IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Science, Engineering, and Applied Research

Vol 1, No 1, 2026

Incommensurable Targets in Machine Learning for Nanomaterials: How Dose-Metric Heterogeneity Imposes an Irreducible Error Floor on Pooled Training Data

IGWE Emmanuel Sunday[1], UDEH Samuel Chibueze[2]

[1] Veristzon Global Research Lab; [2] Department of Pharmaceutical Services, National Obstetric Fistula Centre

ABSTRACT

Nanomaterial datasets are assembled from the published literature rather than generated in a single laboratory, and the studies pooled in this way do not agree on what dose means. Exposure is reported as mass concentration, as particle number, as total or specific surface area, or as particles delivered per cell, and conversion between these metrics requires size, density, porosity and aggregation data that source reports frequently omit. This theoretical and analytical paper treats the resulting heterogeneity not as a reporting inconvenience to be settled by better conventions, but as a measurement-model problem that places a quantifiable ceiling on what any machine-learning model trained on such data can achieve. The study is non-empirical. It derives closed-form relations among the dose metrics for spherical particles of uniform density, showing that mass, surface area and particle number scale with the zeroth, inverse first and inverse third powers of radius, so that conversion factors are functions of particle size rather than constants. From these relations the paper obtains a closed-form condition under which two materials compared at equal mass dose are ranked oppositely by particle number and by surface area, and verifies that condition exhaustively against brute-force computation over two hundred thousand randomly drawn pairs. It then derives the irreducible error floor that dose-metric heterogeneity imposes on a pooled training set, $MSE_{floor} = \beta^2 p^2 \sigma^2 (\ln r) + \sigma^2_{\varepsilon}$, which grows with the square of the metric mismatch, with the dispersion of log particle radius across the corpus, and with the steepness of the dose response. The floor is shown to be independent of model class and of hyperparameter optimisation, and not to recede as further data drawn from the same literature are added. The paper converts the derivation into practice through a six-step procedure for bounding the achievable performance of an existing pooled dataset from its own curation records, together with declaration, partitioning and applicability-domain requirements for training-set construction expressed in the context-of-use idiom of current regulatory guidance.

Keywords: nano-QSAR; machine learning; dose metrics; nanodosimetry; data pooling

Incommensurable Targets in Machine Learning for Nanomaterials: How Dose-Metric Heterogeneity Imposes an Irreducible Error Floor on Pooled Training Data

1. INTRODUCTION

Nanomaterial datasets are assembled rather than generated. Because no single laboratory produces enough dose–response data to train a model of useful breadth, the standard practice is to pool results reported across many published studies, and the curated resources that now underpin nanoinformatics are built on exactly that principle (Afantitis et al., 2020; Chen et al., 2022; Canchola et al., 2025). Pooling solves a sample-size problem. It also quietly creates a measurement problem, because the studies being pooled did not agree on what dose means.

A nanomaterial exposure can be reported as a mass concentration, as a number of particles per unit volume, as a total or specific surface area, or as particles delivered per cell, and all four are in active use. The choice is not arbitrary: there are principled arguments for surface area where the mechanism is interfacial (Mottier et al., 2016; Simkó et al., 2014), for particle number where the mechanism is per-object, and for delivered rather than administered dose wherever sedimentation and diffusion make the two differ substantially (Hinderliter et al., 2010; Cohen et al., 2015). The difficulty is that converting between the metrics requires knowledge of particle size distribution, density, porosity and aggregation state, and source reports frequently omit the quantities the conversion needs.

The consequence is usually described as a standardisation problem, to be addressed by better reporting conventions. That description is correct but incomplete, and the incompleteness matters. When a model is trained on a pooled dose variable whose metric differs between contributing studies, the target variable is not a single well-defined function of exposure. Two rows of the training matrix can carry identical nominal dose values and describe physically different exposures, and the model has no means of telling them apart. What follows is not merely noisy estimation. It is a floor on achievable predictive error that no model class, no hyperparameter search and no additional data drawn from the same literature can cross.

This paper derives that floor. The argument proceeds in three movements. First it sets out the closed-form relations among the dose metrics and shows that, because they scale with different powers of particle radius, the mapping between any two of them is a function of particle size rather than a constant. Second it shows that metric heterogeneity within a pooled training set therefore injects variance into the target whose magnitude is determined by the dispersion of particle sizes across the constituent studies and by the exponent mismatch between the reported and the operative metric, and that this variance is irreducible from within the pooled set. Third it shows that the same mechanism can invert the apparent ranking of two materials, so that pooling can mislead about ordering and not only about magnitude. The paper closes with declaration and harmonisation requirements, and with a procedure by which the error floor of an existing dataset can be bounded.

