Determining Nanoform Similarity via Assessment of Surface Reactivity by Abiotic and In Vitro Assays

Didem Ag Seleci[1], Georgia Tsiliki[2], Kai Werle[1], Derek A. Elam[1], Omena Okpowe[1], Karsten Seidel[1], Xiangyu Bi[1,3], Paul Westerhoff[3], Emma Innes[4], Matthew Boyles[4], Mark Miller[5], Anna Giusti[6], Fiona Murphy[7], Andrea Haase[6], Vicki Stone[7], Wendel Wohlleben[1]

[1]Advanced Materials Research, Dept. of Material Physics and Analytics and Dept. of Experimental Toxicology and Ecology, BASF SE, Ludwigshafen, Germany.

[2]Athena Research Center, Athens, Greece.

[3] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, USA.

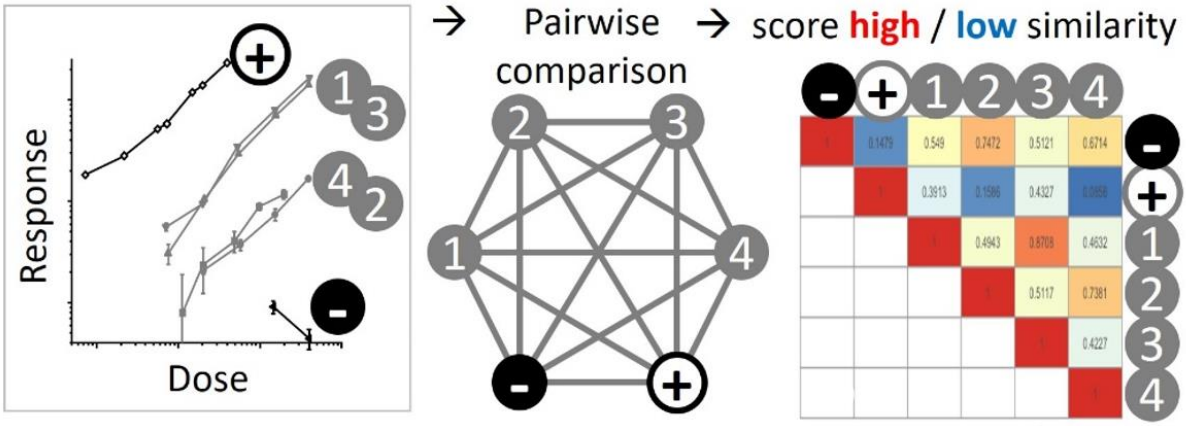
[4] Institute of Occupational Medicine, Edinburgh, United Kingdom.

[5] Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom.

[6] German Federal Institute for Risk Assessment, Department of Chemical and Product Safety, Berlin, Germany.

[7] NanoSafety Group, Heriot-Watt University, Edinburgh, United Kingdom.

TOC figure



Abstract

Grouping of is a method used to streamline hazard and risk assessment. Scientifically evidenced formation of groups provides evidence of similarity. This work reports on justification of grouping of nanoforms (NFs) via similarity of their surface reactivity. Here, four reactivity tests were used for detection of reactive oxygen species (ROS) generated by NFs. Concentration dependent reactivity was tested via the abiotic assays FRAS, EPR and DCFH2-DA, as well as the cellular *in vitro* assay, activation of NRF2/ARE Responsive Luciferase Reporter HEK293 Cell Line. Representative materials (CuO, Mn_2O_3, $BaSO_4$, CeO_2 and ZnO) and three case studies of each several NFs of iron oxides, Diketopyrrolopyrroles (DPP)-based organic pigments and silicas were assessed. A novel similarity assessment algorithm is applied to quantify similarities between pairs of NFs, in a four steps workflow assessing the full concentration-response curves, the individual concentration and response ranges, and finally the representative materials. We found this algorithm to be applicable to all abiotic and *in vitro* assays that were tested. Our findings showed that CuO and $BaSO_4$ were the most and least reactive representative materials respectively and clearly $BaSO_4$/CuO were not found to be similar for all reactivity assays as confirmed by their different NOAECs of *in vivo* studies. However, the similarity outcomes from different reactivity assays are not always in agreement, highlighting the need to generate data by one assay for all materials, comprising representative materials and the candidate group of NFs. Despite low similarity scores *in vitro* some pairs of case study NFs can be accepted as sufficiently similar because the *in vivo* NOAECs are similar for the specific pairs of NFs, highlighting the conservative assessment by the abiotic assays.

Keywords

Nanoforms, grouping, similarity, surface reactivity, dose-response, concentration-response

1. Introduction

Many substances are produced as nanomaterials, each of which can be generated in a variety of nanoforms (NFs) differing in size, crystallinity, shape or surface chemistry [1]. The wide range of nanoforms available provides the possibility to generate a high diversity of commercial nano-enabled products. The different physicochemical characteristics of NFs may influence their toxicological profiles. According to the REACH regulation [2], for each NF of a substance a set of minimum standard information has to be provided, and therefore the cost, duration, and effort of testing may hugely increase. However, this process can be streamlined through the use of alternative approaches such as grouping and read-across. The GRACIOUS Framework provides a logical and science evidenced approach to group similar NFs, allowing read-across of hazard information from source NFs (or non-NFs) with adequate hazard data to target NFs that lack such data [1]. The GRACIOUS Framework supports the user to generate a grouping hypothesis that encompasses the relevant physicochemical characteristics, route of exposure and hazard endpoints. Integrated Approaches to Testing and Assessment (IATAs) are then used to gather the existing information needed to test the grouping hypothesis, and to guide the generation of new data to fill data gaps. The IATAs consist of decision trees, with each decision node posing a question that allows identification of the most relevant information needed to test the grouping hypothesis. Each decision node is supported by a tiered testing strategy (e.g. [3]) to guide the gathering of evidence via the most appropriate, and if possible, standardised methods available. For human health studies, the tiered testing strategy includes simple *in vitro* models at tier 1, as well as a number of alternative more complex multi-cellular *in vitro* models at tier 2 and *in vivo* models at tier 3. The lower tier data provides the evidence to assess similarity of NF physicochemical characteristics (what they are), fate in the environment and toxicokinetics (where they go) and hazards (what they do – including surface reactivity). If sufficiently similar, the data can then be used to support read-across.

Application of grouping of NFs would therefore help to reduce the amount of experimental testing for hazard and risk assessment, thus reducing animal testing [1]. As indicated above, grouping requires methods to assess the similarity of different NFs. A summary of different methods is available in the white paper of this issue[4]. For the case studies investigated in this paper, various NFs were available for study, allowing an assessment of their similarity. As the white paper demonstrates, the use of scalar descriptors (a single value that represents a range of data or a concentration-response curve) is a convenient way to compare similarity. Here, we explore methods to assess similarity using the full concentration-response curve, in order to take into consideration that variations in the shape of the concentration-response curve can lead to loss of information when reduced to a scalar descriptor. To achieve this, reactivity concentration-response curves of the different NFs were evaluated via Bayes Factor (BF) calculations of pairwise (two NFs directly compared at a time) similarity assessments, in addition to assessing their similarity across the separate concentration (x-axis) and reactivity (y-axis)

scales separately. The white paper[4] demonstrates that for scalar descriptors, the BF algorithm was consistent with simpler approaches, but statistically was more robust, and especially well-suited for the comparison of two-dimensional data such as concentration-response curves. The method of assessing similarities between concentration-response curves via BF calculations is originally presented in this issue by Tsiliki et al. [5], however here we introduce a novel similarity assessment approach and integrate information from the similarity assessment between reactivity concentration-response curves and the comparisons of the varying ranges of the concentration and reactivity data available. This integrated information, which we denote by similarity score, quantifies how similar two NFs are and can then be compared to threshold values set by the representative materials to justify grouping.

In line with other scientific approaches and frameworks for grouping of NFs developed [6-8], the GRACIOUS IATA for inhalation[9-10] route of exposure identifies lung deposition, dissolution rate, in vitro inflammation, and surface reactivity as decisive properties to compare NFs. Surface reactivity is also a key parameter to compare the toxicity of NFs in the IATA on oral route of exposure [11]. Moreover, the ECHA guidance recommends justifying a grouping of NFs via similarity of their surface reactivity [2], and the same guidance advises to justify grouping decisions "mainly using physicochemical parameters and/or in vitro screening methods", consistent with tier 1 and tier 2 of the testing strategy of the "reactivity" decision node of the inhalation IATA.[9, 12] Assessment of similarity is key to decide whether different NFs can be included in a group. To achieve this well-defined algorithm to quantify the similarity between two (or more) NFs are required. Until now, there are no harmonized and standardized assays to assess either similarity or the reactivity of NFs. We considered the ROS generation as mode of action of toxicity of NFs; we chose several assays for assessing ROS generation (which could be integrated in a testing strategy) and tested the same NFs in the different assays. Then we developed a procedure to evaluate the similarity between NFs tested in the same assay.

