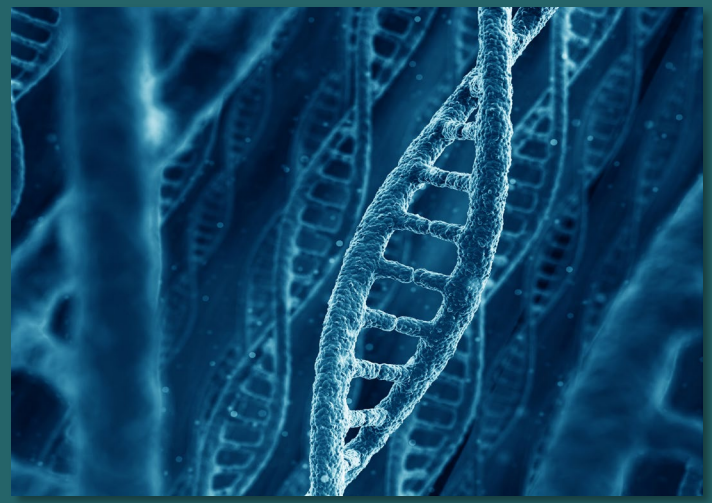
BMD Concordance Analysis with Inter-study Variability

Kelsey Vitense, PhD



The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the U.S. EPA

Introduction

- Concordance between BMD values from short-term transcriptomic studies vs. apical BMD values from chronic rodent bioassays is influenced by inter-study variation in the BMDs