Two questions are addressed. How large is the error floor imposed by dose-metric heterogeneity, and on what does it depend? And what can be done about it, given that the source literature cannot be re-reported? The treatment is analytical and non-empirical. Numerical values appear only to illustrate derived relationships and are stated as assumptions rather than presented as measurements.

2. LITERATURE REVIEW

2.1 CONCEPTUAL REVIEW

Dose in nanotoxicology and nanopharmaceutics is not one concept but several, and the distinctions are well drawn in the dosimetry literature. The administered dose is what the experimenter adds to the well. The delivered dose is the fraction that reaches the cell layer, governed by sedimentation and diffusion and therefore by particle size, density and medium viscosity. The cellular or effective dose is what the cells actually take up. Ahmad Khanbeigi et al. (2012) evaluated administered, delivered and cellular dose in the same system and found the interpretation of particle–cell interaction to depend on which was used. Cohen et al. (2015) reviewed the field and concluded that reliance on administered metrics rather than effective dose is a major obstacle to accurate in vitro models; the integrated methodology they had set out earlier

© IIARJournal. All rights reserved.

makes the point constructively, coupling standardised suspension preparation to the determination of delivered-to-cell metrics (Cohen et al., 2014).

Cutting across that distinction is a second one: the metric in which any of these doses is expressed. Mass concentration dominates because it is what a balance measures. Particle number and surface area are argued to be more mechanistically appropriate, surface area especially where the interaction is interfacial; Simkó et al. (2014) proposed total deposited surface area as the primary metric and set out weighting factors for additional physicochemical properties, and Mottier et al. (2016) argued the surface-area case directly for carbon nanoparticles. Reviews of inhaled aerosol dosimetry continue to press for number and surface area in preference to mass (Ediagbonya et al., 2026). Computational particokinetics has advanced considerably, from the original sedimentation–diffusion treatment of Hinderliter et al. (2010) through models incorporating dissolution (Thomas et al., 2018; Smith et al., 2018) and semi-analytical solutions characterising the governing dimensionless groups (Mahnama et al., 2014), with experimental assessment at single-cell resolution (Kowoll et al., 2018). None of this work disputes that the metric matters; the point of contention is which metric, and for which mechanism.

2.2 THEORETICAL REVIEW

The statistical situation has a recognisable shape but does not fit the standard templates exactly, and the mismatch is worth being precise about.

Classical measurement-error theory treats noise in a predictor, with the familiar consequence that regression slopes attenuate toward zero. Here the corruption enters differently. The nominal dose recorded in the pooled dataset is, in general, accurately reported; what varies is its relationship to the quantity that drives the response. The error is therefore in the correspondence between a correctly recorded predictor and the operative exposure, which propagates into the target as a study-specific offset. Formally this behaves as label noise, and the relevant result is that irreducible target noise sets a lower bound on expected loss which is independent of the hypothesis class. That bound is what the dose-metric problem produces.

The second template is dataset shift, familiar from clinical machine learning, where models degrade because deployment conditions differ from training conditions (Finlayson et al., 2021). Metric heterogeneity differs in being internal to the training set rather than a difference between training and deployment. The pooled set is not drawn from one distribution that later moves; it is a mixture of distributions from the outset, and the mixture components are indexed by a variable (the reporting metric and the particle size on which conversion depends) that the model does not see.

The third relevant body of theory concerns the applicability domain, long established in quantitative structure–activity modelling and increasingly operationalised in nanoinformatics (Li et al., 2022; Li et al., 2026). An applicability domain states where a model's predictions may be trusted, conventionally in terms of the descriptor space spanned by the training data. The analysis below implies that for pooled nanomaterial data the domain must also be stated over the dose-metric and particle-size space, because it is dispersion in those that sets the floor.

2.3 EMPIRICAL REVIEW

Three strands of published evidence support the premises of the derivation.

That the metrics genuinely disagree, and not only in magnitude, is directly documented. Ortenzio et al. (2019) compared indices of silver nanoparticle dose and found that the cell layer contained a greater mass of silver after exposure to 75 nm particles than to 20 nm particles, while expressing the same exposure as particle number or as total surface area reversed the ordering. This is an experimental instance of the metric-dependent ranking inversion derived in Section 4.1, and it establishes that the phenomenon is not a formal curiosity. Tian et al. (2017) likewise reported that airway dosimetry expressed in number, mass and surface area varied significantly across the same exposure, and concluded that a recommendation framed in a single metric is insufficient. The divergence can be large even between two sizes of one material: modelling silver nanoparticle deposition at a fixed airborne mass concentration, Smith and Skinner (2021) found that 20 nm particles deposited nearly elevenfold more total surface area than 110 nm particles, alongside a 2.2-fold difference in deposited mass. The decomposition is instructive. Equation (2) accounts for a factor of 5.5 from the radii alone, through specific surface area; the residual factor of roughly 2.2 comes from the differing deposition efficiency of the two sizes.

