Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2012 *Paper* 90

Hierarchical Rank Aggregation with Applications to Nanotoxicology

Trina Patel[*] Donatello Telesca[†] Robert Rallo[‡]
Saji George[**] Xia Tian[††] Nel Andre[‡‡]

[*]UCLA, Biostatistics, trinapatel@gmail.com
[†]UCLA, Biostatistics, dtelesca@ucla.edu
[‡]Universitat Rovira i Virgili
[**]UCLA
[††]UCLA
[‡‡]UCLA

Hierarchical Rank Aggregation with Applications to Nanotoxicology

Trina Patel, Donatello Telesca, Robert Rallo, Saji George, Xia Tian, and Nel Andre

Abstract

The development of high throughput screening (HTS) assays in the field of nanotoxicology provide new opportunities for the hazard assessment and ranking of engineered nanomaterials (ENM). It is often necessary to rank lists of materials based on multiple risk assessment parameters, often aggregated across several measures of toxicity and possibly spanning an array of experimental platforms. Bayesian models coupled with the optimization of loss functions have been shown to provide an effective framework for conducting inference on ranks. In this article we present various loss function based ranking approaches for comparing ENM within experiments and toxicity parameters. Additionally, we propose a framework for the aggregation of ranks across different sources of evidence while allowing for differential weighting of this evidence based on its reliability and importance in risk ranking. We apply these methods to high throughput toxicity data on 2 human cell lines, exposed to 8 different nanomaterials, and measured in relation to 4 cytotoxicity outcomes.

Hierarchical Rank Aggregation With Applications To Nanotoxicology

TRINA PATEL[1,3], DONATELLO TELESCA[1,3],
ROBERT RALLO[2,3,5], SAJI GEORGE[2,3]
TIAN XIA[2,3], ANDRÉ E. NEL[2,3]

Author's Footnote
[1] UCLA School of Public Health, Department of Biostatistics.
[2] UCLA Department of Medicine, Division of NanoMedicine.
[3] Center for Environmental Implications of Nanotechnology, California NanoSystems Institute.
[4] Informatica i Matematiques, Universitat Rovira i Virgili, Spain.
[5] UCLA Department of Chemical and Biomolecular Engineering

March 15, 2012

[3]For Correspondence

Trina Patel, M.S.

Department of Biostatistics

UCLA School of Public Health

Los Angeles, CA 90095–1772

phone: 714 883 9517

e-mail: trpatel@ucla.edu

Abstract

The development of high throughput screening (HTS) assays in the field of nanotoxicology provide new opportunities for the hazard assessment and ranking of engineered nanomaterials (ENM). It is often necessary to rank lists of materials based on multiple risk assessment parameters, often aggregated across several measures of toxicity and possibly spanning an array of experimental platforms. Bayesian models coupled with the optimization of loss functions have been shown to provide an effective framework for conducting inference on ranks. In this article we present various loss function based ranking approaches for comparing ENM within experiments and toxicity parameters. Additionally, we propose a framework for the aggregation of ranks across different sources of evidence while allowing for differential weighting of this evidence based on its reliability and importance in risk ranking. We apply these methods to high throughput toxicity data on 2 human cell lines, exposed to 8 different nanomaterials, and measured in relation to 4 cytotoxicity outcomes.

Keywords: Bayesian hierarchical models, Loss functions, Hazard Ranking, Nanotoxicology.

1 Introduction

This article considers hazard ranking of engineered nanomaterials (ENM) from high throughput screening (HTS) studies. Nanomaterials are a large class of substances engineered at the molecular level to achieve unique mechanical, optical, electrical, and magnetic properties. Nanotechnology is a rapidly growing field with over 800 consumer products on the market and possibly thousands of engineered nanomaterials under investigation. These developments are expected to have considerable impact on numerous fields such as medicine and technology, but, at the same time, they confer enormous potential for human exposure and environmental release (Tsuji et al. 2005, Society 2004). This scenario is coupled with the fact that the same physical and chemical properties that make ENM so desirable, may relate to interactions that take place at the nano-bio interface, with the potential for these particles to interact with many fundamental molecular and cellular processes that are critical to life (Nel et al. 2009). Given these exposure and hazard concerns, it is becoming increasingly important to make decisions regarding the safety and potential toxicity of these particles to humans and the environment.

Hazard ranking and decision support can be used to develop a framework for the prioritization of extensive in-vivo testing of emerging nanomaterials. Given ethical and economic considerations associated with animal experiments, initial prioritization schemes must indeed rely on high content *in-vitro* screening of a large number of particles (Lilienblum et al. 2008).

Current research in nano-toxicology includes new generation high throughput screening (HTS) assays, which enable the simultaneous observation of multiple cellular injury pathways across an array of doses and times of exposure. These rapid screening approaches include the use of fluorescence-based cellular assays that assess key signals of nanoparticle toxicity in various cell lines (George et al. 2010). HTS assays provide an opportunity for the toxic profiling and hazard ranking of a large number of nanomaterials in order to focus attention and resources on those nanoparticles

with the largest potential risk (Stanley et al. 2008, Maynard et al. 2006).

The statistical challenge associated with hazard ranking from HTS data lies in its richness and heterogeneity as multi-dimensional measurements are often taken over a small number of replicates with relatively low signal. Inferential goals include toxicity ranking of ENM, as well as associated measures of uncertainty, both within and aggregated across the many sources of evidence. A heatmap visualization for a sample HTS data set is provided in Figure 1, where the toxic response of 2 human cell lines, exposed to 8 different nanoparticles, is measured in relation to 4 cellular response outcomes, monitored across an array of doses and durations of exposure.

We propose an approach to ranking particles aimed at achieving three goals: to use a hierarchical dose-response model to rank particles within outcomes and experiments, to derive an aggregate or consensus ranking that summarizes information across outcomes and experiments, and finally to account for the varying levels of reliability and importance of the outcomes and experiments. An aggregate ranking across different outcomes and experiments can aid in decision making for future testing. Although the rankings within outcomes or experiments are expected to be positively correlated, this does not guarantee that the ranked lists of ENM will be in complete agreement. As an example, Figure 2 shows dose response surfaces fit to HTS toxicity data for quantum dots, nano zinc oxide and nano platinum. Nano platinum shows higher responses for mitochondrial superoxide formation and almost no response for membrane damage. Conversely, zinc oxide has a very strong response for membrane damage. Furthermore, different sources of information come with varying levels of reliability and importance in terms of assessing overall toxicity. For example, while membrane damage is viewed as a lethal response to the cell, mitochondrial superoxide formation is only a subletal indicator of cellular oxidative stress. Therefore, the information derived from these outcomes might need to be weighted differently when ranking a material's hazard potential.

