Estimation and Uncertainty analysis of Dose Response in an inter-laboratory experiment

Blaza Toman[1], Matthias Rösslein[2], John T. Elliott[1], Elijah J. Petersen[1]

[1] National Institute of Standards and Technology, US Department of Commerce, Gaithersburg, MD, USA

[2] Swiss Federal Laboratories for Material Science and Technology, St. Gallen, Switzerland

Abstract

An inter-laboratory experiment for the evaluation of toxic effects of NH_2-polystyrene nanoparticles on living human cancer cells was performed with five participating laboratories. Previously published results from nanocytoxicity assays are often contradictory, mostly due to challenges related to producing a reliable cytotoxicity assay protocol for use with nanomaterials. Specific challenges include reproducibility preparing nanoparticle dispersions, biological variability from testing living cell lines, and the potential for nano-related interference effects. In this experiment, such challenges were addressed by developing a detailed experimental protocol and using a specially designed 96-well plate layout which incorporated a range of control measurements to assess multiple factors such as nanomaterial interference, pipetting accuracy, cell seeding density, and instrument performance. Detailed data analysis of these control measurements showed that good control of the experiments was attained by all participants in most cases. The main measurement objective of the study was the estimation of a dose response relationship between concentration of the nanoparticles and metabolic activity of the living cells, under several experimental conditions. The dose curve estimation was achieved by imbedding a three parameter logistic curve in a three level Bayesian hierarchical model, accounting for uncertainty due to all known experimental conditions as well as between laboratory variability in a top-down manner. Computation was performed using Markov Chain Monte Carlo methods. The fit of the model was evaluated using Bayesian posterior predictive probabilities and found to be satisfactory.

1. Introduction

Engineered nanomaterials are currently being used in many commercial applications resulting in the likelihood of human and animal exposure to these materials. It is therefore important to obtain accurate measurements of their potential toxicological effects on living cells. Cell-based toxicity assays can be used to identify potentially hazardous nanomaterials but their use has led to conflicting results from similar nanoparticles in different laboratories [1-4]. This result is most likely due to differences in assay protocols and possibly lack of controls for monitoring assay performance [1,2].

This study illustrated that it is possible to achieve agreement among laboratories performing these measurements. The participants were five national laboratories, all using a 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) viability assay, during which the percentage of live to dead cells is assessed by measuring the number of metabolically active cells in samples exposed to the toxic substance and not exposed. The measurements were of absorbance at 490 nm where the signal is related to the number of metabolically active cells. Two protocols, designed to test nanotoxicity under serum or serum-free conditions, were developed from a cause-and-effect analysis of the MTS assay [see reference 5 for details]. The plate design (Fig 1) contained three replicates of the nanoparticle dose response experiment in Columns 8, 9, and 10, and several system controls to quantify critical sources of variability in the assay. Specifically, Columns 3, 4, and 5 contained three replicates of a chemical control dose experiment. The dose response to the chemical control is well known and thus provides an excellent check of system performance. Column 2 contained no cells and the measurements were of background for the chemical control experiment. Column 6 contained non-treated cells seeded with a single ejection of the pipette and was used to check pipetting variability and general cell counting

procedures. It was very important for initial cell density to be as close to the same as possible in all the wells as there is no way to directly measure the effect of the toxic substance using a measurement of a single well (There cannot be a before- treatment and an after-treatment measurement in a single well). For this reason ratios of absorbance of experimental wells to absorbance of wells in column 6 were used to assess toxicity. The fact that these particular human cancer cells had a 24 hour doubling rate made it especially important to start out with equal numbers in each well. Row B of Columns 3, 4, 5, 8, 9, and 10 was treated with solvent but not the chemical control or nanoparticles and was used to check pipetting variability from multiple pipetting ejections. Column 7 and the wells around the perimeter were used to check instrument background. The participants of the interlaboratory study all used these protocols, a common stock of the nanoparticles, and two specific clones of a human cancer cell line. Other sources of variability such as different cell treatment procedures, sources of serum, culture media, and cell culture plates were intentionally not controlled in order to produce the usual variability among laboratories.

Figure 1. Plate Design

The participants performed several rounds of the assay each on a separate plate under serum conditions and several under serum-free conditions. The number of rounds tested per laboratory ranged from 2 to 6.