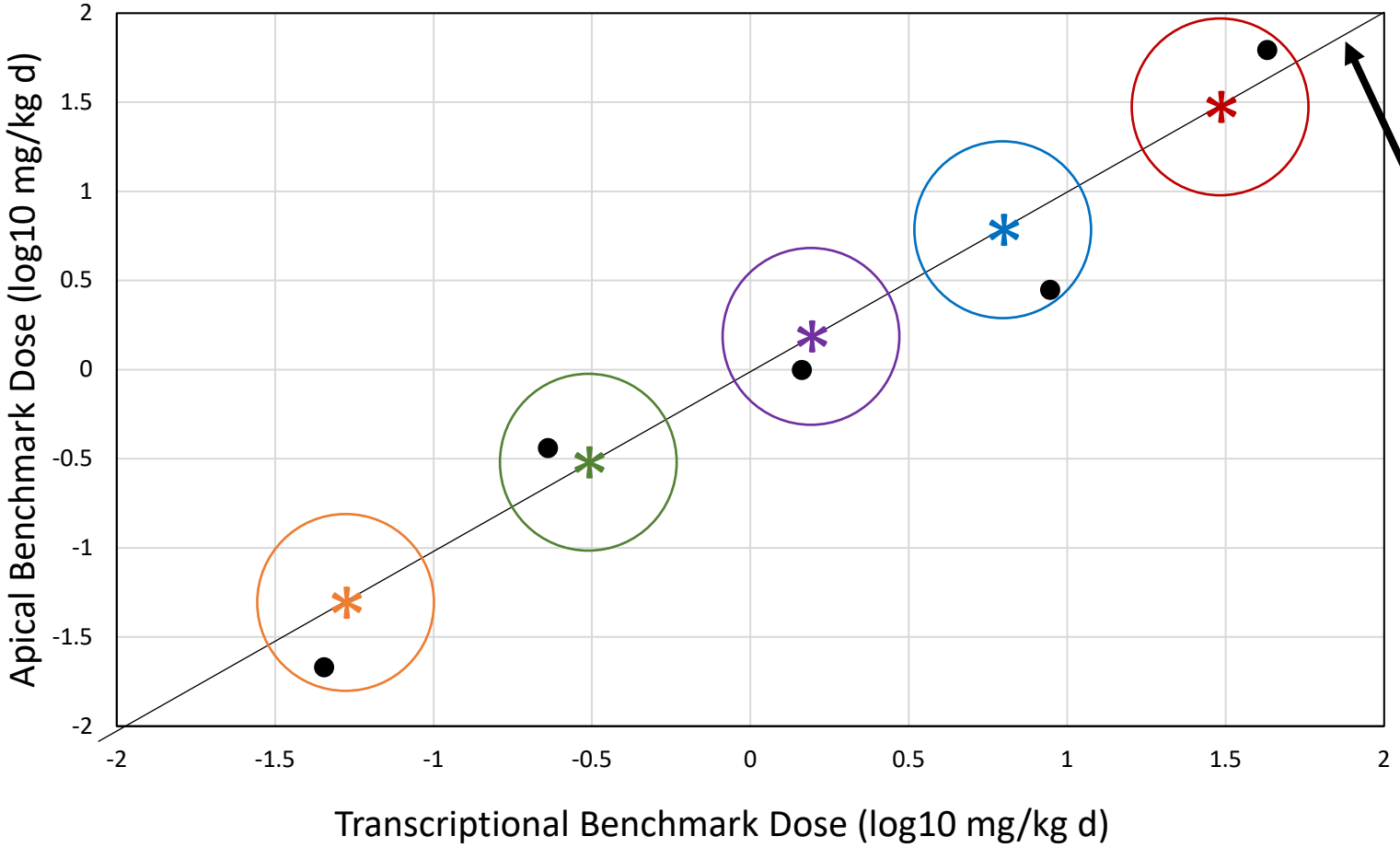
Introduction

- Concordance between BMD values from short-term transcriptomic studies vs. apical BMD values from chronic rodent bioassays is influenced by inter-study variation in the BMDs
- Estimating and considering inter-study variability is important for interpreting concordance metrics and our confidence in application of the ETAP

Introduction

- Concordance between BMD values from short-term transcriptomic studies vs. apical BMD values from chronic rodent bioassays is influenced by inter-study variation in the BMDs
- Estimating and considering inter-study variability is important for interpreting concordance metrics and our confidence in application of the ETAP
- To provide this context, **we estimated the lower bound of expected Mean Squared Difference (MSD) given inter-study variances for comparison with the concordance MSD of the top ETAP model** (i.e., best pre-modeling probe filter, BMD modeling, and gene set summarization parameters)

Inter-study variation impacts apical vs. transcriptomic BMD concordance, even when chemical BMDs are the same on average



Identity line denotes match between apical and transcriptomic BMDs

Derivation of expected MSD lower bound

- We will show that the lower bound of expected MSD is the sum of the transcriptomic and apical BMD variances

$$MSD = \sum_{c=1}^{n} \frac{(x_c - y_c)^2}{n}$$

$$E[MSD] \geq \sigma_X^2 + \sigma_Y^2$$

Derivation of expected MSD lower bound

- Let X_c be the transcriptomic BMD ($\log_{10}$ mg/kg-day) for chemical *c*
- Let Y_c be the apical BMD ($\log_{10}$ mg/kg-day) for chemical *c*

Derivation of expected MSD lower bound

- Let X_c be the transcriptomic BMD ($\log_{10}$ mg/kg-day) for chemical *c*
- Let Y_c be the apical BMD ($\log_{10}$ mg/kg-day) for chemical *c*
- Following Pham et al. 2020, we assume apical BMDs (Y_c) are random variables with:
  - Means dependent on chemical and study design
    - Note: study design is standardized across chemicals in this study
  - Constant variance after accounting for chemical and study design
    - *i.e.*, Common variance across chemicals

Pham et al. 2020. Variability in *in vivo* studies: Defining the upper limit of performance for predictions of systemic effect levels. Computational Toxicology

Derivation of expected MSD lower bound

- Let X_c be the transcriptomic BMD ($\log_{10}$ mg/kg-day) for chemical *c*
- Let Y_c be the apical BMD ($\log_{10}$ mg/kg-day) for chemical *c*
- Following Pham et al. 2020, we assume apical BMDs (Y_c) are random variables with:
  - Means dependent on chemical and study design
    - Note: study design is standardized across chemicals in this study
  - Constant variance after accounting for chemical and study design
    - *i.e.*, Common variance across chemicals
- In the absence of evidence to the contrary, we assume the same for the transcriptomic BMD values (X_c).

