Regulatory Toxicology and Pharmacology 89 (2017) 253–267

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

A quantitative framework to group nanoscale and microscale particles by hazard potency to derive occupational exposure limits: Proof of concept evaluation

Nathan M. Drew [a, *], Eileen D. Kuempel [a], Ying Pei [b], Feng Yang [b]

[a] National Institute for Occupational Safety and Health (NIOSH), Nanotechnology Research Center (NTRC), Cincinnati, OH 45226, USA
[b] West Virginia University, Department of Industrial and Management System Engineering, Morgantown, WV 26506, USA

ARTICLE INFO

Article history:
Received 29 March 2017
Received in revised form
18 July 2017
Accepted 3 August 2017
Available online 5 August 2017

Keywords:
Nanomaterial
Risk assessment
Benchmark dose modeling
Pulmonary inflammation
Hazard potency
Physicochemical properties
Hierarchical clustering
Random forest
Predictive modeling
Occupational exposure limit

ABSTRACT

The large and rapidly growing number of engineered nanomaterials (ENMs) presents a challenge to assessing the potential occupational health risks. An initial database of 25 rodent studies including 1929 animals across various experimental designs and material types was constructed to identify materials that are similar with respect to their potency in eliciting neutrophilic pulmonary inflammation, a response relevant to workers. Doses were normalized across rodent species, strain, and sex as the estimated deposited particle mass dose per gram of lung. Doses associated with specific measures of pulmonary inflammation were estimated by modeling the continuous dose-response relationships using benchmark dose modeling. Hierarchical clustering was used to identify similar materials. The 18 nanoscale and microscale particles were classified into four potency groups, which varied by factors of approximately two to 100. Benchmark particles microscale TiO_2 and crystalline silica were in the lowest and highest potency groups, respectively. Random forest methods were used to identify the important physicochemical predictors of pulmonary toxicity, and group assignments were correctly predicted for five of six new ENMs. Proof-of-concept was demonstrated for this framework. More comprehensive data are needed for further development and validation for use in deriving categorical occupational exposure limits.

Published by Elsevier Inc.

1. Introduction

In this paper, a quantitative framework is described with potential application to the derivation of occupational exposure limits (OELs) for "new" nanomaterials which may have physicochemical property information but limited or insufficient dose-response data. This methodology is consistent with the 21st Century toxicology and risk assessment goals to increase the efficiency and utility of risk assessment for human health risk decision-making (NAS, 2007, 2017). This framework also contributes to risk assessment goals at NIOSH and in the nanotechnology research community to develop methodologies that utilize nanotoxicology and nanomaterials research data to develop occupational exposure limits (OELs) for the large and growing number of engineered nanomaterials (Gordon et al., 2014; Kuempel et al., 2006; Nel et al., 2013; Schulte et al., 2010). Standardized methodology is a major research need for harmonization of OELs for nanomaterials (ISO, 2016; Mihalache et al., 2017) as for OELs in general (Deveau et al., 2015). Categorical methods of data analysis for hazard and risk assessment, such as proposed in this framework, can provide more efficient use of data and support comparative potency, read-across, and other alternative methods (OECD, 2007, 2012, 2014; NAS, 2017; ECHA, 2017). In the absence of sufficient evidence to develop individual OELs for specific ENMs, categorical OELs could be developed and used as occupational exposure bands (OEBs) to inform exposure control decision-making in the workplace (Kuempel et al., 2012; ISO, 2016; NIOSH, 2017). This analysis provides progress towards this goal by providing an initial quantitative framework to categorize ENMs by hazard potency. Following further evaluation of this framework, including application to more comprehensive data, the next step will be to utilize the derived categorical potency estimates in standard risk assessment methods to derive categorical OELs or OEBs for ENMs.

* Corresponding author. National Institute for Occupational Safety and Health, 1090 Tusculum Avenue, Cincinnati, OH 45226-1998, USA.
E-mail address: vom8@cdc.gov (N.M. Drew).

http://dx.doi.org/10.1016/j.yrtph.2017.08.003
0273-2300/Published by Elsevier Inc.

A number of risk assessment frameworks for ENMs have been developed in recent years. In general, these frameworks apply standard risk assessment principles and practice to nanomaterials (OECD, 2012). The standard risk assessment paradigm has four main components: hazard assessment, exposure assessment, dose-response assessment, and risk characterization (NAS, 1983); an update of the framework reaffirmed those key elements and emphasized the role of problem formulation in identifying the risk assessment needs to evaluate the risk management options (NAS, 2009). The current analysis focuses on dose-response assessment of acute pulmonary inflammation data in rodents across a set of nanoscale and microscale particles in various experimental designs. As more data become available, this framework could be evaluated for other endpoints, dose metrics, and particle characteristics.

Other frameworks have proposed grouping ENMs based on hazard and exposure potential (Arts et al., 2015, 2016; Bos et al., 2015; Braakhuis et al., 2016; Cohen et al., 2013; Gebel et al., 2014; Godwin et al., 2015; Oomen et al., 2015; Oosterwijk et al., 2016; Stone et al., 2014; Walser and Studer, 2015), although many of these have not been tested with quantitative data. Case study data have been used with some of these frameworks (Arts et al., 2016; Gkika et al., 2017; Grieger et al., 2017). Exposure scenarios along the ENM "lifecyle" (production to disposal) are included in some of the conceptual frameworks (Arts et al., 2015; Environmental Defense and Dupont, 2007; Shatkin, 2013). Weight of evidence and decision analysis methods have been proposed in other analysis frameworks (Hristozov et al., 2014; Zuin et al., 2011). Several quantitative structure-activity relationship (QSAR) models have been developed, which describe the important factors influencing the toxicity (Munro et al., 1996; Burello and Worth, 2011; Gernand and Casman, 2014, 2016) and allow for hazard grouping and ranking (Liu et al., 2011, 2013; Oh et al., 2016; Zhang et al., 2012); however, these models have not thus far been used in human health risk assessment. A recent quantitative framework proposes a human health risk prioritization based on a margin of exposure method (i.e., estimating the ratio between animal effect level and human exposure) (Hristozov et al., 2016). Still lacking in these currently available frameworks is an integrated methodology to utilize quantitative dose-response data to group ENMs by hazard potency using biological responses and dose metrics that allow for the estimation of human-equivalent concentration (HEC) and development of categorical OELs for ENMs. The current analysis provides progress towards filling this gap in available methods for OEL derivation of a large set of diverse materials.

The objective of this paper is to demonstrate the development and evaluation of a quantitative clustering framework to support evidence-based occupational safety and health decision-making. The overall framework is shown in Fig. 1. This framework utilizes standard methodologies that depend on the amount of data available. If sufficient data are available for a given nanomaterial, an individual OEL can be derived. Such materials can serve as benchmark materials to provide a reference point for interpreting results in experiments with new materials. As shown in the current analysis, empirical dose-response data are used to derive benchmark dose estimates for individual experiments and to develop potency-based groups. A predictive model to estimate the potency-based group using physicochemical information is developed. Physicochemical information of new materials is then used to estimate their group assignments. Acute *in vivo* dose-response data in the lungs of rodents are used in this example. In the future, *in vitro* and long-term *in vivo* data in animals or humans (including biomarker data) could be added to extend and validate the predictive models in the framework (Fig. 1). This strategy is consistent with current guidance on the use of alternative testing strategies in the risk assessment of manufactured nanomaterials, and the need to develop methods for grouping materials by their properties and predicting hazards of new materials (OECD, 2016; NAS, 2017).

2. Methods

2.1. Data analysis plan

The main steps in the data analysis are shown in Fig. 2. A database of nanoscale and microscale particulate materials was constructed to describe the dose-response relationships for pulmonary inflammation in rodents and the physicochemical properties of those materials. Inflammation potency was estimated using benchmark dose modeling, and materials were grouped using hierarchical clustering. Potency is defined in this analysis as the estimated mass deposited lung dose associated with a specific inflammatory response in the lungs (Section 2.3). These responses are considered to be biologically significant and relevant to workers (as described in Section 2.3). A classification random forest model was developed to identify the physicochemical properties that were predictive of the hazard potency group. The model was then tested on a separate dataset of new materials, using only the physicochemical information to predict the hazard potency group and evaluating the predictions against potency estimates derived from the dose-response data.

2.2. Database construction (training: NIOSH/CIIT/ENPRA, testing: NanoGo)

Individual rodent study data on nanoscale and microscale particles from a variety of material types were obtained *ad hoc* from studies identified through research collaborations and from the published literature. Files for 25 *in vivo* rodent studies comprising of data for 1899 unique animals (from 1929 records) were graciously provided by researchers from NIOSH (Porter et al., 2001, 2004, 2013; Roberts et al., 2013; Sager et al., 2013; Xia et al., 2011); CIIT (renamed Hamner Institute) (Bermudez et al., 2002, 2004); and the European Framework 7 Programme on Engineered Nano-Particle Risk Assessment (ENPRA) to which NIOSH was a research partner (ENPRA, 2013). Exposure-response information was available for the individual rodents, and experimental design properties were included (e.g. post-exposure duration, method of exposure).

The duration of exposure varied across materials in this analysis, from single bolus administration (IT or PA) to subchronic inhalation (up to 116 days) (Table 1). The estimated total deposited mass dose of particles in the lungs provides a normalized dose metric over time. It is similar in concept to "Haber's rule" of C x T = K, which assumes that a specific adverse effect (K) would be associated with the cumulative exposure (concentration x time). This dose metric does not account for particle clearance from the lungs, but is considered to be a reasonable assumption in this analysis. Normal clearance of poorly-soluble particles from the lungs is relatively slow (e.g., clearance half-time of 60–90 days in rats) (Pauluhn, 2014); and the subchronic data included in this analysis were poorly-soluble particles (TiO_2 and crystalline silica). In the acute studies, clearance would be minimal at 0–1 day post-exposure.

Some physicochemical information about the materials was provided in the files, but most of this information was gleaned from the resulting publications of the researchers. Six chemicals of various forms were studied: iron oxide (Fe_3O_4) (nano-scale), silver (Ag) (nano-scale), multi-walled carbon nanotubes (MWCNT), crystalline silica (micro-scale), titanium dioxide (TiO_2) (nano- and micro-scale), and zinc oxide (ZnO) (nano-scale). The experimental designs of these studies varied by exposure route, rodent species and strain, and exposure and post-exposure duration. A majority of the studies utilized intratracheal instillation (IT) as the exposure

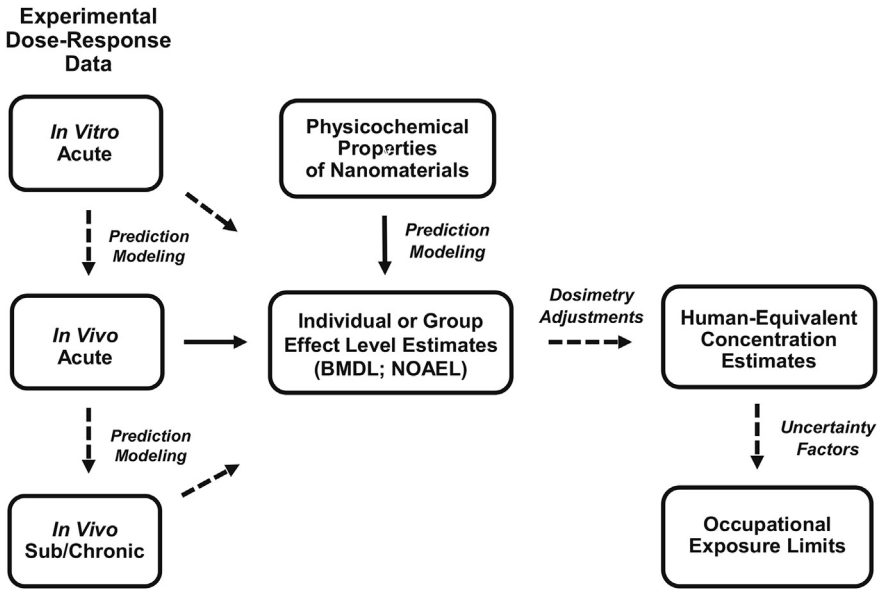

Fig. 1. Framework for Predicting Hazard Potency Groups and Developing Occupational Exposure Limits (OELs) for Engineered Nanomaterials. In the current analysis, rodent data of pulmonary inflammation are used to estimate hazard potency groups for nanoscale and microscale particles; these groups are predicted for other nanomaterials based only on physicochemical properties. Ongoing research involves modeling additional *in vivo* and/or *in vitro* data. After further model validation, these group-based potency estimates will be used as the points of departure (PODs) in risk assessment to estimate human-equivalent concentrations (HECs) and derive occupational exposure limits (OELs). *Key: solid arrows: illustrated in current analysis; hatched arrows: ongoing and future work. Abbreviations: BMDL: Benchmark dose, 95% lower confidence limit estimate; NOAEL: No observed adverse effect level.*

