Variability in *in vivo* Toxicity Studies: Defining the upper limit of predictivity for models of systemic effect levels

Ly Ly Pham, PhD
ORISE Post-Doctoral Fellow
NCCT/EPA
June 19, 2018
ORCID:0000-0001-8467-2645

Hazard information from animals is not available for all chemicals. We need alternative models to predict a point-of-departure.

What are the challenges to acceptance of alternative models?

- Alternative models are often validated against animal studies.
- What is the quantitative uncertainty for a given point-of-departure from animal studies?

Major Objectives of This Work

(1) Provide a quantitative estimate of variance in traditional, systemic effects from regulatory toxicology studies.

- Based on lowest effect levels (LELs) and lowest observable adverse effect levels (LOAELs) from ToxRefDBv2.0 (pre-release).
- Quantify the variance explained by 10 study parameters using multi-linear regression.
- Use two different methods to give variance estimate by chemical and then variance estimate by using an augmented cell means approach.

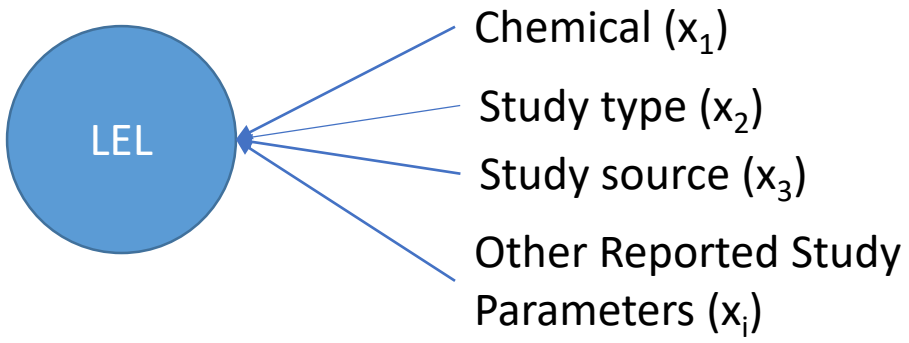
(2) Based on the explained variance, approximated by root mean square error (RMSE), define a reasonable prediction interval for a new approach method (NAM) to approximate a systemic effects point-of-departure dose.

(3) Based on the relationship between unexplained variance, approximated by mean square error (MSE), and R-squared, define an upper limit on the precision for a NAM to predict a systemic effects point-of-departure dose.

Part 1: The models used in this approach

A Conceptual Variance Model

Var(Observed LELs) = Var(Explained by Reported Study Parameters) + Unexplained Variance



Approach 1

Multi-linear regression (MLR) model assumes that each covariate is independent.

Response = Fit + Residual (MSE)

$$Y_i = \alpha + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \epsilon_i$$

Approach 2

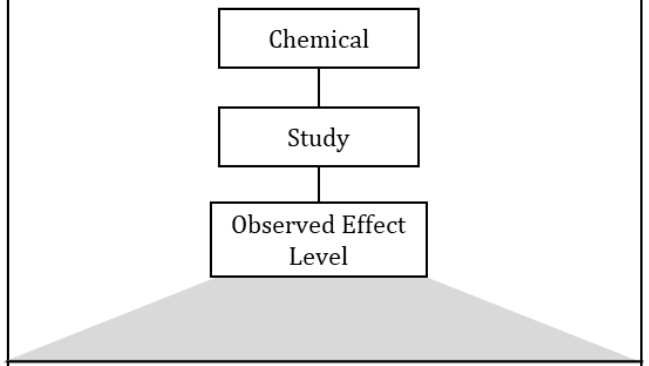
Augmented cell means model (ACM) combines the cells means model with liner regression. The cell means model groups the categorical variables into unique combinations called “cells”. These cells are “truer” replicates.

Cell Means	Factor 1 (Level 1)	Factor 1 (Level 2)
Factor 2 (Level 1)	$\tilde{\mu}_1$	$\tilde{\mu}_2$
Factor 2 (Level 2)	$\tilde{\mu}_3$	$\tilde{\mu}_4$

A Cell ($\tilde{\mu}_i$)

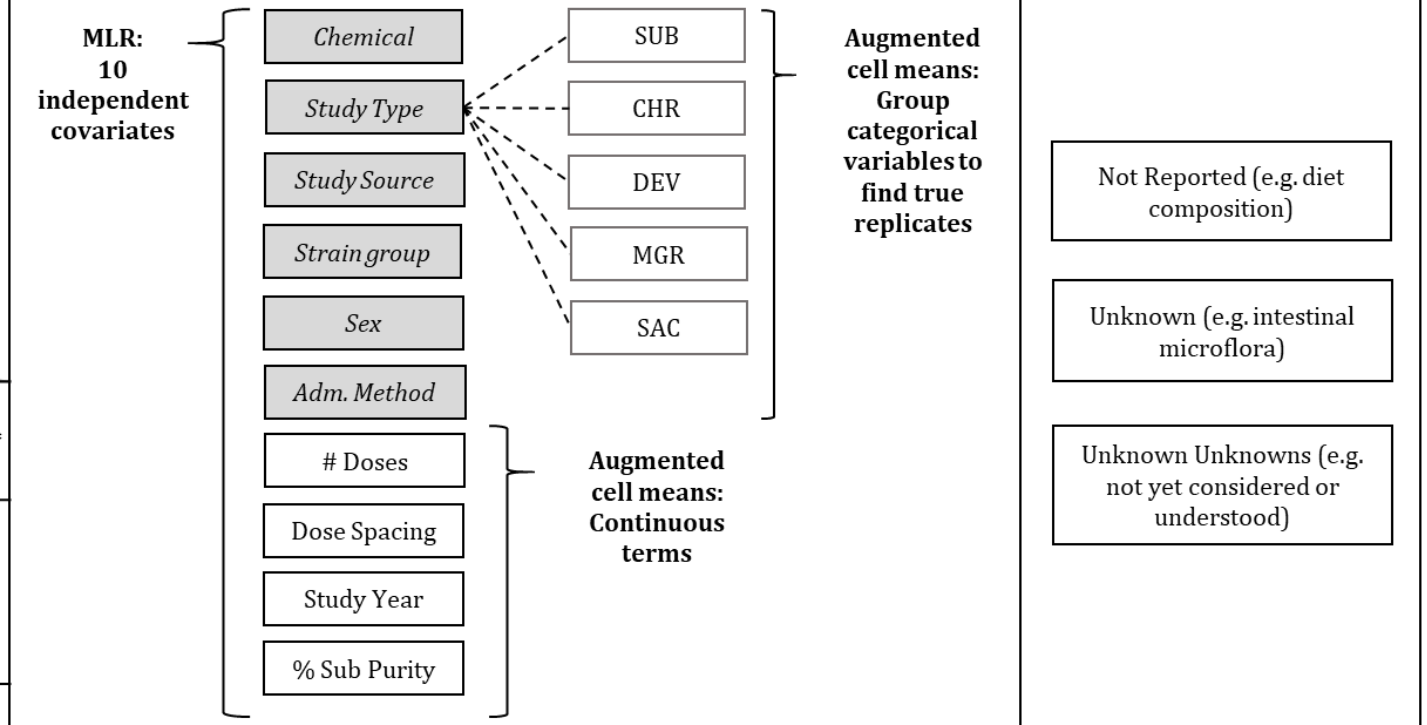
Two approaches to capturing the variance