The main objective of the study was the estimation of response of the living cells to various concentrations of nanoparticles. The relationship was represented by a dose response curve between concentration and relative absorbance at 490 nm. Of particular interest was the EC_{50} value of the dose response which is the concentration that provides an effect (in this case loss of cell viability) in 50 % of the cells. The importance of this number is that it may be used in risk assessment models or in comparisons among studies or among (nano)materials if the toxicity of the substances is ranked. This value could also potentially be used for in vivo to in vitro correlations.

Section 2 presents data analysis of the control variables, estimation of the dose response is discussed in Section 3 and conclusions and discussion are in Section 4.

 2. Data Analysis of control variables

The control measurements in columns 2, 7, 11 and the perimeter wells were quite similar on the average over the various participating laboratories and plates. This indicated that the participants had good control of the pipetting procedures on average. There were some differences in the variability however. For illustration, Table 1 contains means and standard deviations of the control variables for laboratory 2 which was the least variable and laboratory 5 which was the most variable.

	Lab 2 Mean (std) absorbance	Lab 5 Mean (std) absorbance
Col 2 (background Chem Ctrl)	0.0564 (0.0007)	0.057 (0.004)
Col 7	0.0567 (0.0009)	0.058 (0.003)
Col 11 (background NP)	0.20 (0.04)	0.23 (0.04)
wells around the perimeter	0.0558 (0.0005)	0.058 (0.004)

Table 1. Comparison of control measurements of absorbance (optical density or OD)

In addition to these control variables, the absorbance values of the wells containing cells with no treatment were of importance because all dosed cells were expected to have an equivalent or smaller absorbance value. Fig 2 shows means and 95 % confidence intervals of the average absorbance measurements in column 6. Replicates for this calculation are the multiple plates for both clones.
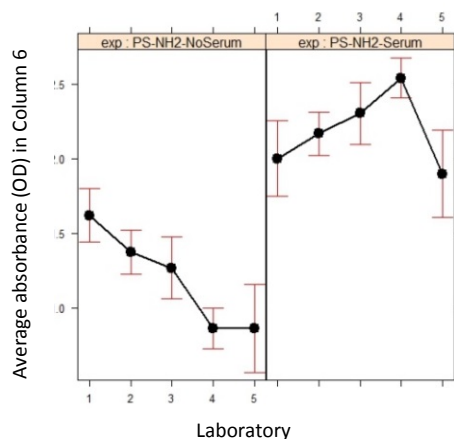

Figure 2. Average absorbance (with 95 % uncertainty intervals) in Column 6, a control containing cells with no treatment. This provides maximum absorbance.

Judging from the uncertainty intervals, it appears that the average maximum absorbance was relatively constant over plates and clones for most laboratories with the exception of laboratory 5. This suggests that initial cell density was in rather good control. Nevertheless, the experiment protocol could be further improved by determining the factors that caused the larger variability for Laboratory 5 and including additional information to address these factors. There did not appear to be differences in the standard deviations of the maximum absorbance between the two serum conditions but the mean maximum absorbance was larger with serum.

3. Analysis of Dose response

3.1 Individual Plate and Laboratory Dose Response Curve

There were 4 distinct experiments in this study defined by the two clones of the cell line (1 or 2) and serum condition (yes or no) which were analyzed separately. Each 96 well plate in one of these experiments produced six columns of data

$$y_{ijkl} = 1 - r_{ijkl}, \quad i = 1,...,5, \ j = 1,...,n_i, \ k = 1,...,3, \ l = 1,...,6$$

where r_{ijkl} is the relative absorbance at the lth concentration x_{ijkl}, of the kth column, of the jth plate, of the ith laboratory. Laboratories measured different number of plates each so this is denoted by n_i. There were three columns for chemical control experiment, and three for the nanoparticle experiment. In all cases a dose response

curve for a single column could be estimated using a three parameter logistic model, that is, the mean of an observation at concentration x_{ijkl} was

$$\mu_{ijkl} = \frac{\gamma_{ijk}}{1+e^{\frac{(\alpha_{ijk}-x_{ijkl})}{\beta_{ijk}}}}.$$