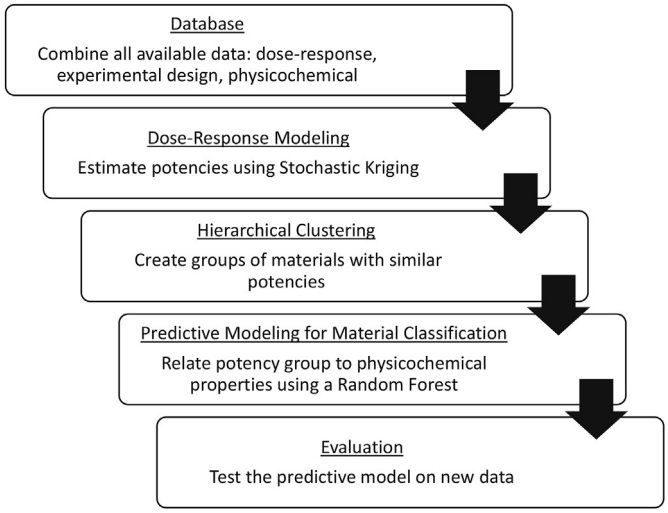

Fig. 2. Schematic of the data analysis methods.

route, while the remainder used inhalation (Inh) or pharyngeal aspiration (PA). Various strains and both sexes of rats (male Sprague-Dawley, male and female F344) and mice (Female C57BL/6N, Male C57BL/6J, Female C57BL/6-Apoetm1) were used across the studies. Acute pulmonary response data (0–1 d post-exposure) were available for most studies, while some studies reported pulmonary response at the end of repeated inhalation exposure (also 0–1 d post-exposure). A summary of these experimental characteristics for the various types of the materials is shown in Table 1. This database (henceforth NIOSH/CIIT/ENPRA) was used for training the predictive models.

A separate database of *in vivo* rodent studies of similar experimental design was constructed from the U.S. National Institute for Environmental Health Studies (NIEHS) NanoGo Consortium (Bonner et al., 2013). This inter-laboratory research program studied three forms each of nanoscale titanium dioxide (TiO_2) and multiwalled carbon nanotubes in 258 unique rats (male Sprague-Dawley, male F344) and 177 mice (male C57BL/6) for a total of 435 unique rodents. Physicochemical property information for these materials was provided in Xia et al. (2013), which included *in vitro* studies of the same nanomaterials studied *in vivo* (Bonner et al., 2013). This database (henceforth NanoGo) was used for validating the predictive framework as these data were received after the NIOSH/CIIT/ENPRA database had been created. A deposited dose metric was created for these data also in order to be commensurate with the NIOSH/CIIT/ENPRA database. A description of the experimental design characteristics of the NanoGo studies is provided in the Supplementary Material (Tables S–6).

The response of interest was the proportion of polymorphonuclear leukocyte cells (PMNs) (a.k.a. neutrophils), which is a common measure of pulmonary inflammation. This response is often reported in particle toxicology studies of the lungs; other endpoints include lactate dehydrogenase (LDH) and fibrosis severity scores. If the proportion of PMNs was not reported in the file, it was calculated from the primary data (ratio of PMN count to total cell count). Toxicology studies often report the PMN percentage, which is simply the proportion x 100. PMN % has also been used to compare results between studies since differences in the methods used for bronchoalveolar lavage fluid (BALF) and cell counting may result in different total cell counts (Bonner et al., 2013). In order to compare the lung doses across the various routes of exposure, a deposited dose (in micrograms) metric was created. The measured particle mass lung doses were used when reported. For studies using IT or PA that did not report the actual lung dose, the deposited mass dose was estimated to be equivalent to the administered mass dose. For inhalation studies that did not report the measured lung doses, the total deposited mass dose was estimated as the product of the airborne exposure concentration x exposure duration x minute ventilation x pulmonary (alveolar) deposition fraction. The exposure concentration and duration were

Table 1
Description of NIOSH/CIIT/ENPRA experimental designs.

Reference[e]	Route[d] & frequency	Species	Strain	Material	Material type	Animals	Dose groups
Porter et al., 2013	PA	Male Mouse	C57BL/6J	TiO_2	Short Nanobelt	115	7.5, 15, 30 (μg)
	Single				Long Nanobelt	140	1.875, 7.5, 15, 30 (μg)
					Nanosphere	92	7.5, 15, 30 (μg)
					Control	54	Dispersion Medium
Xia et al., 2011	IT	Male Rat	Sprague-Dawley	Fe_3O_4	Fe_3O_4 pure	41	0.0375, 0.075, 0.15, 0.3 (mg/rat)
	Single				ZnO 1% Fe	38	0.0375, 0.075, 0.15, (0.3 mg/rat)
					ZnO 10% Fe	42	0.0375, 0.075, 0.15, 0.3 (mg/rat)
					ZnO pure	36	0.0375, 0.075, 0.15, 0.3 (mg/rat)
					Control	21	Dispersion Medium
Roberts et al., 2013	Inh	Male Rat	Sprague-Dawley	Ag	Ionized	12	100 μg/m^3
	5 h				Colloidal	12	1000 μg/m^3
					Control	24	Control Aerosol
Sager et al., 2013	PA	Male Mouse	C57BL/6J	MWCNT	Bare	80	2.5, 10, 40 (μg)
	Single				Carboxylated	82	2.5, 10, 40 (μg)
					Control	61	Dispersion Medium
Porter et al., 2001	Inh	Male Rat	F344	Silica	Crystalline	120	15 (mg/m^3)
	6 h/d, 5 d/wk, 116 d				Control	120	Filtered Air
Porter et al., 2004	Inh	Male Rat	F344	Silica	Crystalline	66	15 (mg/m^3)
	6 h/d, 5 d/wk, 20 d or 40 d or 60 d				Control	66	Filtered Air
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	Ag	Colloidal	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-NRCWE	IT	Female Mouse	C57BL/6N	MWCNT	Long	24	32, 128 (μg/mouse)
	Single				Control	12	Dispersion Medium
ENPRA-NRCWE[c]	IT	Female Mouse	C57BL/6-Apoetm1	MWCNT	Long	26	32, 128 (μg/mouse)
	Single				Control	16	Dispersion Medium
ENPRA-NRCWE[c]	IT	Female Mouse	C57BL/6-Apoetm1	MWCNT	Short	26	32, 128 (μg/mouse)
	Single				Control	16	Dispersion Medium
ENPRA-NRCWE	IT	Female Mouse	C57BL/6N	MWCNT	Short	24	32, 128 (μg/mouse)
	Single				Control	12	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	MWCNT	Long	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	MWCNT	Short	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-UC[a,b]	IT	Female Mouse	C57BL/6N	Silica	Crystalline	8	2500 (μg/mouse)
	Single			MWCNT	Entangled	20	12.5, 25, 50, 100 (μg/mouse)
					Control	7	Dispersion Medium
ENPRA-UC[a,b]	IT	Female Mouse	C57BL/6N	Silica	Crystalline	8	2500 (μg/mouse)
	Single			MWCNT	Entangled	20	12.5, 25, 50, 100 (μg/mouse)
					Control	7	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	ZnO	Coated	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	ZnO	Uncoated	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-UC[a]	IT	Female Mouse	C57BL/6N	Silica	Crystalline	8	2500 (μg/mouse)
	Single			ZnO	Coated	10	12.5, 25, 50, 100 (μg/mouse)
					Control	5	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	TiO_2	Negatively Charged	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	TiO_2	Positively Charged	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	TiO_2	Anatase	39	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	TiO_2	Rutile	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
ENPRA-RIVM	IT	Female Mouse	C57BL/6N	TiO_2	Rutile	21	1, 4, 8, 16, 32, 64, 128 (μg/mouse)
	Single				Control	3	Dispersion Medium
Bermudez et al., 2002	Inh	Female Rat	F344	TiO_2	Fine	75	10, 50, 250 (mg/m^3)
	6 h/d, 5 d/wk, 13 wk				Control	25	Filtered Air
Bermudez et al., 2004	Inh	Female Rat	F344	TiO_2	Ultrafine	75	0.5, 2, 10 (mg/m^3)
	6 h/d, 5 d/wk, 13 wk				Control	25	Filtered Air
					Total Observations	**1929**	

All materials are nano-scale except: Porter et al., 2001; Porter et al., 2004; Bermudez et al., 2002.
[a] The same group of 8 positive control animals (Silica) are used across the 3 studies.
[b] The same group of 7 negative control animals (MWCNT) are used across the 2 studies.
[c] An identical set of 7 negative control animals appear in both studies.
[d] Inh = Inhalation; IT = Intratracheal Instillation; PA = Pharyngeal Aspiration.
[e] Not all studies were used in this analysis (Section 3.2).

as reported in the study. The minute ventilation was derived from allometric equations using body weight (U.S. EPA, 1994). The deposition fractions in rats were estimated from the Multiple-Path Particle Dosimetry (MPPD) model, v. 2.90.1 (ARA, 2011); input values included the particle aerodynamic diameter and geometric standard deviation. In mice, the pulmonary deposition fractions were estimated from Raabe et al. (1988), which is one of the datasets used in testing the mouse lower respiratory tract

deposition model (Asgharian et al., 2014). The mouse model became available in MPPD v. 3.04 [available at: www.ara.com/products/mppd.htm].

Because both mice and rats were used in the toxicology studies, species normalization was required. To account for the differences in size of the animals, the deposited dose was normalized by the wet lung weight of the species. For rats, it was assumed that a typical control lung weight was 0.9 g for F344, 1.3 g for SD; for mice, a typical control lung weight was 0.15 g. These values were based on data from the compiled database, where available, and also from the literature for rodents of the same species, strain, sex and age (Kobayashi et al., 2009; NIOSH, 2013; Porter et al., 2001). The normalized dose metric was the micrograms deposited dose per gram lung (μg/g lung); normalizing particle dose per g lung has been used in other analyses of rodent data (NIOSH, 2011; Schmid and Stoeger, 2016).

During the exploratory data analysis stage to learn about factors that may need to be considered when estimating potency, a random forest regression model found that of the experimental factors, post-exposure duration, (normalized) deposited dose, and material form were most predictive of the PMN proportion across all studies (Fig. 3). Additionally, animal sex, strain, species and exposure route are less predictive, suggesting that the deposited dose metric is working as intended by compensating for these experimental design differences. Variable importance of a particular factor is estimated by measuring the change in mean squared error (MSE), a measure of predictive accuracy, when other values of that factor are permuted across the dataset. Thus, if permuting the values of a factor has a large negative impact on predictions, it is an important predictor and receives a high variable importance score.

Initial efforts to model the post-exposure – dose – response surfaces were unfavorable leading to the decision to stratify the NIOSH/CIIT/ENPRA database by post-exposure duration. Four strata were created: 0–3 days; 7–14 days; 28–60 days; 91 days to 1 year. These strata reflect typical toxicology study designs. A majority of the data and studies were contained in the 0–3 days stratum, which is the focus of this analysis (Table 2).