Variance in Observed Study Level Lowest Effect Level = Variance in Identified Study Parameters (Explained Variance) + Variance in Unidentified Study Parameters and Others (Unexplained Variance)



Study	Effect	Treatment Related Effect* (mg/kg/day)	Critical Effect**
1	Body weight	5†	0
1	Liver	15††	1
1	kidney	20	0
1	heart	10	0

* Dose at which a treatment related effect was observed
** Expert driven designation
† Observed effect level used in set 1
†† Observed effect level used in set 2



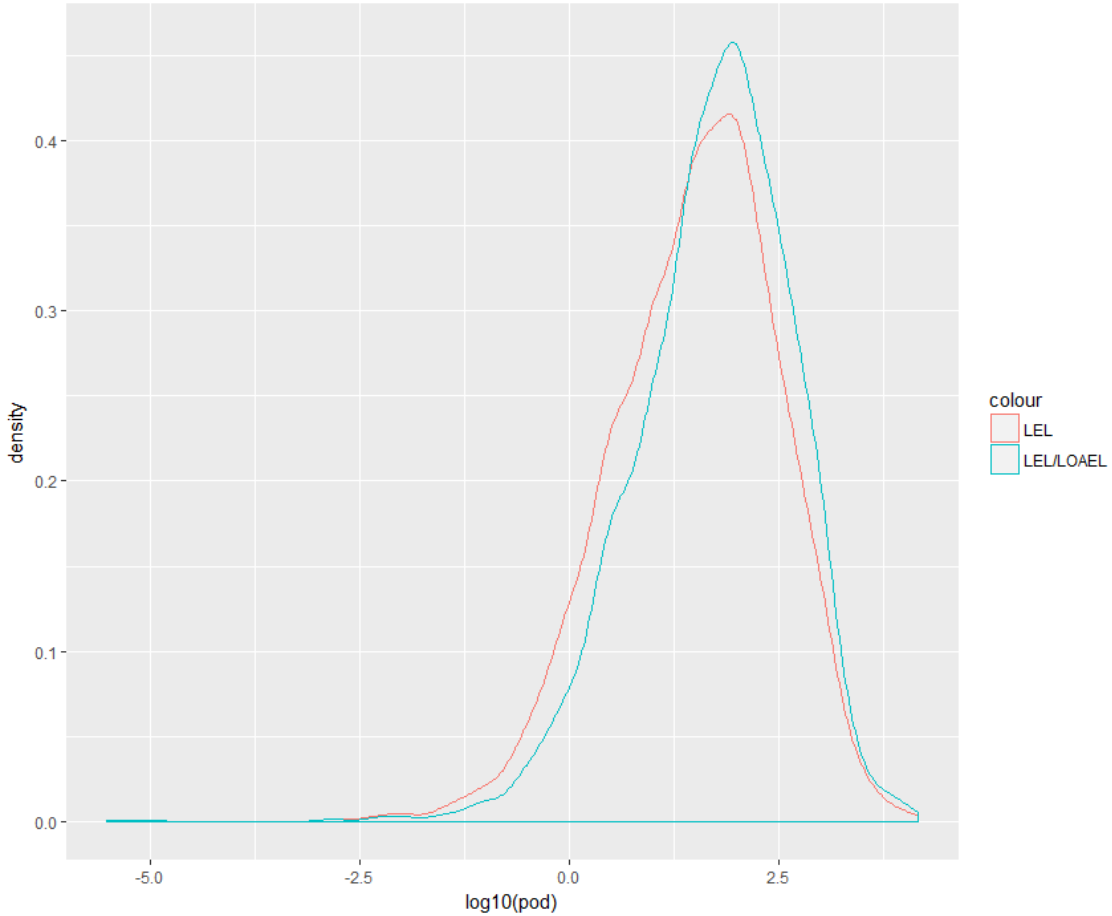
LEL (Set 1)	Log_{10} Lowest dose in which a treatment related effect was observed.
LOAEL (Set 2)	Log_{10} Lowest dose in which an adverse effect was observed.

Data source: ToxRefDB v2.0 (pre-release)

- Initial dataset is fairly heterogeneous and represents the maximum amount of adult systemic effect data available.
  - Life Stage: Adult
  - Admin Route: Oral
  - Effect Category: Systemic
  - Generation: F0
  - At least 2 studies per chemical

- Characteristics
  - 725 chemicals
  - 3439 studies

- This dataset was used first in the MLR, and then a subset of it was used in the augmented cell means (ACM) model.



This frequency distribution demonstrates that ~72.8% of the LELs = LOAELs.

Part 2: The variance estimates from these models

Results: Variance Explained – Main Model (MLR)

Var(Observed LELs) = Var(Explained by Reported Study Descriptors) + Unexplained Variance

Model	Total Variance	Explained Variance	Unexplained Variance	Percent Variance Explained
LEL (Set 1)	1	0.65	0.35	65.4
LOAEL (Set 2)	0.89	0.62	0.27	58.5

The MLR Model has Advantages and Disadvantages

Advantages	Disadvantages
Allows estimation of variance for a very large number of studies.	Has the potential for overestimation of variance because the 10 covariates are likely not independent (potential interactions).
Allows for the inclusion of any study type with recorded effects in adult animals.	Has the potential for overestimation of variance because there are combinations of the 10 covariates that do not occur (matrix sparsity).

So, the MLR may give a good benchmark estimate of variance, but using an augment cell means (ACM) approach may enable a more accurate estimate of variance.

