Available online 5 February 2022
2452-0748/© 2022 Elsevier B.V. All rights reserved.

also non-parametric supervised approaches, such as random forest classification algorithms, were employed to estimate sets of similar NFs depending on the available data and the hazard endpoints of interest (Bahl et al., 2020; Lamon et al., 2019). The European Commission funded Horizon 2020 GRACIOUS project has proposed a range of methods to assess similarity between NFs in terms of predefined grouping hypotheses that specify intrinsic/extrinsic PC characteristics and the toxicity data required via application of Integrated Approaches to Testing and Assessment (IATAs) (Stone et al., 2020). The computational similarity assessment methods suggested in the project are either via a multidimensional analysis or pairwise analyses conducted in a property-by-property manner (Jeliazkova et al. (2021)- this issue). Both approaches have their merits, yet they differ in that, unlike pairwise analysis, multidimensional analysis is directly estimating relationships between groups of NFs. On the other hand, pairwise analysis is a powerful tool when it comes to comparisons to benchmark materials which are largely used to read-across a toxicity endpoint but also assess similarity between well studied NFs to reduce the uncertainty of health risks from new materials (Hund-Rinke et al., 2018). These approaches are also useful for safety assessment of well-established materials such as fillers and pigments (Bahl et al., 2020; Wohlleben et al., 2017).

In this work we suggest estimating similarities between NFs by employing a pairwise similarity approach to dose-response data. This is currently an ongoing research area of particular interest for grouping NFs given its fundamental importance to hazard assessment. We model NFs dose-response curves using statistical distribution functions, and subsequently decide whether pairs of NFs are identically distributed samples from the same statistical distribution or alternatively are derived from different distributions of the same family but with different parameters. Several methods are proposed in the literature for model fitting dose-response data (Ma et al., 2020; Pinheiro et al., 2014). Established methods use mainly nonlinear regression models to describe the curve and rely on likelihood ratio tests for the significance of more complex models relative to the simpler models. Following a similar logic for model comparison, we are suggesting that for each NF, data are derived from a log-normal distribution and Bayesian Factor (BF) analysis is employed to compare dose-response curves of different NFs. BF provides an index of preference for the similarity of a pair of NFs over their dissimilarity.

2. Method

A nested sampling-based Bayesian approach is suggested to infer the parameters and compare dose-response curves for pairs of NFs. A similar approach is described in Baranyi et al. (1993) for comparing bacterial growth curves across different times. According to the literature, the biological dose-response data have a sigmoidal structure and are normally modelled using (non)linear and logistic regression models depending on whether response data are continuous or categorical. Toxicological dose-response modelling is primarily aiming to derive a Benchmark dose (BMD) in human or ecotoxicological risk assessment. The BMD method by Crump (1984), is the most widely applied methodology, however other methods, such as the PROAST software methodology (Slob, 2002), are also popular. Both tools are employing the same set of models ranging from normal to Weibull, and Exponential and Hill family models. Other model-based statistical approaches have been applied to dose-response data, such as Bayesian modelling, however mainly for the purposes of drug-interaction analysis (Hamza et al., 2021; Hennessey et al., 2010). Here, data are assumed to follow a log-normal distribution in a similar fashion as introduced in (Greco et al., 1995; Greco et al., 1990). Uniform prior knowledge is incorporated in the model reflecting a lack of prior information for all parameters, however, there is an option to include prior information from clusters of previously analysed experimental data of similar materials.

Bayesian analysis was preferred to maximum likelihood comparisons and other optimisation techniques to avoid overfitting due to inadequate representation of measurement errors in the data, and more sharply defined parameters than is justifiable given the data (Pullen and Morris, 2014). Using the Bayesian framework, we may capture the full uncertainty of the problem, taking the whole set of parameters space into account to make consistent predictions and make use of the readily available tools for model and hypothesis comparisons, i.e. the BF analysis.

We consider three different models to describe the similarity between a pair of curves. In the first model (M_1) the two curves are considered to be replicates coming from the same log-normal distribution with the same set of parameters (population mean and standard deviation), in the second model (M_2) the two curves have the same curvature but differ in all other parameters, and in the third model (M_3) the curves share no common parameters. The individual models are then compared using BF analysis and the results are interpreted using Jeffreys' scale (Kass and Raftery, 1995). BF_{12} shown in Eq. (1), demonstrates the BF score for comparing model M_1 over model M_2, where high values (above 3 or 0.5 in the log10 scale) are in favour of M_1 against M_2 and low values (less than 1/3 or − 0.5 in the log10 scale) are in favour of M_2 against M_1. For M_1, the joined likelihood function of the combined data from both NF curves is calculated, denoted by $P(D|M_1, I)$, whereas in M_2 the two curves are considered individually and the likelihood function is equal to the product of their two log-normal distributions, denoted by $P(D|M_2, I)$. We denote by D the dose-response data of a pair of NFs, and by I the non-informative prior information currently included in the model. Using similar notation, BF_{13} denotes the BF score for comparing model M_1 over model M_3, i.e. comparing the model of the combined data from both curves, to that of the two curves from two log-normal distributions with different parameters. For the analysis presented in this paper a uniform prior probability is assumed, however Gaussian or Cauchy options could be considered to include existing information from other PC data sets.