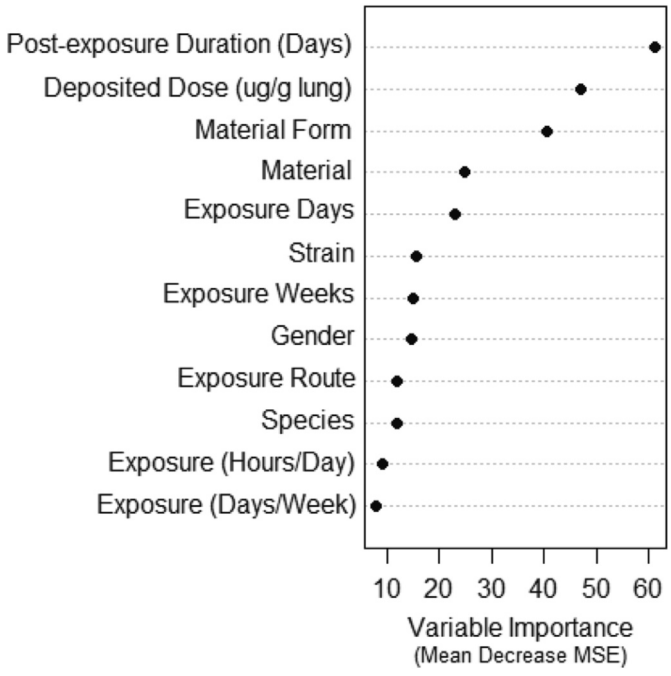

Fig. 3. Important predictors of PMN proportion.

2.3. Benchmark dose modeling and potency estimation

Dose-response modeling was utilized to quantify the potency of a given material, where potency is a benchmark dose (BMD) (Crump, 1984, 2002). The total deposited lung dose (normalized per g lung) is the dose metric used in these analysis. Neutrophilic pulmonary inflammation is the benchmark response (BMR) used in these analyses, and was measured in the rodent studies as increased proportion of PMNs in BALF. The specific BMRs evaluated were either an added 4% PMNs over background (i.e., in the unexposed control animals) or a total of 10% PMNs in BALF, based on biological evidence cited below. Since potency is defined here as the estimated mass deposited lung dose associated with pulmonary inflammation, the lower that dose (i.e., BMD), the greater the potency of the material.

These BMRs were selected as being biologically relevant responses in both rodents and humans (as discussed in NIOSH, 2011). In rats, a response of approximately 4% PMNs in BALF has been associated with particle lung doses at or near overloading of lung clearance in rats exposed to poorly soluble particles (Muhle et al., 1991; Tran et al., 1999; Pauluhn, 2012). Overloading results in a dose-dependent increase in the particle retention in the lungs (Morrow, 1988; Elder et al., 2005; Pauluhn, 2011) and the development of persistent pulmonary inflammation, fibrosis, and lung cancer in rats (Muhle et al., 1991; ILSI, 2000; IARC, 2010). Dose metrics that are most predictive of overloading across microscale and nanoscale particle sizes include particle volume (Pauluhn, 2011, 2014) and particle surface area (Oberdörster et al., 1994; Tran et al., 2000; Morfeld et al., 2015). The background percentage of PMNs in control rats in long-term studies is generally low (<1%), while in the acute studies, higher background % PMNs were observed in animals treated with vehicle control (e.g., Bonner et al., 2013). In humans, approximately 4% PMNs in BALF was associated with respiratory impairment in workers in dusty jobs (Rom, 1991); and 10% PMNs in BALF is considered to be clinically abnormal (Crystal et al., 1981; Martin et al., 1985). This analysis focuses on results for the added 4% PMN response, and detailed results for both the added 4% PMN response and the total 10% PMN response are provided in the Supplemental Materials (Tables S-1 to S-5). The selection of biologically significant BMRs is relevant for risk assessment (U.S. EPA, 2012); these values would not necessarily be statistically significant, which depends on the variability in the observed dose-response relationship. The specific BMR used (i.e., an added 4% PMNs or a total of 10% PMNs) influenced the number of rodent studies with sufficient dose-response data for BMD estimation in this analysis (as explained further in Results).

Benchmark dose estimation for each of the dose-response relationships was performed individually using Stochastic Kriging (SK), which is a flexible modeling method suited for handling the non-linear, heteroskedastic dose-response relationships often seen in toxicology studies (Wang et al., 2014). A wide range of continuous dose-response relationships can be modeled using SK, and the capability to automate the modeling process facilitated the estimation of BMDs from multiple studies, although it still requires a specification of the covariance function. The popular Gaussian covariance function was first used, which should handle all of the two dimensional dose-response shapes and generally creates a smooth curve. However, the model curve fit to the data was not restricted to biologically-relevant shapes and included shapes such as sinusoidal or non-monotonic. If the visual fit was inadequate, the General Exponential covariance function was used, which generally creates a non-smooth curve (e.g. piecewise linear). Only dose-response relationships with statistically significant differences in mean response across dose groups were modeled, and this characteristic was investigated via ANOVA (results not shown). Levene's

Table 2
Distribution of post-exposure durations in the NIOSH/CIIT/ENPRA database.

		Post Exposure Days												
Reference	**Material**	0	1	3	7	14	28	30	36	60	91	112	182	364
Porter et al. 2013	TiO_2		■	■	■		■					■		
Xia et al. 2011	Fe_3O_4, ZnO		■			■		■						
Roberts et al. 2013	Ag		■		■									
Sager et al. 2013	MWCNT		■		■									
Porter et al. 2001	Silica	■												
Porter at al. 2004	Silica	■							■					
ENPRA - RIVM	Ag		■											
ENPRA - NRCWE	MWCNT		■				■							
ENPRA - NRCWE	MWCNT		■				■							
ENPRA - NRCWE	MWCNT		■				■							
ENPRA - NRCWE	MWCNT		■				■							
ENPRA - RIVM	MWCNT		■											
ENPRA - RIVM	MWCNT		■											
ENPRA - UC	MWCNT, Silica									■				
ENPRA - UC	MWCNT, Silica									■				
ENPRA - RIVM	ZnO		■											
ENPRA - RIVM	ZnO		■											
ENPRA - UC	ZnO, Silica									■				
ENPRA - RIVM	TiO_2		■											
ENPRA - RIVM	TiO_2		■											
ENPRA - RIVM	TiO_2		■											
ENPRA - RIVM	TiO_2		■											
ENPRA - RIVM	TiO_2		■											
Bermudez et al. 2002	TiO_2	■					■				■		■	■
Bermudez et al. 2004	TiO_2	■					■				■		■	■

Homogeneity of Variance test was used prior to ANOVA to check the constant variance assumption, and a log transformation of the response was used in cases exhibiting heteroscedasticity. Decisions were made in these two tests using a 5% level of significance.

Dose-response modeling was initially completed using the U.S. EPA Benchmark Dose Software (BMDS), version 2.6 (U.S. EPA, 2015), which offers the choice of several parametric model forms to fit to a single relationship. BMD estimates from BMDS were similar to those from SK (Wang et al., 2014), however modeling many relationships is more time-consuming with the basic BMDS. Variability in the dose-response data in an experiment can result in uncertainty in the estimated potencies; thus, the 95% one-sided confidence limit estimate of the BMD (i.e., BMDL) is calculated in the U.S. EPA BMDS and the SK modeling to provide an estimate of that uncertainty. The BMDL estimate provides 95% confidence that the true BMD is not lower than the BMDL. The BMDL estimates from SK tended to be higher than those from BMDS, indicating higher accuracy for a given confidence level. As a result, BMDs from SK modeling were chosen to represent the potency of the nanoscale or microscale particles. Other published software that allow for the fitting of several BMD model forms to numerous dose-response relationships in succession (Wignall et al., 2014; Shao and Shapiro, 2016) may be useful considerations for future investigations.

2.4. Grouping materials by potency via hierarchical clustering

Once potency estimates are obtained for each of the dose-response relationships in the 0–3 day stratum, similarly behaving materials are identified by comparing those potency estimates. Materials with similar potency estimates are assumed to behave similarly with respect to pulmonary inflammation. Hierarchical clustering was used to create four groups of materials with similar potency estimates. Four groups were chosen to potentially reflect the four broad categories that have been proposed for ENMs based on the biological mode of action and physicochemical properties (Poorly Soluble Low Toxicity, Soluble, High Aspect Ratio, Highly Toxic) (BSI, 2007; Kuempel et al., 2012; BAuA, 2013; Arts et al., 2016). The grouping process involves agglomerative clustering in which each potency estimate first begins as its own cluster. The Euclidean distances (differences) between potency estimates are calculated, and potency estimates (or clusters) nearest to one another are then combined into a set of new clusters. This process repeats until all potency estimates are represented by four clusters. The terms cluster and group are used interchangeably. This process creates groups with descriptively different potency estimates as opposed to statistically different. The BMDs are the best estimates of potency and so were used when creating the groups of materials with similar potencies. Variability in the experimental dose-response data was not considered in creating the groups;

however, variability will be taken into account later by evaluating the distribution of the BMDL estimates within the final groups. BMDL estimates are used as the point of departure (POD) in risk assessment to estimate a safe dose in humans (U.S. EPA, 2012) (Section 4.2).

2.5. Random forest to predict potency group

In order to classify a new material into a potency cluster using that new material's physicochemical properties, a classification random forest model was developed. Note that this is not the same random forest model used in the exploratory data analysis stage (Fig. 3), which was a regression model for identifying experimental design factors that are important in predicting PMN proportion. The classification random forest described next seeks to predict the potency group (1, 2, 3, or 4) of a material using only the available physicochemical property information as predictors. Due to the limited number of materials with a potency cluster label and numerous but sparse physicochemical properties, traditional modeling schemes are not well suited to describe the physicochemical property-cluster space. A non-parametric solution is a classification tree (Breiman et al., 1984); however, a single tree can tend to over-fit the data and have a large amount of variability in its predictions. A classification random forest is a collection of many classification trees that has improved predictive accuracy (Breiman, 2001). The "random" namesake is due to two characteristics: a tree is constructed using a random bootstrap sample of the data; and a random subset of all predictor variables are considered for every branch in the tree. This process is repeated many times, creating many trees, hence a forest. The result is a collection of many decorrelated trees, each of which can provide different information about the relationship of interest. For a given new material, each tree in the forest casts a vote for a potency group, with the final group prediction being the potency group that received a majority of the votes.

Default options were used when constructing the random forest: 500 trees are created, and $\sqrt{p}$ predictors were considered at each branch in the tree. Since many physicochemical properties have missing values, distinct values (−99 if quantitative; "N/A" if qualitative) were used to indicate missingness. Physicochemical properties with missing values for all materials were excluded from the modeling process (Fig. S-1). An examination of pairwise correlation estimates was used to ensure that few, if any, highly correlated physicochemical properties were included in the model, as this can lead to a preference of correlated predictors when the classification trees are created (Strobl et al., 2008).

2.6. Evaluation of model

The six material forms in the NanoGo database, five of which were not used in the model development (anatase/rutile nanospheres are the same type of material used in Bermudez et al., 2004), were considered to be "new materials" and were used to evaluate the random forest models. Using the physicochemical properties of these new materials, a potency cluster was predicted by the random forest model. Pulmonary inflammation potencies were estimated by SK from the individual dose-response data by lab and material. The median potency estimate for each material was used to identify the nearest potency cluster to which that the material would be assigned, and was then used as a comparison to the predicted cluster (Fig. 4).

2.7. Software

Data file formatting, database creation, and tests for differences in mean response were completed using SAS software version 9.3. Exploratory data analyses, hierarchical clustering, and random forest modeling (*randomForest* package) completed using R 3.3.2 (R Core Team, 2014) within the RStudio wrapper (1.0.44). Stochastic Kriging was completed using MATLAB R2016a (MATLAB, 2016).