Derivation of expected MSD lower bound

- That is, define:

$$E[X_c] = \mu_X(c)$$
$$E[Y_c] = \mu_Y(c)$$

where $\mu_X(c)$ and $\mu_Y(c)$ are the mean transcriptomic and apical BMD values for chemical *c*, respectively.

- And:

$$Var(X_c) = \sigma_X^2$$
$$Var(Y_c) = \sigma_Y^2$$

are the inter-study, within-chemical variances for transcriptomic and apical BMD values, respectively.

Derivation of expected MSD lower bound

- Let $Z_c = X_c - Y_c$ be the difference between transcriptomic and apical BMD values for chemical *c*.

Derivation of expected MSD lower bound

- Let $Z_c = X_c - Y_c$ be the difference between transcriptomic and apical BMD values for chemical *c*.
- Then:

$$E[Z_c] = \mu_X(c) - \mu_Y(c) = \mu_Z$$

  - For simplicity, assume constant difference in BMD means across chemicals

Derivation of expected MSD lower bound

- Let $Z_c = X_c - Y_c$ be the difference between transcriptomic and apical BMD values for chemical *c*.
- Then:

$$E[Z_c] = \mu_X(c) - \mu_Y(c) = \mu_Z$$

  - For simplicity, assume constant difference in BMD means across chemicals
- And:

$$Var(Z_c) = \sigma_X^2 + \sigma_Y^2$$

  - X_c and Y_c are conditionally independent given chemical means, so no covariance term is included

Derivation of expected MSD lower bound

- The MSD concordance statistic between X_c and Y_c for n chemicals is:

$$MSD = \sum_{c=1}^{n} \frac{(x_c - y_c)^2}{n} = \sum_{c=1}^{n} \frac{z_c^2}{n}$$

Derivation of expected MSD lower bound

- The MSD concordance statistic between X_c and Y_c for n chemicals is:

$$MSD = \sum_{c=1}^{n} \frac{(x_c - y_c)^2}{n} = \sum_{c=1}^{n} \frac{z_c^2}{n}$$

- MSD is an unbiased estimator of $E[Z_c^2]$:

$$E[MSD] = E\left[\sum_{c=1}^{n} \frac{z_c^2}{n}\right] = \sum_{c=1}^{n} \frac{E[z_c^2]}{n} = \frac{nE[Z_c^2]}{n} = E[Z_c^2]$$

Derivation of expected MSD lower bound

- The variance of Z_c can be decomposed as follows:

$$Var(Z_c) = E[Z_c^2] - \mu_Z^2$$

- Rearranging:

$$E[Z_c^2] = Var(Z_c) + \mu_Z^2$$

- Substituting $E[MSD] = E[Z_c^2]$:

$$E[MSD] = Var(Z_c) + \mu_Z^2$$

Derivation of expected MSD lower bound

- Starting from $E[MSD] = Var(Z_c) + \mu_Z^2$:

$$E[MSD] = \sigma_X^2 + \sigma_Y^2 + \mu_Z^2$$

- If $\mu_Z = 0$ (mean values of X_c and Y_c are equal for each chemical):

$$E[MSD] = \sigma_X^2 + \sigma_Y^2$$

- If $\mu_Z \neq 0$ (mean values of X_c and Y_c differ across chemicals):

$$E[MSD] > \sigma_X^2 + \sigma_Y^2$$

- Thus:

$$E[MSD] \geq \sigma_X^2 + \sigma_Y^2$$

Derivation of expected MSD lower bound

- That is, the lower bound of expected MSD is the sum of the transcriptomic and apical BMD variances:

$$E[MSD] \geq \sigma_X^2 + \sigma_Y^2$$

- MSD is expected to be approximately equal to the sum of the inter-study variances when apical and transcriptomic BMDs are the same on average across chemicals
- Next, we can use estimates of inter-study variances to approximate this lower bound for comparison to our observed MSD

