



## Research Article

# Toward Achieving Harmonization in a Nanocytotoxicity Assay Measurement Through an Interlaboratory Comparison Study

John T. Elliott<sup>1#</sup>, Matthias Rösslein<sup>2#</sup>, Nam Woong Song<sup>3</sup>, Blaza Toman<sup>4</sup>, Agnieszka Kinsner-Ovaskainen<sup>5</sup>, Rawiwan Maniratanachote<sup>6</sup>, Marc L. Salit<sup>1,7</sup>, Elijah J. Petersen<sup>1</sup>, Fatima Sequeira<sup>1</sup>, Erica L. Romsos<sup>8</sup>, Soo Jin Kim<sup>3</sup>, Jieun Lee<sup>3</sup>, Nadia R. von Moos<sup>9</sup>, François Rossi<sup>5</sup>, Cordula Hirsch<sup>2</sup>, Harald F. Krug<sup>10</sup>, Wongsakorn Suchaoin<sup>6</sup> and Peter Wick<sup>2</sup>

<sup>1</sup>Biosystems and Biomaterials Division, Material Measurement Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA; <sup>2</sup>EMPA, Swiss Federal Laboratories for Material Testing and Research, Particles-Biology Interactions Laboratory, St. Gallen, Switzerland; <sup>3</sup>Center for Nanosafety Metrology, Korea Research Institute of Standards and Science (KRIS), Daejeon, Republic of Korea; <sup>4</sup>Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA; <sup>5</sup>European Commission, Joint Research Centre (JRC), Institute for Health and Consumer Protection, Nanobiosciences Unit, Ispra, Italy; <sup>6</sup>National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand; <sup>7</sup>Department of Bioengineering, Stanford University, Stanford, CA, USA; <sup>8</sup>Biomolecular Measurement Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA; <sup>9</sup>Powder Technology Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; <sup>10</sup>EMPA, Swiss Federal Laboratories for Material Testing and Research, International Research Cooperations Manager, St. Gallen, Switzerland

### Summary

Development of reliable cell-based nanotoxicology assays is important for evaluation of potentially hazardous engineered nanomaterials. Challenges to producing a reliable assay protocol include working with nanoparticle dispersions and living cell lines, and the potential for nano-related interference effects. Here we demonstrate the use of a 96-well plate design with several measurement controls and an interlaboratory comparison study involving five laboratories to characterize the robustness of a nanocytotoxicity MTS cell viability assay based on the A549 cell line. The consensus EC<sub>50</sub> values were 22.1 mg/L (95% confidence intervals 16.9 mg/L to 27.2 mg/L) and 52.6 mg/L (44.1 mg/L to 62.6 mg/L) for positively charged polystyrene nanoparticles for the serum-free and serum conditions, respectively, and 49.7 μmol/L (47.5 μmol/L to 51.5 μmol/L) and 77.0 μmol/L (54.3 μmol/L to 99.4 μmol/L) for positive chemical control cadmium sulfate for the serum-free and serum conditions, respectively. Results from the measurement controls can be used to evaluate the sources of variability and their relative magnitudes within and between laboratories. This information revealed steps of the protocol that may need to be modified to improve the overall robustness and precision. The results suggest that protocol details such as cell line ID, media exchange, cell handling, and nanoparticle dispersion are critical to ensure protocol robustness and comparability of nanocytotoxicity assay results. The combination of system control measurements and interlaboratory comparison data yielded insights that would not have been available by either approach by itself.

Keywords: nanotoxicology, MTS assay, interlaboratory comparison, polystyrene nanoparticles, alternatives to animal testing

# authors contributed equally

### Abbreviations

DLS, dynamic light scattering; ENM, engineered nanomaterial; ID, identification; MCMC, Markov Chain Monte Carlo; MTS, 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium; NH<sub>2</sub>-PS, positively charged-polystyrene nanoparticles; NP, nanoparticles; OD, optical density; SEM, scanning electron microscopy; STR, short tandem repeat

Received May 2, 2016;  
Accepted September 20, 2016;  
Epub September 27, 2016;  
<https://doi.org/10.14573/altex.1605021>



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

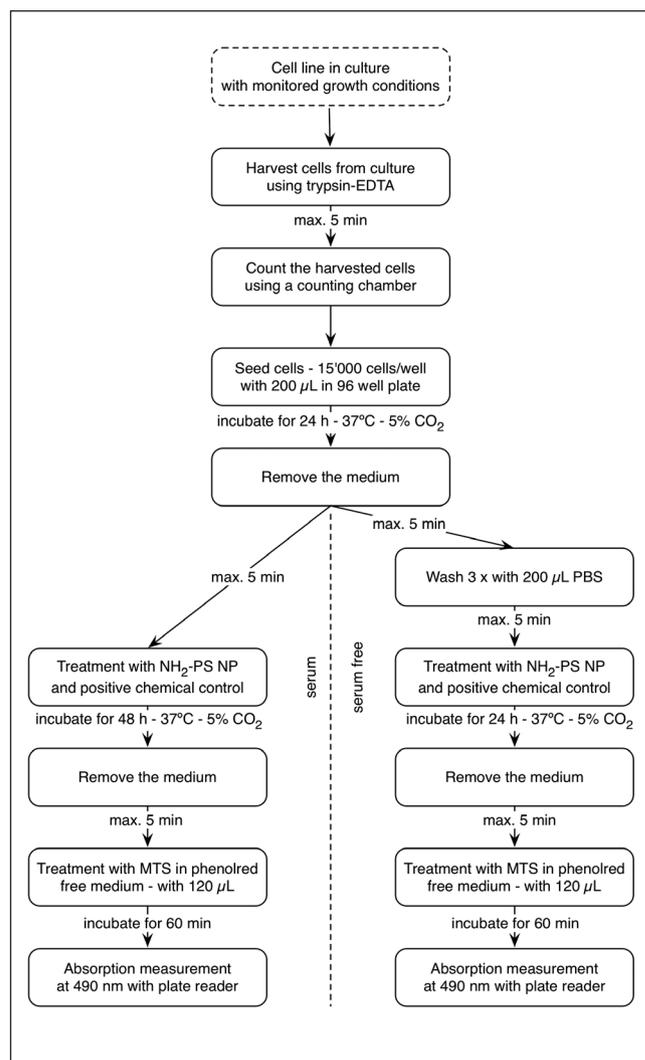


## 1 Introduction

Engineered nanomaterials (ENMs) have unique physicochemical properties due to their small size, high surface to volume ratio and spatially controlled composition. It is expected that sophisticated control of ENM manufacturing will allow the development of advanced materials which have impact in a wide range of fields including energy, textiles, and medicine (De Volder et al., 2013; Wagner et al., 2006; Graetzel et al., 2012). The increasing quantities of manufactured ENM will increase the likelihood of human, animal and organism exposure to these materials (Nel et al., 2006; Schrurs and Lison, 2012; Auffan et al., 2009). Nanotechnology Environmental and Health Safety (nanoEHS) efforts are based on the idea that physicochemical characteristics of ENM may adversely impact components of biological systems. Tests that evaluate the nature of these interactions are important for understanding hazards that may be associated with these materials.

Cell-based toxicity assays can be used as a first tier approach to identify potentially hazardous ENMs (Nel et al., 2013). Advantages of these assays are that they can be rapid, cost effective, mechanistic, used for high-throughput screening, and serve as a pathway for reducing animal testing (Nel et al., 2013; NRC, 2007; Sauer et al., 2013; Horev-Azaria et al., 2013; Clippinger et al., 2016; Landsiedel et al., 2014). However, the use of nanocytotoxicity assays has led to conflicting results from similar ENMs tested in different laboratories (Schrurs and Lison, 2012; Krug and Wick, 2011; Krug, 2014; Kaiser et al., 2011). Undocumented differences in assay protocols, differences in NP dispersions, and inadequate controls for monitoring assay performance are likely responsible for these results (Schrurs and Lison, 2012; Krug and Wick, 2011; Poland et al., 2014; Geys et al., 2010; Monteiro-Riviere et al., 2009). Recognition of these issues has resulted in the request for “standardized” nanocytotoxicity assays (Krug and Wick, 2011; Nel et al., 2013; Landsiedel et al., 2009, 2014). Standardization would indicate that assay results from different operators or laboratories are comparable and that the protocol is robust to small changes in operating conditions.

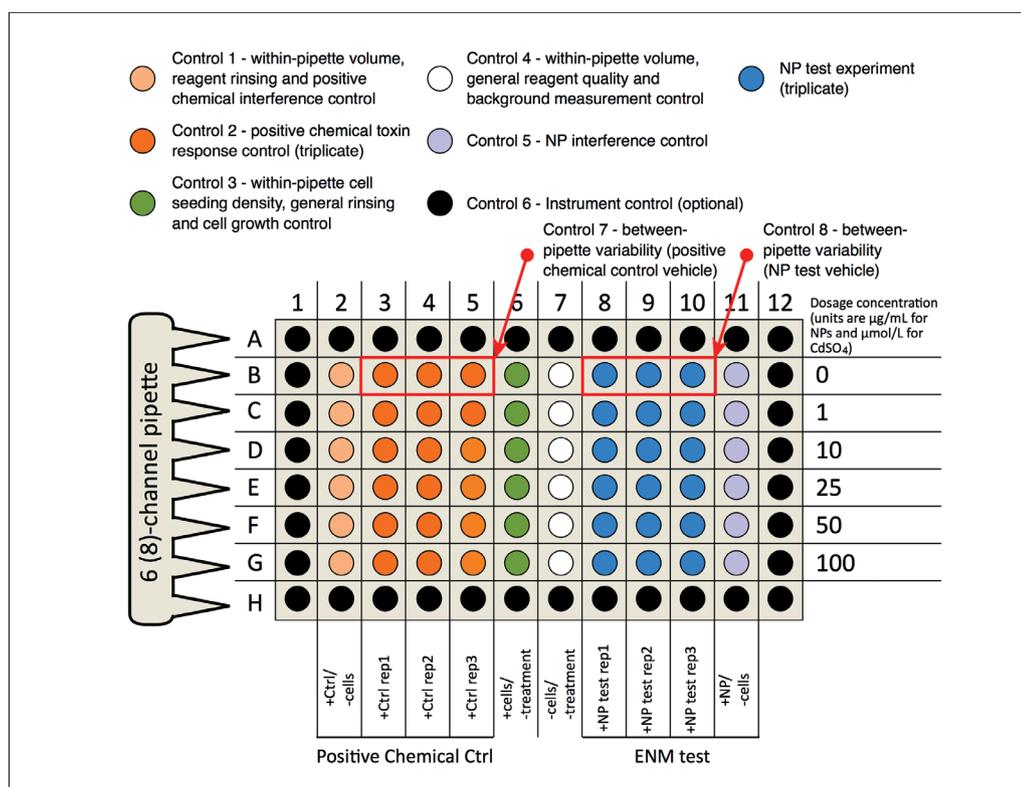
The challenges associated with developing reliable nanocytotoxicity assays are similar to those associated with the reproducibility of biological measurements. In addition to improved statistical analysis, reproduction of the whole biological measurement within a laboratory, and better reporting of method sections (Poland et al., 2014; Miller, 2014), assessing the reproducibility of biological methods across different laboratories is also critical (Plant et al., 2014). In fact, it is impossible to assess all of the sources of variability in an assay (i.e., dark uncertainty (Thompson and Ellison, 2011)) without the aid of a comprehensive interlaboratory study. Measurements from several different laboratories are more likely to produce results that incorporate additional sources of variability including the unknown factors affecting the results of the assay. Thus, results or protocols obtained from a single laboratory, even if they are shown to be reproducible within that laboratory and utilize good quality assurance/quality control practices, are not sufficient for ensuring similar results among multiple laboratories (Warheit



**Fig. 1: Flow chart describing the modified MTS protocol for this study**

and Donner, 2015). Additional experimental design components such as appropriate control experiments, specifications for ensuring valid method performance, and confirmation of the robustness of the method to unintended variation in the experimental protocol are some of the features of interlaboratory studies which enable confidence in comparing test results among the laboratories (Plant et al., 2014). A measurement science approach that includes systematic understanding of the sources of variability in the assay protocol and provides a comprehensive set of system controls to ensure acceptable assay performance is proposed to improve reproducibility in nanotoxicity and other biological assays.