Reactive oxygen species (ROS) cause oxidative stress and cytotoxicity, and can be generated by NFs. Understanding NF reactivity is a key stage towards understanding the toxicology, because the generation of ROS can trigger sub-lethal (e.g. inflammation) and lethal (e.g. apoptosis) effects [13]. NF based ROS production occurs via different mechanisms such as Fenton reaction (in the presence of divalent metal ions such as Fe2+), redox cycling and radical generation. Fenton-like reactions are reported to be the most common mechanism for metal NFs, leading to the generation of hydroxyl radicals [14]. Different assays are available to measure free radicals and ROS, however they all differ in the mechanism of detection, sensitivity and specificity. Since the exact mechanisms of ROS mediated effects of NFs are not well understood, several *in vitro* and abiotic reactivity assays are usually used [13, 15-16]. The ferric

reduction ability of serum (FRAS) assay utilizes human blood serum (HBS), to quantify the total antioxidant depletion induced by NFs as a measure of their oxidative potential. Moreover, this assay has shown potential to separate active from passive NFs [17], and to be specifically useful for grouping purposes because it can distinguish between different amounts of oxidative stress [18] Electron spin resonance (ESR) spectroscopy identifies qualitatively and quantitatively free radical species in abiotic and cellular environments [19]. The ESR spin-trapping technique uses chemical species called spin traps, which react with short-lived free radicals to form relatively stable adducts having a half-life long enough for ESR measurement [20]. Another commonly used assay assesses the oxidation of the non-fluorescent molecule 2'-7'-dichlorodihydrofluorescin diacetate (DCFH2-DA), into a fluorescent form in the presence of ROS [17]. DCFH2-DA was first devised to detect ROS in the absence of cells [21] and more recently it was suggested to be used as a tool to study cellular and abiotic ROS produced in response to nanomaterials [22]. There are also several cellular assays which can be used to measure the impacts of ROS on cells. For example, the nuclear factor erythroid 2-related factor 2 (Nrf2)/ antioxidant response element (ARE) pathway is an important cellular defence system that is activated by various stresses [23]. NRF2/ARE Responsive Luciferase Reporter HEK293 Cell Line can be used as an *in vitro* model for monitoring the activation of antioxidant response pathways triggered by treatment with NFs. The light induction in response to ROS and Nrf2 interaction with the ARE makes this an attractive model to study.

Here, reactivity of representative materials (CuO, Mn_2O_3, $BaSO_4$, CeO_2 and ZnO) plus three case studies (iron oxides, Diketopyrrolopyrroles (DPP)-based organic pigments and silica NFs) were tested via the abiotic assays FRAS, EPR and DCFH2-DA, as well as the cellular *in vitro* assay, activation of NRF2/ARE Responsive Luciferase Reporter HEK293 Cell Line.

In this work we address both the experimental data acquisition and the similarity evaluation. The similarity level of NFs in each reactivity assay and the consistency of similarity when compared with higher-tier (*in vivo*) results were evaluated.

2.Material and Methods

2.1 Materials

ZnO NM110 and CeO_2 NM212 were kindly provided by the NM repository at the Joint Research Center (JRC) repository [24]. $BaSO_4$ NM220 was provided by the NM repository at the

Fraunhofer Institute (IME). Mn_2O_3 was purchased from Skyspring Nanomaterials. CuO was purchased from Plasma Chem. Fe_2O_3 nanoA (small rods, about 15nm), Fe_2O_3 nanoB (rounded particles, about 30nm), Fe_2O_3 larger (irregular particles, above 100nm), FeOOH (small rods, about 15nm) and organic pigments (DPP nano (43 nm), DPP non-nano (233 nm), DPP premixed (230 nm)) were supplied by BASF Colors and Effects BASF Schweiz AG. Silica std (standard silica, 30% in water), Silica Al (Al substituted into the silica surface, 25% in water), Silica silane (silane modified, 28% in water), Silica anis Al (Al substituted into the surface, aggregated nanoparticles, 7% in water) and Silica-anis-std (aggregated silica nanoparticles, 12% in water) were supplied by NOURYON.

The physicochemical properties and TEM images of materials are reported separately (Jeliazkova et al. 2021 data base publication in preparation) and are completely reproduced here in the supplementary information (Table S1 and Figure S1). The reagents employed during each reactivity assay are reported in the supplementary information.

All NFs were tested in all reactivity assays.

2.2 Ferric Reduction Ability of Serum (FRAS)

The SOP, which described a multi-concentration protocol of the FRAS assay and was published in 2017 by Gandon et al. [25], was used for reactivity testing of samples.

Briefly, samples were incubated with human blood serum for 3 h at 37 °C. Before incubation, bath sonication for 1 min was applied to prevent the formation of large agglomerates and allow the reagents to access the whole surface area. NF were separated from HBS via ultracentrifugation (AUC-Beckman XL centrifuge (Brea, CA, USA) at 14,000 G for 150 min). Subsequently, a 100 µL of NF-free HBS supernatant was incubated in the FRAS reagent that contains the Fe^{3+} complex. The total antioxidant depletion, as a measure of the oxidative potential of NF, was determined by assessing the UV-vis spectrum of the iron complex solution. Trolox, a water-soluble analog of vitamin E, was used as an antioxidant to calibrate the FRAS results. Different Trolox concentrations (from 0.001 to 0.1 mg/mL) were tested by the FRAS assay to obtain FRAS absorption signals that were linearly fitted. Finally, the oxidative damage induced by NF was calculated in Trolox equivalent units (TEUs).

Background FRAS signal level is up to 5000 nmol TEU/L and saturation of FRAS signals occurred at the level of about 250,000 nmol TEU/ L, indicating that all antioxidants contained in the human serum are consumed during the incubation. 0.02 to 40 mg/mL concentration range was applied to the representative test materials and case study materials. This range is a two-

sided extension of the range of 0.15 to 10 mg/mL that was used in the extensive screening of 138 nanomaterials, each at an "adjusted" single concentration [26]. We find that CuO and Mn_2O_3 reach saturation of the assay for all concentrations above 0.2 mg/mL.

2.3 Electron Paramagnetic Resonance Spectroscopy (EPR)

EPR can be used to identify and quantify unpaired electron spins, e.g. to characterize the active sites of solid-state catalysts. Two methods have been established to assess the surface-induced reactivity of nanomaterials: Method 1 utilizes the nitrone spin trap 5,5-Dimethyl-1-pyrroline-N-oxide (DMPO), one of the most established spin traps for nanosafety purposes [27]. This method involves trapping reactive short-lived free radical intermediates (e.g. hydroxyl radical) via the creation of the spin adduct DMPO-OH with a characteristic 1:2:2:1 peak pattern and g-value. Method 2 employs the cyclic hydroxylamine spin probe 1-hydroxy-3-carboxy-pyrrolidine (CPH). CPH directly probes/interacts with short-lived reactive oxygen species (e.g. superoxide radical) on the material surface, forming the spin adduct CP· with characteristic 1:1:1 peak pattern and g-value. Both methods are standardized by ISO TS 18827(ISO 2017) and iuta SOP: EPR spectroscopy analysis using the spin probe CPH (by B. Hellack), respectively.

For EPR detection of superoxide anions, the spin trap 1-hydroxyl-2,2,6,6-tetramethyl-4-oxo-piperidine (Tempone-H) was prepared at 100 nM in 0.01 M EDTA, to be used at a final concentration of 1 mM. Test materials were prepared in a phosphate buffer at a starting concentration of 4 mg/mL, with a concentration-response measured between 0.002-4 mg/mL, dependent on material. Pyrogallol, at 32 mM, was used as a positive control. Measurements were taken 60 minutes after addition of Tempone-H, with samples maintained at 37°C during this time. Using a Miniscope MS 200 (Magnettech, Berlin, Germany), the EPR spectrum was obtained with the following parameters: microwave frequency, 9.3–9.55 Hz; microwave power, 20 mW; modulation frequency, 100 kHz; modulation amplitude, 1,500 mG; center field, 3,350 G; sweep width, 55 G; sweep time, 30 sec; number of passes, 1.

2.4 Dichlorodihydrofluorescin diacetate (DCFH2-DA)

Detection of ROS produced using the DCFH2-DA probe was conducted as follows. DCFH2-DA was chemically hydrolysed by incubation with 0.01 M NaOH, neutralized and diluted to form 10 µM DCFH2 in phosphate-buffered saline (PBS). During this reaction, test particles were prepared by suspension in phenol red-free minimum essential medium (MEM) with 2% FCS at a concentration of 40 mg/mL, followed by ultra-sonication in a water bath and serial

dilutions to obtain a range of 5, 10, 20 and 40 mg/mL. Each treatment was then added, in triplicate to a 96-well plate at a volume of 25 µl, followed by addition of 225 µl 10 µM DCFH2 to each well. Final concentrations of 0.5, 1, 2 and 4 mg/mL were obtained, which were incubated at 37°C for 90 minutes. After this time, samples were centrifuged at 3000 x g for 15 minutes, and 100 µl of each well was moved to a black 96-well plate to read fluorescence at ex/em wavelengths of 485/530 nm. To address the potential for interference of particles with the light detection, the same process as above was replicated using particles suspended in solutions of PBS alone (no DCFH2), or with 0.1 µM fluorescein diacetate (FDA). To account for background interference, signals generated with incubation in solutions of PBS alone were removed from signals generated in solutions of DCFH2.

2.5 *In vitro* assay: Nrf2-activation

The detailed SOP for the assessment of Nrf2-activation is published in Giusti et al [28]. The main points are briefly reported hereafter for more clarity.

Preparation of NF dispersion

NF stock suspensions were prepared according to NanoToxClass dispersion protocol, which consists in cup horn sonication of NF in serum free cell medium at a concentration of 0.5 mg/mL (Bandelin Cuphorn UW 2200 for 23 minutes at 100 % power). These suspensions where then diluted to the desired concentrations. For Nrf2 activation the dilution media is the complete cell medium but containing only 1 % FBS; the final NF concentrations are to 0, 1.2, 3.6, 10.7, 32.1 µg/mL.

Measure of Nrf2-activation

A stably transfected cell line encoding a firefly luciferase reporter gene under the control of ARE element (NRF2/ARE Luciferase Reporter HEK293 Stable Cell Line) from Signosis was used. Cells were grown in DMEM cell culture medium (w/o phenol red and L-glutamine, high glucose, PAN Biotech GmbH) supplemented with 584 mg/L L-Glutamin, 0.1 mg/mL Penicillin/Streptomycin, 110 mg/L Sodium Pyruvate, 80 µg/mL Hygromycin B gold and 10 % Fetal Bovine Serum (non-heat inactivated FBS Good from PAN Biotech). The cells were grown in T75 cell culture flasks in an incubator (37 °C, 5 % CO_2 and 90 % humidity) and sub-cultured regularly two times a week at ca. 70 % confluence. Cells were seeded in 96-well white plates Greiner Bio-One P/N 655098 24h before the treatment (10000 cells in 0.1 mL pro well). After 24h incubation at 37°C, 5% CO_2, the cell culture medium is carefully removed and replaced with 0.2 mL treatment cell medium containing the desired final NF concentration (i.e. 0, 0.7,

2.1, 6.3, 18.9 µg/cm2 corresponding to 0, 1.2, 3.6, 10.7, 32.1 µg/mL). Cells were treated for 48 hours.