3. Results

3.1. Analyzed data

From the initial 1899 unique rodents, 844 distinct rodents with a measured response were analyzed across 32 dose-response relationships (a combination of material form and post-exposure duration within each study) from the 22 studies with post-exposure durations between 0 and 3 days:

- 1929 observations for 1899 unique animals
  - 1557 unique animals with a measured PMN response
    - 844 unique animals with a measured PMN response after 0–3 days post-exposure

3.2. Potency estimation and grouping

Of the 32 relationships, 18 were found to have a statistically significant difference in mean response across the dose groups, and Stochastic Kriging was able to estimate a BMD and BMDL for each of those 18 relationships. These potency estimates ranged from 2.1 μg/g lung to 2489.6 μg/g lung. The 14 relationships that were not estimable (e.g., no statistically significant difference in mean response) included the TiO_2 nanospheres (Porter et al., 2013), all nano-scale silver (Roberts et al., 2013; ENPRA, 2013), and various forms of MWCNT, TiO2 and ZnO (ENPRA, 2013). For the absolute 10% PMN response, BMDs were not estimable in an additional six cases as extrapolation would be required due to either a high background response or a low response at the highest dose. Hierarchical clustering created four separate groups of materials with similar potencies, with a majority of materials belonging to the first and most potent group. The microscale TiO_2 was the sole member of the fourth and least potent group. As shown in Fig. 5, the groups are descriptively, and not statistically significantly, different as the variability in the potency estimates were not used when creating the groups (Section 2.4).

The potency estimates within each group are summarized by the minimum, median, and maximum BMD values; uncertainty in the BMD estimates are summarized by the minimum and the 5th percentile BMDL estimates within each group (Table 3). Based on the median BMD estimates, the potency estimates of materials within the first group vary by a factor of about 60; the Group 2 materials are about nine times less potent than those in Group 1; the Group 3 material is roughly half as potent as the materials in Group 2; and the Group 4 material is about six times less potent than the material in Group 3 (Table 3).

Within one material type, TiO_2 potency depended on particle size and shape (Fig. 5). Microscale (fine) TiO_2 was the least potent in eliciting pulmonary inflammation (Group 4); nanoscale (ultrafine) TiO_2 was about six times more potent (Group 3); while TiO_2 nanobelts were the most potent (Group 1). Within Group 1, short nanobelts were less potent than long nanobelts. Three types of ZnO were the most potent materials within Group 1 (Fig. 5). The various types of MWCNTs had varied potencies within Groups 1 and 2. The potency of short and long MWCNT seemed to depend more on the mouse strain than on the shape; and the C57BL/6-Apoetm1 mice appeared less sensitive to either short or long MWCNT (Group 2) than were the C57BL/6N mice for the same MWCNT types (Group

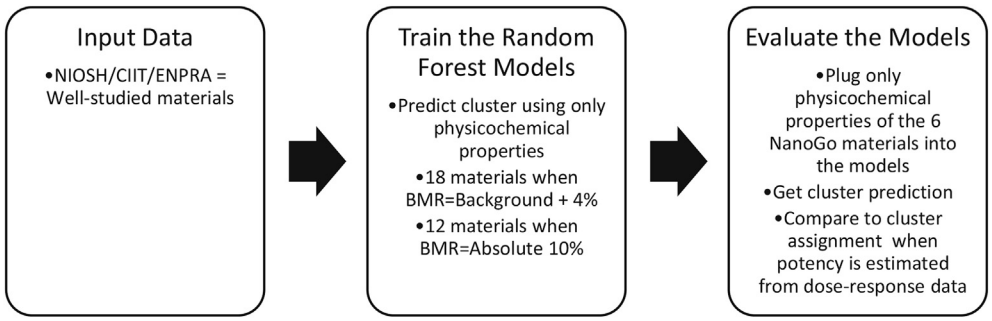

Fig. 4. Process of model development and evaluation.

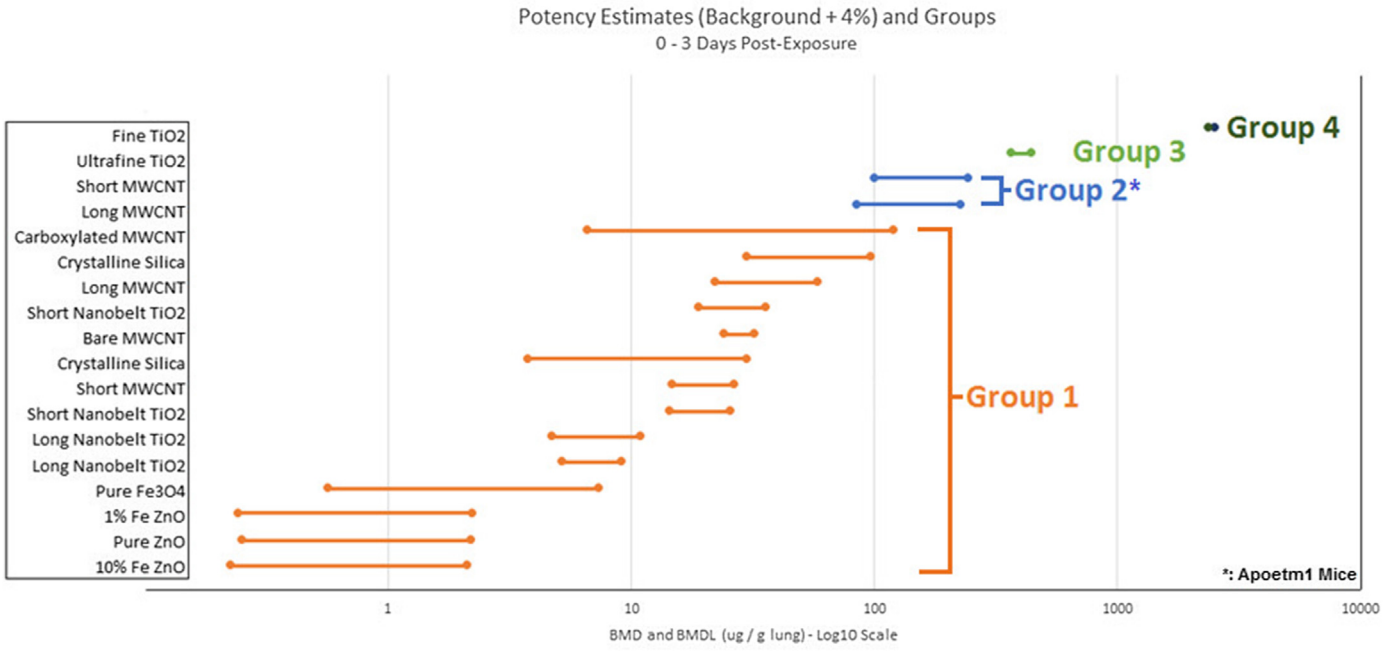

Fig. 5. Visualization of potency estimates, variability and groups.

Table 3
Summary of potency estimates and materials by potency group in NIOSH/CIIT/ENPRA data.

Group	BMDs Min/Median/Max	BMDLs Min/5th Percentile	Material types	
1	2.1/25.8/119.2	0.22/0.23	Fe_3O_4	Pure
			MWCNT	Bare, Carboxylated, Long, Short
			Silica	Crystalline (2)[b]
			TiO_2	Long Nanobelt (2), Short Nanobelt (2)
			ZnO	Pure, 1% Fe_3O_4, 10% Fe_3O_4
2[a]	225.9/233.5/241.1	83.9/84.7	MWCNT	Long, Short
3	440.3	365.3	TiO_2	Ultrafine
4	2489.6	2366.1	TiO_2	Fine[†]

[a] C57BL/6-Apoetm1 mice.
[b] Micro-sized materials.

1). However, the dose-response associations for the C57BL/6-Apoetm1 mice were linear, i.e. responses were measured for the control group and one exposed group. As a result, the BMD estimates for the C57BL/6-Apoetm1 mice were higher than those for the C57BL/6N mice where more exposure groups were available and nonlinear relationships were observed in the data, as the steepest part of the nonlinear curve reached the BMR at an earlier dose than that of the linear model. Carboxylated or bare MWCNT in C57BL/6J mice were also in Group 1. Microscale crystalline silica (Min-U-Sil®) was in the highest potency group (Group 1).

In this framework, a new material is predicted to belong to one of the four groups. In order to derive a potency estimate for that new material, a BMDL from the other materials within that group is used as an estimate for the effect level of the new material. One option is to use the minimum BMDL, but this may be affected by an unusually potent material within a given group, or a material with a

high degree of variability in the experimental data; so an alternative is to use the 5th percentile BMDL. A table of the individual potency estimates is provided in the Supplemental Materials (Tables S–1).

3.3. Training the random forest model

A classification random forest model was trained using the potency group labels and physicochemical properties of these 18 materials. Seventeen physicochemical properties were used during the development of the classification random forest model. Metrics related to particle size – such as density, surface area, and diameter – appeared to be most predictive of the potency group, while properties such as presence of contaminants or modification types were less predictive (Fig. 6). These estimates of variable importance should be interpreted with caution due to the small number of materials used to build the model, as well as the paucity of information contained within the seventeen physicochemical properties. If more information were available, or if other properties were considered, these estimates of importance would likely change. The material (e.g. TiO_2 or MWCNT) is a somewhat important variable. If a new material is a type that was not present in the training data, this covariate could not be used. Thus, a training dataset should cover as many materials as possible for a reliable and predictive model.

3.4. Testing the random forest model

The random forest model was evaluated using the NanoGo data.

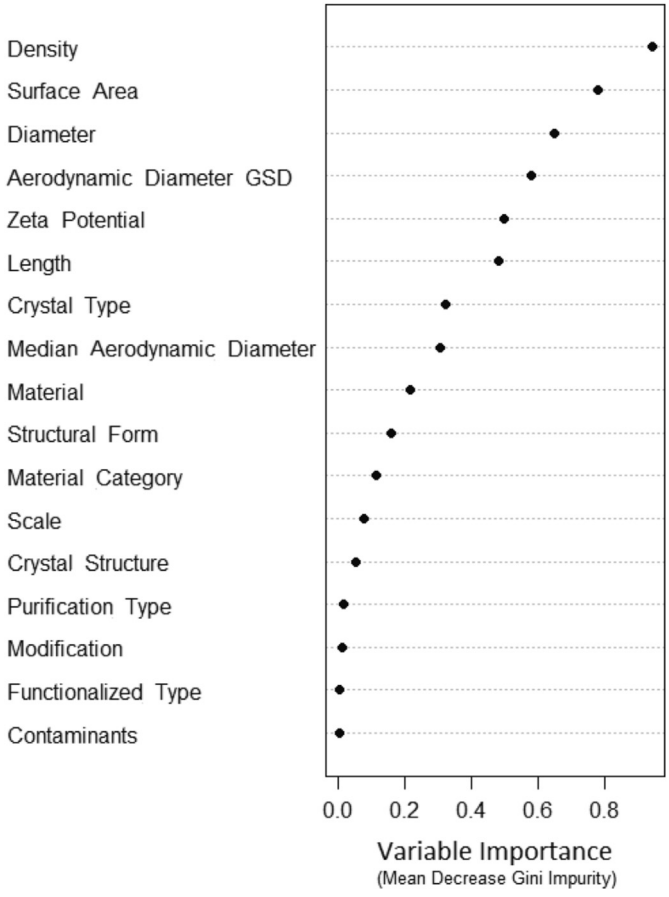

Fig. 6. Physicochemical properties identified as important predictors of potency group.

Each of the NanoGo materials was predicted to belong to Group 1, the most potent cluster, by the random forest model based on their physicochemical properties (Table 4). Chen et al. (2004) have investigated the behavior of random forests when groups are unbalanced, particularly that the largest class tends to receive the most votes, and provide alternatives to improve the predictive accuracy for the minority classes, but these methods were not implemented here. Misclassification of a new material as belonging to a lower potency group could have serious practical implications and the current method appears to be conservative in that sense. Using Stochastic Kriging, a typical potency estimate was found for each of the NanoGo materials by taking the median of the potency estimates across laboratories, where there were up to six potency estimates per material. This median potency estimate was then compared to the four potency groups described in Table 3, and the material was assigned to the nearest potency group. The median potency estimate and nearest potency group (Actual Group) are shown in Table 4. For five of the six materials, the group assignment using the dose-response information matched the predicted group assignment using only the physicochemical properties. An exception was the anatase nanosphere material, which was predicted to belong to a more potent cluster (Group 1) than the potency cluster identified by the dose-response information (Group 2). However, this may be explained by the anatase nanosphere material having only one potency estimate out of the six labs, whereas each of the other materials had at least three estimates of potency, and those estimates tended to vary widely. A summary of the NanoGo potency estimates can be found in the Supplemental Materials (Tables S–2).