© IIARJournal. All rights reserved.

Metric conversion and delivered-dose correction are thus distinct effects that compound, and a pooled dataset which neglects both carries their product.

That the heterogeneity is widespread in the literature available for pooling is equally documented. Secondo et al. (2017) screened over ten thousand studies, retained one hundred and ten meeting eligibility criteria, and reported that little consistency exists in the methodology and dose metrics used, explicitly noting that this prevents interlaboratory comparison and the identification of dose thresholds. Riediker et al. (2019) place dose-metric choice among the core concepts governing particle health effects. Reviews of nanomaterial machine learning identify data heterogeneity and small, high-variance datasets as the field's characteristic obstacle (Barnard et al., 2019).

That pooling is nevertheless the operative practice, and that its products are used, is visible in the datasets themselves. Canchola et al. (2025) compiled data from 83 studies spanning two and a half decades and reported pronounced heterogeneity in materials, surface modifications, sizes from 1 to 1400 nm and zeta potentials from −70 to +70 mV. A size range spanning three orders of magnitude is precisely the condition under which, as shown below, metric mismatch is most costly. Consensus and round-robin approaches to model evaluation have emerged as a means of assessing reliability across modelling groups (Varsou et al., 2024; Zouraris et al., 2025), but they assess agreement between models trained on common data rather than the adequacy of the data's dose representation.

What the literature has not done is quantify what metric heterogeneity costs. The standardisation argument is made qualitatively; the bound it implies has not been derived. That is the gap addressed here.

3. MATERIALS AND METHODS

3.1 RESEARCH DESIGN AND ANALYTICAL STRATEGY

The study is non-empirical, analytical and derivational. Relations among dose metrics are obtained in closed form from the geometry and mass of spherical particles. The statistical consequence is then derived from an explicitly stated dose–response model, and the resulting expressions are verified numerically by simulation from the same stated assumptions. No laboratory work was performed, no new dataset was assembled, and no published dataset was re-analysed. The numerical verification is a check on the algebra, not a source of empirical findings, and every parameter entering it is an assumption declared in the text.

3.2 STATED ASSUMPTIONS

The derivation rests on assumptions that are stated so that each can be examined and, where a reader judges it inapplicable, relaxed. Particles are treated as spheres of uniform density, so that mass, surface area and number are related by geometry; for porous, fractal or strongly anisotropic materials the exponents differ and the argument's structure, though not its arithmetic, is unaffected. The dose–response relationship is taken to be approximately linear in the logarithm of the operative dose over the range spanned by the pooled data, which is a local rather than a global claim. The reporting metric of a study is taken to be independent of the dose level it reports, an assumption discussed in Section 4.2 and relaxed there. And the operative metric, the one the mechanism responds to, is taken to be fixed but unknown, which is the situation of an analyst pooling data without a mechanistic commitment.

3.3 SOURCE IDENTIFICATION AND VERIFICATION

Sources were identified by structured searches of the PubMed and MEDLINE corpus covering nanoparticle dose metrics and dosimetry, in vitro particokinetics, delivered and effective dose, surface-area dose metrics, and data pooling and curation for nanomaterial modelling. Candidates were screened for relevance to a specific claim. Every retained journal source was verified against its indexed publisher record for authorship, title, journal, year, volume, issue, pagination and digital object identifier, and screened for retraction, expression of concern, correction and erratum; no retained source carried such a notice. Where electronic release preceded the assigned issue, the year of the issue of record is used. Verification was performed against the PubMed publisher record rather than by direct query to the Crossref application programming interface, which was unavailable in the working environment; the fields verified are the same. Regulatory and standards instruments are cited in their current official version as issued. Preprints and conference papers are not cited.

3.4 ARTIFICIAL INTELLIGENCE USAGE DISCLOSURE

© IIARJournal. All rights reserved.

A large language model was used to assist with bibliographic searching, reference formatting, numerical checking of the derivations, and the drafting and editing of prose. The derivations, the assumptions, the interpretation of every source and the final text are the authors' own. All bibliographic details were verified against the publisher record. No generative tool was used to produce data or empirical results.

3.5 VALIDITY AND SCOPE

Being derivational, the work is assessed on the correctness of its algebra, the transparency of its assumptions and the checkability of its consequences rather than on measurement validity. The algebraic results were confirmed by independent numerical simulation, including an exhaustive check of the ranking-inversion condition against brute-force evaluation over two hundred thousand randomly drawn material pairs, which produced no disagreement. The scope is pooled dose–response modelling. The argument does not apply to datasets generated under a single declared metric, which is precisely its point.