“Replicate” Studies for MLR model

Obs.	Chem.	Strain Group	Admin Method	Study Type	Sex	Study Source	LEL log10(LEL)	N
1	Chem 1	sprague dawley	Feed	CHR	MF	NTP	1.5	4
2	Chem 1	sprague dawley	Feed	CHR	MF	NTP	2	
3	Chem 1	long evans	Gavage/Intubation	CHR	MF	NTP	3	
4	Chem 1	beagle	Feed	SUB	MF	NTP	2	
5	Chem 2	long evans	Feed	DEV	MF	NTP	1	2
6	Chem 2	fischer	Feed	CHR	F	NTP	1.2	
7	Chem 3	fischer	Gavage/Intubation	CHR	MF	NTP	2.6	1
8	Chem 4	b6c3f1	Feed	SUB	MF	OPP	1.3	2
9	Chem 4	b6c3f1	Feed	SUB	MF	OPP	2.5	

Removed

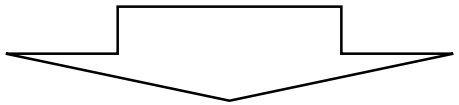
Current filters removed chemicals with less than 2 studies

Defining “Replicates” for the ACM

Obs.	Chem.	Strain Group	Admin Method	Study Type	Sex	Study Source	LEL log10(LEL)	N	Mean log10(LEL)s	Variance	
1	Chem 1	sprague dawley	Feed	CHR	MF	NTP	1.5	2	1.75	0.125	A Cell ($\tilde{\mu}_1$) with 2 replicate
2	Chem 1	sprague dawley	Feed	CHR	MF	NTP	2	2			
3	Chem 1	long evans	Gavage/ Intubation	CHR	MF	NTP	3	1	1.5	0	Removed
4	Chem 1	beagle	Feed	SUB	MF	NTP	2	1	0.9	0	
5	Chem 2	long evans	Feed	DEV	MF	NTP	1	1	1	0	
6	Chem 2	fischer	Feed	CHR	F	NTP	1.2	1	1.2	0	
7	Chem 3	fischer	Gavage/ Intubation	CHR	MF	NTP	2.6	1	1.5	0	
8	Chem 4	b6c3f1	Feed	SUB	MF	OPP	1.3	2	1.9	0.72	A Cell ($\tilde{\mu}_2$) with 2 replicate
9	Chem 4	b6c3f1	Feed	SUB	MF	OPP	2.5	2			

A replicate in the ACM is defined as set of studies performed with the same chemical, study type, strain group, administration method, sex, and source (i.e., all categorical covariates in the MLR).

Obs.	Chem.	Strain Group	Admin Method	Study Type	Sex	Study Source	LEL log10(LEL)	N	Mean log10(LEL)s	Variance	
1	Chem 1	sprague dawley	Feed	CHR	MF	NTP	1.5	2	1.75	0.125	A Cell ($\tilde{\mu}_1$) with 2 replicate
2	Chem 1	sprague dawley	Feed	CHR	MF	NTP	2	2			
3	Chem 1	long evans	Gavage/ Intubation	CHR	MF	NTP	3	1	1.5	0	Removed
4	Chem 1	beagle	Feed	SUB	MF	NTP	2	1	0.9	0	
5	Chem 2	long evans	Feed	DEV	MF	NTP	1	1	1	0	
6	Chem 2	fischer	Feed	CHR	F	NTP	1.2	1	1.2	0	
7	Chem 3	fischer	Gavage/ Intubation	CHR	MF	NTP	2.6	1	1.5	0	
8	Chem 4	b6c3f1	Feed	SUB	MF	OPP	1.3	2	1.9	0.72	A Cell ($\tilde{\mu}_2$) with 2 replicate
9	Chem 4	b6c3f1	Feed	SUB	MF	OPP	2.5	2			



Obs.	Chem.	Strain Group	Admin Method	Study Type	Sex	Study Source	LEL log10(LEL)	N	Mean log10(LEL)s	Variance		New Id
1	Chem 1	sprague dawley	Feed	CHR	MF	NTP	1.5	2	1.75	0.125	A Cell ($\tilde{\mu}_1$) with 2 replicate	1
2	Chem 1	sprague dawley	Feed	CHR	MF	NTP	2	2				
8	Chem 4	b6c3f1	Feed	SUB	MF	OPP	1.3	2	1.9	0.72	A Cell ($\tilde{\mu}_2$) with 2 replicate	2
9	Chem 4	b6c3f1	Feed	SUB	MF	OPP	2.5	2				

We can model/estimate the variance around the LEL

LEL ~ MLR(new id + dose spacing + dose number + study year + substance purity + $\in$)

Results: Variance Explained for Main and Cell Means Model

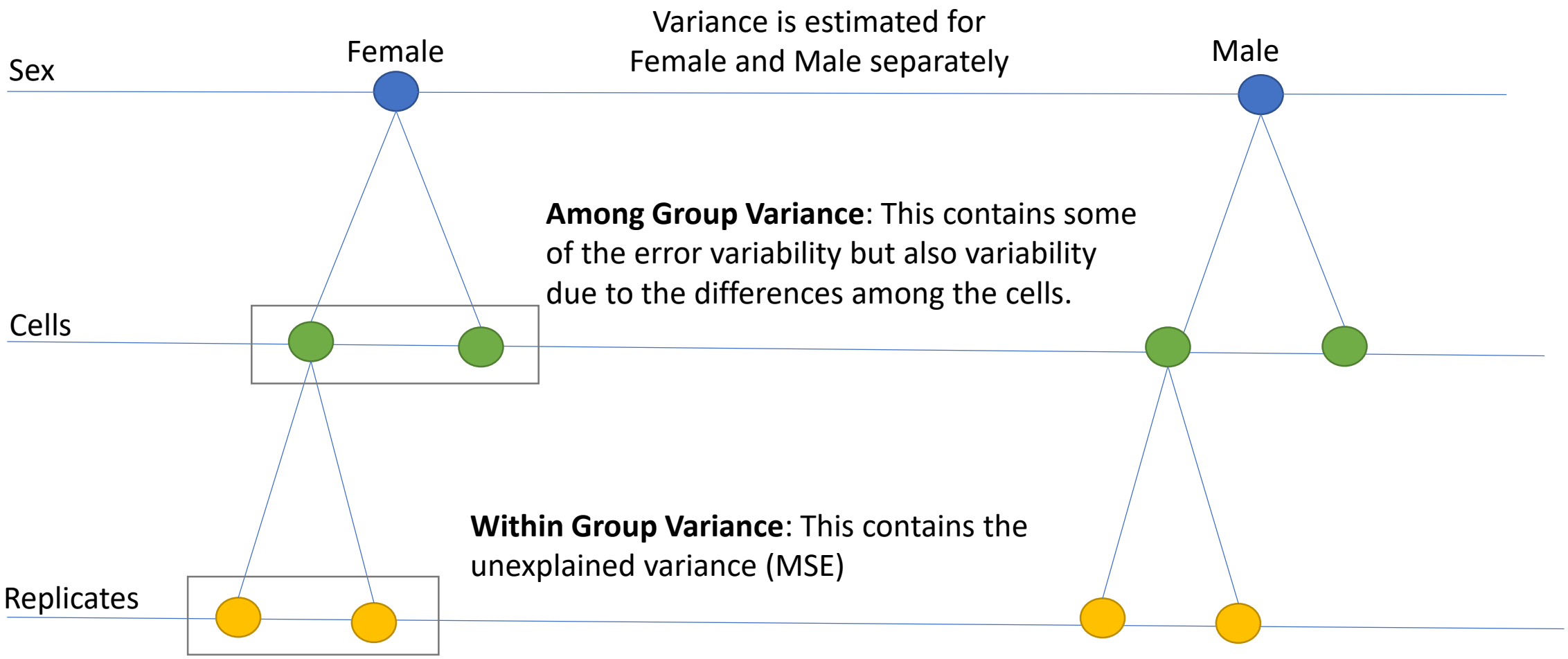
	LEL - Main	LEL - Cell	LOAEL - Main	LOAEL - Cell
Chem (n)	725	151	725	151
Studies(n)	3439	423	3439	423
total var	1	1	0.89	0.89
MSE	0.35	0.22	0.27	0.17
RMSE	0.59	0.47	0.52	0.41
% Explain	65.42	78.11	58.49	70.26

Will we be able to explain more of the variance if we were to subset the data further?