We illustrate the proposed framework analyzing and ranking data on two cell lines exposed to 8 metal and metal oxide nanoparticles, monitored in relation to four cytotoxicity parameters, and

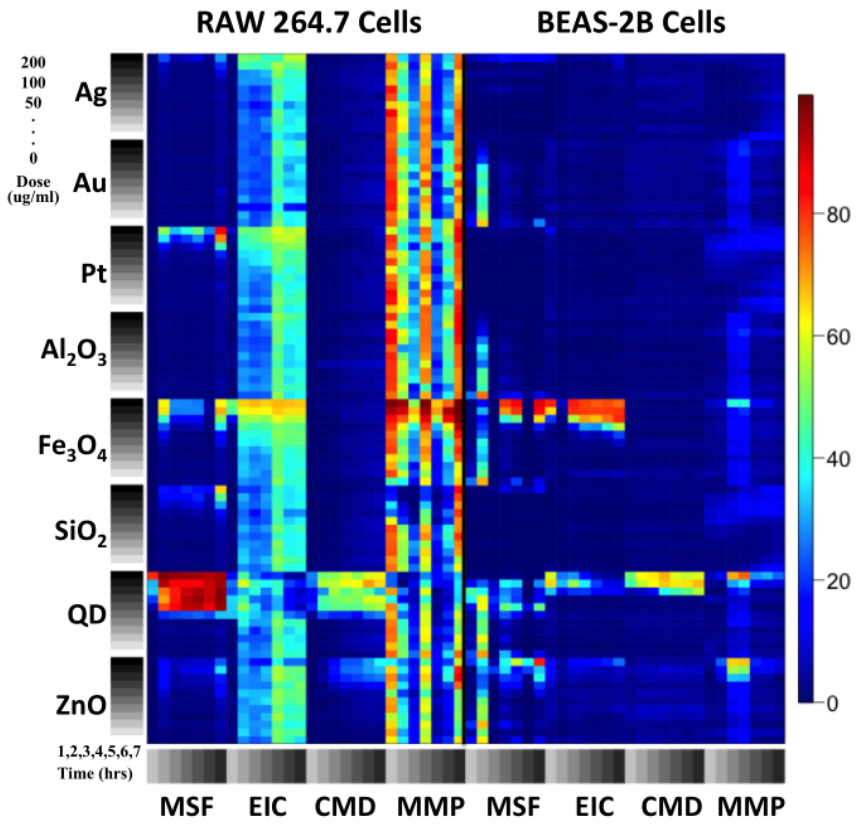

Figure 1: **Heatmap of HTS data assessing the cytotoxicity of 8 different ENM.** The rows and columns correspond to the doses and times of exposure, respectively, for each cell type and cytotoxicity parameter measured. The four cytotoxicity parameters measured include mitochondrial superoxide formation (MSF), loss of mitochondrial membrane potential (MMP), elevated intracellular calcium (EIC), and cellular membrane damage (CMD). Blue colors indicate the least harmful activity, while yellow indicates a high cellular response.

across an array of doses and times of exposure.

The multi-level structure of the data coupled with the non-standard inferential goals can be naturally accounted for under the Bayesian hierarchical framework. Shen and Louis (1998) and Lockwood et al. (2002) describe rank estimation under a Bayesian setting and show that rather than ranking posterior means, estimation based on minimizing a loss function specific to ranks is more appropriate. Lin et al. (2006) explore ranking based on optimizing various loss functions using a

two-stage hierarchical model. More specifically, they review ranking based on square-error loss and extend this to other loss functions that are tuned to application specific goals. For example, when the goal is to identify relatively high or low rankers, a loss function that penalizes classification errors produces estimates which minimize error. Lin et al. (2009) apply various loss functions to ranking health service providers based on standardized mortality ratios estimated from a two-stage hierarchical model. Noma et al. (2010) extend some of these ideas to the analysis of microarray data and develop three empirical Bayes methods for ranking genes based on a hierarchical mixture model for differential expression.

In this article we present various loss-function based ranking methods, including those presented by Shen and Louis (1998) and Lin et al. (2006), and apply them to the ranking of HTS data within outcomes and experiments. Additionally, we build on this approach in order to provide an overall ranking of particles, and thereby propose a framework for the prioritization of further testing of high risk materials.

The remainder of this article is organized as follows. In Section 2 we introduce a general statistical model for toxicity profiling and estimation of various risk assessment parameters. In Section 3 we present various loss function based methods for ranking and discuss the applicability of these methods to nanoparticle toxicology. In Section 4 we apply these methods to the analysis and ranking of 8 metal oxide nanomaterials. Finally, we conclude with a critical discussion of the limitations and possible extensions of these methods in Section 5.

2 Statistical Models of Toxicity

In this section we describe a basic framework for a dose response model for a general HTS study, where we monitor a multivariate continuous outcome y, corresponding to J cytotoxicity parameters, in association with the exposure of a number of cells to I different ENM. We are also interested in combining results across K different experiments, often consisting of multi-outcome HTS assays