4. Discussion

4.1. Overview of the framework

In this study, we illustrate a quantitative framework to evaluate the hazard potency across a set of nanomaterials and use physicochemical property data to predict the hazard potency of other nanomaterials. This is a proof-of-concept evaluation. More comprehensive data are needed to further evaluate this framework across a larger set of materials, dose metrics, and response endpoints. The grouping framework described here can utilize information from data-rich materials to estimate the hazard potency of untested materials with similar properties. Ultimately, the group-based BMDL estimates could be used as the PODs in standard risk assessment methods to derive categorical OELs (Fig. 1).

The current analyses use dose-response data in rodent studies to predict pulmonary inflammation hazard potency groups. The results show that this framework was useful in identifying important physicochemical properties (Fig. 6), which have also been shown in a number of studies to predict pulmonary inflammation and related pulmonary responses (Monteiller et al., 2007; Donaldson et al., 2008; Rushton et al., 2010; Pauluhn, 2011; Porter et al., 2013; Gernand and Casman, 2014). Several standard modeling methods are utilized in this framework to develop hazard potency groups and then to predict the group for a new ENM based on physicochemical properties only. We believe that this is a new contribution to the current scientific literature. The set of methods utilized in this framework is unique compared to currently published frameworks by having the ability to provide quantitative, group-based BMD and BMDL estimates, which could serve as "points of departure" to link to standard risk assessment and OEL derivation methods.

Recent publications which use some of the same methods include Gernand and Casman (2014), who used a random forest model to identify the important physicochemical properties

Table 4
Summary and evaluation of random forest predictions.

Material	Distribution of group votes in the forest				Predicted group	Median BMD	Actual group
	1	2	3	4			
Anatase Nanospheres	409 (82%)	70 (14%)	10 (2%)	11 (2%)	1	204.7	2
Anatase Nanobelt	436 (87%)	44 (9%)	2 (0%)	18 (4%)	1	51.5	1
Anatase/Rutile Nanospheres	421 (84%)	56 (11%)	12 (2%)	11 (2%)	1	37.2	1
Original MWCNT	387 (77%)	102 (20%)	0 (0%)	11 (2%)	1	27.6	1
Purified MWCNT	328 (66%)	162 (32%)	0 (0%)	10 (2%)	1	62.3	1
Functionalized MWCNT	413 (83%)	74 (15%)	2 (0%)	11 (2%)	1	52.6	1

associated with pulmonary toxicity of carbon nanotubes. Oosterwijk et al. (2016) describe a conceptual framework using lung dosimetry models to extrapolate rodent pulmonary responses to humans. Hristozov et al. (2016) used some of the same data (ENPRA) to estimate BMDLs in rodents as IT mass lung doses, then estimated inhalation BMDLs by assuming that the instilled bolus spreads over one volume of breath. The worker-equivalent estimates of these single-day exposures were used in a margin-of-exposure analysis based on model-predicted occupational exposures. Pauluhn (2011, 2014) developed decision criteria and a model for one specific category of particles (low solubility and low toxicity), for which NOAELs for the respirable particle volume associated with rat lung clearance overloading is estimated based on the particle density, ventilation rate, and deposition fraction; OELs would be derived after application of adjustment factors.

Most of the nanotoxicology studies to date are acute *in vivo or in vitro* studies, and this framework provides a systematic method to utilize these data in estimating hazard potency. A growing number of studies have shown concordance between *in vitro* and acute *in vivo* biomarkers for pulmonary inflammation or fibrosis following exposure to engineered nanomaterials (Donaldson et al., 2008; Labib et al., 2016; Monteiller et al., 2007; Rushton et al., 2010; Wang et al., 2015; Wiemann et al., 2016; Zhang et al., 2012). However, these studies have generally not been evaluated for their utility in risk assessment.

Pulmonary inflammation is a well-studied mechanism in the development of chronic lung diseases associated with occupational exposure to respirable particles, including nanoscale and microscale particles (IARC, 2010; NIOSH, 2008, 2011). Pulmonary inflammation is clinically relevant in humans. Some of the materials in this analysis (including TiO_2 and crystalline silica) have been studied in subchronic or chronic *in vivo* studies in rodents and in workers and therefore serve as benchmark materials (Kuempel et al., 2012), providing linkages between early and later biological responses for further predictive modeling.

Although the data available for this initial analysis were limited, the findings were consistent with what is known from the literature. The least potent material tested was microscale (fine) TiO_2 (Group 4), followed in Group 3 by nanoscale (ultrafine) TiO_2 – which scaled by the total particle surface area dose as previously observed (NIOSH, 2011). Microscale crystalline silica was grouped in the most potent cluster (Group 1). A nano-scale equivalent of silica was not available for comparison. All of the other nanoscale metal oxide or carbonaceous nanomaterials were in the higher potency Groups 1 or 2 (Fig. 5). Fine and ultrafine TiO_2 and crystalline silica, which are relatively well-studied materials, are useful benchmark materials (or negative and positive controls, respectively) for comparison to the lung inflammation potency of ENMs. Although measurements of solubility were not available in these data, ZnO is known to be soluble and was the most potent in eliciting acute pulmonary inflammation. These findings are generally consistent with the four *a priori* categories of particles (Poorly Soluble Low Toxicity, Soluble, High Aspect Ratio, Highly Toxic) (Section 2.4). Further grouping or subgrouping of these initial categories may be useful to explore specific models where available (e.g., mechanistic models for metal oxides pertaining to reactivity, protein adsorption, and cell membrane adhesion) (Burello, 2013). This framework could be useful in developing categorical OELs or OEBs for hazard and control banding, which typically include four or five order-of-magnitude bands (Naumann et al., 1996; ISO, 2016; NIOSH, 2017).

4.2. Evaluation of the data and methods

A number of challenges were encountered in the development of this evidence-based framework. First, the toxicology data that were available for these analyses are heterogeneous in several experimental variables, including method of exposure, species/strain/sex, duration of exposure, and biological response measures. Differences in BAL and PMN counting methods across laboratory were addressed by using the PMN proportion response (as reported in Bonner et al., (2013)). Heterogeneity in the method of exposure and the animal species/stain/sex are addressed by normalizing the dose as particle mass per mass of lung tissue (μg/g). It is relevant to note that this dose metric appears to account for these differences at least to some extent given that the estimated deposited dose is an important predictor of pulmonary inflammation (PMN proportion), while strain, sex, species, and method of exposure do not appear to have much influence on the response (Fig. 3).

Evaluation of other dose metrics would be useful in future investigations with more comprehensive data. For example, particle surface area dose was shown to be the best predictor of acute pulmonary inflammation (as % PMN) in rats and mice exposed by IT to various types of nanoparticles (Schmid and Stoeger, 2016); and particle volume dose has been used to predict the NOAEL for overloading of pulmonary clearance in rats exposed to various types of poorly-soluble particles (Pauluhn, 2011, 2014). Information on *in vivo* dissolution of particles and species-specific clearance rates would be useful in future analyses, especially with chronic data of various types of materials. Dose rate has been shown to influence overloading and acute pulmonary inflammatory responses in rats, with greater inflammation at a high dose rate compared to a low dose rate (Baisch et al., 2014). The BMR of an added 4% PMN in this study is similar to that associated with the initiation of overloading and so is biologically relevant without being associated with high lung doses.

Post-exposure duration was shown to be an important variable influencing pulmonary inflammation (Fig. 3). Early attempts to model the complex nonlinear post-exposure duration - dose - response surfaces were not successful due to variability, and so the data were stratified by post-exposure duration (Section 2.2). Post-exposure duration data provide information on the resolution or persistence of the pulmonary inflammation after the end of exposure, and thus to the long-term effects of the material in the lungs (Johnston et al., 2000). Current analyses were limited to the 0–3 day post-exposure duration strata, which included the greatest

number of experiments (about half of the data). Thus, of the 1557 unique individual animals with PMN response across 25 studies available for this analysis, a subset of 844 animals across 22 studies was used for exploring potential groups of materials.

A limitation in these data is that the physicochemical and experimental information was inconsistently available across materials (Fig. S-1). The variables in this compiled dataset include material physicochemical properties, animal descriptors, time factors, and dose-response metrics (Fig. S-1). Except for the ENPRA dataset (which included a table of select physicochemical properties), the physicochemical properties were generally obtained from the publications or from communication with researchers. Some particle properties −solubility, surface charge, volume, and rigidity (of nanotubes) − were of interest *a priori*, based on studies in the literature showing associations with toxicological response, but were not reported in any of the studies included in this dataset. Rodent lung weights were not always available in specific studies and so average estimates were used (Section 2.2).

Although BMDs are the preferred effect levels in quantitative risk assessment (NAS, 2009; U.S. EPA, 2012), methods which compare the entire dose-response relationship may be better suited for identifying similar materials across a range of response endpoints. A BMD estimate represents the lower end of the dose-response relationship at a specified point on the curve. Yet, for a given point, an infinite number of curves could pass through it. This means that while two materials could have similar BMDs, the form of the dose-relationship could be quite different, both in the low dose and high dose regions. It is also clear that dose alone is insufficient to explain the variability in responses across various nanomaterials and experimental designs. Modeling methods that can account for covariates, complex relationships, and heterogeneous data variances are needed to better characterize the dose-response relationships from multiple sources (Pei et al., 2017).

Standard practice in statistical model development is to split the data into training and testing sets. In this analysis, the original dataset (NIOSH, ENPRA, CIIT) assembled was used as the training set, and a subsequently obtained dataset (NanoGo) was used to as the testing set. An alternate method would be to randomly split the combined data (NIOSH, ENPRA, CIIT, and NanoGo) into new training and testing sets, and perhaps all possible combinations therein. Such methods can be considered in the future when a more comprehensive dataset become available for further model development and validation.

In creating the material groups in this analysis, variability in the potency estimate was not taken into account. Rather, the best point estimate, the BMD, was used in the grouping. A method that can account for variability, thereby creating statistically different potency groups, may improve the predictive performance of a classification random forest model. The number of potency groups may also impact predictive performance. In this analysis, a large proportion of materials were classified into one group (Group 1), so using more groups for the classification (e.g., five potency groups instead of four) may contribute to more discernment of the hazard potency across materials. The rationale for the initial four potency groups was to evaluate their relationship with the four biological mode of action categories that have been proposed for ENMs (Section 2.4).

In evaluating the predictive accuracy of the random forest classification model, many algorithms would be expected to be accurate when classifying a new material into Group 1 as 14/18 (~78%) of the training materials were in Group 1, thus one would have a 78% chance to correctly classify a new material by assigning it to Group 1. Creating more potency groups and having more test materials may be helpful in better measuring the predictive accuracy. Expanding the database to include more ENMs representing many different physicochemical properties would provide a more comprehensive database for use in moving beyond the proof-of-concept stage.

The grouping of materials could be sensitive to the biological response endpoint evaluated. The use of other pulmonary endpoints such as LDH or fibrosis severity score may reveal similar or different patterns among materials. Moreover, extending this framework to other endpoints and organ systems would also be useful to more fully evaluate the safety or health hazard potential of ENMs. This framework appears to work well when the BMR was background +4% PMNs in BALF, as 5/6 predictions were correct. However, the experimental data for the grouping was small−18 experiments for the added 4% PMN response, and only 12 experiments for the total 10% PMNs. Using the BMR of 10% total PMNs with this framework was not as successful due to the small number of similar materials in the training set (Tables S−3). Other effect levels such as no observed adverse effect levels (NOAELs) could also be used in this framework, which may be more readily available than BMD estimates that depend on having sufficient dose-response data.

4.3. Minimum data needed for grouping

Essential information in the database for the quantitative analyses in this framework includes the following: (1) sufficient number of dose groups to describe the dose-response relationship and estimate a benchmark dose for endpoint(s) related to human health outcome (as needed for dose-response modeling generally); and (2) a set of physicochemical properties that is sufficient to group particles with regard to biological response. The heterogeneity in experimental design adds variability in describing these relationships but is also reflective of the type of data currently available. Some of these experimental differences can be addressed to some extent in the analysis, for example, by normalizing the lung dose as the estimated deposited mass dose. However, more consistent experimental designs − method of exposure, post-exposure durations, and material descriptors would reduce the "noise" in the data.