4. RESULTS AND DISCUSSION

4.1 RESULTS

Relations among the metrics. For spheres of radius r and density ρ at mass concentration C_m, the particle-number concentration and the total surface-area concentration follow from geometry:

$$C_n = 3C_m / (4\pi\rho r^3) \quad (1)$$

$$C_s = 3C_m / (\rho r) \quad (2)$$

so that the specific surface area per unit mass is 3/(ρr). The three metrics therefore scale with different powers of radius at fixed mass dose: mass as r^0, surface area as r^{-1} and number as r^{-3}. Table 1 collects the exponent p by which any reported metric must be multiplied, as r^p, to recover any operative metric. The entries are not conversion constants; they are functions of particle size, and that is the root of everything that follows.

Table 1. Exponent p in the relation (operative dose) ∝ (reported dose) × r^p for spherical particles of uniform density.

Reported metric	Operative: mass	Operative: surface area
Mass concentration	0	−1
Surface-area concentration	+1	0
Number concentration	+3	+2

Metric-dependent ranking inversion. Because the exponents differ, two materials compared at equal mass dose can be ordered differently by different metrics. Let material A have radius r_A and density ρ_A, material B have r_B and ρ_B, and write s = r_B/r_A and k = ρ_B/ρ_A. From equations (1) and (2), A carries the greater particle number when $k s^3 > 1$ and the greater surface area when k s > 1. These conditions disagree precisely when k lies strictly between 1/s and $1/s^3$:

$$1/s < k < 1/s^3 \text{ for } s < 1, \text{ and } 1/s^3 < k < 1/s \text{ for } s > 1 \quad (3)$$

The band is non-empty for every s ≠ 1, so for any pair of distinct sizes there exists a range of density ratios producing inversion. A worked instance: with r_A = 10 nm, ρ_A = 1000 kg m^{-3}, r_B = 2 nm and ρ_B = 20 000 kg m^{-3}, s = 0.2 and k = 20 fall inside the predicted band (5, 125); at equal mass dose A delivers the greater particle number while B delivers the greater surface area. Brute-force evaluation over two hundred thousand randomly drawn pairs agreed with condition (3) in every case. The experimental reversal reported by Ortenzio et al. (2019) for 20 nm and 75 nm silver particles is an instance of the same effect. The implication for pooled modelling is that a model trained on one metric can rank two materials in an order that a model trained on another metric reverses, from the same underlying exposures.

Metric heterogeneity as injected target variance. Consider a pooled dataset of studies indexed by i. Let x denote the logarithm of the dose as reported and u the logarithm of the operative dose. From Table 1, u = x + p ln r_i + c for a constant c, where p is the exponent mismatch between the metric study i reported and the metric the mechanism responds to. Take the response to be locally linear in operative log-dose:

$$y = \alpha + \beta u + \varepsilon, \varepsilon \sim (0, \sigma^2_\varepsilon) \quad (4)$$

Substituting gives

$$y = \alpha + \beta c + \beta x + \beta p \ln r_i + \varepsilon \quad (5)$$

A model that sees x but not r_i therefore faces an additional term βp ln r_i. If ln r is independent of x across the pooled set with variance σ²(ln r), the variance of y decomposes as

$$Var(y) = \beta^2\sigma^2_x + \beta^2 p^2 \sigma^2(\ln r) + \sigma^2_\varepsilon \quad (6)$$

where σ²_x is the variance of the reported log-dose across the pooled set, σ²(ln r) the variance of log

© IIARJournal. All rights reserved.

particle radius across the contributing studies, and σ^2_ε the intrinsic noise variance of the response.

The error floor. Of the three components in equation (6), a model of x alone can explain only the first. The expected squared error of the best such model is therefore bounded below by

$$MSE_{floor} = \beta^2 p^2 \sigma^2(\ln r) + \sigma^2_\varepsilon \qquad (7)$$

and the attainable coefficient of determination is bounded above by

$$R^2_{max} = \beta^2\sigma^2_x / (\beta^2\sigma^2_x + \beta^2 p^2 \sigma^2(\ln r) + \sigma^2_\varepsilon) \qquad (8)$$

Three features of equation (7) deserve emphasis. The floor grows with the square of the exponent mismatch, so pooling mass-reported data against a number-driven mechanism ($p = -3$) incurs nine times the penalty of the same pooling against a surface-driven mechanism ($p = -1$). It grows with the dispersion of log radius across the pooled studies, which is a property of the corpus rather than of any experiment in it. And it is multiplied by the squared dose–response slope, so steep responses are punished harder than shallow ones. Table 2 gives illustrative magnitudes under the stated assumptions; the values are consequences of equation (7) for the parameters shown and are not measurements.