M1. curves are identical samples of the same distribution.

M2. curves have the same curvature parameter.

M3. curves are samples from different distributions

$$BF_{12} = \frac{P(D|M_1, I)}{P(D|M_2, I)} = \frac{P(M_1|D)}{P(M_2|D)} \cdot \frac{\mathrm{P(M_1, I)}}{\mathrm{P(M_2, I)}} \approx \frac{\mathrm{P(M_1|D)}}{\mathrm{P(M_2|D)}} \quad (1)$$

Since we assume uniform prior probability $P(M_1, I) = P(M_2, I)$, BF_{12} is equal to the ratio of the log-normal posterior probability of model M_1 over model M_2, suggesting by how much data D should update our belief in model M_1 over the competing M_2, tending to accept much simpler models.

The input of the method is a data matrix that includes the two dose-response intrinsic descriptors needed to calculate dose-response curves. Specifically, the data matrix is of the same structure as the one described by Jeliazkova et al. (2021) with a (n x 3) dimension where n is the number of rows corresponding to the number of NFs included in the analysis and 3 is the number of columns, with the first being the 'Names' of NFs, the second one the 'Concentration' descriptor data and the third the 'Response' descriptor data. Since the method can be applied to different purposes risk assessment dose-response similarity analysis, and in order to automate the procedure, the method is expecting as input a data matrix with a header row using the keywords 'Names', 'Concentration','Response'. Data standardization and normalization issues were addressed, to assure that transformed data follow a log-normal distribution. Well-defined benchmark materials (either negative or positive controls) are naturally included in the data matrix and can be used to set threshold values of the similarity score. For some methods or assays the control materials are Representative Test Materials (RTMs) that represent certain biological behaviour, and thus define a biologically relevant range within which materials can be considered similar.

Various alternatives have been considered for adjusting threshold values taking also into account the benchmark NFs similarity score

values to set the biological relevant range of the similarity score. Negative values of the similarity BF score support the belief that the two NFs are derived from different probability distributions and thus cannot be grouped together. Nevertheless, to adjust for the biologically relevant information in the data, the above-mentioned thresholds were informed from similarity values calculated for pairs of positive and negative controls in the data matrix, or when available by RTMs. Specifically, the direction of the similarity relationship agrees with the one mentioned above, positive (similar), negative (dissimilar), and the specific threshold is data-dependent and is set by the most similar RTMs and the most dissimilar RTMs in the data, as indicated by the user. For instance, in the case of the reactivity assay use case presented in the next section, concentration dependent reactivity curves of case study materials were compared with very reactive Mn2O3 and CuO, and non-reactive BaSO4 (Ag Seleci et al., 2022; this issue). Positive similarity score values greater than 0.5 in the log-scale indicate similarity, the similarity score for Mn2O3 and CuO is equal to 2.92 very close to the highest score of 3 when an NF is compared to itself.

The method was applied to single descriptors data by resampling the concentration values from a uniform distribution to generate dose-response curves, and was compared to other similarity assessment methods (Jeliazkova et al., 2021 – this issue). Another application of the method is presented in Ag Seleci et al. 2022 - this issue. The particular application considered a weighted distance metric for the BF score to