2.6 Similarity analysis to compare concentration–response curves

Pairwise similarity analysis was performed in a 3-step manner employing three different criteria, namely assessing the similarities between shapes of reactivity concentration-response curves, similarities between the concentration factor ranges, and similarities between the reactivity factor ranges. The three criteria were quantified with a scalar metric each and aggregated to a unique value, denoted by similarity score, which takes values in the range between 0 and 1, with values close to 1 denoting high similarity. Lastly, representative test materials were used to set the biological relevant range of the assay. Similarities between concentration-response curves for each pair of NF were assessed using BF calculations which can be interpreted as indexes of preference for one model over another, suggesting by how much a data sample should update our belief in one model over a competing one [29]. Given a pair of concentration-response curves, two models are compared, specifically the first model (M_1) assumes the concentration-response curves of the two NFs are identical as opposed to the second model (M_2) which assumes that curves are coming from different distributions [5]. Positive B_{12} values suggest that M_1 is preferable compared to M_2, and the two NFs can be assumed to be similar given the data.

Similarities between ranges of the concentration-response factors, in this case concentration-reactivity data, were also quantified per factor using the Manhattan distance metric in both cases. This was found to be an important adjustment to the BF calculations in order to cope with large differences in the concentration ranges measured.

The final similarity score reported is a weighted average distance metric, which for each pair of NFs, combines the BF value with quantification of the distance between the ranges of the response reactivity values d_R and the distance between the ranges of their concentration d_D. The specific weights shown below were selected after a recurring adaptation procedure to adequately distinguish the active and passive NFs from the representative test materials. Other factors may be possible, but their validity need to be demonstrated on suitable representative test materials.

$$similarity\ score = (0.3 * BF) + (0.5 * d_R) + (0.2 * d_D)$$

3 Results and Discussion

3.1 Results description

Concentration-dependent abiotic and cell-based *in vitro* reactivity assays were applied on representative test materials and case study materials. The following sections first present the concentration-dependent reactivity in mass dose metrics (as required by the IATA to compare potency)[9], and then in surface dose metrics (to check for *qualitative* similarity of the specific reactivity).

3.1.1 Results in mass dose metrics

CuO and Mn_2O_3 induced a concentration dependent reactivity in all reactivity assays (Figure 1A-D). The FRAS, EPR (DMPO) and Nrf2 activation assays showed concentration dependent reactivity of ZnO NF (Figure 1A, D and S2 A). The FRAS assay distinguished very reactive materials (CuO, Mn_2O_3), from the intermediate reactive material ZnO, and from rather passive materials (Figure 1A). ZnO presented the highest reactivity in the Nrf2 activation assay (Figure 1 D). The DCFH2-DA assay and measurement of two spin traps (CPH and Tempone H) by EPR found only low reactivity of ZnO (Figure 1 B, C and S2A); this is not in keeping with the known *in vivo* toxicity of ZnO which induces inflammation (in rats) and ecotoxicity at low concentrations [11, 23]. However, the exact mechanism of these different cases is not discussed here. Only FRAS assay presented concentration dependent responses by CeO_2 (Figure 1A). $BaSO_4$ did not produce any concentration dependent responses in any assay (Figure 1 A-D).

All Fe-based materials induced a concentration dependent increase in reactivity according to FRAS assay and EPR (Tempone H) (Figure 2 A, B). However, Fe_2O_3_nanoA showed an unexplained decrease in reactivity at high concentrations (from 1.0 mg/mL) in EPR assay by using Tempone H spin trap. Except Fe_2O_3_Larger, other Fe-based NFs had concentration dependent reactivity in EPR (DMPO) assay (Figure S3 A). DCFH2-DAand Nrf2 activation assays did not generate any concentration dependent reactivity for Fe-based materials (Figure 2 C, D).

On pigments, the FRAS assay may have suffered from optical interference and flocculation of pigment samples in serum. Colorimetric assays, including the FRAS assay, are not recommended for NM with high absorption coefficient (i.e. where traces of NM generate a large interference with colorimetric readout). Adaptations of the assay for highly absorbing NM were reported recently on the example of graphenes, but were not applied here.[30] All DPP based pigments demonstrated slightly increased reactivity in concentration dependent manner as

assessed by the EPR (Tempone H) assay (Figure 3A), however no concentration response was observed by DCFH2-DA, Nrf2 activation and EPR (DMPO) assays (Figure 3 B, C and S4 A).

Due to low particle concentration of the pristine products Silica-anis-Al (7%) and Silica-anis-std (12%), the FRAS Assay could not applied on these NFs. Silica-std, Silica-Al and Silica-silane showed concentration dependent responses in the FRAS assay (Figure 4A). EPR (Tempone H) assay also generated concentration dependent reactivity for Silica-std, Silica-Al, Silica-silane and Silica-anis-std (Figure 4B). DCFH2-DA, Nrf2 activation and EPR (DMPO) assays did not demonstrate any concentration dependent responses for all silica- based NFs (Figure 4 C, D and S5 A).

3.1.2 Results in surface dose metrics

The reactivity of particles is in fact often refered to as "surface reactivity", because the reaction is thought to occur at the particle surface.[31] If two NFs induce different reactivity in mass-dose metrics, they may still induce similar reactivity after rescaling to surface-dose metrics.[32] This would then indicate a *qualitative* similarity of the reactivity. Examples of qualitative similarity were observed on the Fe-based materials, where all Fe_2O_3 NFs collapse onto one concentration-response curve in the surface-dose representation of FRAS reactivity (Figure S8A), whereas the chemically different FeOOH stands out (Figure S8A). Another example is given by the NF and the non-nano-form of DPP pigment, which collapse onto one concentration-response curve in the surface-dose representation of both EPR and DCFH reactivity (Figure S7A, S7B). In contrast, the different representative test materials maintain a very different concentration response also in surface-dose representation (Figure S6). It has been argued that surface dose allows the best understanding of inhalation toxicity.[33-34] Notwithstanding that systematic understanding -or even based on it- the scaling of effects with specific surface area is one of the reasons for regulators to demand a separate assessment of NFs of a substance. A justification of grouping several NFs with respect to their inhalation hazard must consider that even at the same surface-specific reactivity, particles with a higher specific surface area induce more oxidative damage. In line with the IATA,[9] the following section is devoted to quantitative similarity of mass-dose potency.

3.2 Results of quantitative similarity assessments

Previously, single reactivity descriptors were used for grouping purposes in the nanoGRAVUR framework [35]. For two reasons we are unable to do this here for each assay using concentration response curves: the range in effect was so vastly different between many of the particles, which

makes the EC50 comparison unrealistic, as the method provides a value based on the range restrictions of each individual material. Secondly, some of the treatments have such a low effect (negative for the $BaSO_4$) that a reliable EC50 is unattainable. Therefore, we use a similarity algorithm (BF calculations) for further discussion that facilitates the discussion based on the complete concentration dependency. This enables a direct comparison of the similarity assessment based on different assays.

3.2.1 Representative test materials

Pairwise similarity was analyzed by Bayes factor algorithm using reactivity data from four different assays. The pairwise comparisons result in a triangular similarity matrix (Figure 1) that allows pairwise comparison of NFs by reading along rows and down columns. Colors and similarity score numbers are used in the triangular similarity matrices to indicate the degree of similarity between two NF, with warm colors (yellow /red colors and numbers close to 1.0) indicating a high degree of similarity, and cool colors (blue colors and numbers close to 0.0) at the opposite end of the spectrum representing the NFs that are not similar.

Pairwise similarity scores (numbers) of representative test materials in FRAS, EPR (TemponeH), DCFH2-DA and Nrf2 activation assays were summarized in Table 1 and Table S2. CuO and Mn_2O_3 demonstrated high reactivity in all abiotic reactivity assays, which resulted in similar/very similar similarity as indicated with warm colors (Figure 1). On the other hand, $BaSO_4$ and CeO_2 were non-reactive in all assays and presented very similar similarity with orange color (Figure 1 A-D). However, $BaSO_4$ and CeO_2 are only similar in reactivity while the dissolution data and *in vivo* NOAECs (Table 1) of $BaSO_4$ and CeO_2 are different [36 35]. EPR assays using different spin traps (CPH, DMPO and Tempone H) resulted in different similarity scores compared to the data with TemponeH (as well as DCFH2-DA and Nrf2) (Figure 1B and S2 B, C). All abiotic assays detected CuO and $BaSO_4$ as the most and least reactive materials respectively and clearly $BaSO_4$/CuO were not found to be similar for all reactivity assays. Since CuO with a NOAEC of 0.6 mg/m³ also differed substantially in its *in vivo* response from $BaSO_4$ with a NOAEC of 50 mg/m³ (Table 1), their choice as representative materials was confirmed.

It should be noted here that BF calculations assume log-normally distributed data and sampling from the log-normal distributions of concentration-reactivity data for each of the groups of the materials studied. Parameters of the distributions are estimated from the data. i.e. the group of NFs considered each time, and for that reason BF values vary even though referring to the same pair of NFs.