Here, we show the results of a 5 laboratory (NIST, EMPA, KRISS, JRC, NANOTEC) comparison of a nanocytotoxicity assay and the evaluation of the transferability and reproducibility of the assay procedure. The 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) assay is a colorimetric assay for testing cell viability by measur-



**Fig. 2: 96-well plate layout and 8 control measurements**

Modified and reprinted with permission from Rösslein et al. (2014).

ing the reduction of the tetrazolium dye to insoluble formazan, a process that only occurs in live cells, using a plate reader. The MTS assay was chosen because it is widely used in cytotoxicity studies and the protocol has only a few basic steps. We also evaluated the effect of using serum-free in comparison to serum-containing medium during the incubation of ENM as serum may affect the response of cells to ENM. Figure 1 illustrates the serum and serum-free protocols (Protocols S1<sup>1</sup>, S2<sup>1</sup>) that were derived from the manufacturer's instructions and from the results of a cause-and-effect analysis of the MTS assay (Rösslein et al., 2014). The protocols include 8 system controls to quantify critical sources of variability in the assay (Fig. 2, Tab. 1).

Our study used a single ENM and the human A549 cell line in two variants to demonstrate the value of the measurement science approach in understanding sources of variability in an assay. By not controlling all of the aspects of cell culture, the experimental design of the interlaboratory study mimicked possible sources of variability including differences in cell treatment procedures, sources of serum, culture medium, and cell culture plates (Tab. S1<sup>2</sup>, S2<sup>2</sup>). Some factors were controlled, specifically that all participants had the same stock of  $\text{NH}_2\text{-PS}$  nanoparticles (NP) (i.e., positively charged-polystyrene NP), positive chemical control reagent ( $\text{CdSO}_4$ ) and the two human A549 cell lines. The  $\text{NH}_2\text{-PS}$  NP was chosen as our model system since positively charged NPs have been shown to be toxic to many different cell types (Nemmar et al., 2003; Shen et al.,

2009; Hong et al., 2006) and because the  $\text{NH}_2\text{-PS}$  NP will not release dissolved ions which may cause toxicity unlike, for example, copper oxide NPs. Cadmium sulfate was chosen as a chemical positive control because it is stable in solution, known to be toxic, is highly soluble in aqueous media, and can be readily quantified, which can help ensure that the same concentration is being used across time and among laboratories during the interlaboratory comparison.

This study differs from a recently published interlaboratory comparison (Xia et al., 2013) in that we systematically evaluate the contributions to the total variability from the various steps of the assay and provide numerous specifications that ensure comparability of the assay measurement process. This study was not designed as a comparison of the effect of different cell lines or various ENMs; the objective was to show the power of using an assay design with system control measurements in combination with an interlaboratory comparison experiment to systematically understand the sources of variability in the assay.

## 2 Methods

### Coordination

Tasks were divided between each of the laboratories. EMPA served as sample and data coordinator. EMPA and NIST were responsible for experimental design, statistical analysis, and preliminary data collection. KRIS was responsible for NP

<sup>1</sup> <https://doi.org/10.14573/altex.1605021p>

<sup>2</sup> <https://doi.org/10.14573/altex.1605021s>

**Tab. 1: Control settings deduced from the cause-and-effect analysis and implemented into a 96-well plate layout**

This table is modified and reprinted with permission from Rösslein et al. (2014).

Control Number	Brief Control Description
1	Background correction for the chemical dosing. Allows detecting if increasing concentrations of the reference chemical change the final MTS measurement result.
2	Triplicate reference chemical control. Shows that the assay worked as expected.
3	Assesses <i>within</i> multichannel pipetting variance. Non-treated cells seeded with a single multichannel pipette ejection step. Absolute absorbance measurement provides insight on nominal cell growth. Indicates technical problems with the pipette.
4	These wells contain no cells but MTS reagent (last step of assay procedure). Allows for background absorbance correction and variance correction. Together with outermost wells (control No. 4) assesses possible internal measurement gradients due to instrument malfunction or culture plate variability. Large variations may further indicate issues with the MTS reagent.
5	Background correction for the ENM dosing. Allows detecting if increasing concentrations of the ENM change the final MTS measurement result, i.e., interfere with the final readout.
6	Additional considerations that exceed those described in No. 3 “no cells no treatment”: these wells contain medium from the time of cell seeding on. This helps to circumvent so-called edge effects that might occur during longer incubation times of cells seeded in small volumes in the outermost wells (i.e., evaporation).
7+8	Assesses <i>between</i> multichannel pipetting variance. Solvent treated cells (compare B3-B5 for chemical control, B8-B10 for ENMs) seeded in different ejection steps. This control indicates handling problems of the operator during the seeding procedure and possible effects of the solvent if compared to “no treatment” wells (B6-G6)

characterization in serum-containing and serum-free media. NIST conducted the SEM analysis.

#### *Protocol development and experimental design of the interlaboratory comparison*

The final measurement protocol is a modified version of the MTS manufacturer’s protocol and is contained in the supplementary material (Protocols S1<sup>1</sup>, S2<sup>1</sup>). The protocol flowchart is shown in Figure 1 and is summarized below. The 96-well plate design was based on previous work at EMPA and NIST for studies in the International Alliance of NanoEHS Harmonization (IANH). Figure 2 shows an image of the 96-well plate design including the description of the controls.

All replicate experiments or “rounds” in each lab carried out a full implementation of the protocol. Each round included four 96-well plates; 2 plates contained A549-A cells and 2 plates contained A549-B cells. One plate with each cell line was used with NP dispersed in serum-containing and the other plate with NP dispersed in serum-free medium. Each plate contained both a positive chemical control, i.e., CdSO<sub>4</sub>, a NP test result and several control experiments encoded into the 96-well plate design (Fig. 2, Tab. 1). Laboratories A, B, C, D and E performed 8, 6, 6, 4, and 3 rounds, respectively, before the data was sent to EMPA for statistical analysis.

#### *Reagents and equipment*

NH<sub>2</sub>-PS NP suspended in H<sub>2</sub>O were obtained from Bangs Laboratories Inc. (Fishers, Indiana, US), lot number 10351 and inventory number L120117F, 10% (w/v). CdSO<sub>4</sub>·7H<sub>2</sub>O (Sigma-Aldrich) was dissolved in 18 MΩ water to make a final concentration of 10 mmol/L. Aliquots of both the nanoparticle

suspension (200 μl) and CdSO<sub>4</sub> solution (4 ml) were shipped overnight to each of the laboratories under ambient conditions. MTS reagents (Promega, Madison WI) were purchased by each laboratory. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the NIST nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

The precision of the plate readers was tested by measuring a single plate with cells treated with the MTS reagent multiple times.

#### *Characterization of the NH<sub>2</sub>-PS NP*

The original suspension of NH<sub>2</sub>-PS NP received from EMPA was diluted by a factor of 1000 in 18 MΩ water, cell culture medium (RPMI-1640) without serum, cell culture medium (RPMI-1640) containing 0.1% fetal bovine serum (FBS), or cell culture medium (RPMI-1640) containing 10% FBS. The sample suspensions were measured for size and zeta potential immediately after the dilution and after incubation for 24 and 48 h (37°C, 5% atmospheric CO<sub>2</sub>).

The hydrodynamic diameter was measured by dynamic light scattering (DLS) with a particle size analyzer (ELS-Z, Otsuka electronics Co. Ltd.). Twelve repetitive measurements were performed with 5 sub-runs for each measurement using a 2-sec detection time. To remove potential interference from dust scattering, the six measurement results with the lowest DLS values were selected to calculate the average ENM size and its uncertainty. The scattering intensity of the nanoparticle suspension was ≈ 25 times greater than that of the medium containing



10% serum but without NPs, which suggests that the serum had a negligible impact on the DLS results. The zeta potential of the NH<sub>2</sub>-PS NP in the cell culture media suspensions (without serum, with 1% serum, and with 10% serum) was measured using an electrophoretic mobility analyzer (Zetasizer Nano Z, Malvern Instrument). Values from five independent measurements were used to calculate the average zeta potential and its uncertainty.

Primary particle size was measured via scanning electron microscopy (SEM) using a Zeiss NVision 40 focused ion beam/scanning electron microscope operating at 15 kV.

### Cell lines

Two A549 human lung cancer cell lines were used in this study. One cell line (A549-B) was purchased from ATCC (Manassas, VA) immediately before the study began. Seed and working stock distribution vials were prepared at NIST and working stock vials were sent to EMPA who prepared a working stock and distributed it to the other laboratories. A second A549 cell line (A549-A) purchased from I.A.Z. Institute (Munich, Germany) in 2005 was also distributed to each of the laboratories. The passage number after acquisition of the cell lines from the source was less than 25 for both cell lines for all experiments. The two cell lines were sent to each laboratory at passage number 5 and were cultured for at least 3 passages before being used in the interlaboratory comparison. Information was not given by the providers of the cells about the passage number prior to shipping the cells. Each laboratory then prepared seed and working stocks of the cell lines.

Cells were maintained in Roswell Park Memorial Institute 1640 (RPMI-1640) medium containing 10% fetal bovine serum, L-glutamine and an antibiotic mixture (penicillin and streptomycin at concentrations of 100 IU/mL and 10 µg/mL, respectively). Sources of media, serum, antibiotics and other cell culture items such as plastic ware are minimally specified in the protocols (Protocols S1<sup>1</sup>, S2<sup>1</sup>), and each laboratory obtained the supplies from different vendors to mimic typical laboratory conditions; information about the sources are provided in Table S1<sup>2</sup>.

Cell line characteristics including proliferation rate, average cell volume and DNA short tandem repeats (i.e., cell line ID) were performed on each of the cell stocks used for experiments (Tab. S3<sup>2</sup>) (Yu et al., 2015). Mycoplasma was tested after the cells were received and periodically by several of the laboratories using the ATCC assay and all tests were negative.

### Abbreviated protocols

All culturing of A549-A and A549-B cells was performed at 37°C, 5% CO<sub>2</sub> in humidified air. Maintenance of cells was performed in 75 cm<sup>2</sup> cell culture flasks which were passaged and seeded as described in Table S2<sup>2</sup>. Cell harvesting was performed using trypsin and counting the cells using a hemocytometer in the presence of trypan blue using the procedure described in Protocols S1<sup>1</sup> and S2<sup>1</sup>. The cell suspension was evaluated microscopically to confirm that the cells were single cell suspensions (no clumping) prior to cell counting. Cells were seeded at 1.5 × 10<sup>4</sup> cells per well; additional wells were also prepared using cell culture medium without cells. The plate

was then cultured for 24 h. For the cells exposed in the presence of serum, the medium was removed and fresh complete medium was added in addition to CdSO<sub>4</sub> or NP. For the cells in the serum-free condition, the same procedure was followed except that the cells were first washed three times with PBS to ensure removal of the serum and medium containing no serum was added in addition to CdSO<sub>4</sub> or NP. The exposure period was 24 h in serum-free medium or 48 h in medium containing serum. After the exposure period, the medium was removed from all wells, and the MTS reagent mixed with phenol red-free RPMI-1640 (without L-glutamine or antibiotics) was added to each well. The plates were then incubated for 60 min. Then, absorption measurements were performed with a plate reader at a wavelength of 490 nm.

Microscopic visualization of the cells in the 96-well plate was performed periodically to confirm that the decrease in MTS signal corresponded to a decrease in cell number.

### Statistical analysis

Relative absorbance values were calculated for cells treated with CdSO<sub>4</sub> and NP for each treatment condition by dividing the median of the treatment condition less the median of the no cell background value by the median of the vehicle control condition less the median of the no cell background value.

Conventional statistical analysis such as calculation of median and mean values, 95% confidence intervals and median absolute deviation values for the control experiments grouped by laboratory were performed using the R software package and ggplot2 to produce related figures. Both within-plate for a single laboratory and between-laboratory round statistics were calculated and plotted to identify potential systematic trends and widely dispersed data. Performance specifications for the assay control experiments were generated from data collected in all rounds and laboratories that were not considered to be outliers.