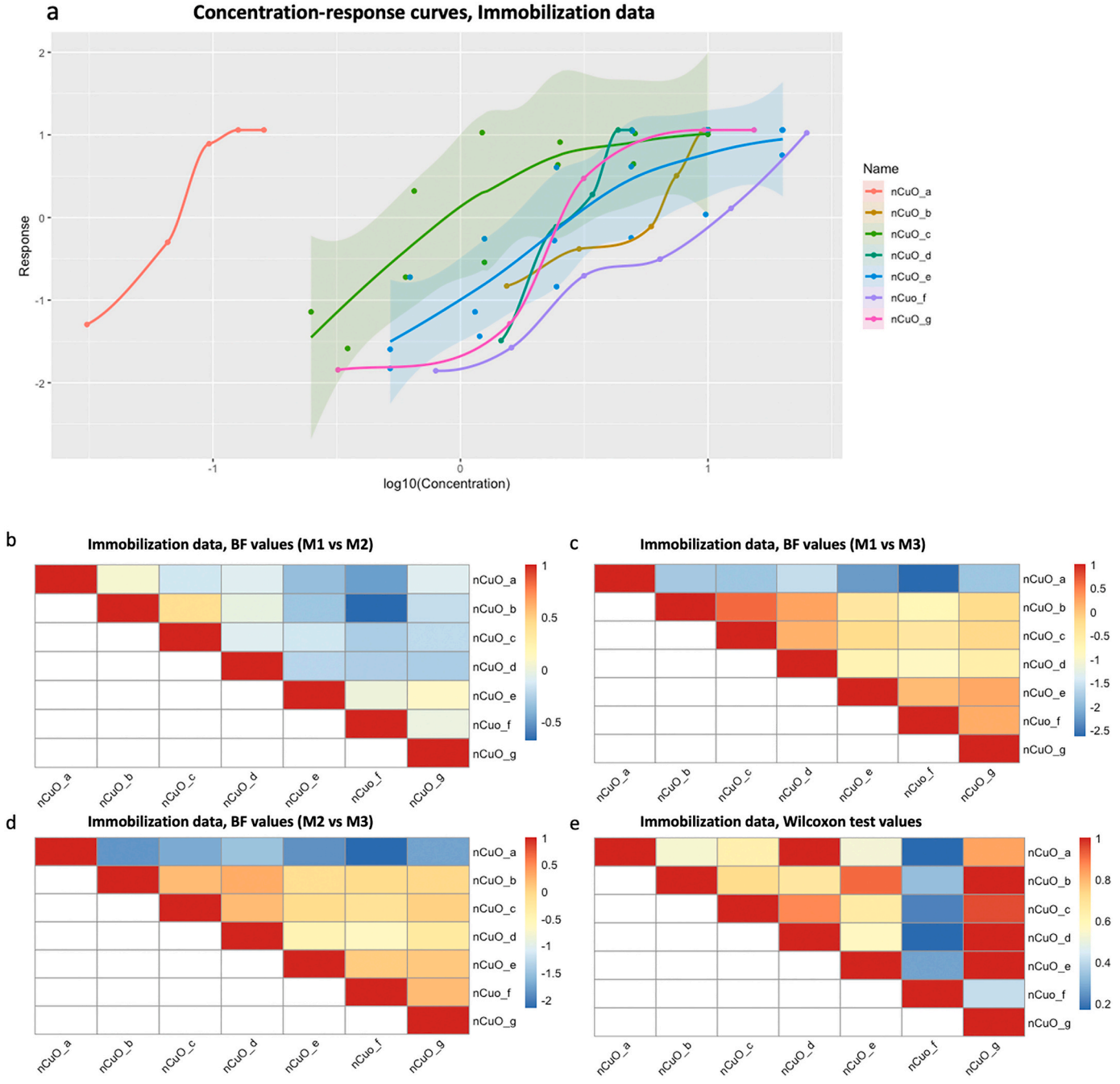

Fig. 1. Similarity analysis for *Daphnia magna* immobilization exposed to seven CuO NFs. a) The raw dose-response data for all seven NFs are shown, together with fitted loess curves. b) M_1 is tested against M_2, c) M_1 is tested against M_3, d) M_2 is tested against M_3. e) *p*-values from the Wilcoxon Mann-Whitney test for comparing the null hypothesis that data are derived from the same distribution against the alternative that they are not. In all graphs, red rectangles correspond to highly similar NFs (log10-scaled BF values). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

combine dose-response curves' similarities and their projections in the one-dimensional space, to reflect the importance of both the dose and response value ranges to characterize two NFs as similar.

All computations and visualizations were performed using the statistical programming language R. A nested sampling algorithm is used to implement BF model comparisons and to generate samples from the posterior distribution (Skilling, 2006). The R source code for analysis, parameter estimation and models comparisons, as well as the use case data are available in github (https://github.com/h2020gracious/BFsimilarity).