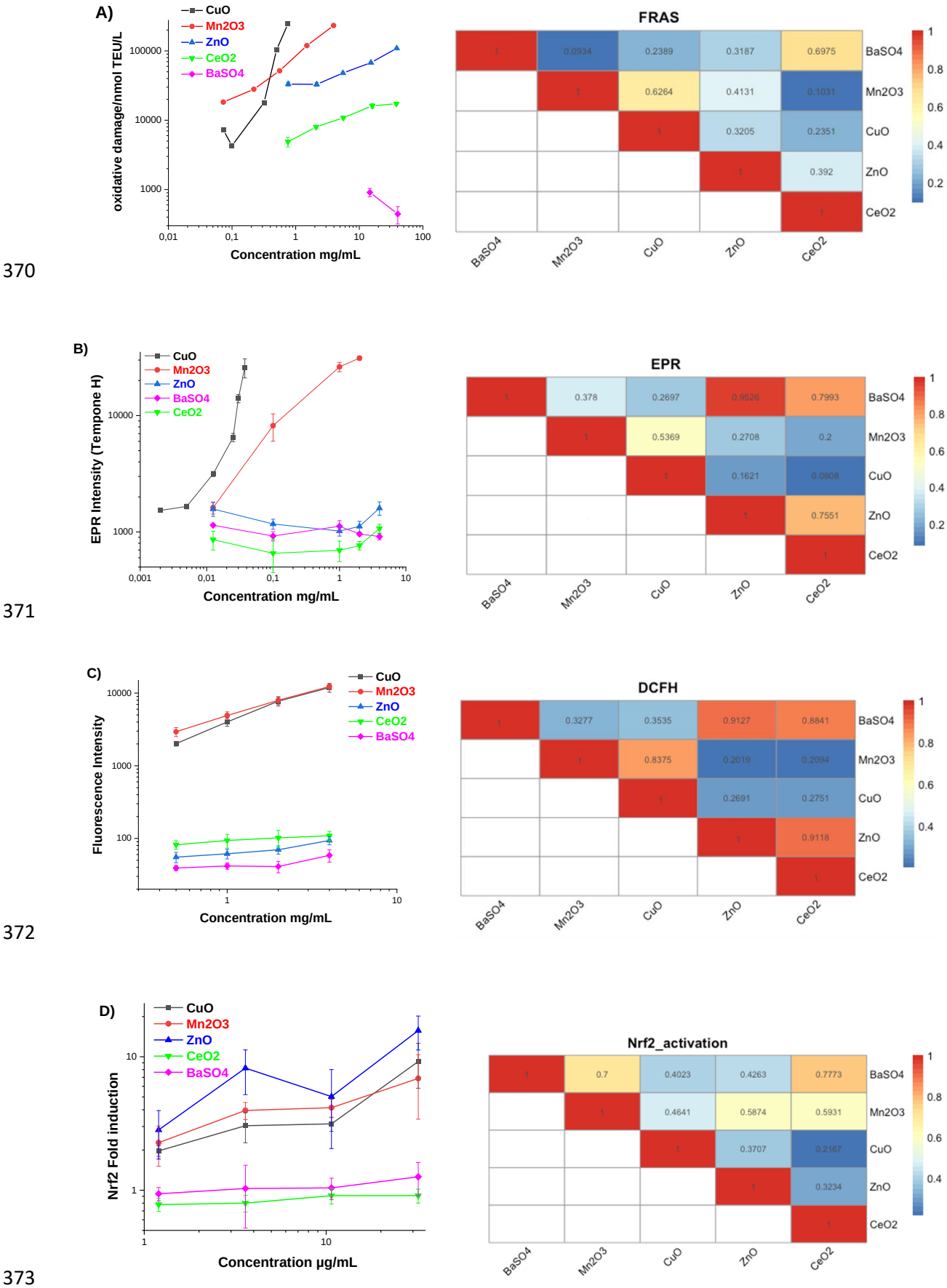


Figure 1. Left side concentration-response curves of FRAS (A), EPR (Tempone H) (B), DCFH2-DA (C), Nrf2 activation (D) assays for representative test materials and on the right side their corresponding

similarity plots for all possibly pairwise comparisons. Ideal similarity has a score of 1 (red), and the lowest similarity has a score of 0 (blue).

Table 1. Pairwise similarity scores for representative test materials in FRAS, EPR (Tempone H), DCFH and Nrf2 activation assays. Ideal similarity has a score of 1, and lowest similarity has a score of 0. Keeping in mind that reactivity is only one of several relevant properties, the results can be compared to the available *in vivo* inhalation NOAEC values for each pair of NFs [35].

Pair of materials	Similarity scores by reactivity assays				NOAEC ($mg\ m^{-3}$)
	FRAS	EPR	DCFH	Nrf2	
$BaSO_4$ / CuO	0.2389	0.2697	0.3535	0.4023	50 / 0.6
$BaSO_4$ / ZnO	0.3187	0.9526	0.9127	0.4263	50 / 0.5
$BaSO_4$ / CeO_2	0.6975	0.7993	0.8841	0.7773	50 / <0.5
CuO / ZnO	0.3205	0.1621	0.2691	0.3707	0.6 / 0.5
CuO / CeO_2	0.2351	0.0808	0.2751	0.2167	0.6 / <0.5
ZnO / CeO_2	0.3920	0.7551	0.9118	0.3234	0.5 / <0.5

3.2.2 Case Studies

Concentration dependent reactivity curves of case study materials were compared with very reactive Mn_2O_3 [8, 18] very reactive CuO[38], and non-reactive $BaSO_4$ [39-40]. These three materials served as representative test materials, and their concentration response was included a) in the graphical presentation for each NM class and each assay (Figures 2 to 4), and b) in the quantitative similarity analysis (Figures 2 to 4), where they represent the biologically relevant range, as recommended by the white paper.[4] Similarity assessment is relevant for NFs with reactivity values in the biologically relevant range. Differences between case study NFs that are very small compared to this range would still allow grouping. For this reason, quantitative similarity assessment of same-substance NFs must always include at least two other substances that represent the biologically relevant range.[4]

3.2.2.1 Fe-based materials

Pairwise similarity scores of Fe-based materials, Mn_2O_3 and $BaSO_4$ in FRAS, EPR (Tempone H), DCFH2-DA and Nrf2 activation assays were compared in Table S3. All Fe-based materials were similar to each other and demonstrated low reactivity, as did $BaSO_4$ in DCFH2-DA and Nrf2 activation assays (Figure 2 C and D). Depending on spin traps, different pairwise similarity scores were presented by EPR assays (Figure 2B and S3). It could not be confirmed that EPR, in general, is a suitable tool for analysing Fe-based materials; different similarity outcomes were found for different spin traps, indicating the importance of understanding the specificity of spin-traps, and possible, but undetermined, sample interferences were observed in EPR assays, including the interrupted concentration-response curve observed after 1 mg/mL measurements of Fe_2O_3_nanoA.

The iron oxides were intermediate in reactivity, and different compared to both negative and positive controls in FRAS assay. Only FRAS assay distinguished two groups of very similar (nano)forms. In the first group, Fe_2O_3_nanoA and FeOOH and in the second one Fe_2O_3_nanoB and Fe_2O_3_larger showed very similar mass-dose reactivity with a score of 0.87 and 0.74 respectively (Figure 2A). The representation of FRAS reactivity in surface-dose metrics (Figure S8A), the response of FeOOH equals the positive control in the FRAS assay (Figure S8A) and is an order of magnitude different from the Fe_2O_3 NFs.

The present results allow us to adjust the decision criteria of the tiered testing strategy such that the grouping decisions made with tier 1 abiotic method are not in conflict with tier 2 in-vitro tests and tier 3 in-vivo testing, considered as gold standard.[3, 9] We maintain that reactivity alone is not predictive of inhalation effects but also the reactivity assessment must not create conflicts between tiers. Similarity scores listed in Table 2 have to accepted as sufficiently similar as Fe_2O_3_nanoA and Fe_2O_3_larger have a similar *in vivo* NOAEC (tested by short-term inhalation screening on rats, Table 1) [35]. In another perspective, the difference in reactivity is a false positive result of the integrated approach to testing and assessment (IATA) that should not prevent grouping, if also the other decision nodes of the IATA indicate sufficient similarity.[10] On the other hand, FeOOH is well known catalyst for Fenton like reactions [41], which is a relevant ROS production mechanism in biological media. Accordingly, detection of higher reactivity in FeOOH was expected and confirmed by the FRAS assay (Figure 2A). Especially in surface dose metrics, the response of FeOOH equals the positive control in the FRAS assay (Figure S8A) and is an order of magnitude different from the Fe_2O_3 NFs.

Table 2. Pairwise similarity scores for Fe_2O_3_nanoA and Fe_2O_3_Larger in FRAS, EPR (Tempone H), DCFH and Nrf2 activation assays, and available NOAEC values for each pair of NFs [4].