Table 2. Error floor and attainable R^2 implied by equation (7) under stated assumptions ($\beta = 1.30$, $\sigma^2_\varepsilon = 0.01$, $\sigma^2_x = 1$).

Metric mismatch	p	SD of ln r
Mass reported, surface operative	−1	0.25
Mass reported, surface operative	−1	0.60
Number reported, surface operative	+2	0.25
Number reported, surface operative	+2	0.60
Mass reported, number operative	−3	0.25
Mass reported, number operative	−3	0.60

The final row is worth reading carefully. A size dispersion of 0.60 in natural logarithm, modest beside the three orders of magnitude spanned by the pooled corpus of Canchola et al. (2025), leaves a ceiling near 0.24 on explained variance when metric and mechanism are mismatched at $p = -3$. A model reporting such a figure would ordinarily be judged weak and attributed to insufficient data or an inadequate learner. Equation (7) says it may instead be at its ceiling.

Irreducibility. The floor is irreducible from within the pooled set, and the reason is structural rather than practical. Additional observations reduce the variance of the estimated coefficients; they do not supply the variable $\beta p \ln r_i$ that is missing. Simulation from the stated assumptions bears this out: with $p = -3$ and $\sigma_{\ln r} = 0.60$, the mean squared error of the best model of x alone moves from 4.42 at 250 observations to 5.43 at 4000 and 5.45 at 64 000, approaching the value 5.486 predicted by equation (7) and remaining there. By contrast, admitting ln r as a predictor returns the error to 0.0098 against an intrinsic noise variance of 0.01. The remedy is declaration, not accumulation.

Bounding the floor of an existing dataset. Equation (7) is usable in practice because its ingredients are estimable from a curated dataset without re-running any experiment. Table 3 sets out the procedure, which yields a bound rather than a point estimate because the operative metric is generally unknown; taking the maximum over plausible p gives a conservative ceiling on what the dataset can support.

Table 3. Procedure for bounding the irreducible error floor of a pooled dataset.

Step	Quantity
1	Reported metric for each contribut study
2	$\sigma^2(\ln r)$, dispersion of log particle
3	Candidate exponent mismatches p
4	β, local dose–response slope
5	Floor and attainable R^2
6	Comparison

Where step 6 shows a model performing at its ceiling, further algorithmic effort is misdirected and the dataset requires partition or re-curation. Where a model performs well below the ceiling, the shortfall is attributable to the model and improvement remains available.

4.2 DISCUSSION

The derivation explains a pattern the field has noticed without explaining. Reviews of nanomaterial machine learning routinely report that

© IIARJournal. All rights reserved.

performance is modest and that data heterogeneity is to blame (Barnard et al., 2019; Chen et al., 2022). Equation (7) converts that diagnosis from an attribution into a quantity, and in doing so changes what should be done about it. If the limitation is data volume, the response is to pool more aggressively. If it is metric incommensurability, aggressive pooling makes matters worse, because $\sigma^2(\ln r)$ grows as the corpus widens. The two prescriptions are opposed, and the distinction between them is testable by the procedure in Table 3.

The ranking result bears on a specific and consequential practice. Models trained on pooled data are increasingly used to prioritise materials for synthesis or for safety follow-up, and prioritisation is an ordering operation. Condition (3) shows that ordering is metric-contingent in an identifiable region of material space, and the experimental reversal of Ortenzio et al. (2019) shows the region is occupied. A hazard ranking or a design shortlist is therefore not metric-neutral even when the model's accuracy is high, and the metric in which a ranking was produced belongs in the statement of the ranking.

Two of the stated assumptions deserve scrutiny. The assumption that reporting metric is independent of dose level is a simplification; in practice laboratories working at high mass concentrations may favour mass reporting, which would correlate the injected term with x and bias the fitted slope in addition to inflating the floor. The derivation is conservative in that case, since correlation adds bias to the variance penalty rather than removing it. The assumption of local log-linearity matters more. Where the true response is strongly non-monotone over the pooled range, the decomposition in equation (6) holds only approximately and β must be read as a local slope. This is a limitation on the numerical magnitudes in Table 2, not on the existence of a floor, which follows from the injected term being invisible to the model.

The sphericity assumption limits the exponents, not the argument. Porous, fractal and high-aspect-ratio materials relate mass, surface and number through different powers, and for aggregates the relevant radius is itself ill-defined, a problem compounded because aggregation state depends on the medium and is frequently unreported. The effect of relaxing sphericity is to widen the set of plausible p and therefore to raise the conservative ceiling from step 5 of Table 3.