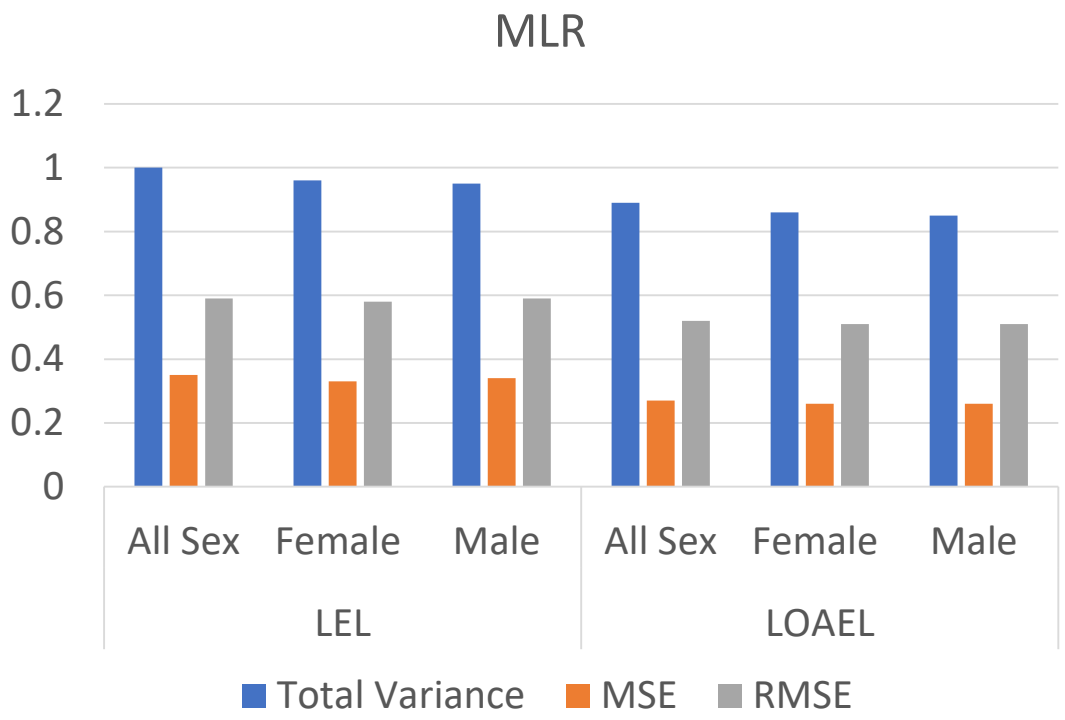
- By Study Type
- By Gender

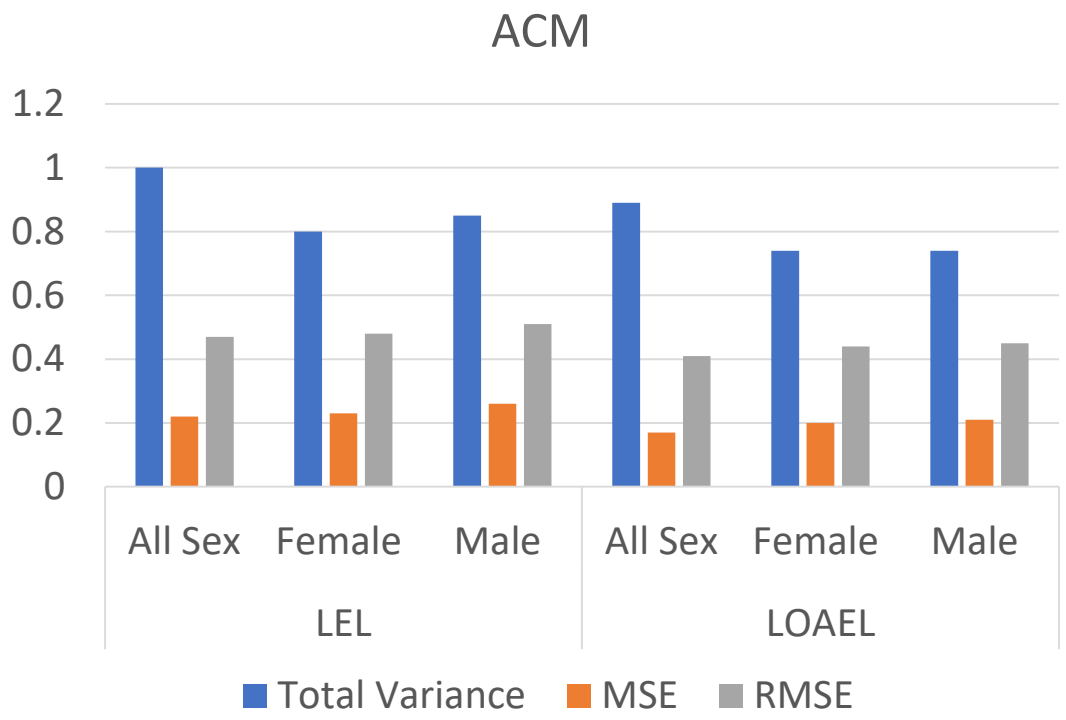
- MSE and RMSE is smaller for the ACM model, but not dramatically
- Consider that the ACM model used significantly less data
- Percent variance explained in the ACM model is larger than the Main model