3. Results

Following testing of the method with simulated data, the suggested methodology was applied to a literature curated data set for *Daphnia magna* immobilization exposed to seven different NFs of copper oxides. The immobilization dose-response data is a literature reviewed data set, now available from the eNanoMapper database, which also includes sixteen intrinsic properties for each of the seven NFs considered (Peinjnenburg, 2021; Zabeo et al., 2021- this issue). Fig. 1 presents all possible pairwise comparisons between the seven CuO NFs also testing all three available models M_1, M_2, M_3 to investigate which one of the three possible comparisons can best capture the grouping of the data (M_1 vs M_2, M_1 vs M_3, or M_2 vs M_3). As a comparison to our methodology, we applied a non-parametric testing approach using the Wilcoxon Mann-Whitney test for comparing the null hypothesis that data are derived from the same distribution against the alternative that pairs of NFs are derived from different statistical distributions. Fig. 1a shows the raw dose-response data for the seven copper oxides, where we can observe differences in the concentration ranges (for instance, nCuO_a and the remaining NFs) as well as differences in the dose-response curves (for instance, nCuO_d and nCuO_e). Fig. 1b shows a.

heatmap of the similarity scores for all possible pairwise comparisons of the seven NFs; each rectangle in the graph corresponds to a BF value, BF_{12}, testing whether NFs are highly similar derived from the same statistical distribution against the alternative model that they are derived from distributions with the same curvature parameters but different location parameters. Red rectangles correspond to highly similar NFs (derived from the same distribution), and blue rectangles to dissimilar NFs. Rectangles in the diagonal are coloured in red showing the highest similarity score when comparing each NF with itself. Negative values in this case would mean that the two NFs are derived from distributions with the same curvature parameters. Fig. 1c shows the BF_{13} similarity score values which quantify whether pairs of NFs are derived from different distributions and negative values indicate dissimilarity between NFs. Fig. 1d shows BF_{23} similarity score values which quantifies how dissimilar NFs are; positive values suggest that NFs are derived from distributions with the same curvature and negative values are again in favour of NFs being derived from different distributions. Depending on the models compared we can understand the different degree of similarity between CuO NFs. It is important to note that across all three model comparisons (Fig. 1b - Fig. 1d) nCuO_a is quite dissimilar to the remaining six CuO NFs which is also supported by the raw data (Fig. 1a). Fig. 1e shows the *p*-values of the Wilcoxon Mann-Whitney test for comparing the null hypothesis that data are derived from the same distribution against the alternative that pairs of NFs are derived from different statistical distributions. Red rectangles in the graph correspond to high similarity (p-value = 1 for comparisons of each NF with itself along the diagonal, suggesting that we cannot reject the null hypothesis that NFs are derived from the same distribution and thus the two NFs are highly similar), and blue rectangles to low similarity (dark blue rectangles corresponding to p-values$\leq$0.01, suggesting that the two NFs are derived from different distributions with confidence 0.99). Overall, Fig. 1c, d agree in which pairs of NFs are dissimilar, also suggesting some similar pairs of NFs (e.g. nCuO_c and nCuO_b, nCuO_g and nCuO_f, or nCuO_d and nCuO_b). In Fig. 1b, the similarity scores are small values around zero, suggesting that this comparison is very sensitive even to small deviations across curves. As only CuO NFs are examined here, we conclude that BF_{13} values (Fig. 1c) are more sensitive to minor differences between NFs and for that reason are preferred compared to other model comparisons.

When comparing the BF_{13} values (Fig. 1c) with the Wilcoxon Mann-Whitney test *p*-values (Fig. 1d), we can see an agreement in only a few cases (e.g. nCuO_a and nCuO_f are declared as dissimilar from both methods) and it seems to suggest some degree of similarity for all NFs apart from nCuO_f and nCuO_d, nCuO_f and nCuO_a (p-values $\leq$ 0.01).

In a second use case, similarity analysis was performed for sixteen NFs tested within the GRACIOUS project to test concentration dependent reactivity similarities via the DCFH2-DA abiotic assay. RTMs (CuO, Mn2O3, BaSO4, CeO2 and ZnO), iron oxide NFs, Diketopyrrolopyrroles (DPP)-based organic pigments and silica NFs were commonly assessed. Surface reactivity is very often recognized as a central parameter when grouping NFs (Arts et al., 2015; Oomen et al., 2015), since several factors may modulate the surface reactivity and thus account for differences in the toxicological potency between different NFs of a substance. Fig. 2a, shows the raw concentration-response data for the sixteen NFs, where we can clearly observe that the highly reactive materials CuO and Mn_2O_3 are forming almost identical curves, with their reactivity values being the highest of the sixteen shown. The low reactive NFs (ZnO, CeO_2, $BaSO_4$) also form a group of very similar curves, however CeO_2 and $BaSO_4$ seem to be closer in terms of their reactivity values. Fig. 2b shows a heatmap of the similarity BF score values in the DCFH reactivity data set including RTMs. Each rectangle in the graph corresponds to a BF value, BF_{13}, quantifying similarities between pairs of NFs. For instance, the blue.