Key material descriptors that are needed in these analyses but are not always reported in toxicology studies include particle size distribution, density, and specific surface area measurements. Such data are needed to estimate other metrics of dose such as particle surface area or volume. Solubility was not reported quantitatively in any of these studies but is likely important in the toxicity and persistence of lung effects, as well as the potential for systemic effects. Material form (chemical composition) was a surrogate predictor for some of the unreported properties. Yet, within material type, other factors such as particle size, shape, and functionalization also influenced this hazard grouping based on pulmonary inflammation potency (Fig. 5). More complete information for these variables would facilitate further analyses.

4.4. Use of grouping results in risk assessment and OEL derivation

Benchmark dose modeling is the preferred method over NOAELs to estimate effect levels for quantitative risk assessment (NAS, 2009; U.S. EPA, 2012). The use of BMD estimates to group materials in this analysis provides a direct linkage to standard quantitative risk assessment methods (Fig. 1). Occupational exposure limits are typically mass-based airborne concentrations, and so these group-based BMD potency estimates based on mass lung dose can be readily extrapolated to estimate a human-equivalent concentration (HEC). The lower confidence limit estimates of the BMDs (i.e., BMDLs) from rodent studies (or eventually *in vitro* studies (Crump et al., 2010; Maier, 2011) could be used as the PODs

in risk assessment and OEL derivation (Fig. 1). The BMDL representing a potency group can be estimated, for example, as the 5th percentile of the distribution of BMDLs for the individual experiments within that group. The 5th percentile BMDL can then be used as a health-protective estimate of a new material's potency while being less sensitive to potentially unusually low minimum BMDL estimates (e.g., due to sparse and highly variable data).

Using standard dosimetry methods, the human-equivalent lung dose to the rodent BMDL estimates can be estimated by accounting for the differences in the factors that influence the particle deposition in the respiratory tract (U.S. EPA, 1994; Jarabek et al., 2005). Clearance and retention kinetics, including particle dissolution, would need to be taken into account in estimating the internal doses in rodents or humans with repeated exposures (Pauluhn, 2014; ICRP, 2015; Kuempel et al., 2015). The choice of dose metric should depend on the knowledge of the biological mechanisms.

Once the human-equivalent lung dose is estimated, the HEC (e.g., as an 8-hr time-weighted average airborne particle mass concentration) can be estimated based on the information on worker ventilation rates, exposure duration, and deposition efficiency of airborne particles in the respiratory tract (Kuempel et al., 2015). To estimate an OEL, uncertainty factors are typically applied to account for variability and uncertainty in the use of experimental animal data to estimate an HEC (Dankovic et al., 2015). In the current analysis, the acute inflammatory responses may be applicable to the derivation of acute exposure limits, e.g., Acute Exposure Guideline Levels, AEGLs (NAS, 2001). The group-based BMDL estimates, as described in this paper, would be used as PODs to derive OELs using standard risk assessment methods.

4.5. Conclusions and next steps

Proof of concept has been demonstrated for a quantitative framework to estimate hazard potency for pulmonary inflammation by utilizing toxicology data in rodents and physicochemical properties of nanoscale and microscale particles. Findings were consistent with the previous literature on benchmark particles TiO_2 and crystalline silica, and with potency of ENMs varying by particle size, shape, and chemical composition. Efforts are underway to compile a more comprehensive toxicology dataset for extension and further testing of this framework. More flexible models will be investigated to make better use of all data as it becomes available. Additional biological responses along the hypothesized adverse outcome pathways in the lungs and other organs will be examined. After further validation, this framework will be used to develop categorical OELs for ENMs without individual occupational exposure limits.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

Acknowledgements

The authors would like to express their appreciation to the researchers who generously provided data for this analysis. We would like to thank the researchers in the E.U. ENPRA consortium (E.U. Grant No. 228789) and in the U.S. NanoGo consortium for generating the standard sets of experimental data that were useful in our model development and testing. NIOSH researchers were U.S. collaborators on ENPRA, including Dr. Eileen Kuempel, coauthor on this paper. The NanoGo consortium grants and principle investigators included: NIEHS grants RC2 ES018772 and R01 ES019311 (J. Bonner, NCSU), RC2 ES018741 (A. Elder, G. Oberdörster, UR), RC2 ES197756 (J. Harkema, MSU), R01 ES016189 and P30 ES007033 (T. Kavanagh, UW), RC2 ES018766 (A. Nel, UCLA), and RC1 ES018232 (K. Pinkerton, UCD). We would like to thank the following individual researchers for providing additional data and information used in this analysis: Dr. Lang Tran and Mr. Peter Ritchie, Institute of Occupational Medicine, Edinburgh, U.K. (ENPRA data); Dr. Alison Elder, University of Rochester, Rochester, New York, and Dr. James Bonner, North Carolina State University, Raleigh, North Carolina (NanoGo data); Drs. Edilberto Bermudez and Harvey Clewell, the Hamner Institutes for Health Sciences (formerly CIIT Centers for Health Research), Research Triangle Park, North Carolina (CIIT data); and Drs. Dale Porter, Tina Sager, Jenny Roberts, NIOSH, Health Effects Laboratory Division, Morgantown, West Virginia (NIOSH data). We would like to thank Mr. Randall Smith, NIOSH, Education and Information Division (EID), Risk Evaluation Branch (REB), Cincinnati, Ohio, for useful discussions on some of the statistical methods in this analysis, as well as Dr. Paul Schulte, NIOSH/EID, Dr. Charles Geraci, NIOSH/NTRC, and others in NIOSH for their feedback on earlier presentations of this work. We would like to thank Dr. Lang Tran, Dr. Alison Elder, and Dr. Steve Bertke, NIOSH, Division of Surveillance, Hazard Evaluations, and Field Studies, and Dr. Christine Whittaker, NIOSH/EID/REB, for their helpful review comments on previous versions of this paper. We also thank the two anonymous reviewers for their constructive comments.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.yrtph.2017.08.003.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.yrtph.2017.08.003.

References

ARA, 2011. Multiple-path Particle Deposition (MPPD 2.1): a Model for Human and Rat Airway Particle Dosimetry. Applied Research Associates, Inc, Raleigh, NC.

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., Wiench, K., Wohlleben, W., Landsiedel, R., 2015. A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). Regul. Toxicol. Pharmacol. 71 (2 Suppl. l), S1–S27.

Arts, J.H., 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., 2016. Case studies putting the decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping) into practice. Regul. Toxicol. Pharmacol. 76, 234–261.

Asgharian, B., Price, O.T., Oldham, M., Chen, L.C., Saunders, E.L., Gordon, T., Mikheev, V.B., Minard, K.R., Teeguarden, J.G., 2014. Computational modeling of nanoscale and microscale particle deposition, retention and dosimetry in the mouse respiratory tract. Inhal. Toxicol. 26 (14), 829–842.

Baisch, B.L., Corson, N.M., Wade-Mercer, P., Gelein, R., Kennell, A.J., Oberdörster, G., Elder, A., 2014. Equivalent titanium dioxide nanoparticle deposition by intratracheal instillation and whole body inhalation: the effect of dose rate on acute respiratory tract inflammation. Part Fibre Toxicol. 11, 5.

BAuA, 2013. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Bekanntmachung zu Gefahrstoffen. Hergestellte Nanomatierialien. – German Federal Institute for Occupational Safety and Health. Announcement regarding hazardous substances. Manufactured nanomaterials. BekGS 527 BMBl 2013, 498–511 (Nr. 25). http://www.baua.de/de/Themen-von-A-Z/Gefahrstoffe/TRGS/Bekanntmachung-527.html.

Bermudez, E., Mangum, J.B., Asgharian, B., Wong, B.A., Reverdy, E.E., Janszen, D.B., Hext, P.M., Warheit, D.B., Everitt, J.I., 2002. Long-term pulmonary responses of three laboratory rodent species to subchronic inhalation of pigmentary titanium dioxide particles. Toxicol. Sci. 70, 86–97.

Bermudez, E., Mangum, J.B., Wong, B.A., Asgharian, B., Hext, P.M., Warheit, D.B., Everitt, J.I., 2004. Pulmonary responses of mice, rats, and hamsters to subchronic inhalation of ultrafine titanium dioxide particles. Toxicol. Sci. 77, 347–357.

Bonner, J.C., Silva, R.M., Taylor, A.J., Brown, J.M., Hilderbrand, S.C., Castranova, V., Porter, D., Elder, A., Oberdörster, G., Harkema, J.R., Bramble, L.A., Kavanagh, T.J., Botta, D., Nel, A., Pinkerton, K.E., 2013. Interlaboratory evaluation of rodent pulmonary responses to engineered nanomaterials: the NIEHS Nano GO Consortium. Environ. Health Perspect. 121 (6), 676–682.

Bos, P.M., Gottardo, S., Scott-Fordsmand, J.J., van Tongeren, M., Semenzin, E., Fernandes, T.F., Hristozov, D., Hund-Rinke, K., Hunt, N., Irfan, M.A., Landsiedel, R., Peijnenburg, W.J., Sánchez Jiménez, A., van Kesteren, P.C., Oomen, A.G., 2015. The MARINA risk assessment strategy: a flexible strategy for efficient information collection and risk assessment of nanomaterials. Int. J. Environ. Res. Public Health 12 (12), 15007–15021.

Braakhuis, H.M., Oomen, A.G., Cassee, F.R., 2016. Grouping nanomaterials to predict their potential to induce pulmonary inflammation. Toxicol. Appl. Pharmacol. 299, 3–7.

Breiman, L., Freidman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman and Hall, New York.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

BSI, 2007. Nanotechnologies, Part 2. PD 6699-2:2007: Guide to Safe Handling and Disposal of Manufactured Nanomaterials. British Standards Institution, London.

Burello, E., 2013. Profiling the biological activity of oxide nanomaterials with mechanistic models. Comput. Sci. Discov. 6, 014009.

Burello, E., Worth, A.P., 2011. QSAR modeling of nanomaterials. Nanomed Nanobiotechnol 3, 298–306.

Chen, C., Liaw, A., Breiman, L., 2004. Using Random Forest to Learn Unbalanced Data. Technical Report 666. University of California at Berkeley, Department of Statistics, Berkeley, CA. http://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf.

Cohen, Y., Rallo, R., Liu, R., Liu, H.H., 2013. In silico analysis of nanomaterials hazard and risk. Acc. Chem. Res. 46 (3), 802–812.

Crump, K.S., 1984. A new method for determining allowable daily intakes. Fund. Appl. Toxicol. 4, 854–871.

Crump, K.S., 2002. Critical issues in benchmark calculations from continuous data. Crit. Rev. Toxicol. 32 (3), 133–153.

Crump, K.S., Chen, C., Louis, T.A., 2010. The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions. Environ. Health Perspect. 118 (10), 1350–1354.

Crystal, R.G., Gadek, J.E., Ferrans, V.J., Fulmer, J.D., Line, B.R., Hunninghake, G.W., 1981. Interstitial lung disease: current concepts of pathogenesis, staging, and therapy. Am. J. Med. 70, 542–568.

Dankovic, D.A., Naumann, B.D., Maier, A., Dourson, M.L., Levy, L.S., 2015. The scientific basis of uncertainty factors used in setting occupational exposure limits. J. Occup. Environ. Hyg. 12 (Suppl. 1), S55–S68.

Donaldson, K., Borm, P.J., Oberdörster, G., Pinkerton, K.E., Stone, V., Tran, C.L., 2008. Concordance between in vitro and in vivo dosimetry in the proinflammatory effects of low-toxicity, low-solubility particles: the key role of the proximal alveolar region. Inhal. Toxicol. 20, 53–62.

Deveau, M., Chen, C.P., Johanson, G., Krewski, D., Maier, A., Niven, K.J., Ripple, S., Schulte, P.A., Silk, J., Urbanus, J.H., Zalk, D.M., Niemeier, R.W., 2015. The global landscape of occupational exposure limits—implementation of harmonization principles to guide limit selection. J. Occup. Environ. Hyg. 12 (Suppl. 1), S127–S144.