Estimates of inter-study transcriptomic variance

- We estimated the transcriptomic BMD variance, σ_X^2, using inter-study replicates from three chemicals
  - Bromodichloroacetic acid, Perfluorooctanoic acid, Furan
  - Three replicates per chemical
  - Each replicate 5-day transcriptomic study performed with same doses, in same contract lab, over several years

Estimates of inter-study transcriptomic variance

- We estimated the transcriptomic BMD variance, σ_X^2, using inter-study replicates from three chemicals
- For replicates *i* and *j* of chemical *c:*

$$E\left[X_{c,i} - X_{c,j}\right] = 0$$

$$Var\left(X_{c,i} - X_{c,j}\right) = 2\sigma_X^2$$

Estimates of inter-study transcriptomic variance

- We estimated the transcriptomic BMD variance, σ_X^2, using inter-study replicates from three chemicals
- For replicates *i* and *j* of chemical *c:*

$$E\left[X_{c,i} - X_{c,j}\right] = 0$$

$$Var\left(X_{c,i} - X_{c,j}\right) = 2\sigma_X^2$$

- Let *k* be the number of chemicals with replicate transcriptomic BMD estimates, let r_c be the number of observed replicates for chemical *c,* and let $I_c = \{1,2, \ldots, r_c\}$. An unbiased estimator of σ_X^2 is:

$$\hat{\sigma}_X^2 = \frac{1}{2} \times \widehat{Var}\left(X_{c,i} - X_{c,j}\right) = \left(2\sum_{c=1}^{k}\binom{r_c}{2}\right)^{-1} \sum_{c=1}^{k}\sum_{i \in I_c}\sum_{j \in I_c; j>i} (x_{c,i} - y_{c,j})^2$$

Estimates of inter-study transcriptomic variance

- We estimated the transcriptomic BMD variance, σ_X^2, using inter-study replicates from three chemicals
- For replicates *i* and *j* of chemical *c:*

$$E\left[X_{c,i} - X_{c,j}\right] = 0$$

$$Var\left(X_{c,i} - X_{c,j}\right) = 2\sigma_X^2$$

- Let *k* be the number of chemicals with replicate transcriptomic BMD estimates, let r_c be the number of observed replicates for chemical *c,* and let $I_c = \{1,2, \ldots, r_c\}$. An unbiased estimator of σ_X^2 is:

$$\hat{\sigma}_X^2 = \frac{1}{2} \times \left[\begin{array}{c} Mean\ squared\ difference\ between\ transcriptomic\ BMD \\ values\ for\ unique\ pairs\ of\ replicates\ for\ each\ chemical \end{array} \right]$$

Estimates of inter-study transcriptomic variance

- We computed transcriptomic BMD variance estimates across all dose-response modeling parameter combinations considered
- Used the min & max of variance estimates to provide a range for σ_X^2:

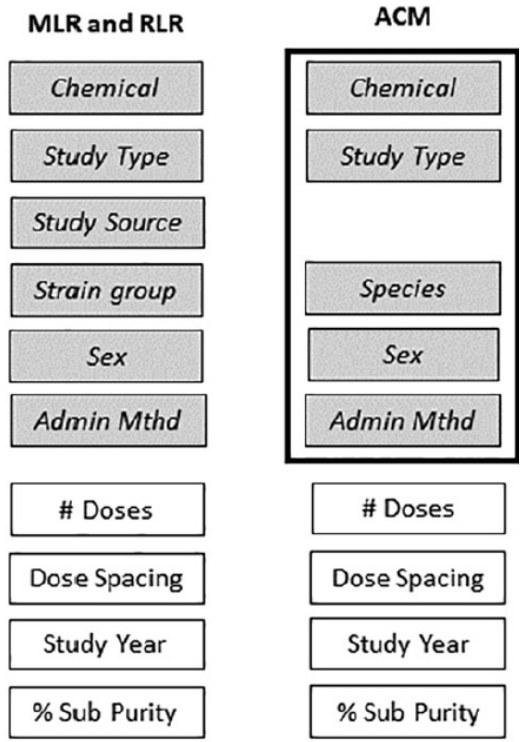
$$\hat{\sigma}_X^2 \approx [0.015, 0.352]$$

$$(\hat{\sigma}_X \approx [0.123, 0.594])$$

Estimates of inter-study apical variance

- We estimated apical BMD variance, σ_Y^2, using mean squared error (MSE) from a multiple regression model (Pham et al. 2020), which estimates inter-study LEL/LOAEL variance after accounting for study descriptors

Study Descriptor	Conditions
Chemical	Identified using CASRN and chemical name
Study Type	CHR, SUB, DEV, MGR, SAC
Study Source	OPP, NTP, Pharma, Open Lit
Strain Group or Species	Species used: mouse, rat, dog, rabbit
Sex	Male, Female, Male & Female
Administration Method	Feed, Capsule, Gavage/Intubation, Oral, Water
Number of Dose Levels	Number of non-control, treatment related doses
Dose Spacing	Average distance between each dose
Study Year	1959 to 2012
% Substance Purity	77% to 100%



Pham et al. 2020. Variability in *in vivo* studies: Defining the upper limit of performance for predictions of systemic effect levels. Computational Toxicology

Estimates of inter-study apical variance

- We estimated apical BMD variance, σ_Y^2, using mean squared error (MSE) from a multiple regression model (Pham et al. 2020), which estimates inter-study LEL/LOAEL variance after accounting for study descriptors

Variance estimation results for subsets by study type.

Regression Type	Data	LEL				LOAEL				N
		Total Variance	MSE	RMSE	% exp.	Total Variance	MSE	RMSE	% exp.	
MLR	SUB	0.879	0.350	0.591	60.2	0.782	0.277	0.527	65.0	705
ACM	SUB	1.013	0.301	0.549	70.3	0.904	0.250	0.500	72.4	92
MLR	CHR	0.952	0.352	0.593	63.1	0.795	0.252	0.502	68.4	1149
ACM	CHR	0.887	0.395	0.629	55.4	0.825	0.265	0.515	68.0	117
MLR	DEV	0.604	0.246	0.496	59.3	0.594	0.217	0.465	63.5	275
ACM	DEV	0.410	0.328	0.573	20.0	0.398	0.316	0.562	20.7	54

Two regression types (MLR = multilinear regression, ACM = augmented cell means) were used to build models using data subset by the study type (SUB = subchronic; CHR = chronic; DEV = developmental) for variance estimation. Total variance and MSE are in units of $(\log10(\text{mg/kg/day}))^2$, whereas RMSE is in log10(mg/kg/day) units just like the dataset. % exp = percent total variance explained. N = number of study records in the dataset.

Pham et al. 2020. Variability in *in vivo* studies: Defining the upper limit of performance for predictions of systemic effect levels. Computational Toxicology

Estimates of inter-study apical variance

- We estimated apical BMD variance, σ_Y^2, using mean squared error (MSE) from a multiple regression model (Pham et al. 2020), which estimates inter-study LEL/LOAEL variance after accounting for study descriptors