rectangle at the right top corner of Fig. 2a, corresponds to the BF_{13} value comparing CuO and Silica-anis-Al NFs and suggests that the two NFs are dissimilar. Dissimilarities are indicated with light yellow to blue colouring. Red rectangles in the graph correspond to areas of high similarity, grouping the low reactive NFs (ZnO, CeO_2, $BaSO_4$), and the highly reactive NFs (CuO, Mn_2O_3). Additionally, we can observe the formation of some groups for NFs of the same material (e.g. iron oxides nano A and nano B, DPP coated organic pigments, and silica NFs). For grouping purposes, we can consider as thresholds of high similarity values corresponding to similar in terms of their reactivity behaviour RTMs (e.g. CeO_2 and $BaSO_4$, or CuO and Mn_2O_3) and as thresholds of highly dissimilarity values that correespond to RTMs with different reactivity behaviour (e.g. $BaSO_4$ and CuO, or $BaSO_4$ and Mn_2O_3). These results align with what is presented for the same set of NFs in Ag Seleci et al., 2022 - this issue, and compared to similarity score values across abiotic and cellular assays. The authors applied the BF similarity assessment method and were able to group RTMs according to their known reactivity behaviour but also provide valuable feedback for the grouping of all NFs per assay.

As previously, we compare our findings with those from the Wilcoxon Mann-Whitney test; Fig. 2c shows the *p*-values of the Wilcoxon Mann-Whitney test for comparing the null hypothesis that data are derived from the same distribution (red rectangles) against the alternative that pairs of NFs are derived from different statistical distributions (blue rectangles). Overall, we can observe that the test fails to distinguish patterns in the data and largely rejects the null hypothesis of similarity (p-values $\leq$ 0.01). Both methods agree that CuO and Mn_2O_3 are not similar to the remaining of the NFs assessed, as well as that low reactive NFs (ZnO, CeO_2, $BaSO_4$) are similar, however the similarity patterns between the highly reactive materials are not evident in Fig. 2b.

The robustness of the method has been tested to various data sets from the GRACIOUS project (Jeliazkova et al., 2021; Ag Seleci et al., 2022; Di Cristo et al., 2021)- this issue, as well as compared to alternatives, namely apart from the Wilcoxon Mann-Whitney test shown here, to the Euclidean distance metric and to an aggregated ordered weighted average (OWA) distance metric (Jeliazkova et al., 2021; Zabeo et al., 2021) - this issue, as well as to the x-fold approach implemented in