The regulatory reading is direct. Model credibility, as both major agencies now frame it, is established not in the abstract but against a declared context of use, and the assurance demanded scales with how much the model influences a decision and how costly a wrong one would be (European Medicines Agency [EMA], 2024; United States Food and Drug Administration [FDA], 2025). A pooled dataset whose dose metrics are undeclared cannot support a context-of-use statement, because the exposure quantity its predictions are conditioned on has not been specified. The procedure in Table 3 produces exactly the kind of bound that a credibility assessment can use: a stated ceiling on achievable performance, derived from declared properties of the training corpus. Work on quality by digital design in nanomedicine development is moving in a compatible direction (Ijjeh et al., 2025).

Future work follows from the propositions. The most direct test is a curation study: take a public pooled dataset, partition it by reported metric, fit within partitions and across the pool, and compare the observed performance gap against the bound from equation (7). A second is a learning-curve study, where the prediction is that metric-homogeneous subsets continue to improve with sample size while the pooled set plateaus early. A third concerns conversion: where source studies report enough to permit conversion to a common metric, the gain from doing so should be measurable and should scale with $p^2\sigma^2(\ln r)$ as derived.

5. CONCLUSION AND RECOMMENDATION

Mass concentration, particle number and surface area are related by powers of particle radius, not by constants. A training set that pools studies reporting in different metrics therefore carries a target variable that is not a single function of exposure, and the mismatch injects variance into the labels that is invisible to the model. The resulting floor on mean squared error is $\beta^2 p^2 \sigma^2(\ln r) + \sigma^2_\varepsilon$: it grows with the square of the metric mismatch, with the dispersion of log radius across the corpus, and with the steepness of the dose response. It is independent of model class and of hyperparameter optimisation, and it does not recede as more data from the same literature is added. The same mechanism makes material rankings metric-contingent, and the condition under which two materials are ordered oppositely by number and by surface area is given in closed form.

© IIARJournal. All rights reserved.

The recommendations are mostly matters of declaration, and none requires new measurement. Investigators assembling pooled datasets should record the reported dose metric as a field of the dataset and flag rather than impute studies that fail to declare one. They should report the dispersion of particle size across the corpus, since it is this, and not the corpus's size, that sets the floor. Where conversion to a common metric is possible from reported quantities, it should be performed and the conversion assumptions recorded; where it is not, the dataset should be partitioned by metric and models fitted within partitions. Applicability domains should be stated over dose-metric and particle-size space as well as descriptor space. Modelling papers should report the bound from equation (7) alongside their performance figures, so that a reader can see whether a model is weak or merely at its ceiling. And those assembling or relying on hazard rankings should state the metric in which the ranking was computed, because condition (3) shows that the ordering is not metric-independent.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

The paper contributes closed-form relations among nanomaterial dose metrics together with the exponent table that governs conversion between any pair, and a closed-form condition, verified exhaustively against brute-force computation, under which two materials compared at equal mass dose are ranked oppositely by particle number and by surface area. It derives the irreducible error floor that dose-metric heterogeneity imposes on pooled training data, $MSE_floor = \beta^2 p^2 \sigma^2 (\ln r) + \sigma^2_\varepsilon$, establishes that the floor is independent of model class and insensitive to sample size drawn from the same literature, and confirms both results by simulation from the stated assumptions. And it converts the derivation into practice by supplying a six-step procedure for bounding the achievable performance of an existing pooled dataset from its own curation records, together with declaration and harmonisation requirements for training-set construction that are expressed in the context-of-use idiom of current regulatory guidance.

7. REFERENCES

Afantitis, A., Melagraki, G., Isigonis, P., Tsoumanis, A., Varsou, D. D., Valsami-Jones, E., Papadiamantis, A., Ellis, L. A., Sarimveis, H., Doganis, P., Karatzas, P., Tsiros, P., Liampa, I., Lobaskin, V., Greco, D., Serra, A., Kinaret, P. A. S., Saarimäki, L. A., Grafström, R., … Lynch, I. (2020). NanoSolveIT project: Driving nanoinformatics research to develop innovative and integrated tools for nanosafety assessment. *Computational and Structural Biotechnology Journal, 18*, 583–602. https://doi.org/10.1016/j.csbj.2020.02.023

Ahmad Khanbeigi, R., Kumar, A., Sadouki, F., Lorenz, C., Forbes, B., Dailey, L. A., & Collins, H. (2012). The delivered dose: Applying particokinetics to in vitro investigations of nanoparticle internalization by macrophages. *Journal of Controlled Release, 162*(2), 259–266. https://doi.org/10.1016/j.jconrel.2012.07.019