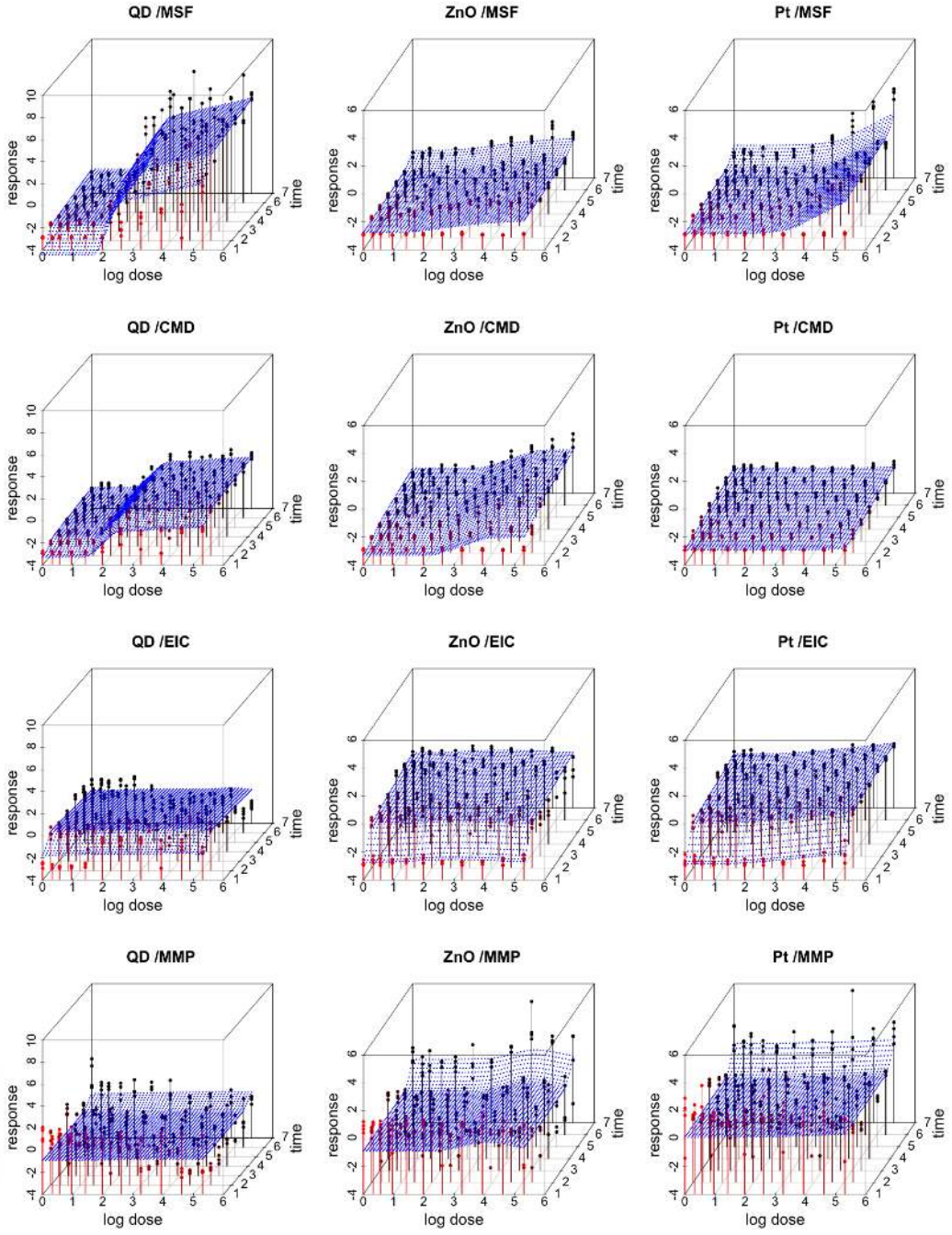

Figure 2: **Fitted response surfaces for the quantum dot (QD), zinc oxide (ZnO) and platinum (Pt) nanomaterials exposed to the macrophage (RAW 264.7) cell line**. Each of the four responses measured include: mitochondrial superoxide formation (MSF), loss of mitochondrial membrane potential (MMP), elevated intracellular calcium (EIC), and cellular membrane damage (CMD).

conducted over K cell lines. More precisely, let $y_{ijk\ell}(d,t)$ denote the response corresponding to ENM i ($i = 1, ..., I$), cytotoxicity parameter j ($j = 1, ..., J$), experiment k ($k = 1, ..., K$) and replicate ℓ ($\ell = 1, ..., L$), for a cell population exposed to dose $d \in [0, D]$ for a duration $t \in [0, T]$. In practice, observations are obtained over a discrete set of doses and durations of exposure. However, for simplicity of notation and without loss of generality, we will assume that doses and times are defined over continuous intervals. Hazard ranking will focus on the expected response $E\{y_{ijk\ell}(d,t)\} = m_{ijk}(d,t)$, where $m_{ijk}(\cdot)$ defines the exposure-duration dynamic for the process under study and may be specified in a parametric or non-parametric fashion. The proposed hazard ranking framework will not depend on specific representations of this quantity, we therefore maintain the discussion on fairly general grounds, leaving specific modeling considerations to the case study of Section 4. A detailed discussion of these issues is also reported in Patel et al. (2011).

For notational convenience we will simply identify the full exposure-duration surface with $m_{ijk} = \{m_{ijk}(d,t) : \ d \in [0,T], \ t \in [0,T]\}$. We assume that the joint distribution of an integrated HTS experiment can be represented by the following multi-stage hierarchical model

$$\begin{array}{lcl} y_{ijk\ell}(d,t) \mid m_{ijk}, \sigma^2_{ijk} & \sim & p(y_{ijk\ell}(d,t) \mid m_{ijk}, \sigma^2_{ijk}) \\ m_{ijk} \mid m_{ij}, \sigma^2_{ij} & \sim & p(m_{ijk} \mid m_{ij}, \sigma^2_{ij}) \\ m_{ij} \mid m_i, \sigma^2_i & \sim & p(m_{ij} \mid m_i, \sigma^2_i) \\ m_i & \sim & p(m_i), \end{array} \tag{1}$$

where m_{ij} and m_i denote exposure-duration surfaces at the integrated particle-by-outcome (ij) and integrated particle (i) levels. The model is completed with priors on variance components and possible nuisance parameters.

Let, $\boldsymbol{m} = \{m_{ijk} : \ i = 1, \ldots, I; \ j = 1, \ldots, J; \ k = 1, \ldots, K\}$ denote the set of all smooth surfaces and $\boldsymbol{\sigma^2} = \{\sigma^2_{ijk} : \ i = 1, \ldots, I; \ j = 1, \ldots, J; \ k = 1, \ldots, K\}$ is the set of all variance parameters. Furthermore if we denote with $\boldsymbol{Y}$ the complete set of response values for all particles, cytotoxicity parameters, and experiments, all evidence available over exposure-duration dynamics and associated variability at the different levels of the hierarchy is contained in the posterior

distribution

$$p(\boldsymbol{m}, \boldsymbol{\sigma} \mid \boldsymbol{Y}) \propto p(\boldsymbol{Y} \mid \boldsymbol{m}, \boldsymbol{\sigma^2}) p(\boldsymbol{m}, \boldsymbol{\sigma^2}). \tag{2}$$

This quantity, or most often Monte Carlo samples from this joint distribution, can be used to calculate the joint posterior distribution of various functionals $g(m_{ijk})$. These functions provide easily interpretable summaries of full exposure-duration dynamics, that may be used as overall measures of hazard. Some examples of risk assessment summaries frequently used in toxicology include, exposure levels or time thresholds below which no significant effect exists, doses or times at which adverse effects rise above some predetermined amount as compared to the background (benchmark doses), summaries of the slope of the dose-response trajectory including the dose or time that produces an $\alpha\%$ inhibitory response (ECα), among others (Edler et al. 2002). These measures summarize different aspects of a response surface, and in most cases, separately for dose and time kinetics. Although, no sufficient univariate summary exists that can synthesize every aspect of the response surface, in this paper we will focus on the area under the response surface as a possible summary measure (Section 4). However, we maintain that the proposed method is directly applicable to any functional of the response surface set $\boldsymbol{m}$.

3 Decision Theoretic Approaches to Hazard Ranking

3.1 Estimating Ranks

Most common ranking methods are based on ordering estimates of target parameters, such as MLEs and posterior means, or on the ordering of statistics testing some null hypothesis of interest. As described, among others by Louis and Shen (1999), Lin et al. (2006), these methods perform poorly when the posterior distributions of the target parameters are not stochastically ordered and, are not invariant under monotone transformation of the target parameters. A more appropriate method of ranking is based on calculating the joint posterior distribution of the ranks, followed by inference

guided by loss functions appropriate for analytic goals.

The rank of some target parameter $g(m_{ijk})$, for each particle i, outcome j, and experiment k can be defined as the $\sum_{h=1}^{I} \mathbf{1}_{\{g(m_{ijk}) \geq g(m_{hjk})\}}$. A direct extension of this definition to the aggregate ranking of particles, for example, across K experiments, might be defined by the ranks of population level parameters $g(m_{ij})$. A caveat of this procedure is that, due to technical and/or biological sources of variability, data on heterogeneous measures of toxicity may support high posterior variances for population level parameters, which are not necessarily reflected in small rank correlations amongst lower level quantities. This procedure may in fact result in the artificial inflation of reported aggregate rank variability.

In order to avoid this paradox, we define aggregate ranks, R_{ij} and R_i as a weighted average of ranks R_{ijk}. More formally, the rank of our target parameter, $g(m_{ijk})$, at each level of hierarchy, can be described as follows:

$$\begin{aligned} R_{ijk} &= rank(g(m_{ijk})) = \sum_{h=1}^{I} \mathbf{1}_{\{g(m_{ijk}) \geq g(m_{hjk})\}} \\ R_{ij} &= \sum_{k=1}^{K} w_k R_{ijk} \qquad \sum_{k=1}^{K} w_k = 1 \\ R_i &= \sum_{j=1}^{J} w_k \sum_{k=1}^{K} w_j R_{ijk} \qquad \sum_{j=1}^{J} w_j = 1 \end{aligned} \tag{3}$$

where, a rank of 1 corresponds to the highest rank, or most hazardous particle. Here, R_{ijk} allow us to determine the rank of each particle within each cytotoxicity outcome and experimental group. R_{ij} is the rank of each particle within each cytotoxicity parameter, but aggregated across all experiments. R_i is an overall summary of the rank of each particle across all cytotoxicity parameters and experimental groups. Finally, w_j and w_k are arbitrary weight functions that allow us to assign a measure of importance to each cytotoxicity parameter j and experiment k.