Subset Data by Sex



Results: Variance Explained – Subset by Sex





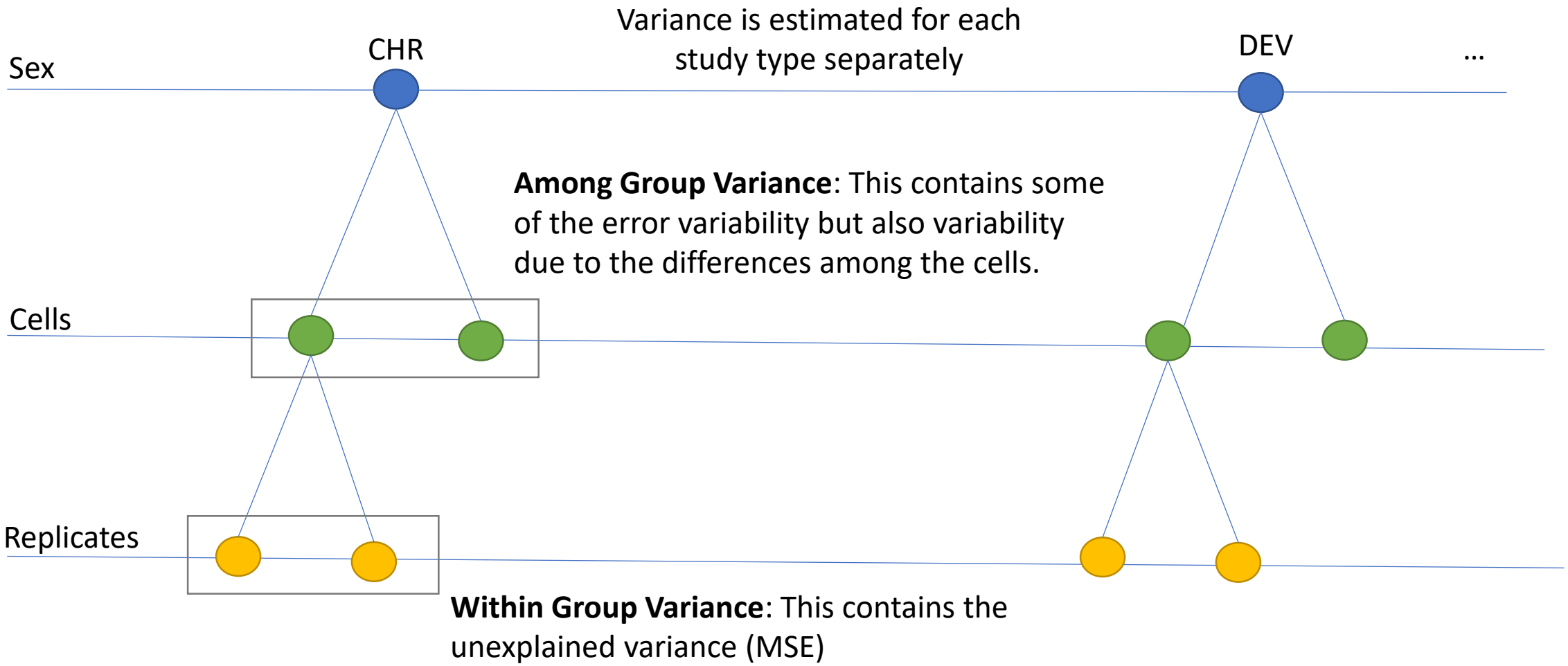
MLR	All Sex	Female	Male
Chem (n)	725	611	611
Studies(n)	3439	2538	2538

ACM	All Sex	Female	Male
Chem (n)	151	100	100
Studies(n)	423	266	266

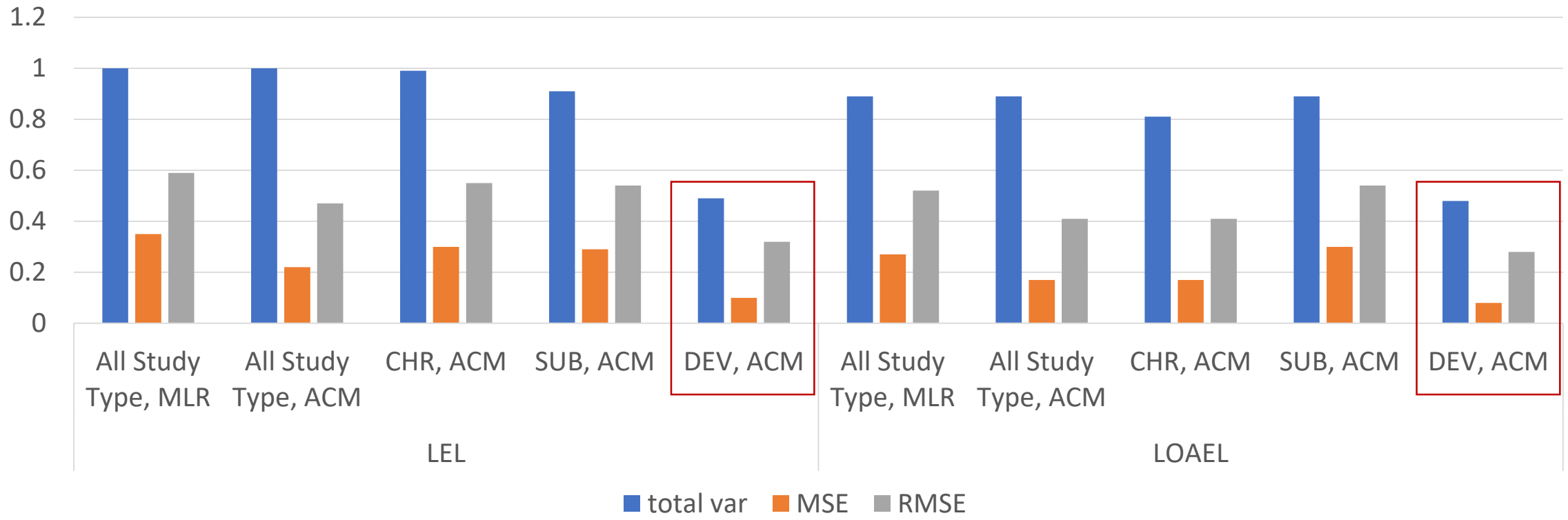
RMSE is in the unit of measurement, whereas variance and MSE are squared units

The MSE and RMSE from the male and Female does not differ much from the full dataset nor from each other.

Subset Data by Study Type



Results: Variance Explained – Subset by Study Type



	MLR	ACM	CHR, AMC	SUB, AMC	DEV, AMC
Chem (n)	725	151	61	41	59
Studies(n)	3439	423	144	101	147

- The MSE and RMSE for when the subset of study types are smaller than the full model
- The MSE and RMSE from the DEV studies are much smaller than both CHR and SUB studies

