**Electronic Supplementary Information**

# Causation or only correlation? Application of causal inference graphs for evaluating causality in nano-QSAR models

Natalia Sizochenko,[a,b] Agnieszka Gajewicz,[a] Jerzy Leszczynski,[b] TomaszPuzyn[*a]

[a]*Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308, Gdansk, Poland*

[b] *Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Jackson State University, 1400 J. R. Lynch Street, P. O. Box 17910, Jackson, MS 39217, USA*

**corresponding author: t.puzyn@qsar.eu.org*

**K-means modeling**

K-means clustering is founded on the approximation of each point in a dataset by the centroid of that point's cluster. K-means clustering in unsupervised learning algorithm. Hence, may be appropriate measure to estimate similarity of nanoparticles. Resulted of k-means clustering in 1D space of Wigner-Seitz radius is presented in Table S1.

**Table S1.** K-means clustering

| **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** |
|---|---|---|---|
| CoO, CuO, NiO, ZnO | $SnO_2$, $TiO_2$, $ZrO_2$, $SiO_2$ | $Al_2O_3$, $Cr_2O_3$, $Fe_2O_3$, $V_2O_3$ | $Bi_2O_3$, $In_2O_3$, $La_2O_3$, $Sb_2O_3$, $Y_2O_3$ |

Cluster 1: $r_w < 0.17$; highest toxicity, MeOx.

Cluster 2: $0.17 < r_w < 0.18$; lowest toxicity, $MeO_2$.

Cluster 3: $0.18 < r_w < 0.20$; intermediate, $Me_2O_3$.

Cluster 4: $r_w < 0.20$; intermediate, $Me_2O_3$.