Samples	**Similarity scores**				**NOAEC (mg m^{-3})**
	FRAS	**EPR**	**DCFH**	**Nrf2**	
Fe_2O_3_nanoA / Fe_2O_3_Larger	0.4632	0.6632	0.8658	0.9531	30 / 30

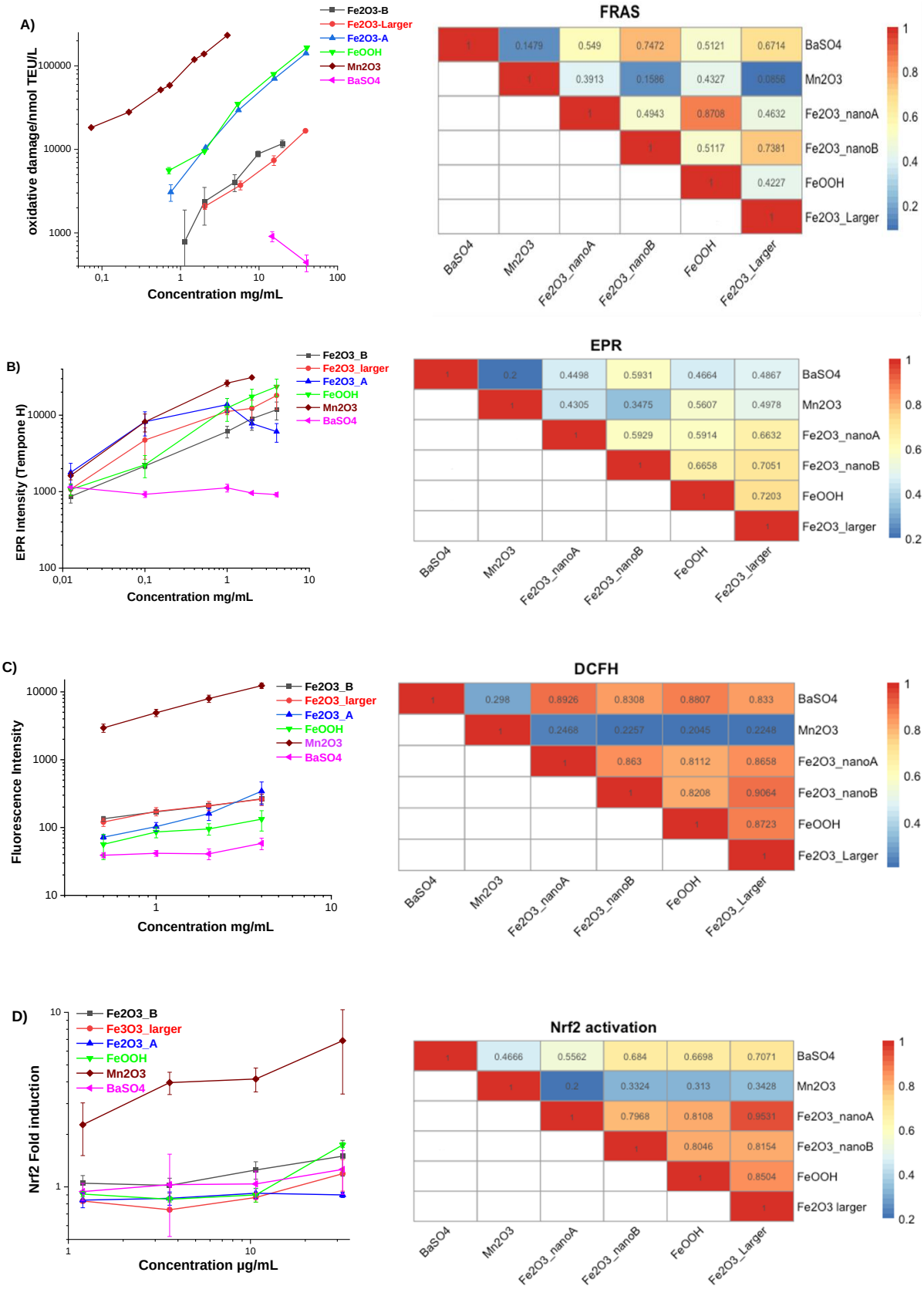
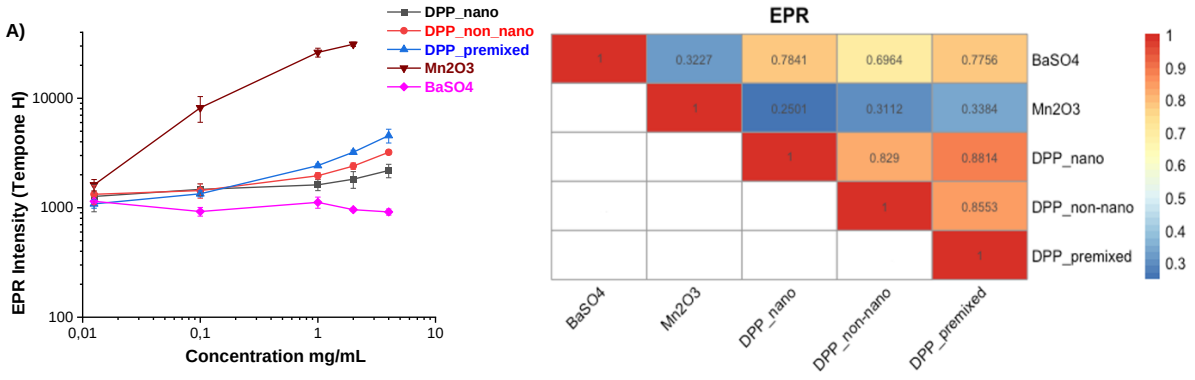


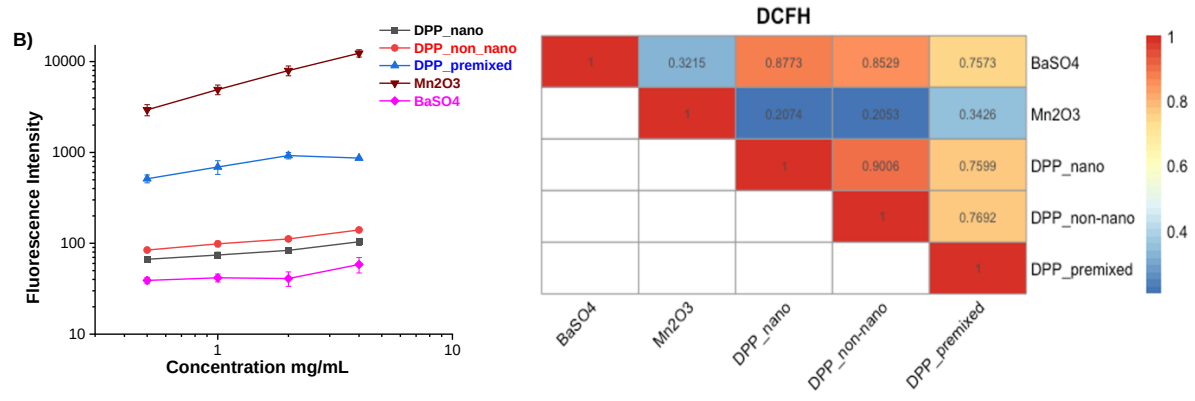
Figure 2. Left side concentration-response curves of FRAS (A), EPR (Tempone H) (B), DCFH2-DA (C), Nrf2 activation (D) assays for Fe-based materials and additional Mn_2O_3 and $BaSO_4$; on the right

side their corresponding similarity plots for all possibly pairwise comparisons. Ideal similarity has a score of 1 (red), and lowest similarity has a score of 0 (blue).

3.2.2.2 DPP-based organic pigments

Pairwise similarity scores of DPP-based materials, and Mn_2O_3 and $BaSO_4$ in EPR (Tempone H), DCFH2-DA and Nrf2 activation assays were compared in Table S4. Except EPR (DMPO), all reactivity assays scored very high pairwise similarity (orange colour) of three pigments (Figure 3 and S4). Moreover, they had low reactivity similar to $BaSO_4$, where the *in vivo* data is available (DPP nano and DPP non-nano), the similarity of low reactivity matches the similarity of low *in vivo* toxicity with inhalation NOEAC at >30mg/m³ (Table 3) [35, 42]. One notes that the similarity of the NF and the non-nano-form of DPP pigment is even higher in surface-dose metrics, where their curves collapse onto one concentration-response curve (Figure S7A, S7B). The comparison to the NOAEC, however, which is a mass-based value, must equally adhere to mass-based similarity assessment.





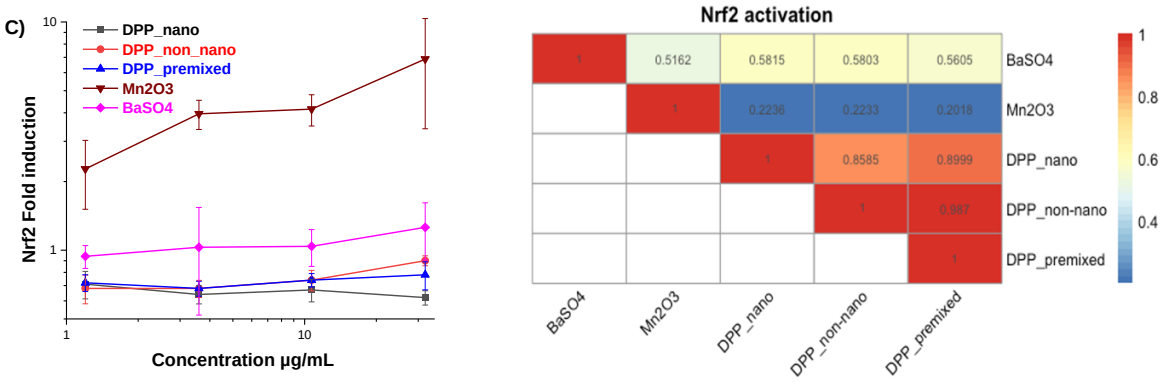


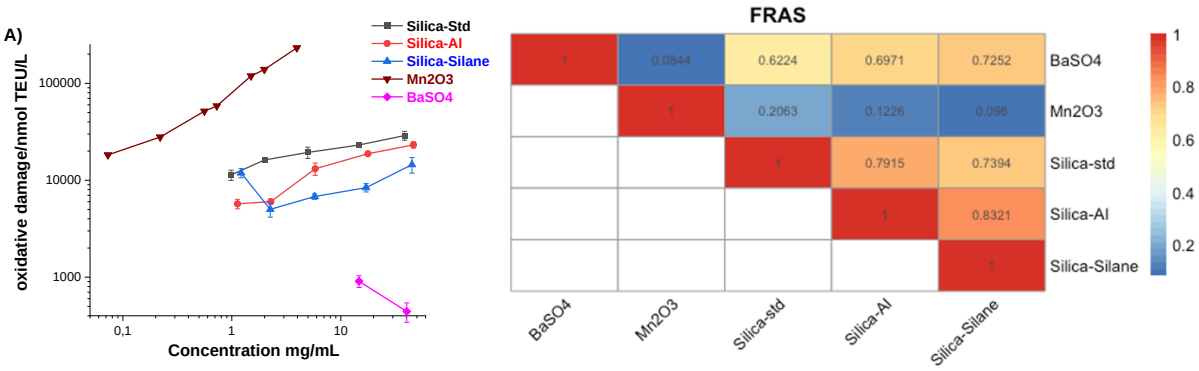
Figure 3. Left side concentration-response curves of EPR (Tempone H) (A), DCFH2-DA (B), Nrf2 activation (C) assays for DPP pigments and additional Mn_2O_3 and $BaSO_4$; on the right side their corresponding similarity plots for all possibly pairwise comparisons. Ideal similarity has a score of 1 (red), and lowest similarity has a score of 0 (blue).