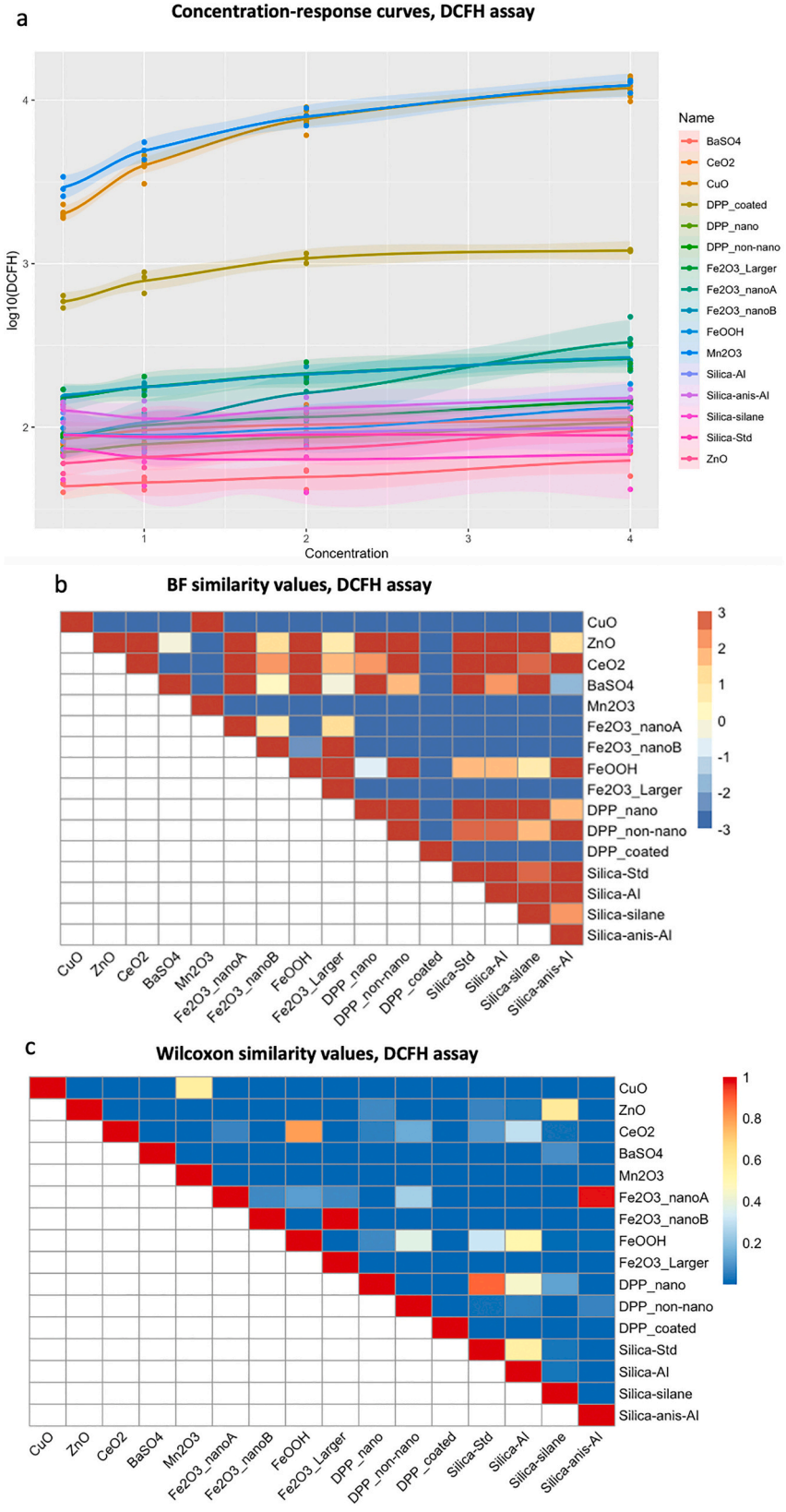

Fig. 2. Similarity analysis for the DCFH assay reactivity-concentration data for sixteen NFs carefully selected by the GRACIOUS project. a) The raw concentration-response DCFH reactivity data for all 16 NFs considered. b) BF similarity score values testing whether each pair of NFs is drawn from the same distribution against the alternative that data are log-normally distributed with different parameters. c) *p*-values from the Wilcoxon Mann-Whitney test for comparing the null hypothesis that data are derived from the same distribution against the alternative that they are derived from different distributions. Red rectangles correspond to highly similar NFs (log10-scaled BF values). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ECETOC NanoApp (Janer et al., 2021). In Jeliazkova et al. (2021)- this issue, a simplified version of the BF analysis was applied to a set of predefined PC and toxicological properties of interest to estimate pairwise similarities between NFs for benchmark materials. In all cases, we found that the BF method was robust across data sets and could group NFs relatively well compared to alternatives.

4. Discussion

This paper presents a methodology that assesses similarity between NFs and therefore can be used to support NFs grouping. We developed a tool to create and justify groups of similar and dissimilar NFs by assessing whether each NF is sufficiently similar to all other NFs. The suggested approach is considering all pairwise BF similarity values to compare pairs of curves of dose-response data and decide whether they can be described by the same underlying distribution. For each pairwise comparison of two dose-response curves, three models of similarity are considered and it was found that the one which assesses if NFs are highly similar as opposed to highly dissimilar is more sensitive to small differences between NFs and therefore was preferred. Although to fully define how NF similarity thresholds can be uniformly used for grouping decisions is beyond the scope of the current study, this work provides useful insights on the interpretation of the similarity scores using RTMs, an issue which is highly critical to the final decision of grouping and the uncertainty related to health risks of innovative new materials. Particularly, BF score values in the range of (−0.5, 0.5) are not informative, a value outside the aforementioned interval of values would suggest similarity (positive side) or dissimilarity (negative side). Pairwise comparisons to benchmark materials when available should be used to comparatively group NFs in terms of their biological relevance and set experiment specific target values of similarity (positive controls) and dissimilarity (negative controls).

Our findings show that the method can identify similar NFs significantly well, suggesting that similarity assessment could be used to verify and strengthen grouping hypotheses as well as the link between nanoform properties and the biological effect considered. The suggested tool could provide a rapid screening result to support grouping and therefore predict hazard in the absence of experimental data. Currently, a uniform prior probability is assumed, however future work will consider utilizing intrinsic data to inform the model of similarities across toxicity data. Additionally, we intend to focus on filtering the data depending on the specificity and complexity of the grouping hypothesis under consideration.