A small supplementary study was performed to assess how repeatable the absorbance measurements were and found that the coefficient of variation was between 0.5 and 1 %. Using the more conservative 1 %, this additional uncertainty was included in the statistical model as

$$Y_{ijkl} \sim N\left(m_{ijkl}, 0.01^2\right), \; m_{ijkl} \sim N\left(\mu_{ijkl}, \sigma_{ijk}^2\right).$$

The estimation of the logistic parameters was performed using a Bayesian model with proper prior distributions for the parameters, that is integrating to 1, but quite vague in the sense that the prior variances were large as

$$a_{ijk} \sim N\left(0, 10^5\right), \; \beta_{ijk} \sim Uniform\left(-1, 30\right), \; \gamma_{ijk} \sim Uniform\left(0,1\right), \text{ and } \sigma_{ijk} \sim Gamma\left(10^4, 10^4\right).$$

The fitting was done using Markov Chain Monte Carlo in OpenBUGS [6]. Sensitivity analysis with respect to the prior distributions showed that the specific parameter values were not important. Visual inspection of plots of the fitted curves, as for example the 18 curves for the chemical control experiment for laboratory 2 in Figure 3, showed very good agreement between curves based on the three columns of a single plate.
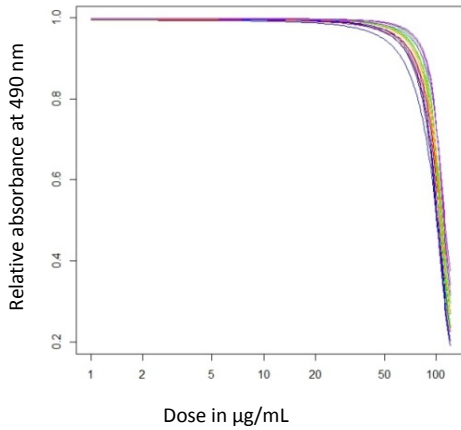

Figure 3. Dose response to chemical control for 6 plates of laboratory 2, that is, all 18 curves.

For this reason it was reasonable to treat the three columns on a single plate as three independent replicates and combine them by assuming that the mean of an observation at concentration x_{ijkl} was

$$\mu_{ijkl} = \frac{\gamma_{ij}}{1+e^{\frac{(\alpha_{ij}-x_{ijkl})}{\beta_{ij}}}}, \quad k=1,2,3.$$

This formed a single response curve per plate and captured the variability among the measurements of the three columns, generally attributed to the variability of separate ejections of the 8 channel pipette used to seed the plate. Reference [5] found pipetting to be the largest source of variability.

Examination of plots of the dose response curves of individual plates showed variability among plates that was somewhat larger than between the columns on each plate. This suggested that between-plate variability should be directly included in the model in terms of additional parameters. A Bayesian hierarchical model can capture this variability by specifying a relationship between the plate parameters of the logistic regression. This is analogous to including a random laboratory effect in the model for the calculation of a consensus value in an interlaboratory comparison with a single constant measurand as discussed for example in [7,8]. Here the individual plate parameters of each laboratory were given Gaussian probability distributions with common means as

$$\alpha_{ij} \sim N\left(a_i, \sigma_{\alpha i}^2\right), \; \log\left(\beta_{ij}\right) \sim N\left(b_i, \sigma_{\beta i}^2\right), \; \log\left(\gamma_{ij}\right) \sim N\left(g_i, \sigma_{\gamma i}^2\right), \; i=1,\ldots,5, \; j=1,\ldots,n_i.$$

The variances of the Gaussian distributions represented the between-plate variability being captured by this

model. Note that the mean for laboratory i, at concentration x was now represented by $\mu_i^C(x) = \dfrac{e^{g_i}}{1 + e^{\frac{(a_i - x)}{e^{b_i}}}}$.

Prior distributions for the additional laboratory parameters were again set to be proper, but quite vague. This model, estimated via MCMC, produced for each experimental condition estimated curves and uncertainty intervals for each plate, and consensus curves for each laboratory. Sensitivity to choice of prior distribution was examined and found to be low.