Provided draws from $p(m_{ijk} \mid \boldsymbol{Y})$ are available via Markov Chain Monte Carlo ($MCMC$) or otherwise, all knowledge about R_{ijk} is easily summarized in the posterior distribution $p(R_{ijk} \mid \boldsymbol{Y})$ and similarly for R_{ij} and R_i. In the following sections we describe in more detail alternative point

estimators of ranks at the different levels of the hierarchy from a decision theoretic perspective, as well as potential substantive strategies to select weights w_j and w_k.

3.2 Rank estimates based on squared-error loss functions

Following Shen and Louis (1998) and Lockwood et al. (2002), natural loss functions may refer to the squared-error associated with posterior ranks. Let R_{ijk}^{est} be a point estimator for the posterior rank of particle i, outcome j, experiment k and similarly for R_{ij}^{est} and R_{i}^{est}. For each particle at each level of hierarchy squared error loss functions can be described as follows:

$$\begin{aligned} L_{jk}^{(SEL)} &= L_{jk}^{(SEL)}\left(R_{ijk}^{est}, R_{ijk}\right) &= \tfrac{1}{I}\sum_{i}\left(R_{ijk}^{est} - R_{ijk}\right)^2 \\ L_{j}^{(SEL)} &= L_{j}^{(SEL)}\left(R_{ij}^{est}, R_{ij}\right) &= \tfrac{1}{I}\sum_{i}\left(R_{ij}^{est} - \sum_{k=1}^{K} w_k R_{ijk}\right)^2 \\ L^{(SEL)} &= L^{(SEL)}\left(R_{i}^{est}, R_{i}\right) &= \tfrac{1}{I}\sum_{i}\left(R_{i}^{est} - \sum_{j=1}^{J} w_j \sum_{k=1}^{K} w_j R_{ijk}\right)^2. \end{aligned} \tag{4}$$

These loss functions are minimized by the following posterior means,

$$\begin{aligned} \bar{R}_{ijk}(\boldsymbol{Y}) &= E_{g(m_{ijk})|\boldsymbol{Y}}[R_{ijk} \mid \boldsymbol{Y}] &= \sum_{h=1}^{I} P(g(m_{ijk}) \geq g(m_{hjk}) \mid \boldsymbol{Y}) \\ \bar{R}_{ij}(\boldsymbol{Y}) &= E_{g(m_{ijk})|\boldsymbol{Y}}[R_{ij} \mid \boldsymbol{Y}] &= \sum_{k=1}^{K} w_k \sum_{h=1}^{I} P(g(m_{ijk}) \geq g(m_{hjk}) \mid \boldsymbol{Y}) \\ \bar{R}_{i}(\boldsymbol{Y}) &= E_{g(m_{ijk})|\boldsymbol{Y}}[R_{i} \mid \boldsymbol{Y}] &= \sum_{j=1}^{J} w_j \sum_{k=1}^{K} w_k \sum_{h=1}^{I} P(g(m_{ijk}) \geq g(m_{hjk}) \mid \boldsymbol{Y}). \end{aligned} \tag{5}$$

The $\bar{R}_i$ are generally not integer values. For optimal integers ranks we use $\hat{R}_i = rank(\bar{R}_i)$, and similarly, for $\bar{R}_{ij}$ and $\bar{R}_{ijk}$. In many situations it is more convenient to define percentiles instead of ranks (Lockwood et al. 2002). Therefore, we let $Q_{ijk} = R_{ijk}/(I+1)$ denote the percentile rank for particle i, outcome j, and experiment k. Again, we can write similar expressions for Q_i and Q_{ij}.

3.3 Rank estimates base on upper $100(1-\gamma)\%$ or lower $100(\gamma)\%$ loss functions

In many situations, interest may focus on identifying some fraction of particles with the highest (or lowest) likelihood of conferring adverse effects. Lin et al. (2006) proposed a loss function which

addresses this goal by penalizing misclassification by an amount that depends on the distance of the estimated percentile from some cut-point γ. Let γ $(0 < \gamma < 1)$ denote the most toxic fraction of total ENM that we would like to identify. Also let Q_{ijk}^{est} be a point estimator for the posterior percentile rank of particle i, outcome j, experiment k and similarly for Q_{ij}^{est} and Q_{i}^{est}. In order to rank ENM at each level of hierarchy, loss functions based on the upper $100(1-\gamma)\%$ classification error can be described as follows:

$$\begin{aligned} L_{jk}^{(PLF)} &= \tfrac{1}{I}\sum_{i}\left(\gamma - Q_{ijk}^{est}\right)^2\left(\mathbf{1}_{\{Q_{ijk}>\gamma, Q_{ijk}^{est}<\gamma\}} + I_{\{Q_{ijk}<\gamma, Q_{ijk}^{est}>\gamma\}}\right) \\ L_{j}^{(PLF)} &= \tfrac{1}{I}\sum_{i}\left(\gamma - Q_{ij}^{est}\right)^2\left(\mathbf{1}_{\{Q_{ij}>\gamma, Q_{ij}^{est}<\gamma\}} + I_{\{Q_{ij}<\gamma, Q_{ij}^{est}>\gamma\}}\right) \\ L^{(PLF)} &= \tfrac{1}{I}\sum_{i}\left(\gamma - Q_{i}^{est}\right)^2\left(\mathbf{1}_{\{Q_{i}>\gamma, Q_{i}^{est}<\gamma\}} + I_{\{Q_{i}<\gamma, Q_{i}^{est}>\gamma\}}\right). \end{aligned} \tag{6}$$

For notational convenience we will assume γI is an integer and let

$$\begin{aligned} \pi_{ijk}(\gamma) &= P(R_{ijk} > \gamma(I+1) \mid \boldsymbol{Y}) &&= \sum_{n=\gamma I+1}^{I} P(R_{ijk} = n \mid \boldsymbol{Y}) \\ \pi_{ij}(\gamma) &= P(\sum_{k=1}^{K} w_j R_{ijk} > \gamma(I+1) \mid \boldsymbol{Y}) &&= \sum_{n=\gamma I+1}^{I} P(\sum_{k=1}^{K} w_j R_{ijk} = n \mid \boldsymbol{Y}) \\ \pi_{i}(\gamma) &= P(\sum_{j=1}^{J} w_k \sum_{k=1}^{K} w_j R_{ijk} > \gamma(I+1) \mid \boldsymbol{Y}) &&= \sum_{n=\gamma I+1}^{I} P(\sum_{j=1}^{J} w_k \sum_{k=1}^{K} w_j R_{ijk} = n \mid \boldsymbol{Y}). \end{aligned} \tag{7}$$

The quantities that minimize the posterior risk induced by the loss function described in (6) can be described by: $\tilde{R}_{ijk}(\gamma) = rank(\pi_{ijk}(\gamma))$ for, particle i, cytotoxicity parameter j, and experiment k. We can write similar expressions for $\tilde{R}_{ij}(\gamma)$ and $\tilde{R}_{i}(\gamma)$. Here we have optimized the classification errors for ranking the most hazardous particles. Ranking based on particles least likely to be hazardous will follow a similar procedure.