CRediT authorship contribution statement

Georgia Tsiliki: Conceptualization, Data curation, Methodology, Formal analysis, Software, Writing - original draft. **Didem Ag Seleci:** Writing - original draft, Validation, Investigation. **Alex Zabeo:** Writing – original draft, Validation. **Gianpietro Basei:** Writing – original draft, Validation. **Danail Hristozov:** Writing – original draft, Validation. **Nina Jeliazkova:** Writing – original draft, Validation. **Matthew Boyles:** Validation, Investigation. **Fiona Murphy:** Validation, Investigation. **Willie Peijnenburg:** Validation, Investigation. **Wendel Wohlleben:** Validation. **Vicki Stone:** Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the GRACIOUS EU H2020 project (grant agreement No 760840). We acknowledge comments and discussion by the entire similarity team of the GRACIOUS consortium.

References

Ag Seleci, D., Tsiliki, G., Okpowe, O., Elam, D., Werle, K., Seidel, K., Bi, X., Westerhoff, P., Innes, E., Boyles, M., Miller, M., Giusti, A., Murphy, F., Haase, A., Stone, V., Wohlleben, W., 2022. Determining nanomaterial similarity via assessment of surface reactivity by abiotic and in vitro assays. NanoImpact, 100390. In this issue.

Arts, J.H., Hadi, M., Irfan, M.A., Keene, A.M., Kreiling, R., Lyon, D., Maier, M., Michel, K., Petry, T., Sauer, U.G., Warheit, D., 2015. A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). Regul. Toxicol. Pharmacol. 71 (2), S1–S27.

Aschberger, K., Asturiol, D., Lamon, L., Richarz, A., Gerloff, K., Worth, A., 2019. Grouping of multi-walled carbon nanotubes to read-across genotoxicity: a case study to evaluate the applicability of regulatory guidance. Comp. Toxicol. 9, 22–35.

Bahl, A., Hellack, B., Wiemann, M., Giusti, A., Werle, K., Haase, A., Wohlleben, W., 2020. Nanomaterial categorization by surface reactivity: a case study comparing 35 materials with four different test methods. NanoImpact 19, 100234.

Baranyi, J., McClure, P.J., Sutherland, J.P., Roberts, T.A., 1993. Modeling bacterial growth responses. J. Ind. Microbiol. Biotechnol. 12 (3–5), 190–194.

Crump, K.S., 1984. A new method for determining allowable daily intakes. Toxicol. Sci. 4 (5), 854–871.

Di Cristo, L., Ude, V.C., Tsiliki, G., Tatulli, G., Romaldini, A., Murphy, F., Wohlleben, W., Oomen, A., Pompa, P.P., Arts, J., Stone, V., Sabella, S., 2021. Integrated Approach to Testing and Assessment Ingested Silica Nanomaterials Showing Gradual Dissolution in the Oro-Gastrointestinal Tract: Grouping Decisional Criteria to Streamline Nanomaterial Risk Assessment (Submitted to NanoImpact – this issue).

Drew, N.M., Kuempel, E.D., Pei, Y., Yang, F., 2017. A quantitative framework to group nanoscale and microscale particles by hazard potency to derive occupational exposure limits: proof of concept evaluation. Regul. Toxicol. Pharmacol. 89, 253–267.

ECHA, 2008. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.6: QSARs and Grouping of Chemicals.

ECHA, 2019a. Appendix for Nanoforms Applicable to the Guidance on Registration and Substance Identification.

ECHA, 2019b. Appendix R.6-1 for Nanoforms Applicable to the Guidance on QSARs and Grouping of Chemicals.

Greco, W.R., Park, H.S., Rustum, Y.M., 1990. Application of a new approach for the quantitation of drug synergism to the combination of Cis-Diamminedichloroplatinum and 1-Beta-D arabinofuranosylcytosine. Cancer Res. 50 (17), 5318–5327. http://www.ncbi.nlm.nih.gov/pubmed/2386940.

Greco, W.R., Bravo, G., Parsons, J.C., 1995. The search for synergy: a critical review from a response surface perspective. Pharmacol. Rev. 47, 331–385.

Hamza, T., Cipriani, A., Furukawa, T.A., Egger, M., Orsini, N., Salanti, G., 2021. A Bayesian dose–response meta-analysis model: a simulations study and application. Stat. Methods Med. Res. 30 (5), 1358–1372. https://doi.org/10.1177/096228022098264.

Hennessey, V.G., Rosner, G.L., Bast Jr., R.C., Chen, M.Y., 2010. A bayesian approach to dose–response assessment and synergy and its application to in vitro dose–response studies. Biometrics 66 (4), 1275–1283.