To illustrate the results, Figure 4 shows the individual measurements in blue, the 6 plate curves in green, and the consensus laboratory curve with 95 % uncertainty bounds in black for the chemical control experiment with clone 1, for laboratory 2.
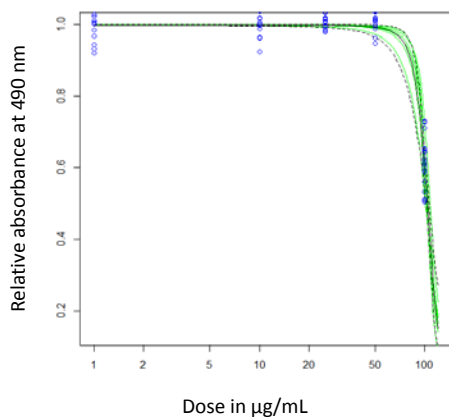

Figure 4. Individual measurements (blue circles), plate curves (green lines), and consensus laboratory curve with 95 % uncertainty bounds (black dotted lines) for the chemical control experiment for laboratory 2, with clone 1, serum-free.

3.2 Overall Consensus Dose Response Curve

In order to summarize all relevant information from the 5 laboratories and to enable examination of their agreement it is useful to estimate an overall consensus dose response curve. This was accomplished by adding another layer to the hierarchy of the statistical model, that is, specifying that

$$a_i \sim N\left(\mu_a, \sigma_a^2\right), \; b_i \sim N\left(\mu_b, \sigma_b^2\right), \; g_i \sim N\left(\mu_g, \sigma_g^2\right),$$

with prior distributions for the new parameters set to be proper but quite vague as $\mu_a \sim N(0,10^5)$, $\mu_b \sim Uniform(-1, 30)$, $\mu_g \sim Uniform(-1, 2)$, and $\sigma_a \sim Gamma(10^4, 10^4)$, $\sigma_b \sim Uniform(0,10)$, $\sigma_g \sim Uniform(0,10)$, $\sigma_i \sim Gamma(10^4, 10^4)$. The consensus mean at concentration x was given by

$$\mu^C(x) = \frac{e^{\mu_g}}{1 + e^{\frac{(\mu_a - x)}{e^{\mu_b}}}}.$$

This model produced curves and uncertainty bounds for plates, consensus curves for laboratories and an overall consensus curve for each of the four experimental conditions. Figure 5 shows the plot of the laboratory dose response curves for clone 1, the nanoparticle experiment. It also shows the consensus curve, and the 95 % uncertainty bounds for the consensus curve. Agreement among the 5 laboratories is remarkably good for a nanocytoxicity experiment, especially for the serum-free condition. Degree of agreement can also be assessed by the EC$_{50}$ values, that is, the a_i (and the μ_a). Table 2 shows these values and their standard uncertainty for the 4 experimental conditions of the nanoparticle experiment. Not all laboratories used clone 2 and so are missing these values.
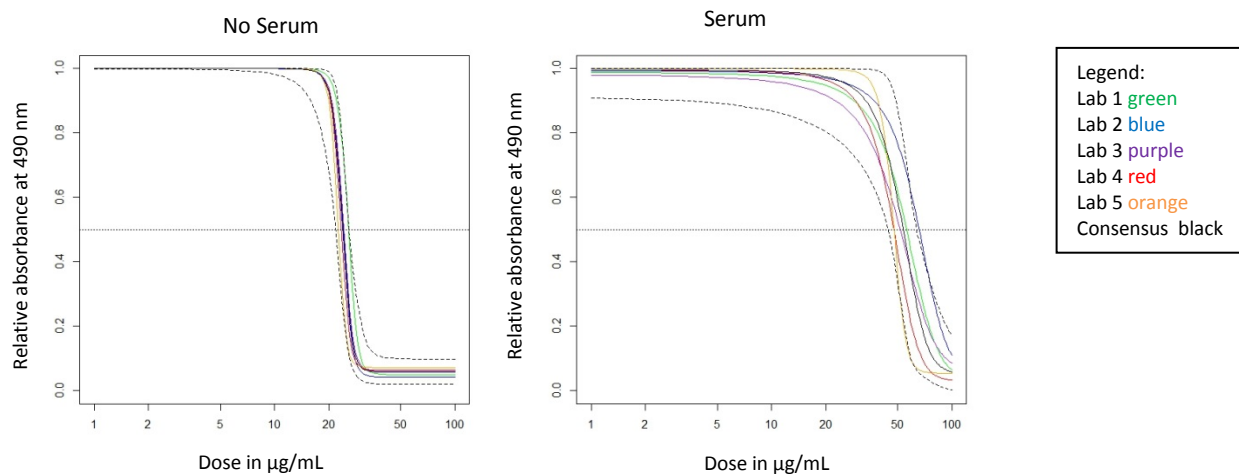

Figure 5. Nanoparticle dose response curves with 95 % uncertainty bounds for clone 1. Horizontal dashed line shows relative absorbance of 50%, the corresponding value on the horizontal axis is the EC$_{50}$.