- The min & max of chronic apical LOAEL variance estimates from Pham et al. 2020 used to approximate the apical BMD variance, σ_Y^2, were:

$$\hat{\sigma}_Y^2 \approx [0.252, 0.265]$$

$$(\hat{\sigma}_Y \approx [0.502, 0.515])$$

Expected MSD lower bound estimate

- Min & max of transcriptomic BMD variance estimates:

$$\hat{\sigma}_X^2 \approx [0.015, 0.352]$$

- Min & max of chronic apical BMD variance estimates:

$$\hat{\sigma}_Y^2 \approx [0.252, 0.265]$$

Expected MSD lower bound estimate

- Min & max of transcriptomic BMD variance estimates:

$$\hat{\sigma}_X^2 \approx [0.015, 0.352]$$

- Min & max of chronic apical BMD variance estimates:

$$\hat{\sigma}_Y^2 \approx [0.252, 0.265]$$

- Sum provides lower bound estimate for expected MSD:

$$E[MSD] \geq [0.267, 0.617]$$

- Lower bound provides an estimate of what we would expect MSD to be if the apical and transcriptomic BMDs are the same on average but inter-study variation exists for both BMDs

MSD of top transcriptomic model compared to estimated lower bound for E[MSD]

- MSD of the top combination of transcriptomic model parameters computed using mean BMD values for chemicals with replicates was:

$$0.567^2 = 0.321$$

MSD of top transcriptomic model compared to estimated lower bound for E[MSD]

- MSD of the top combination of transcriptomic model parameters computed using mean BMD values for chemicals with replicates was:

$$0.567^2 = 0.321$$

- However, using mean BMD values for only some chemicals violates the assumption of equal variance across chemicals used to derive the lower bound of expected MSD.

MSD of top transcriptomic model compared to estimated lower bound for E[MSD]

- MSD of the top combination of transcriptomic model parameters computed using mean BMD values for chemicals with replicates was:

$$0.567^2 = 0.321$$

- However, using mean BMD values for only some chemicals violates the assumption of equal variance across chemicals used to derive the lower bound of expected MSD.
- For fair comparison with the lower bound estimate, the MSD of the top model was computed using all combinations of single replicates per chemical, with the following MSD min & max:

$$[0.285, 0.386]$$

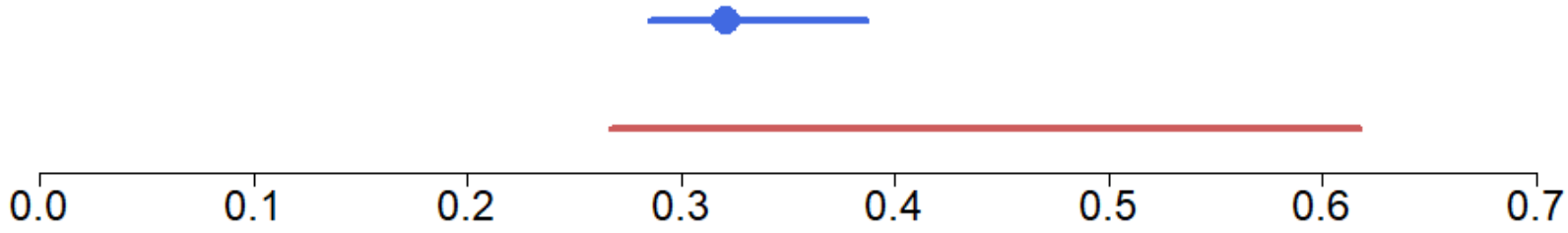
MSD of top transcriptomic model compared to estimated lower bound for E[MSD]

The min & max MSD values computed using single chemical replicates

[0.285, 0.386]

fall within the range of lower bound estimates for expected MSD

[0.267, 0.617]



Conclusion

- The error associated with the concordance between the transcriptomic BMD values vs. apical BMD values is approximately equivalent to the combined inter-study variability associated with the 5-day transcriptomic study and the two-year rodent bioassay
- Thus, transcriptomic and apical BMD values are highly concordant in the context of inter-study variation in BMDs