Hund-Rinke, Kerstin, Schlich, Karsten, Kühnel, Dana, Hellack, Bryan, Kaminski, Heinz, Nickel, Carmen, 2018. Grouping concept for metal and metal oxide nanomaterials with regard to their Ecotoxicological effects on algae, Daphnids and fish embryos. NanoImpact 9. https://doi.org/10.1016/j.impact.2017.10.003.

Janer, G., Landsiedel, R., Wohlleben, W., 2021. Rationale and decision rules behind the ECETOC NanoApp to support registration of sets of similar nanoforms within REACH. Nanotoxicology 15 (2), 145–166. https://doi.org/10.1080/17435390.2020.1842933.

Jeliazkova, N., Bleeker, E., Cross, R., Haase, A., Janer, G., Peijnenburg, W., Pink, M., Rauscher, H., Svendsen, C., Tsiliki, G., Zabeo, A., Hristozov, D., Stone, V., Wohlleben, W., 2021 Nov 20. How can we justify grouping of nanoforms for hazard assessment? Concepts and tools to quantify similarity. NanoImpact 25, 100366.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795.

Knudsen, K.B., Berthing, T., Jackson, P., Poulsen, S.S., Mortensen, A., Jacobsen, N.R., Skaug, V., Szarek, J., Hougaard, K.S., Wolff, H., Wallin, H., 2019. Physicochemical predictors of multi-walled carbon nanotube–induced pulmonary histopathology and toxicity one year after pulmonary deposition of 11 different multi-walled carbon nanotubes in mice. Basic Clin. Pharmacol. Toxicol. 124 (2), 211–227.

Lamon, L., Aschberger, K., Asturiol, D., Richarz, A., Worth, A., 2019. Grouping of nanomaterials to read-across hazard endpoints: a review. Nanotoxicology 13 (1), 100–118. https://doi.org/10.1080/17435390.2018.1506060.

Ma, J., Bair, E., Motsinger-Reif, A., 2020. Nonlinear dose–response modeling of high-throughput screening data using an evolutionary algorithm. Dose-Response 18 (2) p.1559325820926734.

OECD, 2014. Guidance on grouping of chemicals. In: Series on Testing & Assessment no 194. second edition.

Oomen, A.G., Bleeker, E.A.J., Bos, P.M.J., van Broekhuizen, F., Gottardo, S., Groenewold, M., Hristozov, D., Hund-Rinke, K., Irfan, M.A., Marcomini, A., et al., 2015. Grouping and read-across approaches for risk assessment of nanomaterials. Int. J. Environ. Res. Public Health 12 (10), 13415–13434.

Oomen, A.G., Steinhäuser, K.G., Bleeker, E.A., van Broekhuizen, F., Sips, A., Dekkers, S., Wijnhoven, S.W., Sayre, P.G., 2018. Risk assessment frameworks for nanomaterials: scope, link to regulations, applicability, and outline for future directions in view of needed increase in efficiency. NanoImpact 9, 1–13.

Peinjnenburg, W., March 8, 2021. personal communication.
Pinheiro, J., Bornkamp, B., Glimm, E., Bretz, F., 2014. Model-based dose finding under model uncertainty using general parametric models. Stat. Med. 33 (10), 1646–1661.
Pullen, N., Morris, R.J., 2014. Bayesian model comparison and parameter inference in systems biology using nested sampling. PLoS One 9 (2), e88419.
Skilling, J., 2006. Nested sampling for general Bayesian computation. Bayesian Anal. 1 (4), 833–859.
Slob, Wout, 2002. Dose-response modeling of continuous endpoints. Toxicol. Sci. 66 (2) https://doi.org/10.1093/toxsci/66.2.298.
Stone, V., Gottardo, S., Bleeker, E.A., Braakhuis, H., Dekkers, S., Fernandes, T., Haase, A., Hunt, N., Hristozov, D., Jantunen, P., Jeliazkova, N., 2020. A framework for grouping and read-across of nanomaterials-supporting innovation and risk assessment. Nano Today 35, 100941.
Wohlleben, W., Punckt, C., Aghassi-Hagmann, J., Siebers, F., Menzel, F., Esken, D., Drexel, C.-P., Zoz, H., Benz, H.U., Weier, A., et al., 2017. Nanoenabled products: categories, manufacture, and applications. In: Metrology and Standardization of Nanotechnology, pp. 409–464.
Zabeo, A., Basei, G., Tsiliki, G., Peinjnenburg, W., Hristozov, D., 2021. OWA Based grouping of nanomaterials with Arsinh and Dose Response similarity models. NanoImpact 100370, 2021 Nov 26.