Laboratory	Clone 1, No serum	Clone 1,Serum	Clone 2, No serum	Clone 2, Serum
1	25.6(0.4)	55.2(2.4)	25.0(0.6)	66.9(8.5)
2	23.7(0.5)	64.4(3.2)	21.3(0.5)	63.1(4.8)
3	23.9(0.5)	49.8(1.7)	21.0(0.8)	46.7(2.5)
4	23.2(0.7)	47.1(0.7)		52.2(0.8)
5	22.3(1.2)	47.3(1.0)		
Consensus	23.7(0.9)	52.7(4.8)	22.5(3.8)	57.1(8.7)

Table 2. EC$_{50}$ (in µg/mL) for the nanoparticle dose response, means and standard uncertainties

Table 2 shows good agreement among laboratories for clone 1 and more variability for clone 2. Interestingly, the mean EC_{50} values for the serum condition for both clones were about 30 µg/mL higher than for the serum-free condition, and the uncertainty was also substantially higher. This is surprising as Figure 2 showed that serum did not introduce additional uncertainty to the control variables. An explanation is suggested by Figure 6 which contrasts the laboratory consensus curves for the serum and serum-free conditions for laboratory 2, clone 1.
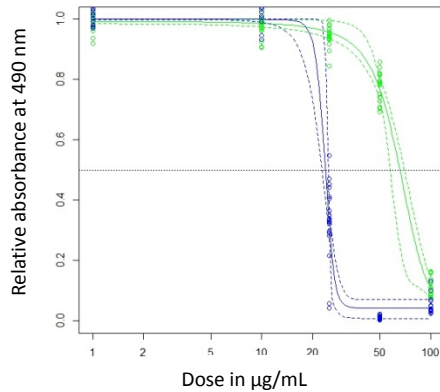

Figure 6. Laboratory consensus curves with 95 % uncertainty bounds for nanoparticle dose response, serum (green) and serum-free (blue) for clone 1, laboratory 2. Dashed horizontal line shows 50% relative absorbance.

The uncertainty bounds of the consensus curves were the narrowest around the experimental concentration that was close to the inflection point of the curve. For the serum-free condition this was 25 µg/mL which is close to the EC_{50} value of 23.7 µg/mL thus making the uncertainty of this parameter small compared to that of the serum condition where the EC_{50} value was 64.4 µg/mL and the bounds were narrowest at 50 µg/mL.Thus the larger variability of the serum EC_{50} value was due to the experimental design and not a response of the cells to the serum.

3.3 Checking Model Fit
For any estimation of parameters based on a statistical model, it is important to check the fit of the model to the data. In a Bayesian analysis, one method of checking model fit is to compute posterior predictive probabilities $P(Y_p > y_{obs})$, where Y_p are predicted values according to the fitted Bayesian statistical model and y_{obs} are the observed measurements [9]. These probabilities, usually called Bayesian posterior predictive *p*-values, measure how likely it is to obtain the value y_{obs} given the model and the observed data (These differ from conventional p-values in that they are computed using the posterior predictive distribution of the Y_p, not the sampling distribution). Figure 7 shows means and uncertainty bounds of the predictive p-values computed over the replicates and plates, for the clone 1 experiment.