ECHA, 2017. Guidance on Information Requirements and Chemical Safety Assessment. Appendix R.6–1 for Nanomaterials Applicable to the Guidance on QSARs and Grouping of Chemicals. Reference No. ECHA-17-G-17-EN. European Chemicals Agency, Helsinki, Finland, Version 1.0. .

Elder, A., Gelein, R., Finkelstein, J.N., Driscoll, K.E., Harkema, J., Oberdörster, G., 2005. Effects of subchronically inhaled carbon black in three species. I. Retention kinetics, lung inflammation, and histopathology. Toxicol. Sci. 88 (2), 614–629.

ENPRA, 2013. Final Report Summary—ENPRA (Risk Assessment of Engineered Nanoparticles). April 29, 2013. Engineered NanoParticle Risk Assessment. European Commission Community Research and Development Information Service. http://cordis.europa.eu/result/rcn/56040_en.html. (Accessed 23 January 2017).

Environmental Defense and Dupont, 2007. Nano Risk Framework. Environmental Defense Fund, Washington, DC. http://business.edf.org/projects/featured/past-projects/dupont-safer-nanotech/.

Gebel, T., Foth, H., Damm, G., Freyberger, A., Kramer, P.J., Lilienblum, W., Röhl, C., Schupp, T., Weiss, C., Wollin, K.M., Hengstler, J.G., 2014. Manufactured nanomaterials: categorization and approaches to hazard assessment. Arch. Toxicol. 88 (12), 2191–2211.

Gernand, J.M., Casman, E.A., 2014. A meta-analysis of carbon nanotube pulmonary toxicity studies—how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal. 34 (3), 583–597.

Gernand, J.M., Casman, E.A., 2016. Nanoparticle characteristic interaction effects on pulmonary toxicity: a random forest modeling framework to compare risks of nanomaterial variants. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng. 2, 021002-1 to 021002-13.

Gkika, D.A., Nolan, J.W., Vansant, E.F., Vordos, N., Kontogoulidou, C., Mitropoulos, A.C., Cool, P., Braet, J., 2017. A framework for health-related nanomaterial grouping. Biochim. Biophys. Acta 1861 (6), 1478–1485.

Godwin, H., Nameth, C., Avery, D., Bergeson, L.L., Bernard, D., Beryt, E., Boyes, W., Brown, S., Clippinger, A.J., Cohen, Y., Doa, M., Hendren, C.O., Holden, P., Houck, K., Kane, A.B., Klaessig, F., Kodas, T., Landsiedel, R., Lynch, I., Malloy, T., Miller, M.B., Muller, J., Oberdorster, G., Petersen, E.J., Pleus, R.C., Sayre, P., Stone, V., Sullivan, K.M., Tentschert, J., Wallis, P., Nel, A.E., 2015. Nanomaterial categorization for assessing risk potential to facilitate regulatory decision-making. ACS Nano 9 (4), 3409–3417.

Gordon, S.C., Butala, J.H., Carter, J.M., Elder, A., Gordon, T., Gray, G., Sayre, P.G., Schulte, P.A., Tsai, C.S., West, J., 2014. Workshop report: strategies for setting occupational exposure limits for engineered nanomaterials. Regul. Toxicol. Pharmacol. 68 (3), 305–311.

Grieger, K.D., Redmon, J.H., Money, E.S., Widder, M.W., van der Schalie, W.H., Beaulieu, S.M., Womak, D., 2015. A relative ranking approach for nano-enabled applications to improve risk-based decision making: a case study of Army materiel. Environ. Syst. Decis. 35, 42–53.

Hristozov, D.R., Zabeo, A., Foran, C., Isigonis, P., Critto, A., Marcomini, A., Linkov, I., 2014. A weight of evidence approach for hazard screening of engineered nanomaterials. Nanotoxicology 8 (1), 72–87.

Hristozov, D., Zabeo, A., Alstrup Jensen, K., Gottardo, S., Isigonis, P., Maccalman, L., Critto, A., Marcomini, A., 2016. Demonstration of a modelling-based multi-criteria decision analysis procedure for prioritization of occupational risks from manufactured nanomaterials. Nanotoxicology 10 (9), 1215–1228.

IARC, 2010. IARC monographs on the evaluation of carcinogenic risks to humans. In: Carbon Black, Titanium Dioxide, and Talc. International Agency for Research on Cancer, World Health Organization, vol. 93. WHO Press, Geneva, Switzerland.

ICRP, 2015. Occupational intakes of radionuclides: Part 1. The international commission on radiological protection (ICRP) publication, 130 Ann. ICRP 44 (2).

ILSI (International Life Sciences Institute), 2000. The relevance of the rat lung response to particle overload for human risk assessment: a workshop consensus report. Inhal. Toxicol. 12, 1–17.

ISO, 2016. Nanotechnologies — Overview of Available Frameworks for the Development of Occupational Exposure Limits and Bands for Nano-objects and Their Aggregates and Agglomerates (NOAAs). International Organization for Standardization (ISO) Technical Report. ISO/TR 18637, published Nov. 21. ISO, Geneva, Switzerland.

Jarabek, A.M., Asgharian, B., Miller, F.J., 2005. Dosimetric adjustments for interspecies extrapolation of inhaled poorly soluble particles (PSP). Inhal. Toxicol. 17 (7–8), 317–334.

Johnston, C.J., Driscoll, K.E., Finkelstein, J.N., Baggs, R., O'Reilly, M.A., Carter, J., Gelein, R., Oberdörster, G., 2000. Pulmonary chemokine and mutagenic responses in rats after subchronic inhalation of amorphous and crystalline silica. Toxicol. Sci. 56 (2), 405–413.

Kobayashi, N.1, Naya, M., Endoh, S., Maru, J., Yamamoto, K., Nakanishi, J., 2009. Comparative pulmonary toxicity study of nano-TiO(2) particles of different sizes and agglomerations in rats: different short- and long-term post-instillation results. Toxicology 264 (1–2), 110–118.

Kuempel, E.D., Tran, C.L., Castranova, V., Bailer, A.J., 2006. Lung dosimetry and risk assessment of nanoparticles: evaluating and extending current models in rats and humans. Inhal. Toxicol. 18 (10), 717–724.

Kuempel, E.D., Castranova, V., Geraci, C.L., Schulte, P.A., 2012. Development of risk-based nanomaterial groups for occupational exposure control. J. Nanopart Res. 14, 1029.

Kuempel, E.D., Sweeney, L.M., Morris, J.B., Jarabek, A.M., 2015. Advances in inhalation dosimetry models and methods for occupational risk assessment and exposure limit derivation. J. Occup. Environ. Hyg. 12 (Suppl. 1), S18–S40.

Labib, S., Williams, A., Yauk, C.L., Nikota, J.K., Wallin, H., Vogel, U., Halappanavar, S., 2016. Nano-risk science: application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. Part Fibre Toxicol. 13, 15.

Liu, R., Rallo, R., George, S., Ji, Z., Nair, S., Nel, A.E., Cohen, Y., 2011. Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles. Small 7 (8), 1118–1126.

Liu, R., Zhang, H.Y., Ji, Z.X., Rallo, R., Xia, T., Chang, C.H., Nel, A., Cohen, Y., 2013. Development of structure-activity relationship for metal oxide nanoparticles. Nanoscale 5 (12), 5644–5653.

Maier, M.S., 2011. Setting occupational exposure limits for unstudied pharmaceutical intermediates using an in vitro parallelogram approach. Toxicol. Mech. Methods 21 (2), 76–85.

MATLAB, 2016. MATLAB and Statistics Toolbox, Release 2016a. The MathWorks, Inc, Natick, MA.

Martin, T.R., Ganesh, R., Maunder, R.J., Springmeyer, S.C., 1985. The effects of chronic bronchitis and chronic air-flow obstruction on lung cell populations recovered by bronchoalveolar lavage. Am. Rev. Respir. Dis. 132, 254–260.

Mihalache, R., Verbeek, J., Graczyk, H., Murashov, V., van Broekhuizen, P., 2017. Occupational exposure limits for manufactured nanomaterials, a systematic review. Nanotoxicology 9, 1–13.

Monteiller, C.1, Tran, L., MacNee, W., Faux, S., Jones, A., Miller, B., Donaldson, K., 2007. The pro-inflammatory effects of low-toxicity low-solubility particles, nanoparticles and fine particles, on epithelial cells in vitro: the role of surface area. Occup. Environ. Med. 64 (9), 609–615.

Morfeld, P., Bruch, J., Levy, L., Ngiewih, Y., Chaudhuri, I., Muranko, H.J., Myerson, R., McCunney, R.J., 2015. Translational toxicology in setting occupational exposure limits for dusts and hazard classification - a critical evaluation of a recent approach to translate dust overload findings from rats to humans. Part Fibre Toxicol. 12 (3).

Morrow, P.E., 1988. Possible mechanisms to explain dust overloading of the lungs. Fund. Appl. Toxicol. 10, 369–384.

Muhle, H., Bellmann, B., Creutzenberg, O., Dasenbrock, C., Ernst, H., Kilpper, R., MacKenzie, J.C., Morrow, P., Mohr, U., Takenaka, S., Mermelstein, R., 1991. Pulmonary response to toner upon chronic inhalation exposure in rats. Fund. Appl.