3.4 Assigning Weights

Different sources of information come with varying levels of reliability and importance in terms of hazard ranking. Our definition of aggregate ranks include weights w_j, averaging across J possible outcomes and w_k, averaging across K experiments. These quantities may be used directly in the

definition of aggregate ranks, in a way that reflects possible differing measures of importance to be assigned to each outcome and experiment.

The possibility to define differentiated aggregate hazard measures is especially important in HTS cytotoxicity studies. A typical HTS assay will, in fact, include outcomes measuring only sublethal effects, often used as hypothesis generation and confirmation tools by biologists, as well as outcomes measuring lethal effects, which are more directly related to cytotoxicity. Intuitively, we might want to give a higher weight to particles that induce lethal effects relative to particles that only induce sublethal effects without cytotoxicity. For example, in cytotoxicity studies of metal oxide nanomaterials, PI uptake, a measure of cellular membrane damage, is the outcome most often chosen to carry out risk ranking (George et al. 2011). Similar considerations are valid for the differential weighting of different experiments. For example George et al. (2010) report an experiment carried out on two cell lines: bronchial epithelial cells and a macrophage cell line. In this case, one possibility is that one cell line may prove more relevant for the prioritization of animal inhalation toxicity experiments then the other. Clearly, uniform weights are always applicable, when no particular experiment or outcome are considered to be more or less relevant as a measure of hazard.

While differential weighting is, in principle, arbitrary and can be determined using expert's criteria, we found that it is often easier to obtain expert opinion on the ordering of different outcomes and experiments. If we let $(\pi_1, ...\pi_k...\pi_K)$ be the ordering of the K experiments corresponding to the importance of each experiment k, then one possibility for assigning weights is to find each $w_1, ...w_{\pi_K}$, such that

$$w_{\pi_{k-1}} = \delta w_{\pi_k}, \qquad \sum_{k=1}^{K} \delta^{k-1} w_{\pi_1} = 1, \quad \text{for } k = 1, \ldots, K; \;\; \delta \geq 1; \tag{8}$$

leading to $w_k = \delta^{k-1} / \sum_{h=1}^{K} \delta^{h-1}$. Aggregate weight determination across J outcomes can be obtained in the same fashion. In this formulation, differential weighting depends solely on a one

dimensional, easily interpreted parameter δ. A value of δ equal to 1 assigns equal importance to each experiment, while a value of δ equal to .5 assigns twice the importance to experiment π_{k-1} as experiment π_k. Alternatively, weights w_j and w_k can be given arbitrary values, for example based on expert elicitation, as long as they satisfy the constraints $\sum_{j=1}^{J} w_j = 1$ and $\sum_{k=1}^{K} w_k = 1$.

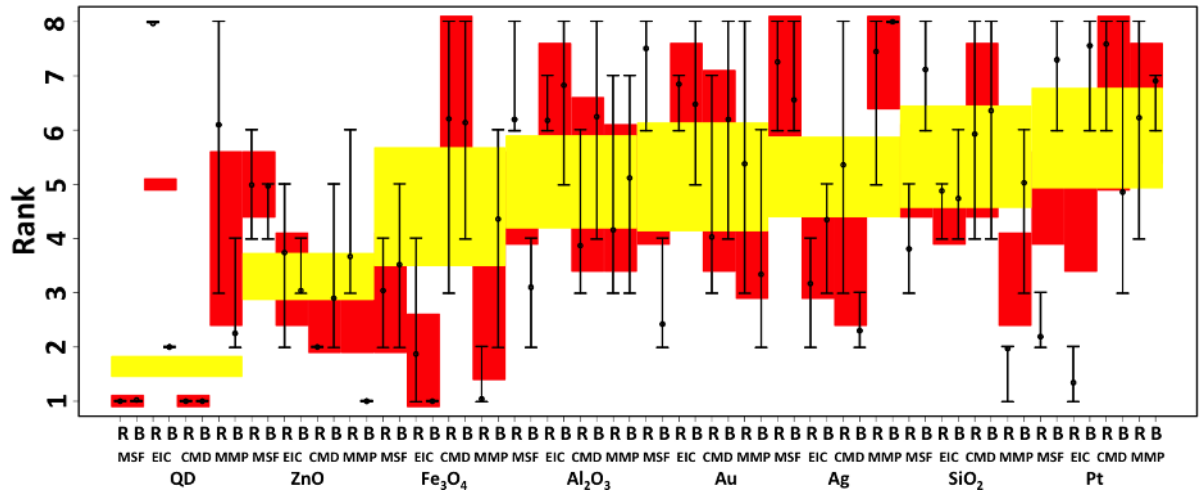

Figure 3: **Rankings based on squared error loss.** Posterior expected ranks and 95% posterior intervals computed by minimizing squared error loss. Each cell-line is given equal weight, and each outcome is given weights $(.54, .27, .12, .07)$, in the order (CMD, MSF, MMP, EIC). Ranks are displayed individually within outcomes and experiments (black), aggregated across experiments, but within outcomes (red), and aggregated across all outcomes and experiments (yellow).

4 A Case Study in Nanotoxicology

4.1 Background

We analyze and rank data on cells exposed to eight different metal and metal oxide nanoparticles and, monitored in relation to four cytotoxicity parameters. All four outcomes are measured over a grid of ten doses $(0, .375, .750, 1.6, 3.12, 6.25, 12.5, 25, 50, 100, 200\ \mu g/ml)$, seven times (hours) of exposure, and four replicates at every dose-time combination (see Figure 2). Cytotoxicity screening is based on the hierarchical oxidative stress model described in Xia et al. (2006). More specifically,

a multi-parametric, epiflourescence assay was used to measure four responses relating to toxic oxidative stress. Three of the responses measured sublethal effects to the cell including, mitochondrial superoxide formation (MSF), loss of mitochondrial membrane potential (MMP), and elevated intracellular calcium (EIC), and the last outcome measured, cellular membrane damage (CMD), a lethal effect to the cell. The nanoparticles measured include: silver (Ag), gold (Au), platinum (Pt), iron oxide (Fe_3O_4), aluminum oxide (Al_2O_2), silicon dioxide (SiO_2), zinc oxide (ZnO), and quantum dot (QD). Experiments were conducted in two different cell lines, both related to inhalation toxicity, including bronchial epithelial cell lines (BEAS-2B) and macrophage (RAW 264.7) cell lines.