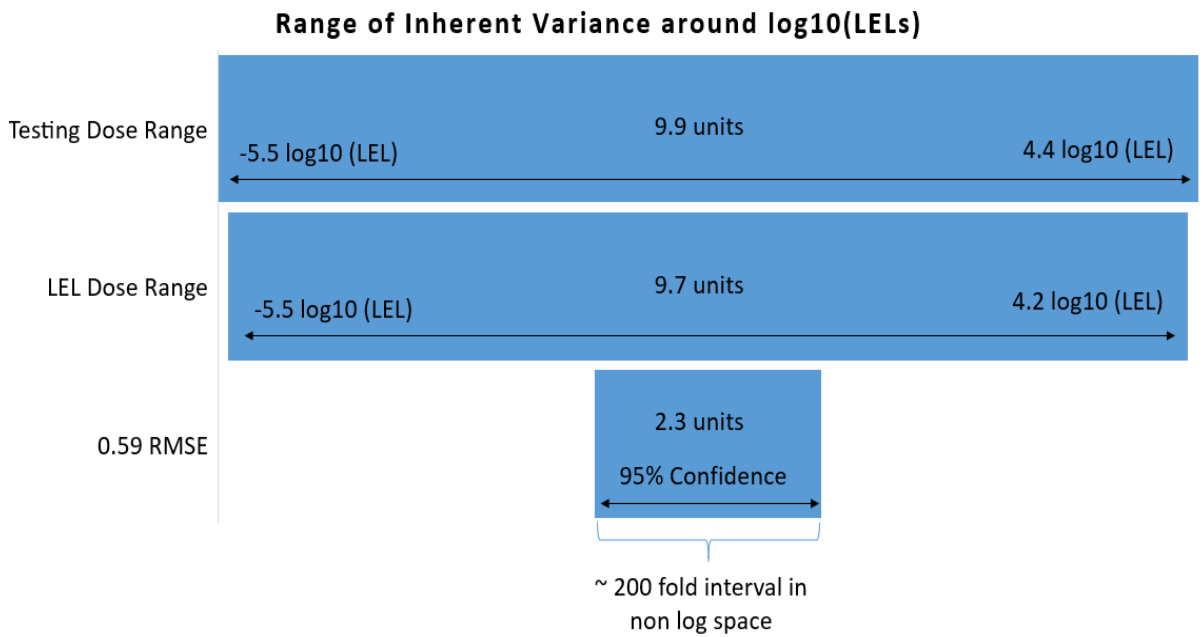
Part 3: Translating variance estimates to meaning for predictive models

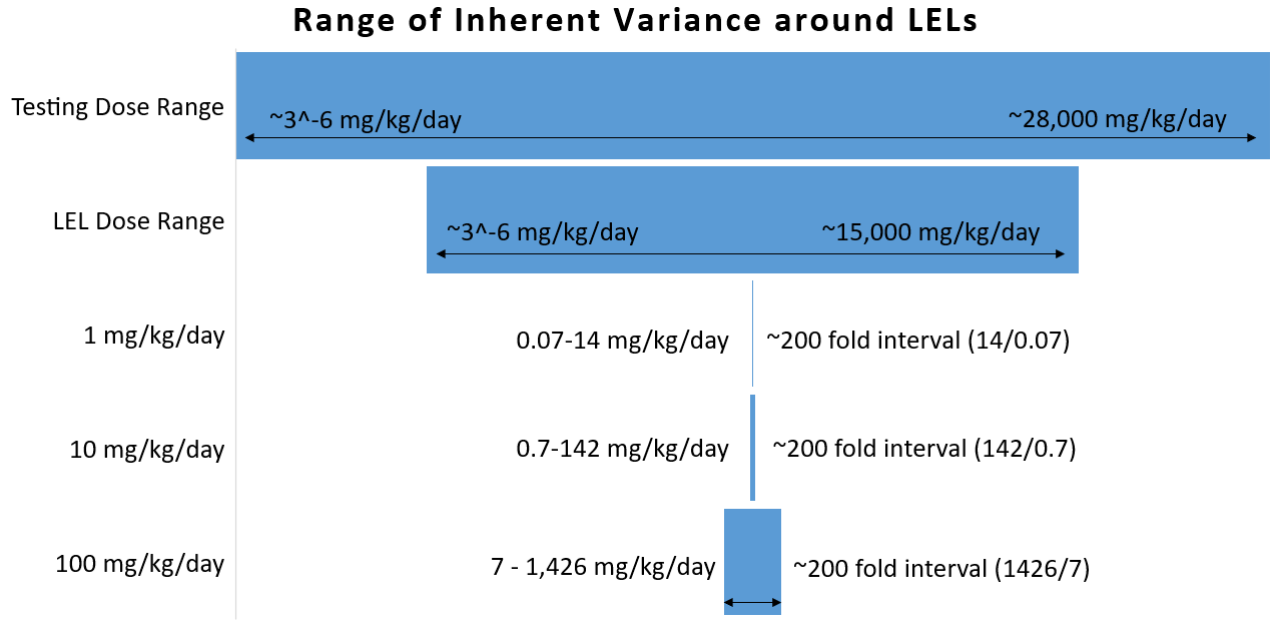
What does this mean for our dataset? - MLR

Inherent Uncertainty around $\log_{10}$ LEL

$$MSE \approx 0.35;\ RMSE \approx \pm 0.59\ log_{10}LEL \quad (RMSE = \sqrt{MSE})$$