Table 3. Pairwise similarity scores for DPP_nano and DPP_non-nano in EPR (Tempone H), DCFH and Nrf2 activation assays, and available NOAEC values for each pair of NFs [4].

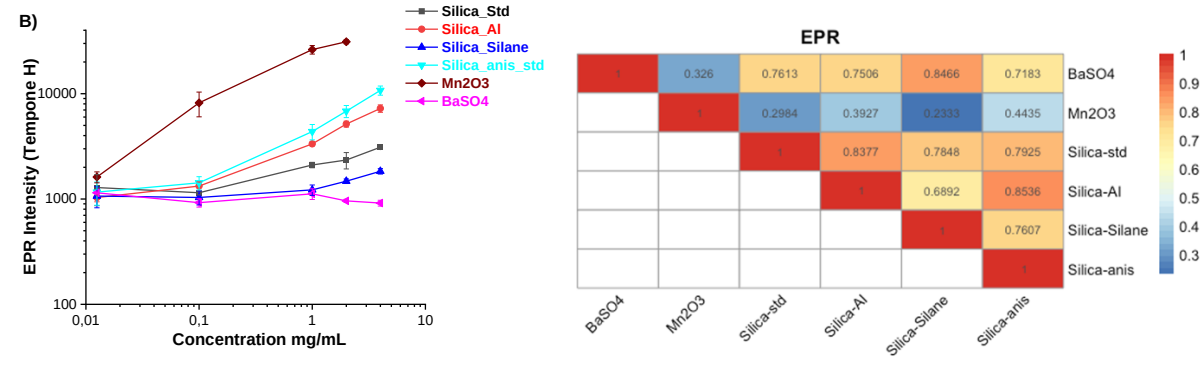
Samples	**Similarity scores**			**NOAEC (mg m^{-3})**
	EPR	**DCFH**	**Nrf2**	
DPP_nano / DPP_non-nano	0.6632	0.8658	0.9531	>30 / >30

3.2.2.3 Silica-based materials

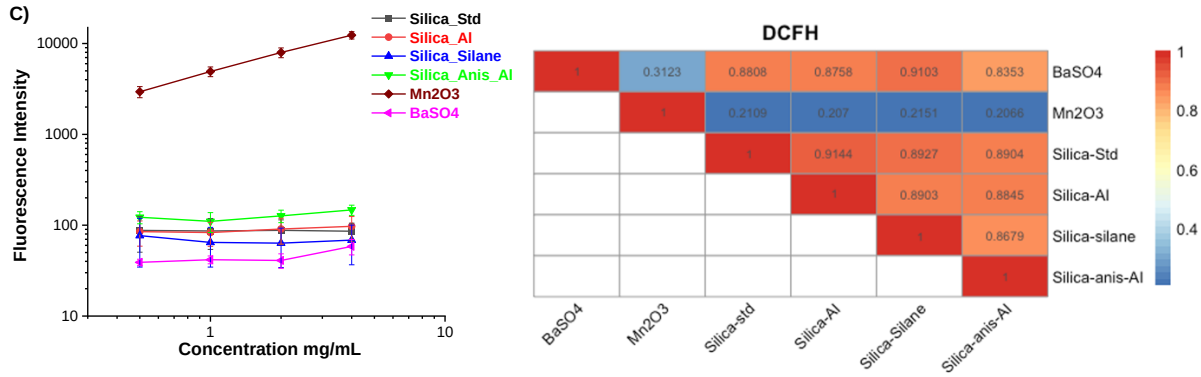
Comparing between silica NFs, all abiotic and *in vitro* reactivity assays indicated a high similarity within all samples (orange and red color) (Figure 4). Pairwise similarity scores of silica materials, and Mn_2O_3 and $BaSO_4$ in FRAS, EPR (Tempone H), DCFH2-DA and Nrf2 activation assays were compared in Table S5. All assays demonstrated low reactivity of all silica samples which is very similar to $BaSO_4$.



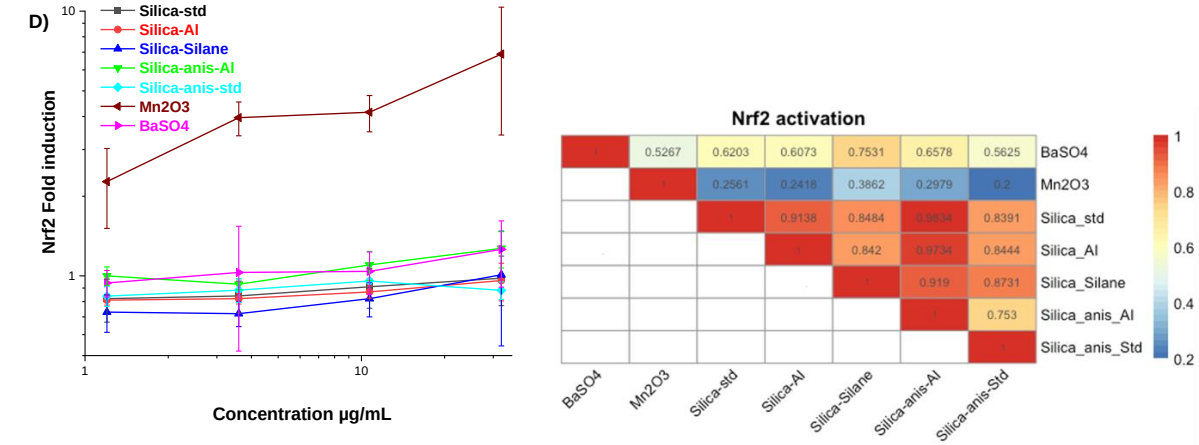
474



475



476



477

478 **Figure 4.** Left side concentration-response curves of FRAS (A), EPR (Tempone H) (B), DCFH2-DA
479 (C), Nrf2 activation (D) assays for silica-based materials and additional Mn_2O_3 and $BaSO_4$; on the

right side their corresponding similarity plots for all possibly pairwise comparisons. Ideal similarity has a score of 1 (red), and lowest similarity has a score of 0 (blue).

3.3 Sensitivity of assays for the reactivity induced by specific substances

By means of several commonly used assays for the assessment of NM reactivity, we have identified a number of important factors to consider. One consideration is that specific assays can be observed as being sensitive (or insensitive) to specific material classes. This was observed here on numerous occasions, and was evident for representative test materials and case study substances. For example, CuO and Mn_2O_3 consistently induced concentration dependent reactivity across all reactivity assays, while the reactivity of CeO_2 and ZnO was particularly confounded: Only FRAS assay demonstrated a concentration-dependent response to CeO_2, and when ZnO was assessed, FRAS, EPR (using DMPO) and Nrf2 activation assays demonstrated clear concentration-dependent reactivity, while by $DCFH_2$-DA and EPR using either CPH or Tempone H, no such concentration-response was observed; these general responses to all substances tested can be seen in Table 4 as a simple portrayal of whether a concentration response was observed or not. These inconsistences raise an issue of how much understanding there needs to be in specific assay parameters and the interpretation of simple reactivity endpoints. Should we consider the low reactivity of ZnO in certain assays as a false-negative affect as ZnO is known to be hazardous *in vivo*? Probably not, it just means that the mode-of-action of ZnO is better represented by certain assays than others. These findings were reflected with the use of statistical analysis and quantification of similarity. The robust three-parameter assessment model used for the statistical analysis confirmed the similarity and high reactivity of substances, such as CuO and Mn_2O_3 in all assays, and furthermore identified $BaSO_4$ and CeO_2 as being non-reactive in all assays. The specificity of abiotic assays to certain modes-of-action is especially versatile,[43-44] e.g. via different EPR probe molecules, of which CPH and DMPO spin probes (respectively, spin traps) are included in the ISO standard,[45] and have been used to group nanomaterials by surface-induced oxidative damage.[32, 35] Also the simplified assays such as the FRAN[14] or FRAP[46-47] versions of the FRAS assay, using individual probes instead of entire human serum, are less sensitive but more specific. The lack of "realism" may thus be seen as an advantage for targeted investigations,[31] but was less of a focus here.

Table 4. Comparison of concentration-response observations according to different reactivity assays; x = concentration-response observed, o = no concentration-response observed, xo = some but not all case study NFs provided a concentration-response, - = not measured in this assay.

	FRAS	$DCFH_2$-DA	EPR-Tempone-H	Nrf2 activation
Representative test materials				
CuO	x	x	x	x
Mn2O3	x	x	x	x
ZnO	x	o	o	x
CeO2	x	o	o	o
BaSO4	o	o	o	o
Case studies				
Fe-based	x	xo	x	o
Pigments	-	o	x	o
Silica	x	o	x	o

4. Conclusions

Assessment of reactivity is only one of the decision nodes on a complete IATA to justify grouping of NFs, while proving similarity in each of these decision nodes is required to justify a grouping decision; conversely other considerations, such as ranking of effects will help to inform source and target for read-across decisions. Although to fully define what contributions similarity in NF reactivity can play in grouping decisions are beyond scope of the current study, this work provides useful insights in how similarity in reactivity assessment can be assessed, using representative test materials and numerous NF case studies.