The EC<sub>50</sub> values (e.g., concentration that causes a 50% change in the assay readout) and the uncertainty of the EC<sub>50</sub> values were calculated using Markov Chain Monte Carlo (MCMC) fitting procedures (Cornfield and Mantel, 1950) on the following statistical model. The data were the responses  $r_i$  observed at  $i = 1, \dots, 6$  increasing dosing concentrations. The function of concentration versus  $1 - r_i$  was used to form a dose-response curve, specifically, a logistic regression curve of the following form

$$E(1 - r_{ij}) = \frac{\gamma_i}{1 + e^{(\alpha_i - x_j)/\beta_i}} \text{ for } i = 1, \dots, 5 \text{ and } j = 1, \dots, 6 \quad (1)$$

$$\alpha_i \sim N(\alpha, \text{sig}\alpha) \quad (2)$$

$$\log(\beta_i) \sim N(\beta, \text{sig}\beta) \quad (3)$$

$$\log(\gamma_i) \sim N(\gamma, \text{sig}\gamma) \quad (4)$$

Here  $E$  is the expected value of the response,  $N(a, b)$  stands for Gaussian distribution with mean  $a$  and variance  $b$ . In this model,  $\alpha_i$  is the EC<sub>50</sub> for the  $i^{\text{th}}$  lab. In this hierarchical statistical model (Gelman et al., 2008) the  $\alpha_i$ ,  $\beta_i$ , and  $\gamma_i$  parameters of (1) in the top hierarchy determine the dose-response curves for each lab and in the second hierarchy the  $\alpha$ ,  $\beta$ , and  $\gamma$  of expressions (2),



**Tab. 2: Comparison of within lab uncertainty to total uncertainty for determination of the EC<sub>50</sub> values**

Condition	Total uncertainty (span of 95% confidence interval for consensus values)	Within lab uncertainty (average span of 95% confidence interval for each laboratory <sup>a</sup> )	Proportion of within lab uncertainty to total uncertainty <sup>b</sup>
NP serum-free A cell (mg/l)	11.9	2.1	18%
NP serum-free B cell (mg/l)	10.3	3.6	35%
NP serum A cell (mg/l)	24	14.3	59%
NP serum B cell (mg/l)	18.5	6.1	33%
CD serum-free A cell (μmol/l)	4.3	2.9	67%
CD serum-free B cell (μmol/l)	4.0	3.9	98%
CD serum A cell (μmol/l)	32	32	100%
CD serum B cell (μmol/l)	45.1	45.1	100%

<sup>a</sup> Outliers identified elsewhere were excluded in this analysis.

<sup>b</sup> Outliers identified elsewhere were excluded in this analysis. A value of 100% indicates that all of the total uncertainty in the test result can be explained by the within lab uncertainty. This would indicate a negligible contribution from between lab variability.

(3), and (4) represent the parameters of a common consensus curve. This random effects interlaboratory model (Toman and Possolo, 2009) is generalized to logistic regression. The model takes full advantage of the additional between-laboratory variability information produced by the interlaboratory study and estimates the variability that is due to unknown factors which differ among laboratories (Thompson and Ellison, 2011).

The parameters were estimated using Bayesian MCMC methods (Congdon, 2001) with non-informative prior distributions implemented using the freeware OpenBUGS (Lunn et al., 2009). The Bayesian analysis resulted in point estimates of the parameters as well as asymmetric 95% probability confidence intervals.

In cases where the logistic regression model resulted in a poor fit (i.e., the positive chemical control in the serum containing experiments, Fig. 10) other regression models such as a straight-line model were used to fit the data. The resulting EC<sub>50</sub> values were not statistically different and thus a logistic regression model was used for all fitting. Correlation plots between control statistics and the EC<sub>50</sub> values were used to identify potential control measurements whose value directly affects the measured EC<sub>50</sub> value. Additional details about the MCMC approach used in this study are provided in a recent publication (Toman et al., 2016). In some cases, the 95% confidence interval of the dose-response curve extended below 0 absorbance units. This is an artifact of the fitting procedure and does not have physical significance.

Two different approaches were used to further study the variability in the EC<sub>50</sub> values. Multivariate analysis of variance (ANOVA) of the EC<sub>50</sub> values consistently revealed significant differences ( $p < 0.05$ ) among laboratories under all conditions (both cell types; serum-containing and serum-free medium; and for the CdSO<sub>4</sub> and PS-NH<sub>2</sub> NP). However, the ANOVA approach only compared the EC<sub>50</sub> values and did not

include the substantial uncertainty in these values from the logistic curve fitting procedure. The second approach was a comparison of the 95% confidence intervals for the EC<sub>50</sub> values for each laboratory (excluding outliers) to the 95% confidence interval for the consensus EC<sub>50</sub> values (Tab. S5<sup>2</sup>, S9<sup>2</sup>). These confidence intervals were produced as part of the curve fitting and thus accounted for all uncertainty, including the between-laboratory uncertainty, and as a consequence show much better agreement among the measured EC<sub>50</sub> values. It is important to note that this approach is capable of calculating these values even though the number of rounds performed varied among laboratories. In summary, Table 2 shows the total uncertainty, average within-lab uncertainty, and the proportion of within-laboratory uncertainty to total uncertainty for all of the conditions tested. When the average within-laboratory uncertainty was comparable to the total uncertainty, the data was considered to be harmonized. When the between-laboratory uncertainty was substantially larger than the within-laboratory uncertainty, control measurements were used to identify protocol steps that contributed to the larger between-laboratory variability.

#### *In vitro sedimentation, diffusion and dosimetry model for nanomaterials*

Target cell dose as a function of time and exposure conditions, i.e., the effectively deposited dose at the bottom of the exposure well, was computationally estimated by the *In vitro* Sedimentation, Diffusion and Dosimetry model known as ISDD. The model numerically solves a partial differential equation for both diffusion and sedimentation. It is available as Matlab code and as a Windows Executable from its developers (Hinderliter et al., 2010). It can be downloaded from <http://nanodose.pnnl.gov/ModelDownload.aspx>. Further details and modeling parameters can be found in the supplementary information<sup>2</sup>.

### 3 Results

#### 3.1 Cell line characterization

Two A549 cell lines from different sources were used in the interlaboratory study. While the mean generation times of the two cell lines were statistically identical ( $22.6 \text{ h} \pm 2.2 \text{ h}$  for A549-A and  $22.5 \text{ h} \pm 2.5 \text{ h}$  for A549-B, Tab. S3<sup>2</sup>), genotyping of the genomic composition of the cell lines using commercial human-specific STR assays revealed that the A549-A cell line exhibited a dropout of the 12<sup>th</sup> allele at the CSF1PO position of chromosome 5 (Fig. S1<sup>2</sup>). All other 23 markers were identical between the two cell lines and were consistent with the STR profile determined by ATCC. The short tandem repeat (STR) DNA analysis was repeated on the initial parent stock cell lines with the same results, indicating the A549-A cells arrived from the original vendor with the missing allele. No mycoplasma contamination was detected in the stock solutions sent out to the laboratories or in a receiving laboratory.

#### 3.2 The effect of serum on NH<sub>2</sub>-PS NP

Incubation of the NH<sub>2</sub>-PS NP in 10% FBS but not in the serum-free medium (Tab. S4<sup>2</sup>) led to substantial agglomeration

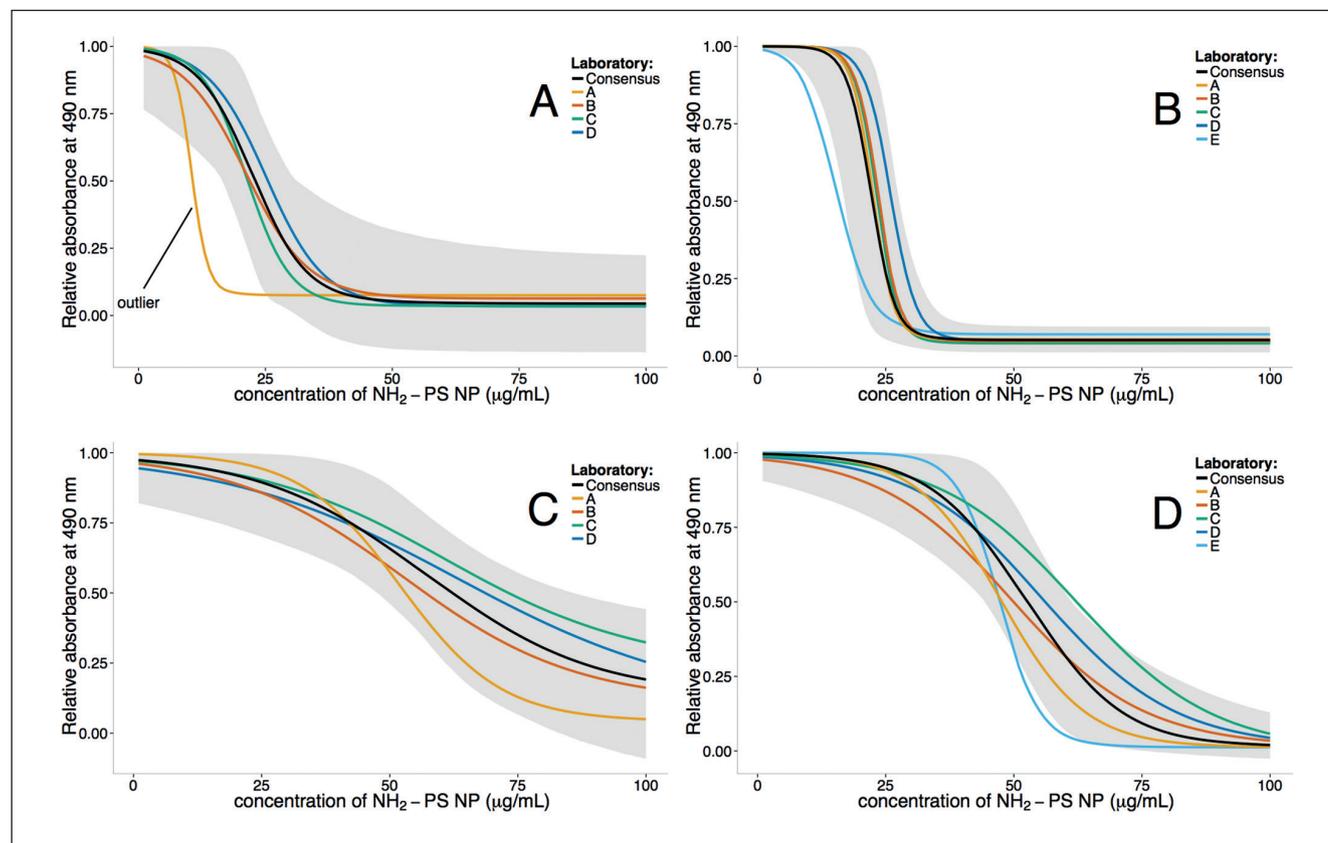
(56 nm to 1258 nm) and change in zeta potential (48.8 mV to -10 mV). The size distribution of the primary particles determined by scanning electron microscopy was  $51 \text{ nm} \pm 9 \text{ nm}$  (uncertainty indicates standard deviation value of 200 particles; Fig. S2<sup>2</sup>) (Hanna et al., 2016).

Therefore, it was decided to test the effects of NH<sub>2</sub>-PS NP in serum-containing and serum-free conditions to assess the effect of agglomeration on the variability in the assay results.