Reasonable prediction interval for LELs





Upper Bounds of Performance for a Systemic Effect Model

$$R^2 = \frac{Total\ Variance - MSE}{Total\ Variance}$$

Example:
Variance in this dataset is 1

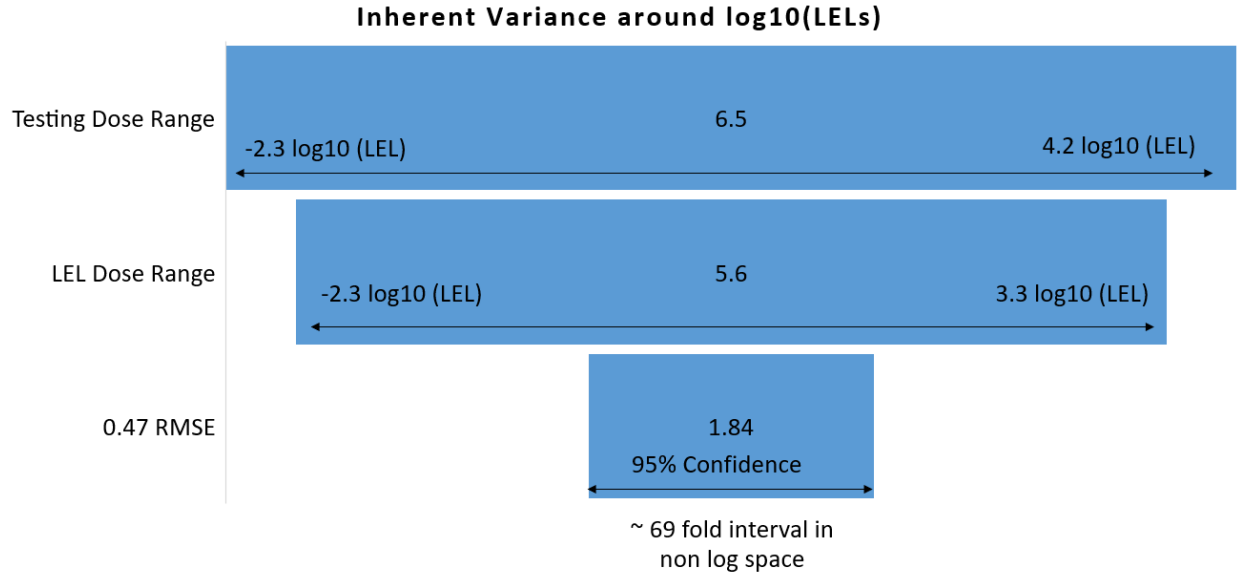
$$R^2 = \frac{1-0.35}{1} = 0.65$$

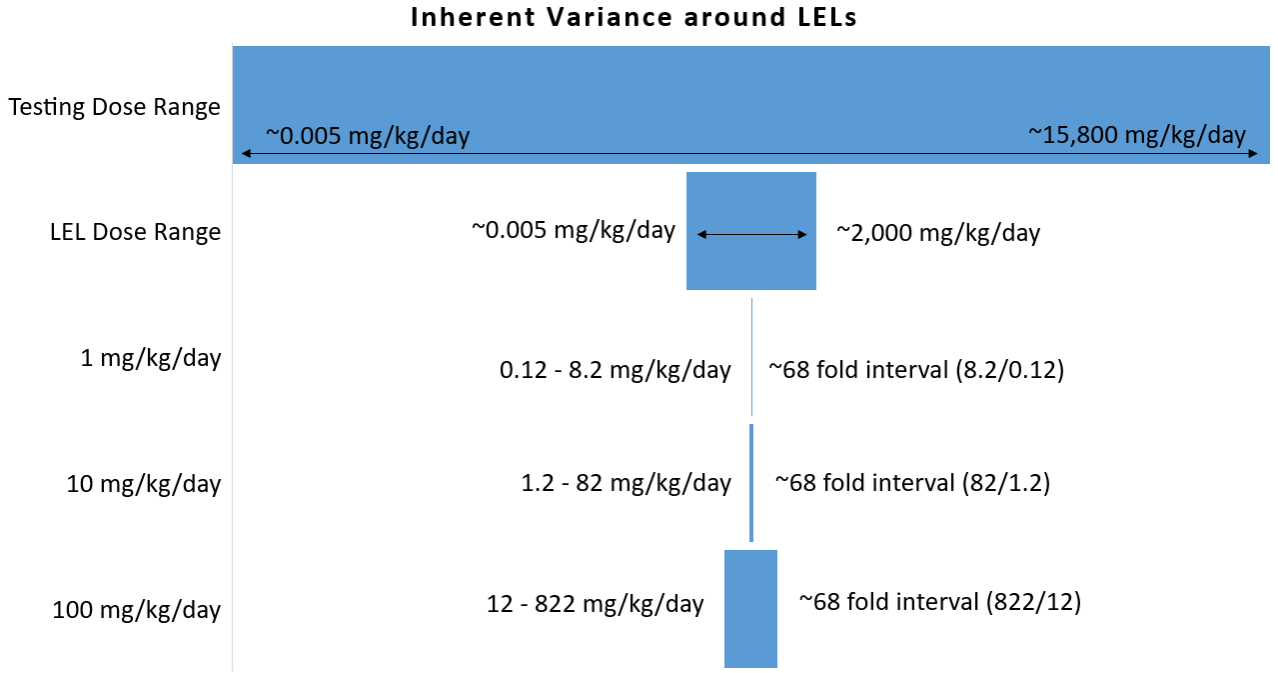
What does this mean for our dataset? - ACM

Inherent Uncertainty around $\log_{10}$ LEL

$$MSE \approx 0.22;\ RMSE \approx \pm 0.47\ log_{10}LEL \quad (RMSE = \sqrt{MSE})$$

Reasonable prediction interval for LELs





Upper Bounds of Performance for a Systemic Effect Model

$$R^2 = \frac{Total\ Variance - MSE}{Total\ Variance}$$

Example:
Variance in this dataset is 1

$$R^2 = \frac{1-0.22}{1} = 0.78$$

Conclusions

- Two different methods of partitioning the data from ToxRefDB were used to estimate the total variance in systemic effects in adults, a MLR model and an ACM model.
- Both models suggest that the total variance for LELs approaches 1.0, and that roughly 20-35% of this is unexplained by study descriptors (MSE approaches 0.2-0.35, depending on the model). The percent explained was slightly greater using the ACM (78%) than MLR (65%).
- The RMSE is useful for characterizing a reasonable prediction interval for a NAM for systemic effects modeling.
  - The RMSE for MLR suggests a prediction interval that is roughly 2.3 log10(mg/kg/day) wide.
  - The RMSE for ACM suggests a prediction interval that is roughly 1.8 log10(mg/kg/day) wide.
  - Interestingly, the results for CHR and SUB studies alone are similar. Predicting effects in adults from the DEV study along may have a narrower prediction interval, perhaps in part due to unique parameters of that study type.
- The MSE is useful for setting an expectation of the R-squared for a predictive model for systemic effects.
  - The MSE for the MLR suggests an upper limit on R-squared of approximately 60-65%.
  - The MSE for the ACM suggests an upper limit on R-squared of approximately 70-78%. Note that this is a much smaller dataset than the MLR.