4.2 Analysis and Results

We model response surfaces for the metal oxide data set using the model described in Section 1. Specifically, we model the data in a generalized additive fashion and parameterized using linear basis spline functions. Details for a similar reduced experiment were published in Patel et al. (2011). To summarize briefly, the response surface $m_{ijk}(d,t)$, corresponding to ENM i ($i = 1, ..., I$), cytotoxicity parameter j ($j = 1, ..., J$), and replicate k ($k = 1, ..., K$) at dose $d \in [0, D]$ and time $t \in [0, T]$, was modeled as follows:

$$m_{ijk}(d,t) = \alpha_{ijk} + \mathcal{B}(d, \boldsymbol{\phi}_{ijk})'\boldsymbol{\beta}_{ijk} + \mathcal{B}(t, \boldsymbol{\psi}_{ijk})'\boldsymbol{\gamma}_{ijk}. \tag{9}$$

where, $\mathcal{B}(d, \boldsymbol{\phi}_{ijk})$ and $\mathcal{B}(d, \boldsymbol{\psi}_{ijk})$ denote two 4-dimensional B-spline basis with interior knots $\boldsymbol{\phi}_{ijk} = (\phi_{ijk1}, \phi_{ijk2})'$ and $\boldsymbol{\psi}_{ijk} = (\psi_{ijk1}, \psi_{ijk2})'$. Also, $\boldsymbol{\beta}_{ij} = (\beta_{ij1}, .., \beta_{ij4})'$ and $\boldsymbol{\gamma}_{ij} = (\gamma_{ij1}, ..., \gamma_{ij4})'$ are two 4-dimensional vectors of spline coefficients. Identifiability restrictions are implemented by fixing $\beta_{ijk1} = 0$ and $\gamma_{ijk1} = 0$, allowing us to interpret α_{ij} as the background response level for each particle, outcome, and experiment. We also fix $\beta_{ij2} = 0$ and $\gamma_{ij2} = 0$, assuming no effect before ϕ_{ij1} and ψ_{ij1}. See Figure 2 for an example of surfaces fit to the data on RAW 264.7 cells exposed to the zinc oxide (ZnO), quantum dot (QD), and platinum (Pt) nanomaterials and measured

Equal Weights	
ENM	Rank
QD	2.79 (2.38,3.12)
ZnO	3.29 (3,3.75)
Fe3O4	3.4 (2.75,4)
SiO2	4.98 (4.38,5.5)
Al2O3	5.21 (4.62,5.75)
Au	5.28 (4.62,5.88)
Pt	5.5 (5,6)
Ag	5.56 (5.12,6)

$w_j = (.37, .27, .21, .15)$	
ENM	Rank
QD	2.27 (1.92,2.54)
ZnO	3.25 (2.96,3.69)
Fe3O4	3.95 (3.09,4.73)
Al2O3	5.08 (4.43,5.75)
Au	5.15 (4.41,5.94)
SiO2	5.2 (4.48,5.88)
Ag	5.47 (4.92,6.04)
Pt	5.63 (5.03,6.37)

$w_j = (.54, .27, .12, .07)$	
ENM	Rank
QD	1.66 (1.46,1.82)
ZnO	3.18 (2.88,3.72)
Fe3O4	4.65 (3.51,5.67)
Al2O3	5 (4.2,5.89)
Au	5.09 (4.15,6.13)
Ag	5.12 (4.41,5.87)
SiO2	5.55 (4.58,6.44)
Pt	5.74 (4.94,6.77)

Table 1: **Overall rankings based on squared error loss.** Aggregated ranks across each outcome and cell-line. Posterior expected ranks and 95% posterior intervals computed by minimizing squared error loss. Each cell-line is given equal weight and each outcome (CMD,MSF,MMP,EIC) is given varying weights w_j.

MSF		EIC		CMD		MMP	
ENM	Rank	ENM	Rank	ENM	Rank	ENM	Rank
QD	1.01 (1,1)	Fe3O4	1.43 (1,2.5)	QD	1 (1,1)	ZnO	2.34 (2,3.5)
Fe3O4	3.28 (2,4)	ZnO	3.39 (2.5,4)	ZnO	2.45 (2,3.5)	Fe3O4	2.7 (1.5,4)
Al2O3	4.65 (4,5.5)	Ag	3.76 (3,4.5)	Ag	3.83 (2.5,5)	SiO2	3.5 (2.5,4)
Pt	4.75 (4,5.5)	Pt	4.45 (3.5,5)	Al2O3	5.06 (3.5,6.5)	QD	4.17 (2.5,5.5)
Au	4.96 (4,6)	SiO2	4.81 (4,5.5)	Au	5.12 (3.5,7)	Au	4.36 (3,6)
ZnO	4.98 (4.5,5.5)	QD	4.99 (5,5)	SiO2	6.14 (4.5,7.5)	Al2O3	4.64 (3.5,6)
SiO2	5.47 (4.5,6)	Al2O3	6.5 (5.5,7.5)	Fe3O4	6.18 (4,8)	Pt	6.57 (5.5,7.5)
Ag	6.91 (6,8)	Au	6.66 (6,7.5)	Pt	6.23 (5,8)	Ag	7.72 (6.5,8)

Table 2: **Rankings within outcomes based on squared error loss.** Aggregated ranks across each each outcome and aggregated across cell-lines. Posterior expected ranks and 95% posterior intervals computed by minimizing squared error loss. Each cell-line is given equal weight.

across all four cytotoxicity outcomes. Our inferences are based on 20,000 MCMC samples from the posterior distribution in (2), after discarding a conservative 60,000 iterations for burn-in. MCMC sampling was performed in R version 2.10.0, and convergence diagnostics were performed using the package CODA (Convergence Diagnostics and Output Analysis), (Plummerm et al. 2006).

We define our target parameter, $g(m_{ijk})$, as the area under the response surface, excluding any background response, for each particle, outcome, and experiment. Any number of classical summaries can be derived from this model and used for risk assessment such as benchmark doses (BMD), effective concentrations (ECα), no observable adverse effect level (NOAEL), among others. These measures summarize different aspects of the response surface and typically disjointly for dose and time kinetics. In fact, there is still disagreement in the HTS setting, on the best measures of

risk (Stern and McNeil 2008 and Maynard et al. 2006). Furthermore, some of these summaries can become even more problematic in the nanotoxicology setting due to issues such as dosimetry, where the administered doses are confounded by different particle bioavailability. The area under the response surface is an overall summary of the entire dose and duration response dynamic. Although no sufficient one-dimensional summary of the dose and duration response profile exists, the area under the surface may be a more comprehensive summary of risk. Using the model described in (9), we define area under the surface as follows:

$$\begin{aligned} AUS &= \int_0^T \int_0^D m_{ijk}(d,t) - \alpha_{ijk}\,\mathrm{d}d\,\mathrm{d}t \\ &= \int_0^T \int_0^D \boldsymbol{\mathcal{B}}(d,\boldsymbol{\phi}_{ijk})'\boldsymbol{\beta}_{ijk} + \boldsymbol{\mathcal{B}}(t,\boldsymbol{\psi}_{ijk})'\boldsymbol{\gamma}_{ijk}\,\mathrm{d}d\,\mathrm{d}t \\ &= T[\tfrac{1}{2}(D-\phi_{ijk1})\beta_{ijk3} + \tfrac{1}{2}(D-\phi_{ijk2})\beta_{ijk4}] + D[\tfrac{1}{2}(T-\psi_{ijk1})\gamma_{ijk3} + \tfrac{1}{2}(T-\psi_{ijk2})\gamma_{ijk4}] \end{aligned} \tag{10}$$