#### 3.3 The effect of NH<sub>2</sub>-PS NP on A549 cell response

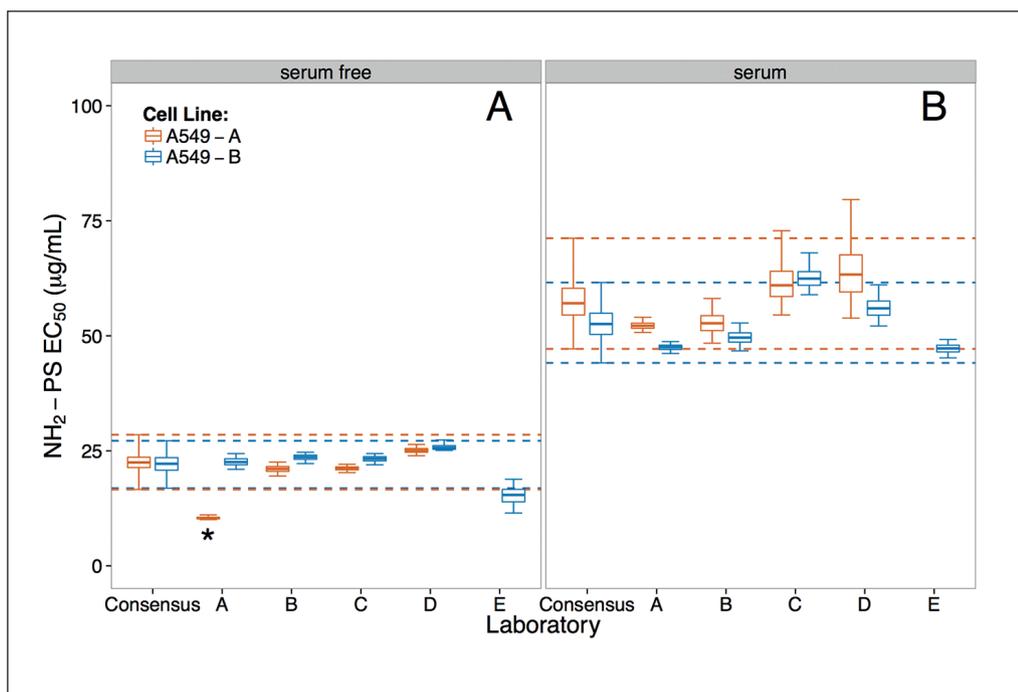
Figure 3 shows the dose-response curves obtained for A549-A and A549-B cells treated with NH<sub>2</sub>-PS NP in serum-free and serum-containing media in the 5 laboratories using Protocols S1<sup>1</sup> and S2<sup>1</sup>, respectively. Figure 4 shows the estimated EC<sub>50</sub> values for each of the average NP dose-response curves over all the rounds for each laboratory as determined by fitting the data with a MCMC simulated logistic curve.

Under serum-free conditions (Fig. 4A), the consensus EC<sub>50</sub> values (excluding the laboratory A outlier) for the NH<sub>2</sub>-PS NP on the A549-A and A549-B cells were both approximately  $22 \text{ } \mu\text{g}/\text{mL}$  (see Tab. S5<sup>2</sup> for exact values), indicating a similar effect of the NP on both cell lines. Within-laboratory variabil-



**Fig. 3: NH<sub>2</sub>-PS NP dose response curves for A549-A cells (A, C) and A549-B cells (B, D) conducted in serum-free (A, B) and serum-containing (C, D) conditions**

Relative absorbance value calculations are described in Section 2. All rounds from each laboratory except those identified as outliers were used in a Bayesian statistical model to generate the dose-response curves. For part A, Laboratory A was identified as an outlier under these conditions. The grey area represents the 95% confidence interval of the consensus curve as determined by Bayesian statistical modeling of all data.



**Fig. 4: NH<sub>2</sub>-PS EC<sub>50</sub> values for A549-A and A549-B cells in serum-free (A) and serum-containing (B) conditions**

Whisker error bars represent 95% confidence interval from all rounds from each laboratory. Consensus values were generated by Bayesian modeling of all rounds of data from all of the laboratories that were not considered outliers. Only one set of data (serum-free condition, A549-A cells, Lab A) was considered an outlier. This data set was marked with an asterisk.

ity for the EC<sub>50</sub> value for the A549-A and A549-B cell lines is only 18% and 35%, respectively, of the total laboratory variability represented by the confidence intervals on the consensus value (Tab. 2). This suggests that despite the use of a detailed protocol, differences among the laboratories, such as reagents and experimental technique, contribute to greater than 65% of the total variability associated with the consensus value. Our experimental design did not allow the further separation of the variability that results specifically from differences in reagents or experimental techniques among laboratories.

The consensus EC<sub>50</sub> values for NH<sub>2</sub>-PS NP incubated in the presence of serum with the A549-A or A549-B cells (Fig. 4B, Tab. S5<sup>2</sup>) were 57 µg/mL and 53 µg/mL, respectively, and exhibited overlapping 95% confidence intervals. These intervals were significantly larger than those observed under serum-free conditions. The within-laboratory variability for the A549-A and A549-B cell lines was 59% and 33%, respectively, of the total variability observed in the consensus value (Tab. 2). Testing NH<sub>2</sub>-PS NP in the presence of serum thus increased the within-laboratory variability, suggesting that parts of the protocol involved in the addition of NP to serum-containing media increase variability in the test result even when repeated within the same laboratory. The variability between laboratories was at least 41% of the total variability, indicating that differences between reagents and experimental technique among the laboratories were also significant contributors to the total variability in the consensus value.

The EC<sub>50</sub> values reported in Table S5<sup>2</sup> are based on the nominal mass concentration of NP in the cell culture media. Modified dosing metrics that account for the diffusion and sedimentation properties of the nanomaterials in the media were generated using the ISDD model (see Supplemental Materials<sup>2</sup> for additional

details and discussion). This model estimates the delivered dose of the NP to the surface of the cells based on the measured properties of the NP and the respective medium (see Tab. S6<sup>2</sup> for the delivered doses and Tab. S8<sup>2</sup> for modeling parameters). The corrected EC<sub>50</sub> mass dose for the NP in both serum-containing and serum-free medium was approximately 2.3 µg/cm<sup>2</sup> (Tab. S7<sup>2</sup>). Although this value is reported in mass per surface area, the similarity in the value between the two conditions suggests that the increased measured EC<sub>50</sub> value for NP in serum is due to a reduction in available NP that can interact with cells due to agglomeration. Further studies are required to understand the mechanisms responsible for the difference in EC<sub>50</sub> values between the serum-containing and serum-free conditions.

### 3.4 System control experiments

Several system control experiments were incorporated in the 96-well plate design and the protocol to further evaluate potential sources of variability in the measurement process.

#### *Within-column cell density control (control 3)*

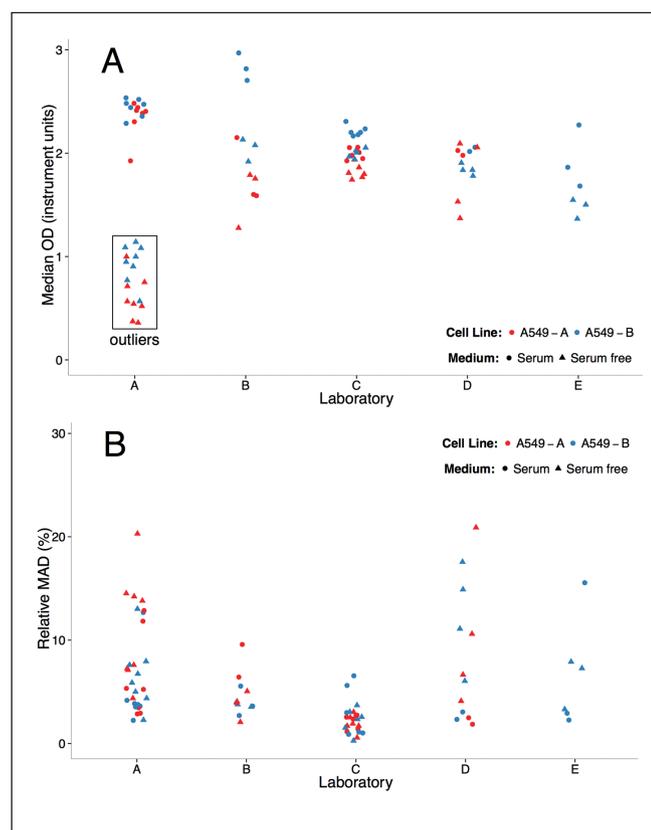
Protocols S1<sup>1</sup> and S2<sup>1</sup> specified that a single multi-channel pipetter is used to seed cells into each designated column of a 96-well plate. The nominal number of cells remaining in a well after rinsing and the variability in the cell density along the wells in the column were assessed by control 3 (Tab. 1, Fig. 2). The median absolute optical density of this control in both serum-containing and serum-free conditions for each round was on average 1.8 with a CV of < 10% (Fig. 5A). Laboratory B exhibited a CV larger than 10% but the highest OD values were only observed for the A549-B cells under serum-containing conditions. Laboratory A showed a cluster of data points from all the serum-free experiments with an average OD value of 1. This

outlier data cluster was lower than in the control experiments for the serum condition in that laboratory and the results from the other laboratories and correlated with the outlier  $EC_{50}$  value data shown in Figure 3A and 4A, suggesting that the experimental steps that this laboratory used during the execution of the serum-free Protocol S1<sup>1</sup> resulted in both a low value for this control and a low  $EC_{50}$  value for the NP toxicity experiment.

Sensitivity analysis was used to determine if the estimated  $EC_{50}$  values for the  $NH_2$ -PS NP in all experiments were correlated with the cell densities measured in the control 3 wells. A correlation plot between the median OD values of control 3 and the  $EC_{50}$  data from each round from all of the laboratories (Fig. 6A) indicated that significant correlations were only observed under the serum-free conditions when including the outlier data cluster from laboratory A. This result suggests that an OD value below 1 in control 3 wells can decrease the  $EC_{50}$  value of the test result. There is no correlation with the OD value in control

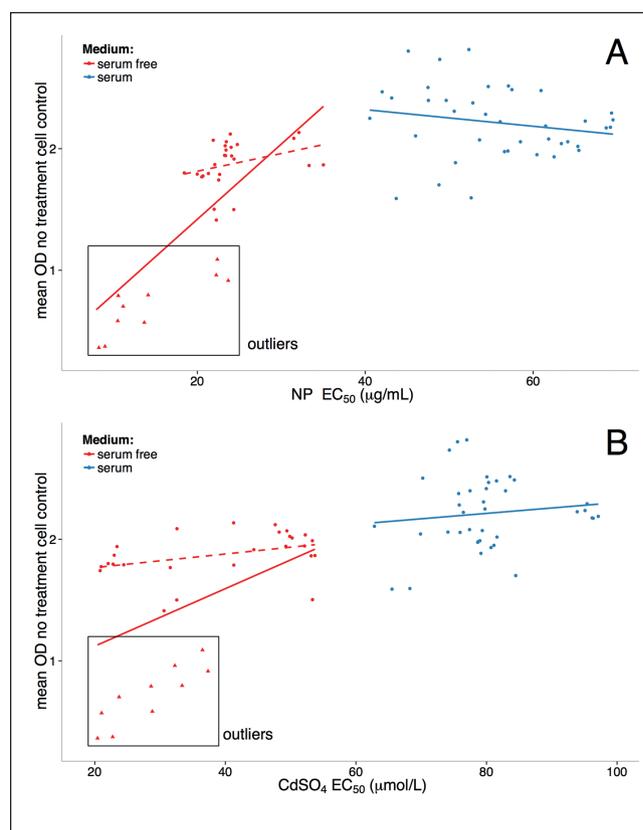
3 wells and the  $NH_2$ -PS NP  $EC_{50}$  value if the control wells have an OD value between approximately 1.4 and 2.7. This analysis helps to explain the outlier dose-response curve shown for A549-A cells after NP treatment in serum-free conditions observed for Laboratory A in Figure 3A because the OD values for control 3 wells were below this range.

The wells in control 3 also provide a measure of within-column well-to-well variability in cell density. This control was prepared with a single multi-channel pipette ejection, and thus, large standard deviations between the wells can indicate pipette malfunction, inhomogeneous filling of the pipettes with cells, or aggressive rinsing techniques that dislodge cells from the surface inhomogeneously. The within-column well variability was as high as 20% in some cases but most of the rounds showed variability below 10% (Fig. 5B). Only laboratories B and C exhibited within-column well-to-well variability of less than 10% for all rounds.