Barnard, A. S., Motevalli, B., Parker, A. J., Fischer, J. M., Feigl, C. A., & Opletal, G. (2019). Nanoinformatics, and the big challenges for the science of small things. *Nanoscale, 11*(41), 19190–19201. https://doi.org/10.1039/c9nr05912a

Canchola, A., Li, K., Chen, K., Borboa-Pimentel, A., Chou, C., Dela Rama, R., Chen, C. Y., Chen, X., Strobel, M., Riviere, J. E., Monteiro-Riviere, N. A., Wang, M., Zhang, F., Lin, Z., & Chou, W. C. (2025). Meta-analysis and machine learning prediction of protein corona composition across nanoparticle systems in biological media. *ACS Nano, 19*(43), 37633–37650. https://doi.org/10.1021/acsnano.5c08608

Chen, C., Yaari, Z., Apfelbaum, E., Grodzinski, P., Shamay, Y., & Heller, D. A. (2022). Merging data curation and machine learning to improve nanomedicines. *Advanced Drug Delivery Reviews, 183*, 114172. https://doi.org/10.1016/j.addr.2022.114172

Cohen, J. M., Teeguarden, J. G., & Demokritou, P. (2014). An integrated approach for the in vitro dosimetry of engineered nanomaterials. *Particle and Fibre Toxicology, 11*, 20. https://doi.org/10.1186/1743-8977-11-20

Cohen, J. M., Derk, R., Wang, L., Godleski, J., Kobzik, L., Brain, J., & Demokritou, P. (2015). A critical review of in vitro dosimetry for engineered nanomaterials. *Nanomedicine, 10*(19), 3015–3032. https://doi.org/10.2217/nnm.15.129

© IIARJournal. All rights reserved.

Ediagbonya, T. F., Akanbi, S. A., & John, J. O. (2026). Dosimetry of inhaled aerosols: From exposure to target tissue dose in risk assessment and drug delivery. *Journal of Applied Toxicology*. Advance online publication. https://doi.org/10.1002/jat.70327

European Medicines Agency. (2024). *Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle* (EMA/CHMP/CVMP/83833/2023). European Medicines Agency.

Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I. S., & Saria, S. (2021). The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine, 385*(3), 283–286. https://doi.org/10.1056/NEJMc2104626

Hinderliter, P. M., Minard, K. R., Orr, G., Chrisler, W. B., Thrall, B. D., Pounds, J. G., & Teeguarden, J. G. (2010). ISDD: A computational model of particle sedimentation, diffusion and target cell dosimetry for in vitro toxicity studies. *Particle and Fibre Toxicology, 7*(1), 36. https://doi.org/10.1186/1743-8977-7-36

Ijjeh, Y., Alsarayreh, N., Rifai, A., Abdelnabi, H., Al-Mahamid, S., Alqudah, D. A., Nsairat, H., Lafi, Z., Mousa, M. A., Al-Hiari, Y., & Alshaer, W. (2025). Quality by digital design for accelerated sustainable nanomedicine development. *European Journal of Pharmaceutical Sciences, 213*, 107239. https://doi.org/10.1016/j.ejps.2025.107239

Kowoll, T., Fritsch-Decker, S., Diabaté, S., Stöger, T., Weiss, C., & Gerthsen, D. (2018). Assessment of in vitro particle dosimetry models at the single cell and particle level by scanning electron microscopy. *Journal of Nanobiotechnology, 16*(1), 100. https://doi.org/10.1186/s12951-018-0426-2

Li, J., Wang, C., Yue, L., Chen, F., Cao, X., & Wang, Z. (2022). Nano-QSAR modeling for predicting the cytotoxicity of metallic and metal oxide nanoparticles: A review. *Ecotoxicology and Environmental Safety, 243*, 113955. https://doi.org/10.1016/j.ecoenv.2022.113955

Li, K., Canchola, A., Zhang, F., & Chou, W. C. (2026). Determinants of protein corona adsorption and abundance revealed by interpretable machine learning across nanoparticle systems. *Scientific Reports, 16*(1), 55258. https://doi.org/10.1038/s41598-026-55258-6

Mahnama, A., Tafazzoli-Shadpour, M., Geramizadeh, B., & Mehrabi, M. (2014). Semi-analytical solution for the in-vitro sedimentation, diffusion and dosimetry model: Surveying the impact of the Peclet number. *Colloids and Surfaces B: Biointerfaces, 122*, 324–331. https://doi.org/10.1016/j.colsurfb.2014.06.063