Predictive p-values should be around 0.5 [9], that is, about half the time the predicted value should be larger than the observed and half the time smaller. It appears that the data from the serum experiment had a better fit to the model than the serum-free data, except possibly for lab 5. The serum-free data exhibited somewhat poor fit for laboratories 3, 4, and 5 in the low concentrations of 1 and 10 µg/mL. On further examination these are data points that exceeded 1.0 due to the adjustments by background and maximum absorbance. The poor fit was due to the fact that such values were not allowed by the model. But even for the serum-free condition the fit close to the EC_{50} concentration (around level 4) was quite good.
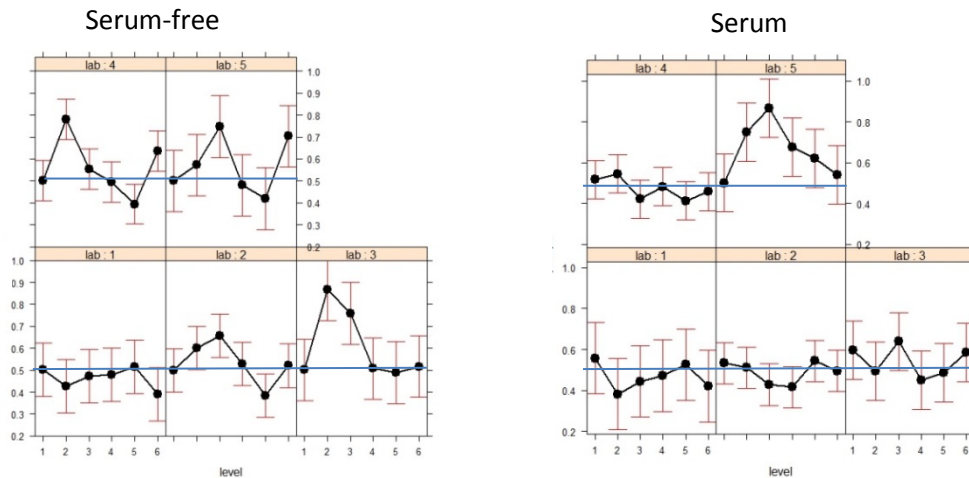
Figure 7. Predictive p-values (with 95 % uncertainty intervals) by laboratory and concentration level (#1=0 µg/mL, #2=1 µg/mL, #3=10 µg/mL, #4=25 µg/mL, #5=50 µg/mL, #6=100 µg/mL) for clone 1.

4.0 Discussion

Analysis of this inter-laboratory experiment showed that it is possible to achieve agreement among laboratories making biological assay measurements. It is also clear that careful monitoring of the protocol is necessary. Uncertainty analysis of the dose response function was accomplished in a top-down fashion using a Bayesian statistical model and MCMC estimation. The statistical model included an uncertainty component for the absorbance measurement itself, showing that it is feasible to augment the top-down model with further uncertainty evaluations in a bottom-up manner.

The plate design was a crucial element of this study, providing critical information as to the quality of the measurements. A potential experimental design improvement is to include concentrations close to the anticipated EC_{50} value.

5.0 References

[1]   Grainger, D. W., and Castner, D. G. (2008) Nanobiomaterials and nanoanalysis: opportunities for improving the science to benefit biomedical technologies. Adv. Mater. 20, 867–877.

[2]   Bhattacharya, S., Zhang, Q., Carmichael, P. L., Boekelheide, K., and Andersen, M. E. (2011) Toxicity testing in the 21 century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. PLoS One 6, e20887.

[3]   Hartung, T. (2009) Toxicology for the twenty-first century. Nature 460, 208–212.

[4]   Rushton, E. K., Jiang, J., Leonard, S. S., Eberly, S., Castranova, V., Biswas, P., Elder, A., Han, X., Gelein, R., Finkelstein, J., and Oberdörster, G. (2010) Concept of assessing nanoparticle hazards considering nanoparticle dosemetric and chemical/biological response metrics. J. Toxicol. Environ. Health A 73, 445–461.

[5]   Rösslein, M., Elliott, J., Salit, M., Petersen, E., Hirsch, C., Krug, H., and Wick, P. (2015) Use of Cause-and-Effect Analysis to Design a High-Quality Nanocytotoxicology Assay. *Chem. Res. Toxicol.*, 28, 21-30.

[6]   Lunn DJ, Spiegelhalter D, Thomas A, Best N. (2009) The BUGS project: Evolution, critique and future directions (with discussion), Statistics in Medicine 28, 3049–3082.

[7]   Toman and Possolo (2009) *Accreditation and Quality Assurance* 14, 553 – 563.

[8]   Toman and Possolo (2010) *Accreditation and Quality Assurance* 15, 653 – 654.

[9]   Gelman A, Meng XL, Stern H. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). Statistica Sinica 6, 733–807.