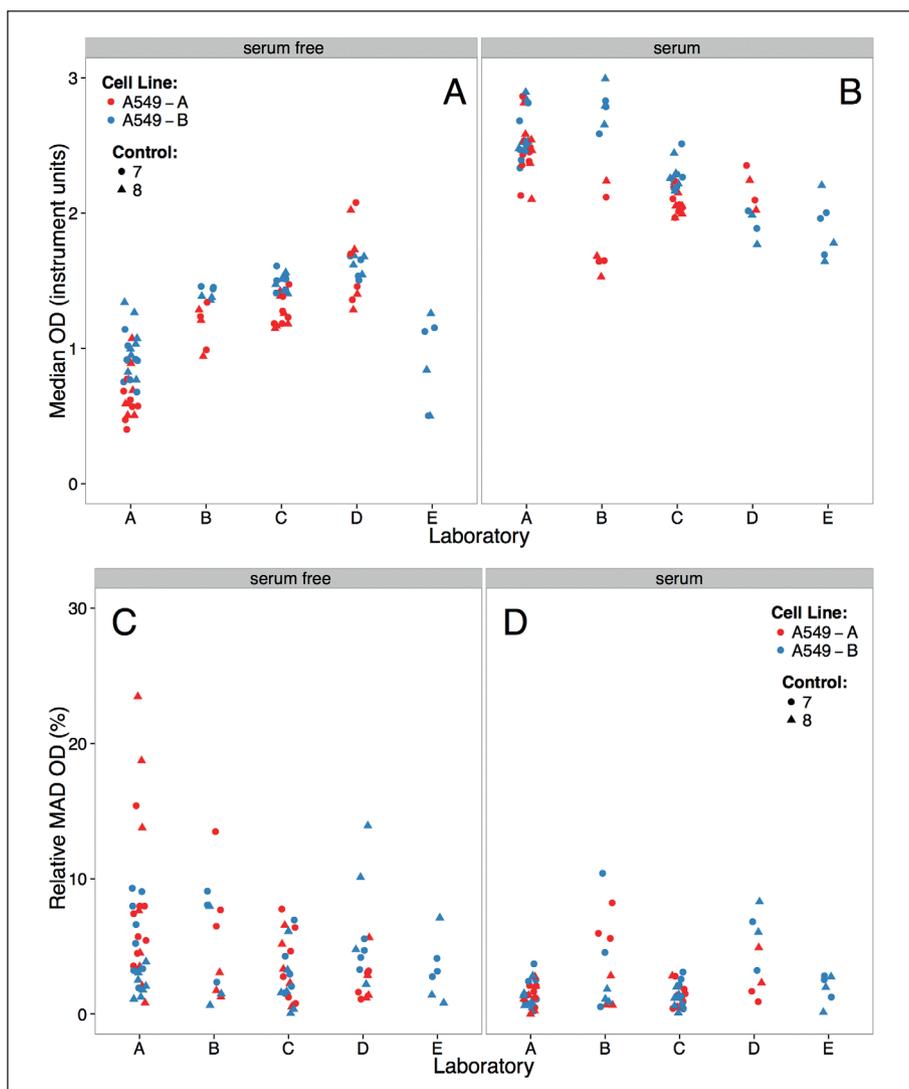
**Fig. 5: Median values (A) and relative median average deviation (MAD) (B) of control 3 – within pipette cell seeding density of non-treated cells**

Values for serum-free conditions included a rinsing step. An outlier set of data (red box) was observed in laboratory A. Each data point is the median or MAD value of a column of a plate for each round from a laboratory. Different levels of variability observed between laboratories suggests that there may be additional day to day variability in the rinsing step.



**Fig. 6: Correlation of  $NH_2$ -PS NP  $EC_{50}$  values (A) and  $CdSO_4$   $EC_{50}$  values (B) with mean OD no treatment cell control values for both cell lines (control 3)**

Outlier data is indicated by red boxes. The solid line is the linear regression fit with outliers, while the dotted line is the fit without outliers. When outliers are included for the serum-free treatments for the  $CdSO_4$  and the  $NH_4$ -PS NP (OD below 1.0), the slopes are statistically different from 0, indicating that the  $EC_{50}$  value is correlated with control 3 values. However,  $EC_{50}$  values are not correlated with the non-treatment mean OD when these outliers are removed or for the serum treatment condition.



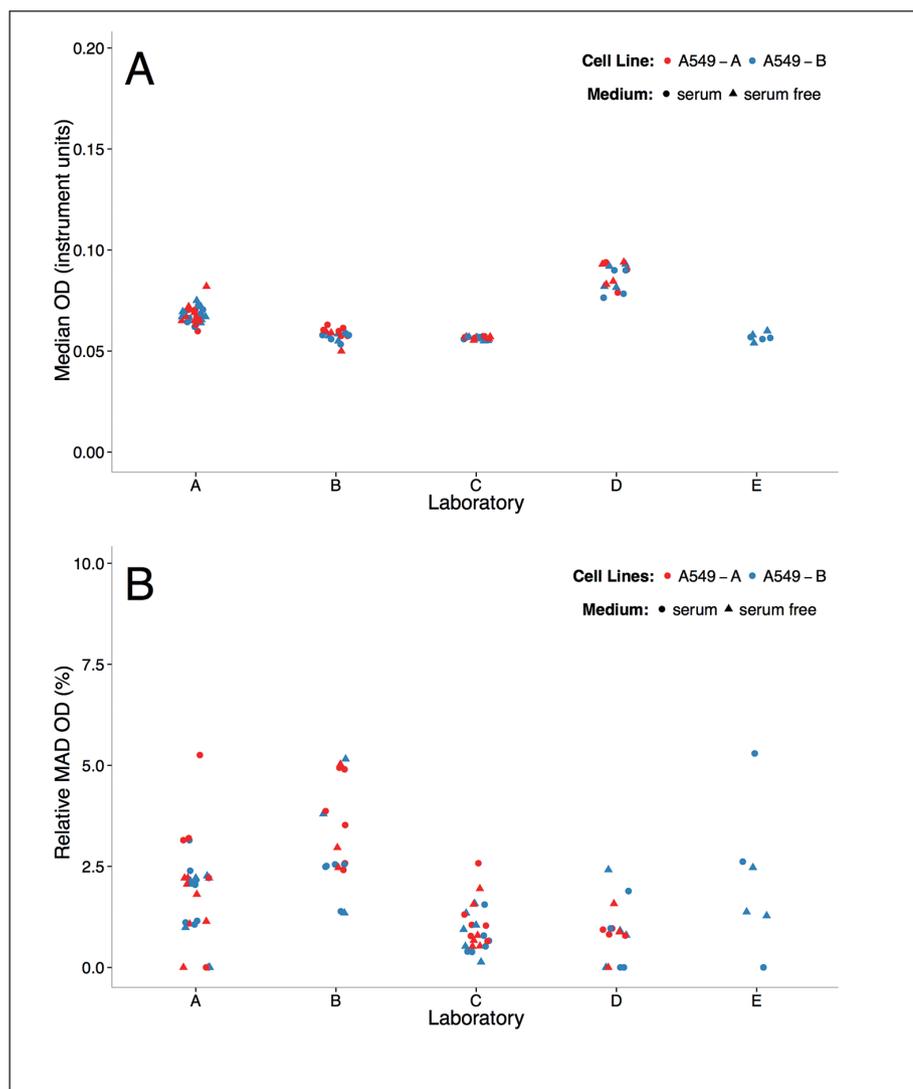
**Fig. 7: Median values (A, B) and relative median average deviation (MAD) (C, D) of controls 7 + 8 – between pipette variability in serum-free (A, C) and serum (B, D) conditions**

Values for serum-free conditions included a rinsing step. Each data point is the median or MAD value of a plate for each round from a laboratory. Higher MAD values are generally observed in the serum-free condition, which includes a rinsing step. The similar values observed here as compared to Figure S4<sup>2</sup> indicate minimal vehicle effect.

#### *Between-column cell density control*

Controls 7 and 8 are a second set of pipetting controls that evaluate cell density variability between pipetting steps. They also report on the impact of different solvent vehicles for the positive chemical control and NP on a cell density measurement (Fig. 2). The average median OD values for both controls were significantly lower in serum-free conditions compared with serum-containing conditions (Fig. 7A,B). This could be due to a combination of longer incubation and fewer rinsing steps for the serum condition. The variability in these control values is similar between the laboratories (approximately 10%), but there is a systematic shift of the cluster average dependent on the laboratory. This suggests that reagents and protocol techniques specific to each laboratory such as cell resuspension during seeding, rinsing or cell counting before seeding steps introduced a systematic bias. An exception to this result was observed in laboratory B where clear separation of the A549-A and A549-B data occurred under the serum conditions (Fig. 7B). The cause of this systematic difference is unknown. The relative median

average deviation, a robust measurement of the variability of a univariate sample, between the wells in control 7 and 8 from all laboratories in both serum-containing and serum-free conditions is approximately 10% (Fig. 7C,D). Using medium containing serum, laboratories A, C and D showed variability of less than 5%, which is less than the average variability of 7% from all the laboratories. This suggests that pipetting or cell resuspension techniques in laboratory B and E should be compared to laboratories A, C, and D to identify possible differences in executing the cell seeding steps. The average within-laboratory variability between the control wells under serum-free conditions is significantly larger (approximately 12%) than under serum-containing conditions. This suggests that the additional rinsing step in the serum-free protocol may introduce additional variability. Correlations between either the median OD or the relative median average deviation values of these controls and the EC<sub>50</sub> value for the NH<sub>2</sub>-PS NP and CdSO<sub>4</sub> were not detected, suggesting that the assay can tolerate the level of variability shown in Figure 7 for controls 7 and 8.



**Fig. 8: Median values (A) and relative median average deviation (MAD) (B) of control 4 – within pipette volume**

### Non-cellular controls

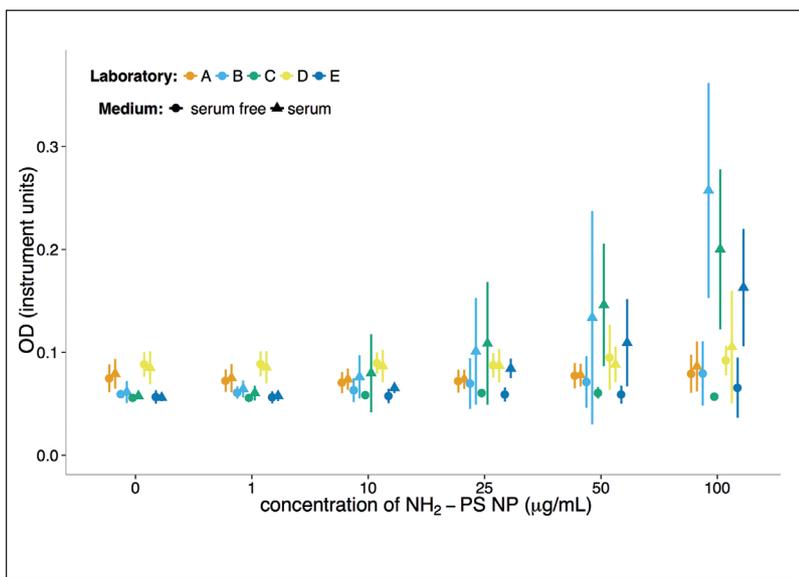
The 96-well plate design has three control columns that do not contain cells. Control 4 (within pipette volume control), which is used to assess MTS background absorbance, showed tightly clustered median OD values of less than 6% from all laboratories (Fig. 8A). Laboratory C demonstrated a within-laboratory variability consistently less than 3%, suggesting that there is a procedure difference in the other laboratories that introduces up to 50% more variability in this control between rounds. Laboratory D exhibited a systematic upward bias in the median OD value. This suggests that the MTS reagent or pipetting volume setting in laboratory D introduced a systematic bias. The relative median average distribution of the wells in the control 4 column for all rounds and all laboratories is less than 5% (Fig. 8B). Laboratory C and D showed variations of less than 2.5%, indicating that larger variability between the pipetting tip volumes in the multichannel pipette occurred in the other laboratories.