Mottier, A., Mouchet, F., Pinelli, É., Gauthier, L., & Flahaut, E. (2016). Surface area of carbon nanoparticles: A dose metric for a more realistic ecotoxicological assessment. *Nano Letters, 16*(6), 3514–3518. https://doi.org/10.1021/acs.nanolett.6b00348

Ortenzio, J., Degn, L., Goldstein-Plesser, A., McGee, J. K., Navratilova, J., Rogers, K., Zucker, R. M., & Boyes, W. K. (2019). Determination of silver nanoparticle dose. *NanoImpact, 14*, 100156. https://doi.org/10.1016/j.impact.2019.100156

Riediker, M., Zink, D., Kreyling, W., Oberdörster, G., Elder, A., Graham, U., Lynch, I., Duschl, A., Ichihara, G., Ichihara, S., Kobayashi, T., Hisanaga, N., Umezawa, M., Cheng, T. J., Handy, R., Gulumian, M., Tinkle, S., & Cassee, F. (2019). Particle toxicology and health - where are we? *Particle and Fibre Toxicology, 16*(1), 19. https://doi.org/10.1186/s12989-019-0302-8

Secondo, L. E., Liu, N. J., & Lewinski, N. A. (2017). Methodological considerations when conducting in vitro, air–liquid interface exposures to engineered nanoparticle aerosols. *Critical Reviews in Toxicology, 47*(3), 225–262. https://doi.org/10.1080/10408444.2016.1223015

Simkó, M., Nosske, D., & Kreyling, W. G. (2014). Metrics, dose, and dose concept: The need for a proper dose concept in the risk assessment of nanoparticles. *International Journal of Environmental Research and Public Health, 11*(4), 4026–4048. https://doi.org/10.3390/ijerph110404026

Smith, J. N., Thomas, D. G., Jolley, H., Kodali, V. K., Littke, M. H., Munusamy, P., Baer, D. R., Gaffrey, M. J., Thrall, B. D., & Teeguarden, J. G. (2018). All that is silver is not toxic: Silver

© IIARJournal. All rights reserved.

ion and particle kinetics reveals the role of silver ion aging and dosimetry on the toxicity of silver nanoparticles. *Particle and Fibre Toxicology, 15*(1), 47. https://doi.org/10.1186/s12989-018-0283-z

Smith, J. N., & Skinner, A. W. (2021). Translating nanoparticle dosimetry from conventional in vitro systems to occupational inhalation exposures. *Journal of Aerosol Science, 155*, 105771. https://doi.org/10.1016/j.jaerosci.2021.105771

Thomas, D. G., Smith, J. N., Thrall, B. D., Baer, D. R., Jolley, H., Munusamy, P., Kodali, V., Demokritou, P., Cohen, J., & Teeguarden, J. G. (2018). ISD3: A particokinetic model for predicting the combined effects of particle sedimentation, diffusion and dissolution on cellular dosimetry for in vitro systems. *Particle and Fibre Toxicology, 15*(1), 6. https://doi.org/10.1186/s12989-018-0243-7

Tian, L., Shang, Y., Chen, R., Bai, R., Chen, C., Inthavong, K., & Tu, J. (2017). A combined experimental and numerical study on upper airway dosimetry of inhaled nanoparticles from an electrical discharge machine shop. *Particle and Fibre Toxicology, 14*(1), 24. https://doi.org/10.1186/s12989-017-0203-7

United States Food and Drug Administration. (2025). *Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: Draft guidance for industry and other interested parties* (Docket No. FDA-2024-D-4689). U.S. Department of Health and Human Services.

Varsou, D. D., Banerjee, A., Roy, J., Roy, K., Savvas, G., Sarimveis, H., Wyrzykowska, E., Balicki, M., Puzyn, T., Melagraki, G., Lynch, I., & Afantitis, A. (2024). The round-robin approach applied to nanoinformatics: Consensus prediction of nanomaterials zeta potential. *Beilstein Journal of Nanotechnology, 15*, 1536–1553. https://doi.org/10.3762/bjnano.15.121

Zouraris, D., Mavrogiorgis, A., Tsoumanis, A., Saarimäki, L. A., Del Giudice, G., Federico, A., Serra, A., Greco, D., Rouse, I., Subbotina, J., Lobaskin, V., Jagiello, K., Ciura, K., Judzinska, B., Mikolajczyk, A., Sosnowska, A., Puzyn, T., Gulumian, M., Wepener, V., … Afantitis, A. (2025). CompSafeNano project: Nanoinformatics approaches for safe-by-design nanomaterials. *Computational and Structural Biotechnology Journal, 29*, 13–28. https://doi.org/10.1016/j.csbj.2024.12.024

© IIARJournal. All rights reserved.