Using the ranking methods described in Section 3, we rank the eight nanoparticles within cell lines and outcomes, within outcomes but aggregated across cell-lines, and aggregated across cell-lines and outcomes. Table 1 (*left panel*) provides posterior expected ranks, and associated 95% posterior intervals ranks, for each particle, aggregated across all cytotoxicity outcomes and cell-lines, with each outcome and cell-line weighted equally. Based on knowledge about oxidative stress pathways and the assays used to measure these outcomes, it is believed that the outcomes measured can be ranked in order of importance as follows: (CMD, MSF, MMP, EIC). Using the weight function described in (8) and a value of $\delta = .75$, we can derive weights $(.37, .27, .21, .15)$, for the four outcomes. Similarly, using a slightly more aggressive $\delta = .5$, we can derive weights $(.54, .27, .12, .07)$. Table 1 (*middle panel* and *right panel*) provides overall summaries, aggregated across outcomes and cell-lines, for the posterior ranks using these different weight functions. Additionally, Table 2 and 3 provide rankings for each particle, within outcomes but across cell lines, and within outcomes and experiments, respectively. In each of these cases, we compute expected ranks and 95% posterior intervals by minimizing squared error loss. Figure 3 provides a graphical summary of the expected posterior ranks and associated 95% posterior intervals, at each level of

	MSF		EIC		CMD		MMP	
Cell-line	ENM	Rank	ENM	Rank	ENM	Rank	ENM	Rank
RAW 264.7	QD	1 (1,1)	Pt	1.34 (1,2)	QD	1 (1,1)	Fe3O4	1.04 (1,2)
	Pt	2.19 (2,3)	Fe3O4	1.87 (1,4)	ZnO	2 (2,2)	SiO2	1.97 (1,2)
	Fe3O4	3.04 (2,4)	Ag	3.17 (2,4)	Al2O3	3.87 (3,6)	ZnO	3.67 (3,6)
	SiO2	3.81 (3,5)	ZnO	3.74 (2,5)	Au	4.03 (3,7)	Al2O3	4.16 (3,7)
	ZnO	4.99 (4,6)	SiO2	4.88 (4,5)	Ag	5.36 (3,8)	Au	5.38 (3,8)
	Al2O3	6.2 (6,8)	Al2O3	6.18 (6,7)	SiO2	5.93 (4,8)	QD	6.1 (3,8)
	Ag	7.26 (6,8)	Au	6.85 (6,7)	Fe3O4	6.21 (3,8)	Pt	6.23 (4,8)
	Au	7.51 (6,8)	QD	7.97 (8,8)	Pt	7.59 (6,8)	Ag	7.45 (5,8)
BEAS-2B	QD	1.02 (1,1)	Fe3O4	1 (1,1)	QD	1 (1,1)	ZnO	1 (1,1)
	Au	2.42 (2,4)	QD	2 (2,2)	Ag	2.3 (2,3)	QD	2.25 (2,4)
	Al2O3	3.1 (2,4)	ZnO	3.04 (3,4)	ZnO	2.9 (2,5)	Au	3.34 (2,6)
	Fe3O4	3.52 (2,5)	Ag	4.35 (3,5)	Pt	4.86 (3,8)	Fe3O4	4.36 (2,6)
	ZnO	4.97 (4,5)	SiO2	4.74 (4,6)	Fe3O4	6.14 (4,8)	SiO2	5.03 (3,6)
	Ag	6.56 (6,8)	Au	6.48 (5,8)	Au	6.2 (4,8)	Al2O3	5.12 (3,7)
	SiO2	7.12 (6,8)	Al2O3	6.83 (5,8)	Al2O3	6.25 (4,8)	Pt	6.91 (6,7)
	Pt	7.3 (6,8)	Pt	7.56 (6,8)	SiO2	6.36 (4,8)	Ag	8 (8,8)

Table 3: **Rankings based on squared error loss.** Posterior expected ranks and 95% posterior intervals computed by minimizing squared error loss. Each cell-line is given weight and each outcome (CMD, MSF, MMP, EIC) is given weights: $(.37, .27, .21, .15)$.

hierarchy. The ranks are computed by minimizing squared error loss and weighing the information from each cell line equally but, differentially weighing each outcome using a $\delta = .5$, as described above.

In Figure 3, the posterior intervals reflect the uncertainty about the hazard rankings and provide an illustration of how the proposed technique combines information and provides more precise estimates of ranks when aggregated across outcomes and cell lines. Although it is difficult to distinguish between more or less toxic particles when looking at individual ranks, we can discern that, aggregated across outcomes and cell lines, quantum dot nanoparticles show significantly higher responses than all other particles. Zinc oxide nanoparticles also show a significantly higher toxic response than nano platinum, silver, aluminum oxide, silicon dioxide, and gold particles. Table 2 (*column 3*) provides rankings for cellular membrane damage, aggregated across the 2 cell lines. In terms of cellular membrane damage, quantum dot and zinc oxide show significantly higher cytotoxic responses then the remaining 6 nanoparticles. This supports what has previously been demonstrated in conventional assays that QD nanoparticles stabilized by toluene are capable of inducing tier 2

and 3 oxidative stress responses induced by the toluene (George et al. 2011). Similarly, it has been demonstrated that ZnO nanoparticles are capable of inducing tier 2 and 3 oxidative stress responses through Zn_2^+ release (George et al. 2010). In contrast, while platinum, silver, aluminum oxide, silicon dioxide, and gold particles nanoparticles have been shown to trigger sublethal responses, they do not increase PI uptake, an indicator of cellular membrane damage, leading to cell death (George et al. 2011).

5 Discussion

In this article we present various loss function based ranking approaches and apply them to the hazard ranking of nanomaterials, using multivariate toxicity data obtained from HTS assays. Furthermore, we extend these methods to the aggregation of ranks across different sources of evidence. We account for the multivariate nature of the data using a Bayesian hierarchical framework, and coupled with a loss function, are thereby able to derive a rank estimate and its associated uncertainty. The proposed methodology accounts for the variability in the scale of the response across cytotoxicity measures and experimental platforms and allows for the differential weighting of these measures in the estimation of an aggregate rank distribution.

As described by Louis and Shen (1999), Lin et al. (2006), when the posterior distributions of the target parameters are stochastically ordered and are invariant under monotone transformation, the choice of ranking method does not matter. However, in many cases there is a clear benefit involved with using an optimal procedure which is clearly defined by inferential goals. In this paper we present the most commonly used squared error loss function which optimizes the overall ranking of all particles. We also present the upper $100(1-\gamma)\%$ (or lower $100(\gamma)\%$) loss function which is useful when the goal is identifying the most (or least) toxic fraction of particles. Many other loss functions can be considered. For example, Lin et al. (2006) suggest the use of a weighted combination of several loss functions in order to broaden the class of all loss functions.

One advantage of a ranked list of ENM is that it allows for the prioritization of further testing of these particles, especially when resources are limited. Although, loss function based ranking methods are optimal, in many cases the results may not be conclusive as somewhat indicated by the large, often overlapping, confidence intervals around the rank estimates. Aggregated ranks across different sources of information can be used to combine information and reduce the uncertainty of the results. Although the ranked lists are expected to be correlated, this does not guarantee that the ranked lists of ENM will be in complete agreement. In some cases, this disagreement may in fact increase the uncertainty of the results.