Comparison of the values from control 4 to those from control 1 and 5 assesses the impacts of the positive chemical control or NP,

respectively, on the assay readout (i.e., absorption) in the absence of cells. The absorbance values from control 1 were statistically identical to those from control 4 for all rounds (data not shown), indicating that the positive control  $\text{CdSO}_4$  did not interfere with the MTS assay readout system. However, comparing control 4 to control 5 results indicated that the NPs can interfere with the assay readout in the presence of serum but not in serum-free conditions (Fig. 9). In the presence of serum, three of the laboratories (B, C and E) show an absorbance change from 0.05 to up to 0.3 OD units (data average from all rounds) as the NP dosing concentration increases. Laboratories A and D showed no significant changes in absorbance value over these same concentrations. This control indicates that procedural differences between the laboratories when using NP in serum conditions can result in a NP interference effect during the measurement of the OD value.

### Positive chemical control

The assay plate design incorporates positive control measurements (control 2) that serve to confirm that the complete as-



**Fig. 9: Nanoparticle interference control results (control 5)**

Absorption values for all rounds in each laboratory were averaged and plotted for the Nanoparticle Interference Control column. Dose-dependent increases in absorption at 490 nm were observed in 3 laboratories (A, C and E) when the NH<sub>2</sub>-PS NP were dispersed in serum containing culture media. This effect was not observed for any of the laboratories when the NH<sub>2</sub>-PS NP were dispersed in serum-free cell culture media. Light scattering experiments indicated that the NP agglomerated in serum containing media (Tab. S4<sup>2</sup>), suggesting this increase in absorption is related to absorbance from NP remaining in the well. This effect was not observed in all laboratories, suggesting that differences in laboratory media removal protocols may play a role in minimizing this effect. Error bars represent two times the standard deviation values.

**Tab. 3: System specifications for the MTS assay as defined by the interlaboratory comparison for A549 cells**

Meeting these control specifications is critical for achieving measurement assurance in the MTS nanocytotoxicity assay.

Control	Serum-free			Serum		
	target value	range	variability	target value	range	variability
Control 1 <sup>a</sup> (background chemical control)	0.06 OD	0.05-0.09 OD	< 6%	see serum-free values		
Control 2 <sup>b</sup> (CdSO <sub>4</sub> control (EC <sub>50</sub> value))	49.9 µmol/l	47.5-51.5 µmol/l		77.2 µmol/l	54.3-99.4 µmol/l	
Control 3 (within pipette variability)	1.8 OD	1.5-2.0 OD	< 10%	2.0 OD	1.8-2.3	< 7%
Control 4 (no cell and no treatment control)	0.06 OD	0.05-0.09 OD	< 6%	see serum-free values		
Control 5 <sup>c</sup> (background NP B11-G11)	see Fig. 9					
Control 6	not tested					
Control 7+8 (between pipette variability)	1.5 OD	1.3-1.8 OD	< 12%	2.2 OD	1.8-2.8	< 7%

<sup>a</sup> If no additional background from the chemical reaction control is observed.

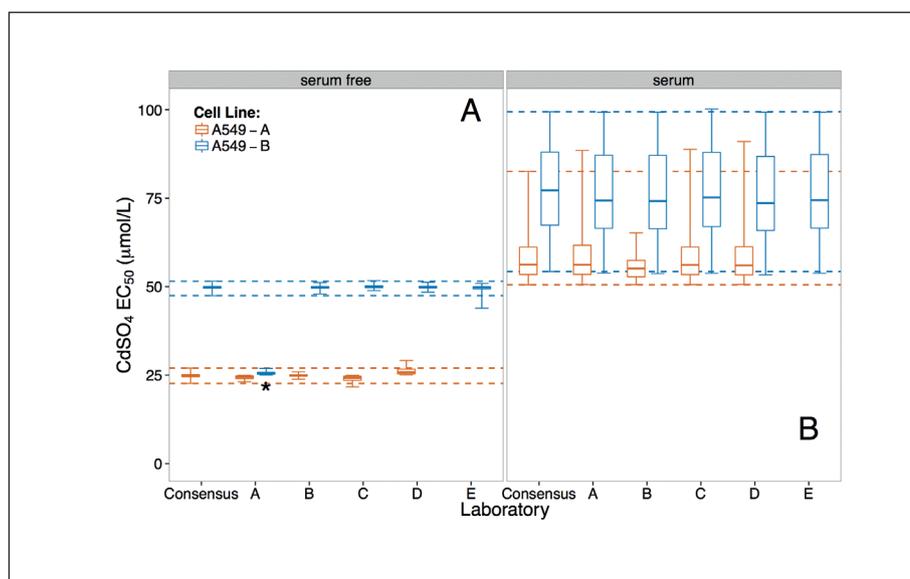
<sup>b</sup> Values of the A549-B cell line are given. They were fresh out of storage from ATCC and were passaged once and then sent to the participants.

<sup>c</sup> No values given, because some of the laboratories observed a background signal while others did not.

say system is operating as expected. There was a high level of agreement between the laboratories for the CdSO<sub>4</sub> control under serum-free conditions (Fig. 10) but the results significantly ( $p < 0.001$ ) differed for the two cell lines (Tab. S9<sup>2</sup> shows exact EC<sub>50</sub> values). For each cell line, the within-laboratory variability is comparable to the total variability, suggesting that overall the protocol techniques used in the different laboratories can result in a well harmonized EC<sub>50</sub> value for a chemical response

(Tab. 2). Interestingly, the NP dose-response curves showed no significant sensitivity to the two A549 cell lines used in this study (see Section 3.3, Fig. 4).

We observed a high level of agreement between each of the laboratories for the CdSO<sub>4</sub> EC<sub>50</sub> values in the presence of serum for both cell lines, but also a large uncertainty associated with the EC<sub>50</sub> estimation (Fig. 10). Similar to the results from the serum-free condition, there appeared to be a difference in the



**Fig. 10: CdSO<sub>4</sub> EC<sub>50</sub> values for A549-A and A549-B cells in serum-free (A) and serum (B, D) conditions**

Whisker error bars represent 95% confidence interval from all rounds from each laboratory. Consensus values were generated by Bayesian modeling of all rounds of data from all of the laboratories that were not considered outliers. Only one set of data (serum-free condition, A549-B cells, Lab A) was considered an outlier. This data set was marked with an asterisk.

response between the A549-A and A549-B cells, but it was not statistically significant due to the considerable overlap of the 95% confidence intervals. Evaluation of the sigmoidal curve fitting process for EC<sub>50</sub> determination suggests that the large uncertainty for this condition was not entirely due to assay performance, but also due to the EC<sub>50</sub> value falling between the last two dosing intervals (50 μmol/L and 100 μmol/L) of the assay. By having few dosing concentrations around the EC<sub>50</sub> value, the error due to curve fitting the EC<sub>50</sub> value is increased.

#### Performance specifications

The results from the system control experiments (Tab. 1) for all laboratories and all rounds excluding outliers were consolidated and summarized to form a set of system performance specifications (Tab. 3).

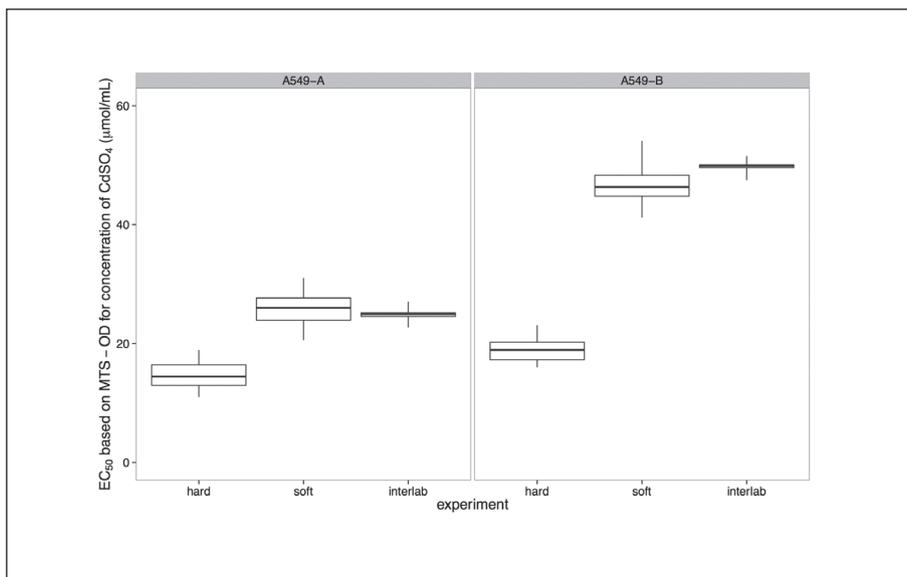
## 4 Discussion

The major focus of the study was the use of an assay plate design and an interlaboratory comparison study to evaluate the robustness of the steps of an MTS assay protocol. Although the biological effects of the NH<sub>2</sub>-PS NP on the A549 cell line are interesting, further investigation into the mechanism of these effects was not considered in this study. We also did not consider how the use of different cell lines and nanoparticles would influence the values that were measured in this study. Although the protocols described here may need to be modified for a particular cell line and nanoparticle, the general concepts introduced in this study should be applicable to most nanocytotoxicity tests.

The characterization of the NPs in the absence or presence of serum revealed substantial agglomeration and a drop in the zeta potential in the presence of serum. While many studies

have shown decreased agglomeration of NPs in media with proteins (Schulze et al., 2008; Sager et al., 2007; Tantra et al., 2010; Kwon et al., 2014), increased agglomeration from serum proteins has been observed in other studies (Murdock et al., 2008; Caracciolo et al., 2014). In the serum-containing media, the lower zeta potential reduced electrostatic stabilization while stabilization by steric repulsion depends on the protein corona which in turn depends on the protein loading on the surface, composition of the proteins in the serum, and the surface chemistry of the particles, which in this case was not sufficient to stabilize the NH<sub>2</sub>-PS NP. This may be at least partly attributable to the high ionic strength in the serum-containing medium ( $\approx 1.5$  mol/L) causing compression of the protein coating to 1 nm at which point van der Waals forces would become dominant.

An advantage of expressing *in vitro* exposures based on surface area dose metrics is that they can be compared to *in vivo* exposure results, but differences in the method of nanomaterial exposures should be considered. For example, *in vivo* studies are often carried out with lower doses per lung surface (excluding overload experiments). For various carbonaceous materials, it has been demonstrated that at similar doses, carbon black and graphite have no effect (0.1 or 0.03 μg/cm<sup>2</sup>, respectively; an estimate of the rat lung surface area 0.3 m<sup>2</sup> was used in this calculation (Brown et al., 2005)), but graphene and multiwall carbon nanotubes do have an effect (0.03 or 0.01 μg/cm<sup>2</sup>, respectively (Ma-Hock et al., 2013)). Moreover, *in vivo* investigations have shown that both the dosing rate and the delivered dose are responsible for toxicological effects. In a recent study, it could be demonstrated that at the same applied dose of approximately 160 μm TiO<sub>2</sub> per rat lung, the time of application was important for the adverse effect. At high dose rates (> 105 μg/minute via instillation), the severe inflammatory effects were observed whereas at more typical dose rates of < 1 μg/min via inhalation, the nano-titanium dioxide had



**Fig. 11: CdSO<sub>4</sub> EC<sub>50</sub> values for A549-A and A549-B cells in serum-free media using “hard” or “soft” rinsing procedures from Laboratory A and comparing to the interlaboratory consensus values (“interlab”)**

Whisker error bars represent 95% confidence intervals. Laboratory A performed 3 rounds for both the “hard” and “soft” rinsing procedures. When the “soft” rinsing procedure was performed, the results from Laboratory A aligned with those from the interlaboratory comparison consensus values.

no effect in the lungs of the rats (Baisch et al., 2014). Further studies may be required to clarify which physiologically relevant parameters of an *in vivo* system are captured in actual *in vitro* experiments.

Each laboratory performed three to eight rounds of the experiment which allowed assessment of both the within-laboratory variability and the between-laboratory variability of the MTS assay result. Comparison of the within-laboratory variability to the between-laboratory variability revealed apparent harmonization of the positive chemical control CdSO<sub>4</sub> EC<sub>50</sub> results while only moderate harmonization was observed for the NH<sub>2</sub>-PS NP treatment. Evaluation of the system controls and comparing the system controls under serum-containing and serum-free conditions provides insights into which steps of the protocol may need further refinement to improve interlaboratory comparability of this assay especially when used with NPs.