This study demonstrates that a similarity assessment of NFs can be compiled via use of well-defined reactivity assays, as long as the limitations of such an assessment are understood. Here, the BF calculations were applied to compare concentration-dependent reactivity over different concentration ranges from four different reactivity assays. The strength in this analysis partially comes from the robust nature in which multiple concentration-response parameters are considered within one model, including the assessment of three distinct opportunities to address similarity: the shape of the reactivity concentration-response curve, the concentration factor ranges, and the reactivity factor ranges. We found the algorithm used to be applicable to all abiotic and cell-based *in vitro* assays that were tested. This similarity assessment can serve as decision criterion in an IATA, where reactivity is one of several criteria on a data matrix of NFs and control materials. However, in this comparison, the same analytical method should be used for all NFs and control materials.

We observed several examples of *qualitative* similarity, where materials of different shape and size (but same composition) collapse onto one concentration-response curve in the surface-dose representation reactivity. However, the scaling of effects with specific surface area is one of the reasons for regulators to demand a separate assessment of NFs of a substance, and the justification of grouping must respect that even at same surface-specific reactivity, particles with a higher specific surface area induce more oxidative damage. The *quantitative* similarity analysis should thus be performed on concentration-response data provided as mass-metric representation.

We have used comparisons of *in vivo* NOAEC data here is provide a biological relevance to these reactivity measurements, and in doing so have again found correlations and disparities. With similarity scores for $BaSO_4$ and CuO, for example, being low in all assays and likewise were considerably different in their *in vivo* NOAEC. However, this was not always the case, as reflected in various examples of the representative test materials in Table 1, but also in assessment of case study substances such as the Fe-based materials; differences between Fe-based materials were observed in the FRAS assay which were not portrayed in relation to *in vivo* results. If we set the acceptable limit to a reactivity similarity score e.g. above 0.6, the Fe-based materials would not be justified for grouping by tier 1 reactivity methods (Table 2), and the GRACIOUS framework would require to exclude certain NFs from the candidate group, or to use another tier 1 assay (with justification), or to escalate to higher tier testing, where in the specific case tier 3 (in vivo, Table 2) confirms grouping.

These observations lead us to conclude that although our use of *in vivo* NOAEC values can provide some level of assurance that the similarity confirmations made have merit and justify an implication of a potential hazard, it should be stressed that any resulting *in vivo* NOAEC may be a culmination of many contributing factors, while our similarity assessment is implicit to one, that being reactivity (in this case specifically ROS generation). This illustrates the earlier discussion of how a reactivity provides just one decision node of a complex IATA, with other consideration being important. For example, we observed $BaSO_4$ and CeO_2 as statistically similar and both non-reactive in all assays, however, they differ considerably in their *in vivo* NOAEC. There must be a reason to this, and another decision node of the IATA prevents the grouping of these two materials, since they are not similar in their dissolution rate.[48-49]

In case study assessments we have also directly included the biologically relevant range of reactivity through use of high and low reactivity representative test materials (Mn_2O_3 and $BaSO_4$, respectively). In general, this allowed for a strong agreement in how the data from each

of the different reactivity assays was interpreted, with Fe-based substances being consistently found (in FRAS, DCFH2-DA and Nrf2 activation) similar to $BaSO_4$ and dissimilar to Mn_2O_3; only EPR was in disagreement. For the other case studies (pigments and silica particles) there was also a good level of correlation found across the different assays, when using this range of high to low reactivity. Table 1 highlights that the least similar pairs of NFs reach similarity scores around 0.2, whereas Tables 2 and 3 highlight that pairs of NFs which actually have similar *in vivo* NOAEC values -and thus should be accepted as being similar- score between 0.46 and 0.95 in their similarity of reactivity, depending on the chosen assay.

When *excluding* from the similarity analysis the least similar pair of representative materials, which represent the biologically relevant range, it was possible to tease out sensitive details *within* individual case studies. For example, when considering individual Fe-based substances the extent of similarity of two Fe particles (Fe_2O_3_nanoA and Fe_2O_3_Larger) was shown to differ considerably across different assays, with a high level of similarity shown in the $DCFH_2$-DA and Nrf2 assays, slightly lower level of similarity in EPR, and what can be considered as closer to dissimilar in the FRAS assay. With these observations in mind we would suggest that there are different purposes of conducting such analysis under both these conditions: i) assessment of similarity for regulatory purposes must include the representative materials of high and low levels of reactivity to align findings to known benchmarking values; ii) for mechanistic studies, or to identify trends that can guide Safer-by-Design optimisations, one may decide to assess NFs independently from these benchmark values to allow more sensitive assessment.

The Bayesian similarity algorithm could also be used for *in vivo* dose response to quantify similarity of the tier 3 results. The reactivity similarity assessment calibration would then be more robust, and our methodology could be transferred for use with other data.

In summary, this work demonstrates that the grouping of candidate NFs with regard to the similarity of their surface reactivity can be justified or rejected by well-established, partially ISO-standardised assays and by a novel but transparent, easily reproduced algorithm. The data matrix must include materials that represent high and low reactivity –typically two NFs of other substances– and must be filled by only one assay for all candidate NFs and the representative materials.

Acknowledgements:

The GRACIOUS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 760840. We thank Veronica Di Battista for support in data representation.

References

1. Stone, V.; Gottardo, S.; Bleeker, E. A.; Braakhuis, H.; Dekkers, S.; Fernandes, T.; Haase, A.; Hunt, N.; Hristozov, D.; Jantunen, P., A framework for grouping and read-across of nanomaterials-supporting innovation and risk assessment. *Nano Today* **2020,** *35*, 100941.
2. Appendix, R.6-1 for Nanoforms Applicable to the Guidance on QSARs and Grouping of Chemicals (ECHA), E. C. A., Ed. Helsinki, Finland, 2019; Vol. ED-04-19-681-EN-N.
3. Verdon, R.; Gillies, S. L.; Brown, D. M.; Henry, T.; Tran, L.; Tyler, C. R.; Rossi, A. G.; Stone, V.; Johnston, H. J., Neutrophil activation by nanomaterials in vitro: comparing strengths and limitations of primary human cells with those of an immortalized (HL-60) cell line. *Nanotoxicology* **2021,** *15* (1), 1-20.
4. 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., How can we justify grouping of nanoforms for hazard assessment? Concepts and tools to quantify similarity. *NanoImpact* **2022,** *25*, 100366.
5. G. Tsiliki, D. A. S., A. Zabeo, G. Basei, D. Hristozov, N. Jeliazkova, M. Boyles, F. Murphy, W. Peijnenburg, W. Wohlleben, V Stone, Bayesian based grouping of nanomaterials and Dose Response similarity models. *in preparation NanoImpact special issue* **2021**.
6. Worth, A.; Aschberger, K.; Asturiol, D.; Bessems, J.; Gerloff, K.; Graepel, R.; Joossens, E.; Lamon, L.; Palosaari, T.; Richarz, A., Evaluation of the availability and applicability of computational approaches in the safety assessment of nanomaterials. *Publications Office of the European Union, Luxembourg* **2017**.
7. Giusti, A.; Atluri, R.; Tsekovska, R.; Gajewicz, A.; Apostolova, M. D.; Battistelli, C. L.; Bleeker, E. A.; Bossa, C.; Bouillard, J.; Dusinska, M., Nanomaterial grouping: Existing approaches and future recommendations. *NanoImpact* **2019,** *16*, 100182.
8. Arts, J. H.; Hadi, M.; Irfan, M.-A.; Keene, A. M.; Kreiling, R.; Lyon, D.; Maier, M.; Michel, K.; Petry, T.; Sauer, U. G., A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). *Regulatory Toxicology and Pharmacology* **2015,** *71* (2), S1-S27.
9. Braakhuis, H. M.; Murphy, F.; Ma-Hock, L.; Dekkers, S.; Keller, J.; Oomen, A. G.; Stone, V., An Integrated Approach to Testing and Assessment to Support Grouping and Read-Across of Nanomaterials After Inhalation Exposure. *Applied In Vitro Toxicology* **2021**.
10. Murphy, F.; Dekkers, S.; Braakhuis, H.; Ma-Hock, L.; Johnston, H.; Janer, G.; di Cristo, L.; Sabella, S.; Jacobsen, N. R.; Oomen, A. G., An integrated approach to testing and assessment of high aspect ratio nanomaterials and its application for grouping based on a common mesothelioma hazard. *NanoImpact* **2021,** *22*, 100314.
11. Di Cristo, L.; Oomen, A. G.; Dekkers, S.; Moore, C.; Rocchia, W.; Murphy, F.; Johnston, H. J.; Janer, G.; Haase, A.; Stone, V.; Sabella, S., Grouping Hypotheses and an Integrated Approach to Testing and Assessment of Nanomaterials Following Oral Ingestion. *Nanomaterials* **2021,** *11* (10), 2623.
12. Echa, Appendix R.6-1 for nanoforms applicable to the Guidance on QSARs and Grouping of Chemicals. Helsinki, 2019.