The problems of comparing lists and rank aggregation have also been considered in the computer science and bioinformatics literature in-relation to meta-search (Dwork et al. 2001; Fagin et al. 2003). Fagin et al. (2003) defined a set of distance measures that could be used to quantify dissimilarities between lists. In the context of rank aggregation, they are interested in finding the aggregation that has the minimum total distance with respect to the given lists. Dwork et al. (2001) considered the problem of aggregating across lists using a Markov process approach. First Pairwise majority preferences are summarized across lists and then the matrix of pairwise preferences are used to produce a Markov Chain (MC) transition matrix. The aggregate ranking can then be defined according to the stationary distribution of this MC. DeSemet et al. (2002) used a similar approach to model the aggregate behavior of a large number of decision makers. These techniques, however, are designed to work with summaries of previously analyzed data sets and, to our knowledge are not directly applicable to a comprehensive data-fusion exercise.

The ranking of chemicals has also been extensively studied, especially using partial order techniques and multi-criteria analysis (see Lerche et al. (2002) for a recent comparison of these methods). For example, Lerche and Sørensen (2003) consider the ranking of objects using partial order theory and random linear extensions, and Lerche et al. (2004) apply these ideas to the ranking of chemicals. These methods are extensions of simple scoring methods, and again, are designed to work

with summaries of previously analyzed data sets. Their direct applicability to the general HTS setting, described earlier, is therefore limited.

Another important consideration when ranking particles is the use of a proper estimate of toxicity. In this paper, we focus on one summary, such as the area under the response surface, that we believe is an adequate measure of risk. An alternative formulation would be to construct multiple ranked lists of particles, using more then one summary of risk, followed by the aggregation of these lists. Furthermore, in this paper we have focused on ENM hazard as the sole factor in determing risk. In toxicology, risk assessment involves the characterization of hazard as well as the potential for exposure. Currently, there is not enough exposure information available to perform a traditional risk assessment, but an important area of future research involves the aggregation of hazard and exposure rankings (Maynard et al. 2006).

References

DeSemet, Y., J. Sprinagael, and P. Kunsch (2002). Towards statistical multicriteria decision modeling: A first approach. *Journal of Multi-Criteria Decision Analysis 11*(6), 305–313.

Dwork, C., R. Kumar, M. Naor, and D. Sivakumar (2001). Rank aggregation methods for the web. *Proceedings of the 10th International Conference on World Wide Web*, 613–622.

Edler, L., K. Poirier, M. Dourson, J. Kleiner, B. Mileson, H. Nordmann, A. Renwick, W. Slob, K. Walton, and G. Würtzen (2002). Mathematical modelling and quantitative methods. *Food and Chemical Toxicology 40*, 283–326.

Fagin, R., R. Kumar, and D. Sivakumar (2003). Comparing top-k lists. *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms 17*(1), 134–160.

George, S., S. Pokhrel, T. Xia, B. Gilbert, Z. Ji, M. Schowalter, A. Rosenauer, R. Damoiseaux, K. Bradley, L. Madler, and A. Nel (2010). Use of a rapid cytotoxicity screening approach to engineer a safer zinc oxide nanoparticle through iron doping. *American Chemical Society 4*(1), 15–29.

George, S., T. Xia, R. Rallo, Y. Zhao, Z. Ji, S. Lin, X. Wang, H. Zhang, B. France, D. Schoenfeld, R. Damoiseaux, R. Liu, S. Lin, K. Bradley, Y. Cohen, and A. Nel (2011). Use of a high-throughput

screening approach coupled with in vivo zebrafish embryo screening to develop hazard ranking for engineered nanomaterials. *ACS Nano 5*(3), 1805–1817.

Lerche, D., R. Brüggemann, P. Sørensen, L. Carlsen, and O. Nielsen (2002). A comparison of partial order technique with three methods of multi-criteria analysis for ranking of chemical substances. *Journal of Chemical Information and Computer Sciences 42*, 1086– 1098.

Lerche, D., S. Matsuzaki, P. Sørensen, L. Carlsen, and O. Nielsen (2004). Ranking of chemical substances based on the japanese pollutant release and transfer register using partial order theory and random linear extensions. *Chemosphere 55*, 1005– 1025.

Lerche, D. and P. Sørensen (2003). Evaluation of the ranking probabilities for partial orders based on random linear extensions. *Chemosphere 53*, 981– 992.

Lilienblum, W., W. Dekant, H. Foth, T. Gebel, J. Hengstler, R. Kahl, P. Kramer, H. Schweinfurth, and K. Wollin (2008). Alternative methods to safety studies in experimental animals: Role in the risk assessment of chemicals under the new european chemicals legislation (reach). *Archives of Toxicology 82*(4), 211– 236.

Lin, R., T. Louis, S. Paddock, and G. Ridegeway (2006). Loss function based ranking in two-stage, hierarchical models. *Bayesian Analysis 1*(4), 915–946.

Lin, R., T. Louis, S. Paddock, and G. Ridegeway (2009). Prior distributions for variance parameters in hierarchical models. *Health Services Outcomes Reseach Methodology 4*, 515–533.

Lockwood, J., T. Louis, and D. McCaffrey (2002). Uncertainty in rank estimation: Implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics 27*(3), 255–270.

Louis, T. and W. Shen (1999). Innovations in bayes and empirical bayes methods: estimating parameters, populations and ranks. *Statistics in Medicine 18*, 24932505.

Maynard, A., R. Aitken, T. Butz, V. Colvin, K. Donaldson, G. Oberdörster, M. Philbert, J. Ryan, A. Seaton, V. Stone, S. Tinkle, L. Tran, N. Walker, and D. Warheit (2006). Safe handling of nanotechnology. *Nature Biotechnology 444*, 267–268.

Nel, A., L. Mädler, D. Velegol, T. Xial, E. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, and M. Thompson (2009). Understanding biophysicochemical interactions at the nano−bio interface. *Nature Materials 8*, 543–557.

Noma, H., M. Matsui, T. Omori, and T. Sato (2010). Bayesian ranking and selection methods using hierarchical mixture models in microarray studies. *Biostatistics 11*(2), 281–289.

Patel, T., D. Telesca, S. George, and A. Nel (December 2011). Toxicity profiling of engineered nanomaterials via multivariate dose response surface modeling. *COBRA Preprint Series Article 88*.

Plummerm, M., N. Best, K. Cowles, and K. Vines (2006, March). CODA: Convergence diagnosis and output analysis for MCMC. *R News 6*(1), 7–11.

Shen, W. and T. Louis (1998). Triple-goal estimates in two-stage, hierarchical models. *Journal of the Royal Statistical Society 60*, 455–471.

Society, R. (2004). Nanoscience and nanotechnologies: Opportunities and uncertainties. *The Royal Society, Science Policy Section, London, England*, 24942509.

Stanley, S., E. Westly, M. Pittet, A. Subramanian, S. Schreiber, and R. Weissleder (2008). Pertubational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences 105*(21), 7387–7392.

Stern, S. and S. McNeil (2008). Nanotechnology safetly concerns revisited. *Toxicological Sciences 101*(1), 4–21.

Tsuji, J., A. Maynard, P. Howard, J. James, C. Lam, D. Warheit, and A. Santamaria (2005). Research strategies for safety evaluation of nanomaterials, part iv: Risk assessment of nanoparticles. *Toxicological Sciences 89*(1), 4250.

Xia, T., M. Kovochich, J. Brant, M. Hotze, J. Sempf, T. Oberley, C. Sioutas, J. I. Yeh, M. Wiesner, and A. Nel (2006). Comparison of the abilities of ambient and manufactured nanoparticles to induce cellular toxicity according to an oxidative stress paradigm. *Nano Letters 6*(8), 1794–1807.