When evaluating the within pipette cell seeding density of non-treated cells (control 3; Fig. 5A), the values from laboratory A clearly indicate seeding of a lower cell concentration under serum-free conditions than in the other laboratories and these lower cell concentrations resulted in lower EC<sub>50</sub> values (Fig. 6). These results agree with findings which were obtained in a previous study demonstrating that the Min-U-Sil particles, a colloidal crystalline silica particle, caused a greater decrease in viability with the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay performed at lower cell densities (Geys et al., 2010). However, the cell concentrations in this previous study were intentionally varied while those in this interlaboratory comparison study occurred unintentionally as a result of unexpected differences in rinsing procedures. Discussion among participating laboratories revealed that laboratory A used a more rigorous cell rinsing aspiration technique than the other laboratories, which appears to have resulted in cell detachment. Confirmatory experiments (Fig. 11) comparing

the rigorous cell rinsing technique (vacuum aspiration) with a gentler rinsing technique (pipetting) demonstrated this to be the cause of the lower optical density in Figure 6 and the likely cause of the outlier data for laboratory A shown in Figures 3A, 4A and 5A. These data indicate that the rinsing procedure used to remove media from the cells should be more clearly specified in the protocol. The inclusion of the triple rinsing step in only the serum-free protocol was intended to minimize the number of assay steps and it was not anticipated prior to the interlaboratory study that the rinsing procedure might impact the assay results.

Evaluation of the NP interference control (control 5; Fig. 9) reveals that the rinsing procedures for removing media containing NP also need to be more clearly specified in the protocol. The increase in the optical density for control 5 at the two highest dosing concentrations indicates that laboratory B, C and E did not fully remove the NP before adding the MTS reagents; this could have increased variability in the measured NP EC<sub>50</sub> value. Discussion between the laboratories did not fully identify a cause for this effect, but a likely reason could be related to the laboratories incompletely removing the media from the cell culture wells prior to the addition of the MTS reagent. This issue was also observed in previous ENM interlaboratory comparison using the MTS assay. Xia et al. added a centrifugation step to their protocol to remove NP interference effects (Xia et al., 2013). Testing the impact of a centrifugation step or modified rinsing steps to decrease ENM interference will be important for optimizing the MTS protocol described here for better harmonization between laboratories.

Results from the within pipette volume control (control 4) show harmonization between laboratories with respect to their ability to pipette volumes within a pipetting step (Fig. 8). Figure 8B shows there is less than 5% variability between pipette volumes within a pipette step over all rounds performed

in all laboratories during this experiment. This suggests that pipetting a volume of liquid is not a major contributor to the variability in the assay results. When this data is compared to pipetting with cells (control 3), the variability between rounds for a single laboratory increases to as high as 20%, and there is more variability among the laboratories. This reveals that the combination of cell counting, cell seeding, and cell rinsing is a significant cause of the variability in the assay results. This suggests that the methods that are used for these steps may need to be more clearly defined in the protocol. Additional experiments may be required to establish the best cell handling techniques.

Variability of  $EC_{50}$  values for the  $CdSO_4$  positive control (control 2) and the  $NH_2$ -PS NP material within each of the laboratories increased in the order of serum-free  $CdSO_4$  < serum-free NP < serum  $CdSO_4$   $\approx$  serum NP (Fig. 4, 5). This suggests that the use of NP introduces additional variability to the assay for the serum-free condition, possibly due to the more complicated behaviors of the dispersed NP compared to well-dissolved chemicals (Cohen et al., 2015; Teeguarden et al., 2007).

Overall, serum does appear to have a protective effect for both cell lines exposed to both toxicants as the consensus  $EC_{50}$  are significantly increased in the presence of serum (Fig. 4, 5). The source of this effect is unclear as the serum could affect both the cell physiology and the NP agglomeration state and surface charge (Tab. S4<sup>2</sup>) and may interact with Cd. Serum is known to induce a protein corona around NP, which can insulate the cell from NP surface effects (Cedervall et al., 2007; Lundqvist et al., 2008). These factors may also explain the increased variability observed for NP under the serum condition. The NP dispersion step of the protocol is not precisely defined and the timing between the dispersion step and the dosing step is not specified, which can influence the extent of agglomeration during the exposure or the structure/composition of the protein corona. If variable precipitation of the NP agglomerates occurred during the separate rounds, the cells may have been effectively exposed to different dosing conditions (Petersen et al., 2014). Improved dispersion procedures or the use of more automation in preparing the dosing plate could reduce the variability observed within the laboratory when using  $NH_2$ -PS NP and serum. This factor may also contribute to the variability of the  $EC_{50}$  value for the  $NH_2$ -PS NP material between laboratories. The experimental design used here does not provide specific insight on this source of variability as several reagents including the sources of serum (Tab. S1<sup>2</sup>) differ between the laboratories. The different sources of serum could biologically influence the cellular response to both the  $CdSO_4$  and  $NH_2$ -PS NP treatments. Further experiments are needed to evaluate how these factors influence the test response. If serum is shown to induce significant variability in the test results, a recommendation would be to share the serum between the laboratories to improve the comparability of test results. However, legal regulations involving bovine serum products may prevent sharing of serum between laboratories in different countries. Alternatives such as chemically defined cell culture media should be investigated for future studies.

An interesting source of variability that was clearly observed in the  $CdSO_4$  positive control (control 2) only under serum-free conditions was the source of the A549 cell line (Fig. 10). Although the stocks for both cell lines were at early passage times (less than passage 25), the  $EC_{50}$  value for  $CdSO_4$  was 2-fold larger for the A549-B cell line when compared to the A549-A cell line despite the cell lines having statistically identical growth rates (Tab. S3<sup>2</sup>). This difference may be attributable to the genetic modification detected when comparing the STR typing profiles between the A549 cell lines (Fig. S1<sup>2</sup>). The  $CdSO_4$  positive control was added to the plate layout as a system control that could verify the nominal operation of the MTS assay. In this study, it revealed how cell lines with the same name but from a different source can influence a toxicity result. These results suggest that it is critical that laboratories share a genetically identical cell line to ensure the comparability of their results. Furthermore, genetic identification should be established routinely to ensure the cell line is not changing with time (Chatterjee, 2007; Almeida et al., 2011).

The dosing intervals can affect the 95% confidence interval predicted for  $EC_{50}$  values due to limitations in modeling the cell response between the dose interval values (Robinson et al., 2009). In this study, the calculated 95% confidence intervals in the  $EC_{50}$  value for both  $NH_2$ -PS NP and  $CdSO_4$  in the presence of serum for all laboratories were several-fold larger than those calculated for the treatments in serum-free conditions (Tab. S5<sup>2</sup>, S9<sup>2</sup>, Fig. 4, 10). Although the experimental details are responsible for some fraction of the observed variability, it is important to note that both of the  $EC_{50}$  values for  $NH_2$ -PS NP and  $CdSO_4$  in the presence of serum occurred in the last dosing interval resulting in additional apparent variability during the fitting process. Improved dosing intervals for NP toxicity experiments will reduce variability due to curve fitting uncertainty. Pilot studies could be used to estimate the  $EC_{50}$  value and dosing intervals could be designed around these values (Robinson et al., 2009). Additional studies are needed to determine what fraction of the variability in the  $EC_{50}$  value is derived from curve fitting with selected dosing intervals.

An important use of the system control data obtained from this interlaboratory comparison is to set the ranges for system specifications to ensure comparability in the assay process and confidence in the assay test result (Plant et al., 2014). Table 3 shows several performance specifications for the MTS assay protocol. Correlation analysis of these control values and the  $EC_{50}$  value of the  $NH_2$ -PS NP test results suggest that these ranges do not directly influence the measured  $EC_{50}$  values (Fig. 6, results for other controls not shown). The test result would be considered “valid” or at least comparable to data described here if each of these specifications is met. Each of the specifications provides unique information about the protocol steps, and failure to meet a specification can allow troubleshooting of the assay procedure to identify possible improvements in executing the protocol. Charting the system control measurements over time enables the observation of trends indicative of changes in assay performance. Developing and using assay



protocols that provide performance specifications and using control charts will aid in establishing quality management of a cell-based assay and improve confidence in short- and long-term comparability of assay results within a laboratory and between laboratories.

In conclusion, the combination of system control measurements and interlaboratory comparison data yielded insights that would not have been available by either approach by itself. While an interlaboratory comparison, without the control measurements and common assay design, would provide measures of within- and between-laboratory variability, it would not reveal the specific causes of variability associated with different steps of the assay. Similarly, a single laboratory assessing an assay with system control measurements would provide insight about the relative sources of variability for the assay within that laboratory. However, this approach would not provide information about the unknown sources of variability among different laboratories and how different interpretations of steps of the protocol or differing precision in a single step among laboratories affect the test results. Our study revealed that cell line ID, rinsing procedures, cell handling, NP dispersion, and choice of dosing intervals can significantly influence the assay results. The approach described here can be used in future studies to test a broader array of assays with a larger number of cell lines. It may be necessary to modify the protocol before use with other cytotoxicity readouts (e.g., MTT and XTT assay) to tailor the control measurements for assay-specific sources of variability which may differ from the MTS assay. This approach is designed to produce robust assays which will enable better decision-making during risk analysis of engineered nanomaterials.

## References

- Almeida, J., Hill, C. and Cole, K. (2011). Authentication of African green monkey cell lines using human short tandem repeat markers. *BMC Biotech* 11, 102. <https://doi.org/10.1186/1472-6750-11-102>
- Auffan, M., Rose, J., Bottero, J. Y. et al. (2009). Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective. *Nat Nanotechnol* 4, 634-641. <https://doi.org/10.1038/nnano.2009.242>
- Baisch, B. L., Corson, N. M., Wade-Mercer, P. et al. (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. <https://doi.org/10.1186/1743-8977-11-5>
- Brown, J. S., Wilson, W. E. and Grant, L. D. (2005). Dosimetric comparisons of particle deposition and retention in rats and humans. *Inhal Toxicol* 17, 355-385. <https://doi.org/10.1080/08958370590929475>
- Caracciolo, G., Caputo, D., Pozzi, D. et al. (2014). Size and charge of nanoparticles following incubation with human plasma of healthy and pancreatic cancer patients. *Colloids Surf B* 123, 673-678. <https://doi.org/10.1016/j.colsurfb.2014.10.008>
- Cedervall, T., Lynch, I., Lindman, S. et al. (2007). Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. *Proc Natl Acad Sci U S A* 104, 2050-2055. <https://doi.org/10.1073/pnas.0608582104>
- Chatterjee, R. (2007). Cell biology. Cases of mistaken identity. *Science* 315, 928-931. <https://doi.org/10.1126/science.315.5814.928>
- Clippinger, A. J., Ahluwalia, A., Allen, D. et al. (2016). Expert consensus on an in vitro approach to assess pulmonary fibrogenic potential of aerosolized nanomaterials. *Arch Toxicol* 90, 1769-1783. <https://doi.org/10.1007/s00204-016-1717-8>
- Cohen, J. M., DeLoid, G. M. and Demokritou, P. (2015). A critical review of in vitro dosimetry for engineered nanomaterials. *Nanomedicine* 10, 3015-3032. <https://doi.org/10.2217/nmm.15.129>
- Congdon, P. (2001). *Bayesian Statistical Modeling*. John Wiley & Sons.
- Cornfield, J. and Mantel, N. (1950). Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve. *J Am Stat Assoc* 45, 181-210. <https://doi.org/10.2307/2280677>
- De Volder, M. F., Tawfick, S. H., Baughman, R. H. et al. (2013). Carbon nanotubes: Present and future commercial applications. *Science* 339, 535-539. <https://doi.org/10.1126/science.1222453>
- Gelman, A., Carlin, J., Stern, H. et al. (2008). *Bayesian Data Analysis*. 2<sup>nd</sup> edition. Chapman & Hall.
- Geys, J., Nemery, B. and Hoet, P. H. M. (2010). Assay conditions can influence the outcome of cytotoxicity tests of nanomaterials: Better assay characterization is needed to compare studies. *Toxicol In Vitro* 24, 620-629. <https://doi.org/10.1016/j.tiv.2009.10.007>
- Graetzel, M., Janssen, R. A. J., Mitzi, D. B. et al. (2012). Materials interface engineering for solution-processed photovoltaics. *Nature* 488, 304-312. <https://doi.org/10.1038/nature11476>
- Hanna, S. K., Cooksey, G. A., Dong, S. et al. (2016). Feasibility of using a standardized *Caenorhabditis elegans* toxicity test to assess nanomaterial toxicity. *Environ Sci: Nano* 5, 1080-1089. <https://doi.org/10.1039/C6EN00105J>
- Hinderliter, P. M., Minard, K. R., Orr, G. et al. (2010). ISDD: A computational model of particle sedimentation, diffusion and target cell dosimetry for in vitro toxicity studies. *Part Fibre Toxicol* 7, <https://doi.org/10.1186/1743-8977-7-36>
- Hong, S. P., Leroueil, P. R., Janus, E. K. et al. (2006). Interaction of polycationic polymers with supported lipid bilayers and cells: Nanoscale hole formation and enhanced membrane permeability. *Bioconjugate Chem* 17, 728-734. <https://doi.org/10.1021/bc060077y>
- Horev-Azaria, L., Baldi, G., Beno, D. et al. (2013). Predictive toxicology of cobalt ferrite nanoparticles: Comparative in-vitro study of different cellular models using methods of knowledge discovery from data. *Part Fibre Toxicol* 10, 32. <https://doi.org/10.1186/1743-8977-10-32>
- Kaiser, J. P., Roesslein, M., Buerki-Thurnherr, T. et al. (2011). Carbon Nanotubes – curse or blessing. *Curr Med Chem* 18, 2115-2128. <https://doi.org/10.2174/092986711795656171>