Toxicol. 17, 280–299.
Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 1996. Correlation of structural class with No-Observed-Effect levels: a proposal for establishing a threshold of concern. Food Chem. Toxicol. 34, 829–867.
NAS, 1983. Risk Assessment in the Federal Government: Managing the Process. Committee on the Institutional Means for Assessment of Risks to Public Health, Commission on Life Sciences, National Research Council, National Academy of Sciences. National Academies Press, Washington, DC.
NAS, 2001. Standing Operating Procedures for Developing Acute Exposure Guideline Levels for Hazardous Chemicals. National Academy of Sciences. National Academies Press, Washington, DC.
NAS, 2007. Toxicity Testing in the 21st Century: a Vision and a Strategy. National Academy of Sciences. National Academies Press, Washington, DC.
NAS, 2009. Science and Decisions: Advancing Risk Assessment. Committee on Improving Risk Analysis Approaches Used by the U.S. EPA, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council, National Academy of Sciences. National Academies Press, Washington, DC.
NAS, 2017. Using 21st Century Science to Improve Risk-related Evaluation. National Academy of Sciences. National Academies Press, Washington, DC.
Naumann, B.D., Sargent, E.V., Starkman, B.S., Fraser, W.J., Becker, G.T., Kirk, G.D., 1996. Performance-based exposure control limits for pharmaceutical active ingredients. Am. Ind. Hyg. Assoc. J. 57, 33–42.
Nel, A.E., Nasser, E., Godwin, H., Avery, D., Bahadori, T., Bergeson, L., Beryt, E., Bonner, J.C., Boverhof, D., Carter, J., Castranova, V., Deshazo, J.R., Hussain, S.M., Kane, A.B., Klaessig, F., Kuempel, E., Lafranconi, M., Landsiedel, R., Malloy, T., Miller, M.B., Morris, J., Moss, K., Oberdorster, G., Pinkerton, K., Pleus, R.C., Shatkin, J.A., Thomas, R., Tolaymat, T., Wang, A., Wong, J., 2013. A multi-stakeholder perspective on the use of alternative test strategies for nanomaterial safety assessment. ACS Nano 7 (8), 6422–6433.
NIOSH, 2008. Work-related Lung Disease Surveillance Report. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Morgantown, WV. DHHS (NIOSH) Publication No. 2008–2143.
NIOSH, 2011. Current Intelligence Bulletin 63: Occupational Exposure to Titanium Dioxide. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, OH. DHHS (NIOSH) Publication No. 2011–2160.
NIOSH, 2013. Current Intelligence Bulletin 65: Occupational Exposure to Carbon Nanotubes and Nanofibers. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, OH. DHHS (NIOSH) Publication No. 2013–2014.
NIOSH, 2017. Draft Current Intelligence Bulletin: the Occupational Exposure Banding Process: Guidance for the Evaluation of Chemical Hazards. Department of Health and Human Services (HHS), Centers for Disease Control and Prevention (CDC), National Institute for Occupational Safety and Health (NIOSH). March 15, 2017.
Oberdörster, G., Ferin, J., Soderholm, S., Gelein, R., Cox, C., Baggs, R., Morrow, P.E., 1994. Increased pulmonary toxicity of inhaled ultrafine particles: due to lung overload alone? Ann. Occup. Hyg. 38, 295–302.
OECD, 2007. Guidance on Grouping of Chemicals. Series on Testing and Assessment, No. 80. ENV/JM/MONO(2007)28. Organization for Economic Cooperation and Development, Environmental Health and Safety Publications, Paris.
OECD, 2012. Important Issues on Risk Assessment of Manufactured Nanomaterials. Series on the Safety of Manufactured Nanomaterials, No. 33. ENV/JM/MONO(2012)8. Organization for Economic Cooperation and Development, Environmental Health and Safety Publications, Paris, France.
OECD, 2014. Guidance on Grouping of Chemicals. Series on Testing and Assessment, No. 194. ENV/JM/MONO(2014)4, second ed. Organization for Economic Cooperation and Development, Environmental Health and Safety Publications, Paris, France.
OECD, 2016. Alternative Testing Strategies in Risk Assessment of Manufactured Nanomaterials: Current State of Knowledge and Research Needs to Advance Their Use. Series on the Safety of Manufactured Nanomaterials, No. 80. ENV/JM/MONO(2016)63. Organization for Economic Cooperation and Development, Environmental Health and Safety Publications, Paris, France.
Oh, E., Liu, R., Nel, A., Gemill, K.B., Bilal, M., Cohen, Y., Medintz, I.L., 2016. Meta-analysis of cellular toxicity for cadmium-containing quantum dots. Nat. Nanotechnol. 11 (5), 479–486.
Oomen, A.G., Bleeker, E.A., Bos, P.M., van Broekhuizen, F., Gottardo, S., Groenewold, M., Hristozov, D., Hund-Rinke, K., Irfan, M.A., Marcomini, A., Peijnenburg, W.J., Rasmussen, K., Jiménez, A.S., Scott-Fordsmand, J.J., van Tongeren, M., Wiench, K., Wohlleben, W., Landsiedel, R., 2015. Grouping and read-across approaches for risk assessment of nanomaterials. Int. J. Environ. Res. Public Health 12 (10), 13415–13434.
Oosterwijk, M.T., Feber, M.L., Burello, E., 2016. Proposal for a risk banding framework for inhaled low aspect ratio nanoparticles based on physicochemical properties. Nanotoxicology 10 (6), 780–793.
Pauluhn, J., 2011. Poorly soluble particulates: searching for a unifying denominator of nanoparticles and fine particles for DNEL estimation. Toxicology 279 (1–3), 176–188.
Pauluhn, 2012. Subchronic inhalation toxicity of iron oxide (magnetite, Fe_3O_4) in rats: pulmonary toxicity is determined by the particle kinetics typical of poorly soluble particles. J. Appl. Toxicol. 32, 488–504.
Pauluhn, J., 2014. Derivation of occupational exposure levels (OELs) of low-toxicity isometric biopersistent particles: how can the kinetic lung overload paradigm be used for improved inhalation toxicity study design and OEL-derivation? Part Fibre Toxicol. 11, 72.
Pei, Y., Yang, F., Chen, X., Wu, N., Wang, K., 2017. Kriging-based design of experiments for multi-source exposure-response studies in nanotoxicology. ACS Sustain. Chem. Eng. Article ASAP. http://dx.doi.org/10.1021/acssuschemeng.6b02981.
Porter, D.W., Ramsey, D., Hubbs, A.F., Battelli, L., Ma, J., Barger, M., Landsittel, D., Robinson, V.A., McLaurin, J., Khan, A., Jones, W., Teass, A., Castranova, V., 2001. Time course of pulmonary response of rats to inhalation of crystalline silica: histological results and biochemical indices of damage, lipidosis, and fibrosis. J. Environ. Pathol. Toxicol. Oncol. 20 (Suppl. 1), 1–14.
Porter, D.W., Hubbs, A.F., Mercer, R., Robinson, V.A., Ramsey, D., McLaurin, K., Khan, A., Battelli, L., Brumbaugh, K., Teass, A., Castranova, V., 2004. Progression of lung inflammation and damage in rats after cessation of silica inhalation. Toxicol. Sci. 79, 370–380.
Porter, D.W., Wu, N., Hubb, A.F., Mercer, R.R., Funk, K., Meng, F., Li, J., Wolfarth, M.G., Battelli, L., Friend, S., Andrew, M., Hamilton, R., Sriram, K., Yang, F., Castranova, V., Holian, A., 2013. Differential mouse pulmonary dose and time course responses to titanium dioxide nanospheres and nanobelts. Toxicol. Sci. 131 (1), 179–193.
R Core Team, 2014. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Raabe, O.G., Al-Bayati, M.A., Teague, S.V., Rasolt, A., 1988. Regional deposition of inhaled monodisperse coarse and fine aerosol particles in small laboratory animals. Ann. Occup. Hyg. 32, 53–63.
Roberts, J.R., McKinney, W., Kan, H., Krajnak, K., Frazer, D.G., Thomas, T.A., Waugh, S., Kenyon, A., MacCuspie, R.I., Hackley, V.A., Castranova, V., 2013. Pulmonary and cardiovascular responses of rats to inhalation of silver nanoparticles. J. Toxicol. Environ. Health A 76 (11), 651–668.
Rom, W.L., 1991. Relationship of inflammatory cell cytokines to disease severity in individuals with occupational inorganic dust exposure. Am. J. Indus Med. 19, 15–27.
Rushton, E.K., Jiang, J., Leonard, S.S., Eberly, S., Castranova, V., Biswas, P., Elder, A., Han, X., Gelein, R., Finkelstein, J., Oberdörster, G., 2010. Concept of assessing nanoparticle hazards considering nanoparticle dosemetric and chemical/biological response metrics. J. Toxicol. Environ. Health A 73 (5), 445–461.
Sager, T.M., Wolfarth, M.W., Andrew, M., Hubbs, A., Friend, S., Chen, T., Porter, D.W., Wu, N., Yang, F., Hamilton, R.F., Holian, A., 2013. Effect of multi-walled carbon nanotube surface modification on bioactivity in the C57BL/6 mouse model. Nanotoxicology 8 (3), 317–327.
Schmid, O., Stoeger, T., 2016. Surface area is the biologically most effective dose metric for acute nanoparticle toxicity in the lung. J. Aerosol Sci. 99, 133–143.
Schulte, P.A., Murashov, V., Zumwalde, R., Kuempel, E.D., Geraci, C.L., 2010. Occupational exposure limits for nanomaterials: state of the art. J. Nanopart Res. 12 (6), 1971–1987.
Shao, K., Shapiro, A., 2016. A next generation benchmark dose computation system. Poster presentation at Society of Toxicology 55th Annual Meeting, March 13-16, New Orleans, LA. Abstract #2165 Toxicology 150 (1), 274.
Shatkin, J. (Ed.), 2013. Nanotechnology Health and Environmental Risks, second ed. CRC Press, Taylor & Francis, Boca Raton, FL.
Stone, V., Pozzi-Mucelli, S., Tran, L., Aschberger, K., Sabella, S., Vogel, U.B., Poland, C., Balharry, D., Fernandes, T., Gottardo, S., Hankin, S., Hartl, M.G., Hartmann, N., Hristozov, D., Hund-Rinke, K., Johnston, H., Marcomini, A., Panzer, O., Roncato, D., Saber, A.T., Wallin, H., Scott-Fordsmand, J.J., 2014. ITS-NANO – prioritizing nanosafety research to develop a stakeholder driven intelligent testing strategy. Part Fibre Toxicol. 11, 9.
Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinforma. 9, 307.
Tran, C.L., Cullen, R.T., Buchanan, D., Jones, A.D., Miller, B.G., Searl, A., Davis, J.M.G., Donaldson, K., 1999. Investigation and prediction of pulmonary responses to dust. Part II. In: Investigations into the Pulmonary Effects of Low Toxicity Dusts. Parts I and II. Health and Safety Executive, Suffolk, UK. Contract Research Report 216/1999.
Tran, C.L., Buchanan, D., Cullen, R.T., Searl, A., Jones, A.D., Donaldson, K., 2000. Inhalation of poorly soluble particles. II. Influence of particle surface area on inflammation and clearance. Inhal. Toxicol. 12 (12), 1113–1126.
U.S. EPA, 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. EPA/600/8–90/066F. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. http://nepis.epa.gov/EPA/html/Pubs/pubtitleORD.html.
U.S. EPA, 2012. Benchmark Dose Technical Guidance. EPA/100/R-12/001. U.S. Environmental Protection Agency, Washington, DC.
U.S. EPA, 2015. Benchmark Dose Software (BMDS). U.S. Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC, version 2.6. .
Walser, T., Studer, C., 2015. Sameness: the regulatory crux with nanomaterial identity and grouping schemes for hazard assessment. Regul. Toxicol. Pharmacol. 72 (3), 569–571.
Wang, K., Chen, X., Yang, F., Porter, D.W., Wu, N., 2014. A new stochastic kriging method for modeling multi-source exposure-response data in toxicology studies. ACS Sustain Chem. Eng. 2, 1581–1591.
Wang, X., Duch, M.C., Mansukhani, N., Ji, Z., Liao, Y.P., Wang, M., Zhang, H., Sun, B., Chang, C.H., Li, R., Lin, S., Meng, H., Xia, T., Hersam, M.C., Nel, A.E., 2015. Use of a

pro-fibrogenic mechanism-based predictive toxicological approach for tiered testing and decision analysis of carbonaceous nanomaterials. ACS Nano 9 (3), 3032–3043.

Wiemann, M., Vennemann, A., Sauer, U.G., Wiench, K., Ma-Hock, L., Landsiedel, R., 2016. An in vitro alveolar macrophage assay for predicting the short-term inhalation toxicity of nanomaterials. J. Nanobiotechnol 14, 16.

Wignall, J.A., Shapiro, A.J., Wright, F.A., Woodruff, T.J., Chiu, W.A., Guyton, K.Z., Rusyn, I., 2014. Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. Environ. Health Perspect. 122 (5), 499–505.

Xia, T., Zhao, Y., Sager, T., George, A., Pokhrel, S., Li, N., Schoenfeld, D., Meng, H., Lin, S., Wang, X., Wang, M., Ji, Z., Zink, J.I., Madler, L., Castranova, V., Lin, S., Nel, A.E., 2011. Decreased dissolution of ZnO by iron doping yields nanoparticles with reduced toxicity in the rodent lung and zebrafish embryos. ACS Nano 5 (2), 1223–1235.

Xia, T., Hamilton, R.F., Bonner, J.C., Crandall, E.D., Elder, A., Fazlollahi, F., Girtsman, T.A., Kim, K., Mitra, S., Ntim, S.A., Orr, G., Tagmount, M., Taylor, A.J., Telesca, D., Tolic, A., Vulpe, C.D., Walker, A.J., Wang, X., Witzmann, F.A., Wu, N., Xie, Y., Zink, J.I., Nel, A., Holian, A., 2013. Interlaboratory evaluation of in vitro cytotoxicity and inflammatory responses to engineered nanomaterials: the NIEHS Nano GO Consortium. Environ. Health Perspect. 121 (6), 683–690.

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., Pokhrel, S., Lin, S., Wang, X., Liao, Y.P., Wang, M., Li, L., Rallo, R., Damoiseaux, R., Telesca, D., Mädler, L., Cohen, Y., Zink, J.I., Nel, A.E., 2012. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. ACS Nano 6 (5), 4349–4368.

Zuin, S., Micheletti, C., Critto, A., Pojana, G., Johnston, H., Stone, V., Tran, L., Marcomini, A., 2011. Weight of evidence approach for the relative hazard ranking of nanomaterials. Nanotoxicology 5 (3), 445–458.