13. Zhao, J.; Riediker, M., Detecting the oxidative reactivity of nanoparticles: a new protocol for reducing artifacts. *Journal of Nanoparticle Research* **2014,** *16* (7), 2493.
14. Bi, X.; Westerhoff, P., Ferric reducing reactivity assay with theoretical kinetic modeling uncovers electron transfer schemes of metallic-nanoparticle-mediated redox in water solutions. *Environmental Science: Nano* **2019,** *6* (6), 1791-1798.
15. Angelé-Martínez, C.; Nguyen, K. V.; Ameer, F. S.; Anker, J. N.; Brumaghim, J. L., Reactive oxygen species generation by copper(II) oxide nanoparticles determined by DNA damage assays and EPR spectroscopy. *Nanotoxicology* **2017,** *11* (2), 278-288.
16. Eom, H. J.; Choi, J., Oxidative stress of CeO2 nanoparticles via p38-Nrf-2 signaling pathway in human bronchial epithelial cell, Beas-2B. *Toxicol Lett* **2009,** *187* (2), 77-83.
17. Pal, A. K.; Hsieh, S.-F.; Khatri, M.; Isaacs, J. A.; Demokritou, P.; Gaines, P.; Schmidt, D. F.; Rogers, E. J.; Bello, D., Screening for oxidative damage by engineered nanomaterials: a comparative evaluation of FRAS and DCFH. *Journal of Nanoparticle Research* **2014,** *16* (2), 2167.
18. Arts, J. H. E.; Irfan, M.-A.; Keene, A. M.; Kreiling, R.; Lyon, D.; Maier, M.; Michel, K.; Neubauer, N.; Petry, T.; Sauer, U. G.; Warheit, D.; Wiench, K.; Wohlleben, W.; Landsiedel, R., Case studies putting the decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping) into practice. *Regulatory Toxicology and Pharmacology* **2016,** *76*, 234-261.
19. He, W.; Liu, Y.; Wamer, W. G.; Yin, J.-J., Electron spin resonance spectroscopy for the study of nanomaterial-mediated generation of reactive oxygen species. *Journal of Food and Drug Analysis* **2014,** *22* (1), 49-63.
20. Buettner, G. R., Spin trapping: ESR parameters of spin adducts. *Free Radic Biol Med* **1987,** *3* (4), 259-303.
21. Brandt, R.; Keston, A. S., Synthesis of diacetyldichlorofluorescin: A stable reagent for fluorometric analysis. *Analytical Biochemistry* **1965,** *11* (1), 6-9.
22. Wilson, M. R.; Lightbody, J. H.; Donaldson, K.; Sales, J.; Stone, V., Interactions between Ultrafine Particles and Transition Metals in Vivo and in Vitro. *Toxicology and Applied Pharmacology* **2002,** *184* (3), 172-179.
23. Niture, S. K.; Khatri, R.; Jaiswal, A. K., Regulation of Nrf2—an update. *Free Radical Biology and Medicine* **2014,** *66*, 36-44.
24. JRC Nanomaterials Repository. Available online:. https://ec.europa.eu/jrc/en/scientific-tool/jrc-nanomaterials-repository.
25. Gandon, A.; Werle, K.; Neubauer, N.; Wohlleben, W., Surface reactivity measurements as required for grouping and read-across: An advanced FRAS protocol. *Journal of Physics: Conference Series* **2017,** *838*, 012033.
26. Hsieh, S. F.; Bello, D.; Schmidt, D. F.; Pal, A. K.; Stella, A.; Isaacs, J. A.; Rogers, E. J., Mapping the biological oxidative damage of engineered nanomaterials. *Small* **2013,** *9* (9-10), 1853-65.
27. Hellack, B.; Nickel, C.; Albrecht, C.; Kuhlbusch, T. A. J.; Boland, S.; Baeza-Squiban, A.; Wohlleben, W.; Schins, R. P. F., Analytical methods to assess the oxidative potential of nanoparticles: a review. *Environmental Science: Nano* **2017,** *4* (10), 1920-1934.
28. Guisti, A. D., Nils; Haase, Andrea;, SOP: Determining Nrf2 Activation. *https://zenodo.org/record/5084750* **2021**.
29. Faulkenberry, T. J., Computing Bayes factors to measure evidence from experiments: An extension of the BIC approximation. *Biometrical Letters* **2018,** *55* (1), 31-43.
30. Achawi, S.; Feneon, B.; Pourchez, J.; Forest, V., Assessing biological oxidative damage induced by graphene-based materials: An asset for grouping approaches using the FRAS assay. *Regulatory Toxicology and Pharmacology* **2021,** *127*, 105067.
31. Hellack, B.; Nickel, C.; Albrecht, C.; Kuhlbusch, T. A. J.; Boland, S.; Baeza-Squiban, A.; Wohlleben, W.; Schins, R. P. F., Analytical methods to assess the oxidative potential of nanoparticles: a review. *Environmental Science: Nano* **2017,** *4*, 1920-1934.
32. Bahl, A.; Hellack, B.; Wiemann, M.; Giusti, A.; Werle, K.; Haase, A.; Wohlleben, W., Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods. *NanoImpact* **2020,** *19*, 100234-100234.

33. Schmid, O.; Stoeger, T., Surface area is the biologically most effective dose metric for acute nanoparticle toxicity in the lung. *Journal of Aerosol Science* **2016,** *99*, 133-143.
34. Oberdörster, G.; Kuhlbusch, T. A. J., In vivo effects: Methodologies and biokinetics of inhaled nanomaterials. *NanoImpact* **2018,** *10* (Supplement C), 38-60.
35. Wohlleben, W.; Hellack, B.; Nickel, C.; Herrchen, M.; Hund-Rinke, K.; Kettler, K.; Riebeling, C.; Haase, A.; Funk, B.; Kühnel, D.; Göhler, D.; Stintz, M.; Schumacher, C.; Wiemann, M.; Keller, J.; Landsiedel, R.; Broßell, D.; Pitzko, S.; Kuhlbusch, T. A. J., The nanoGRAVUR framework to group (nano)materials for their occupational, consumer, environmental risks based on a harmonized set of material properties, applied to 34 case studies. *Nanoscale* **2019,** *11* (38), 17637-17654.
36. Johannes G. Keller, M. P., Philipp Müller, Lan Ma-Hock, Kai Werle; Josje Arts, Robert Landsiedel, Wendel Wohlleben, Variation in dissolution behavior among different nanoforms and its implication for grouping approaches in inhalation toxicity. *nanoImpact* **2021**.
37. N. Jeliazkova, E. B., R. Cross, A. Haase , G. Janer ,W. Peijnenburg, M. Pink, H. Rauscher, C. Svendsen, G. Tsiliki, A. Zabeo, D. Hristozov, V. Stone, W. Wohlleben, How can we justify hazard assessment of nanoforms by grouping in order to reduce animal testing? Concepts and usable tools to quantify similarity. *in preparation NanoImpact special issue* **2021**.
38. Gosens, I.; Cassee, F. R.; Zanella, M.; Manodori, L.; Brunelli, A.; Costa, A. L.; Bokkers, B. G. H.; de Jong, W. H.; Brown, D.; Hristozov, D.; Stone, V., Organ burden and pulmonary toxicity of nano-sized copper (II) oxide particles after short-term inhalation exposure. *Nanotoxicology* **2016,** *10* (8), 1084-1095.
39. Buesen, R.; Landsiedel, R.; Sauer, U.; Wohlleben, W.; Groeters, S.; Strauss, V.; Kamp, H.; van Ravenzwaay, B., Effects of SiO2, ZrO2, and BaSO4 nanomaterials with or without surface functionalization upon 28-day oral exposure to rats. *Archives of Toxicology* **2014,** *88* (10), 1881-1906.
40. Landsiedel, R.; Ma-Hock, L.; Hofmann, T.; Wiemann, M.; Strauss, V.; Treumann, S.; Wohlleben, W.; Gröters, S.; Wiench, K.; van Ravenzwaay, B., Application of short-term inhalation studies to assess the inhalation toxicity of nanomaterials. *Part Fibre Toxicol* **2014,** *11*, 16-16.
41. Li, X.; Huang, Y.; Li, C.; Shen, J.; Deng, Y., Degradation of pCNB by Fenton like process using α-FeOOH. *Chemical Engineering Journal* **2015,** *260*, 28-36.
42. Hofmann, T.; Ma-Hock, L.; Strauss, V.; Treumann, S.; Rey Moreno, M.; Neubauer, N.; Wohlleben, W.; Gröters, S.; Wiench, K.; Veith, U.; Teubner, W.; van Ravenzwaay, B.; Landsiedel, R., Comparative short-term inhalation toxicity of five organic diketopyrrolopyrrole pigments and two inorganic iron-oxide-based pigments. *Inhal Toxicol* **2016,** *28* (10), 463-79.
43. Lakshmi Prasanna, V.; Vijayaraghavan, R., Insight into the Mechanism of Antibacterial Activity of ZnO: Surface Defects Mediated Reactive Oxygen Species Even in the Dark. *Langmuir* **2015,** *31* (33), 9155-9162.
44. Angelé-Martínez, C.; Nguyen, K. V. T.; Ameer, F. S.; Anker, J. N.; Brumaghim, J. L., Reactive oxygen species generation by copper(II) oxide nanoparticles determined by DNA damage assays and EPR spectroscopy. *Nanotoxicology* **2017,** *11* (2), 278-288.
45. ISO, Nanotechnologies — Electron spin resonance (ESR) as a method for measuring reactive oxygen species (ROS) generated by metal oxide nanomaterials. 2017; Vol. ISO/TS 18827.
46. Thaipong, K.; Boonprakob, U.; Crosby, K.; Cisneros-Zevallos, L.; Hawkins Byrne, D., Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. *Journal of Food Composition and Analysis* **2006,** *19* (6–7), 669-675.
47. Benzie, I. F. F.; Strain, J. J., The Ferric Reducing Ability of Plasma (FRAP) as a Measure of "Antioxidant Power": The FRAP Assay. *Analytical Biochemistry* **1996,** *239* (1), 70-76.
48. Keller, J.; Persson, M.; Müller, P.; Ma-Hock, L.; Werle, K.; Arts, J.; Landsiedel, R.; Wohlleben, W., Variation in dissolution behavior among different nanoforms and its implication for grouping approaches in inhalation toxicity. *NanoImpact* **2021**, 100341.
49. Keller, J. G.; Graham, U. M.; Koltermann-Jülly, J.; Gelein, R.; Ma-Hock, L.; Landsiedel, R.; Wiemann, M.; Oberdörster, G.; Elder, A.; Wohlleben, W., Predicting dissolution and transformation of inhaled nanoparticles in the lung using abiotic flow cells: The case of barium sulfate. *Scientific Reports* **2020,** *10* (1), 458.

753