- Krug, H. F. and Wick, P. (2011). Nanotoxicology: An interdisciplinary challenge. *Angew Chem Int Edit* 50, 1260-1278. <https://doi.org/10.1002/anie.201001037>
- Krug, H. F. (2014). Nanosafety research – are we on the right track? *Angew Chem Int Edit* 53, 12304-12319. <https://doi.org/10.1002/anie.201403367>
- Kwon, D., Park, J., Park, J. et al. (2014). Effects of surface-modifying ligands on the colloidal stability of ZnO nanoparticle dispersions in in vitro cytotoxicity test media. *Int J Nanomedicine* 9, 57-65. <https://doi.org/10.2147/ijn.s57924>
- Landsiedel, R., Kapp, M. D., Schulz, M. et al. (2009). Genotoxicity investigations on nanomaterials: Methods, preparation and characterization of test material, potential artifacts and limitations – many questions, some answers. *Mut Res* 681, 241-258. <https://doi.org/10.1016/j.mrrev.2008.10.002>
- Landsiedel, R., Sauer, U. G., Ma-Hock, L. et al. (2014). Pulmonary toxicity of nanomaterials: A critical comparison of published in vitro assays and in vivo inhalation or instillation studies. *Nanomedicine* 9, 2557-2585. <https://doi.org/10.2217/nnm.14.149>
- Lundqvist, M., Stigler, J., Elia, G. et al. (2008). Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. *Proc Natl Acad Sci U S A* 105, 14265-14270. <https://doi.org/10.1073/pnas.0805135105>
- Lunn, D., Spiegelhalter, D., Thomas, A. et al. (2009). The BUGS project: Evolution, critique and future directions. *Stat Med* 28, 3049-3067. <https://doi.org/10.1002/sim.3680>
- Ma-Hock, L., Strauss, V., Treumann, S. et al. (2013). Comparative inhalation toxicity of multi-wall carbon nanotubes, graphene, graphite nanoplatelets and low surface carbon black. *Part Fibre Toxicol* 10, <https://doi.org/10.1186/1743-8977-10-23>
- Miller, G. W. (2014). Improving reproducibility in toxicology. *Toxicol Sci* 139, 1-3. <https://doi.org/10.1093/toxsci/kfu050>
- Monteiro-Riviere, N. A., Inman, A. O. and Zhang, L. W. (2009). Limitations and relative utility of screening assays to assess engineered nanoparticle toxicity in a human cell line. *Toxicol Appl Pharm* 234, 222-235. <https://doi.org/10.1016/j.taap.2008.09.030>
- Murdock, R. C., Braydich-Stolle, L., Schrand, A. M. et al. (2008). Characterization of nanomaterial dispersion in solution prior to in vitro exposure using dynamic light scattering technique. *Toxicol Sci* 101, 239-253. <https://doi.org/10.1093/toxsci/kfm240>
- Nel, A., Xia, T., Madler, L. et al. (2006). Toxic potential of materials at the nanolevel. *Science* 311, 622-627. <https://doi.org/10.1126/science.1114397>
- Nel, A. E., Nasser, E., Godwin, H. et al. (2013). A multi-stakeholder perspective on the use of alternative test strategies for nanomaterial safety assessment. *ACS Nano* 7, 6422-6433. <https://doi.org/10.1021/nn4037927>
- Nemmar, A., Hoylaerts, M. F., Hoet, P. H. M. et al. (2003). Size effect of intratracheally instilled particles on pulmonary inflammation and vascular thrombosis. *Toxicol Appl Pharm* 186, 38-45. [https://doi.org/10.1016/s0041-008x\(02\)00024-8](https://doi.org/10.1016/s0041-008x(02)00024-8)
- NRC (2007). *Toxicity Testing in the 21<sup>st</sup> Century: A Vision and a Strategy*. The National Academies Press. [http://www.nap.edu/openbook.php?record\\_id=11970](http://www.nap.edu/openbook.php?record_id=11970)
- Petersen, E. J., Henry, T. B., Zhao, J. et al. (2014). Identification and avoidance of potential artifacts and misinterpretations in nanomaterial ecotoxicity measurements. *Environ Sci Technol* 48, 4226-4246. <https://doi.org/10.1021/es4052999>
- Plant, A. L., Locascio, L. E., May, W. E. et al. (2014). Improved reproducibility by assuring confidence in measurements in biomedical research. *Nat Methods* 11, 895-898. <https://doi.org/10.1038/nmeth.3076>
- Poland, C., Miller, M., Duffin, R. et al. (2014). The elephant in the room: Reproducibility in toxicology. *Part Fibre Toxicol* 11, 42. <https://doi.org/10.1186/s12989-014-0042-8>
- Robinson, S., Chapman, K., Hudson, S. et al. (2009). Guidance on dose level selection for regulatory general toxicology studies for pharmaceuticals. *Laboratory Animal Science Association and National Centre for the Replacement, Refinement and Reduction of Animals in Research*.
- Rösslein, M., Elliott, J. T., Salit, M. L. et al. (2014). The use of cause-and-effect analysis to design a high quality nano-cytotoxicology assay. *Chem Res Toxicol* 27, 1877-1884. <https://doi.org/10.1021/tx500327y>
- Sager, T. M., Porter, D. W., Robinson, V. A. et al. (2007). Improved method to disperse nanoparticles for in vitro and in vivo investigation of toxicity. *Nanotoxicology* 1, 118-129. <https://doi.org/10.1080/17435390701381596>
- Sauer, U. G., Vogel, S., Hess, A. et al. (2013). In vivo-in vitro comparison of acute respiratory tract toxicity using human 3D airway epithelial models and human A549 and murine 3T3 monolayer cell systems. *Toxicol In Vitro* 27, 174-190. <https://doi.org/10.1016/j.tiv.2012.10.007>
- Schurs, F. and Lison, D. (2012). Focusing the research efforts. *Nat Nano* 7, 546-548. <https://doi.org/10.1038/nnano.2012.148>
- Schulze, C., Kroll, A., Lehr, C. M. et al. (2008). Not ready to use – overcoming pitfalls when dispersing nanoparticles in physiological media. *Nanotoxicology* 2, 51-U17. <https://doi.org/10.1080/17435390802018378>
- Shen, M. W., Wang, S. H., Shi, X. Y. et al. (2009). Polyethyleneimine-mediated functionalization of multiwalled carbon nanotubes: Synthesis, characterization, and in vitro toxicity assay. *J Phys Chem C* 113, 3150-3156. <https://doi.org/10.1021/jp809323e>
- Tantra, R., Tompkins, J. and Quincey, P. (2010). Characterisation of the de-agglomeration effects of bovine serum albumin on nanoparticles in aqueous suspension. *Colloids Surf B* 75, 275-281. <https://doi.org/10.1016/j.colsurfb.2009.08.049>
- Teeguarden, J. G., Hinderliter, P. M., Orr, G. et al. (2007). Particokinetics in vitro: Dosimetry considerations for in vitro nanoparticle toxicity assessments. *Toxicol Sci* 95, 300-312. <https://doi.org/10.1093/toxsci/kfl165>
- Thompson, M. and Ellison, S. L. R. (2011). Dark uncertainty. *Accredit Qual Assur* 16, 483-487. <https://doi.org/10.1007/s00769-011-0803-0>
- Toman, B. and Possolo, A. (2009). Laboratory effects models for interlaboratory comparisons. *Accredit Qual Assur* 14,



- 553-563. <https://doi.org/10.1007/s00769-009-0547-2>
- Toman, B., Rösslein, M., Elliott, J. T. et al. (2016). Estimation and uncertainty analysis of dose response in an inter-laboratory experiment. *Metrologia* 53, S40. <https://doi.org/10.1088/0026-1394/53/1/S40>
- Wagner, V., Dullaart, A., Bock, A.-K. et al. (2006). The emerging nanomedicine landscape. *Nat Biotech* 24, 1211-1217. [https://doi.org/http://www.nature.com/nbt/journal/v24/n10/suppinfo/nbt1006-1211\\_S1.html](https://doi.org/http://www.nature.com/nbt/journal/v24/n10/suppinfo/nbt1006-1211_S1.html)
- Warheit, D. B. and Donner, E. M. (2015). How meaningful are risk determinations in the absence of a complete dataset? Making the case for publishing standardized test guideline and 'no effect' studies for evaluating the safety of nanoparticles versus spurious 'high effect' results from single investigative studies. *Sci Technol Adv Mat* 16, 034603. <https://doi.org/10.1088/1468-6996/16/3/034603>
- Xia, T., Hamilton, R. F., Bonner, J. C. et al. (2013). Interlaboratory evaluation of in vitro cytotoxicity and inflammatory responses to engineered nanomaterials: The NIEHS nano GO consortium. *Environ Health Perspect* 121, 683-690. <https://doi.org/https://doi.org/10.1289/ehp.1306561>
- Yu, M., Selvaraj, S. K., Liang-Chu, M. M. Y. et al. (2015). A resource for cell line authentication, annotation and quality control. *Nature* 520, 307-311. <https://doi.org/10.1038/nature14397>

#### Conflict of interest

The authors declare that they have no conflict of interest.

#### Acknowledgements

This project was supported by the Nano Material Technology Development Program (NRF-2011-0020504 and NRF-2014M3A7B6020163) of MSIP/NRF, the NanoScreen Materials Challenge co-funded by the Competence Centre for Materials Science and Technology (CCMX), and the Bilateral Korean-Swiss S&T program in 2012. Research performed in part at the NIST Center for Nanoscale Science and Technology.

#### Author contributions

JTE, MR, MLS, HFK, NWS, and PW designed the experimental protocols. JTE, MR, NWS, AKO, RM, FS, JL, SJK, FR, CH, and WS conducted the cytotoxicity experiments. ELR conducted the STR analysis. BT conducted the Markov Chain Monte Carlo fitting simulation and the comparison of the within-laboratory variability to total variability. JTE, MR, and EJP contributed to the data analysis and writing the manuscript. All coauthors reviewed the manuscript.

#### Correspondence to

John T. Elliott  
Biosystems and Biomaterials Division  
Material Measurement Laboratory  
National Institute of Standards and Technology  
100 Bureau Dr.  
Gaithersburg, MD 20899, United States  
Phone: +1 301 975 8551  
e-mail: john.elliott@nist.gov