Acknowledgements

Katie Paul-Friedman

Woodrow Setzer

Matthew Martin

Sean Watford

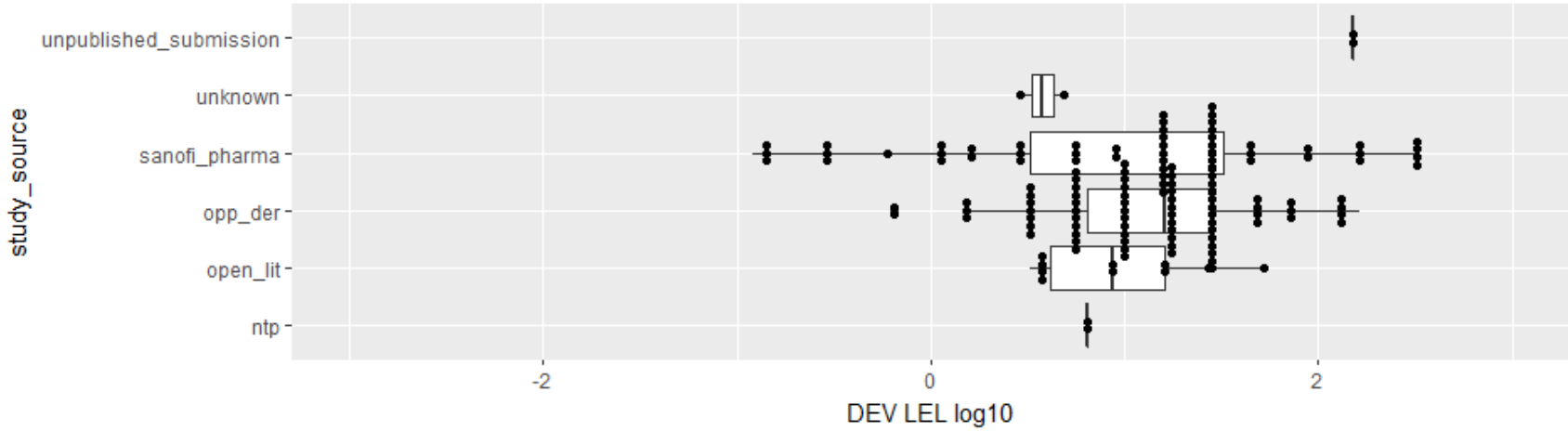
Angelica Adrian

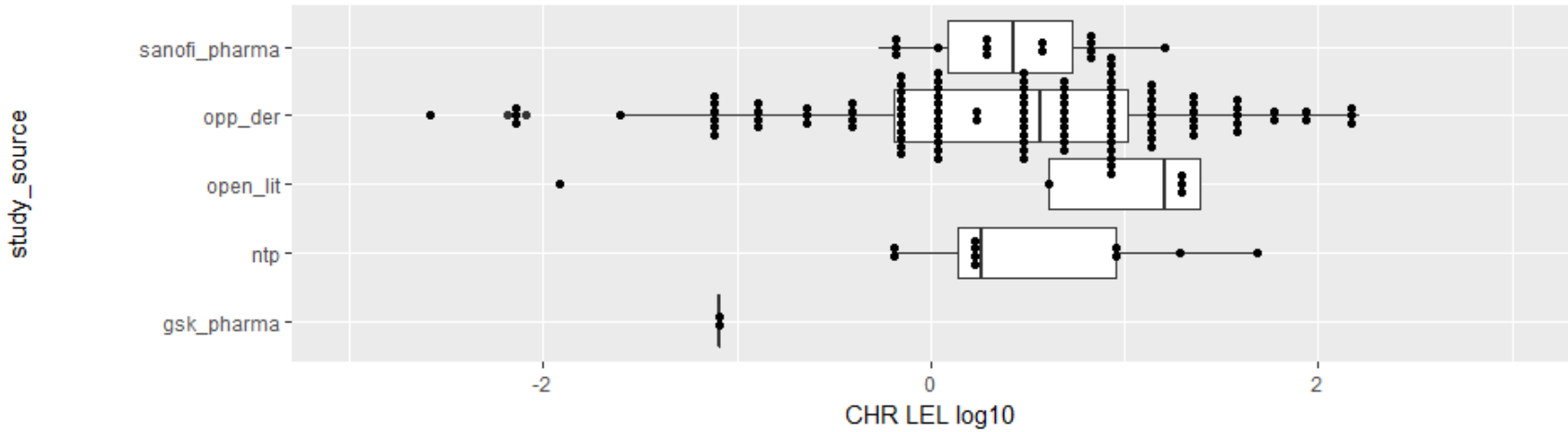
Josephine Brown

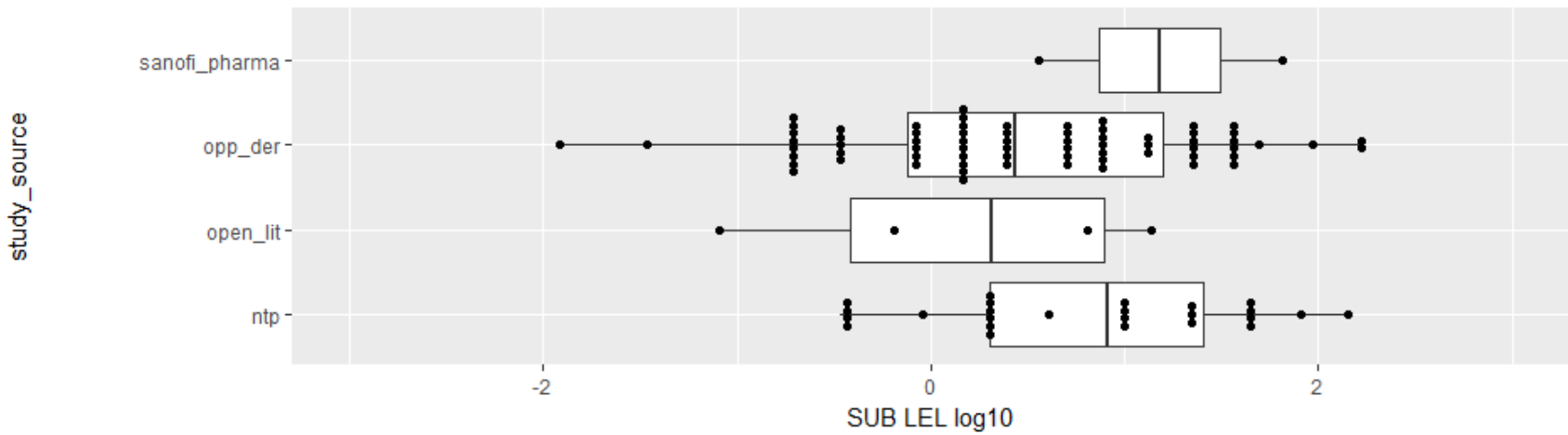
+ Many others

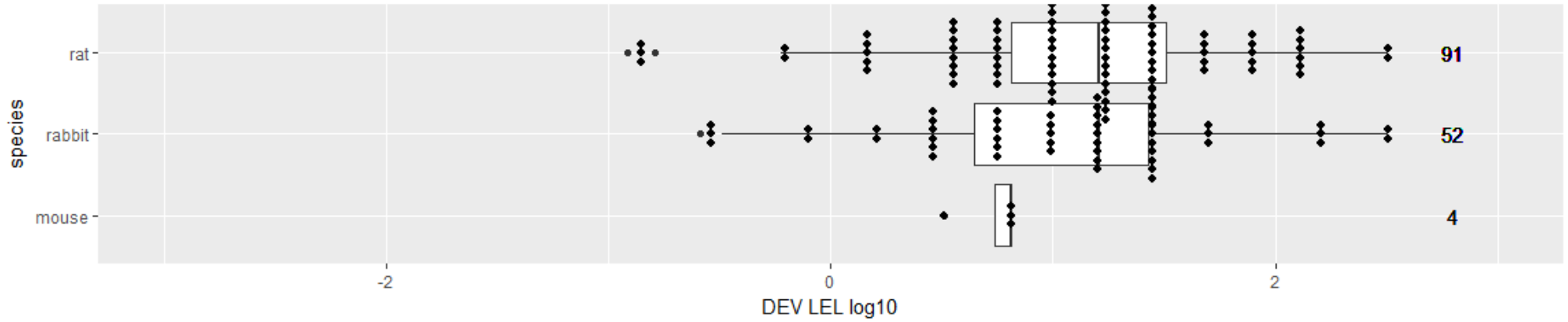
Questions?

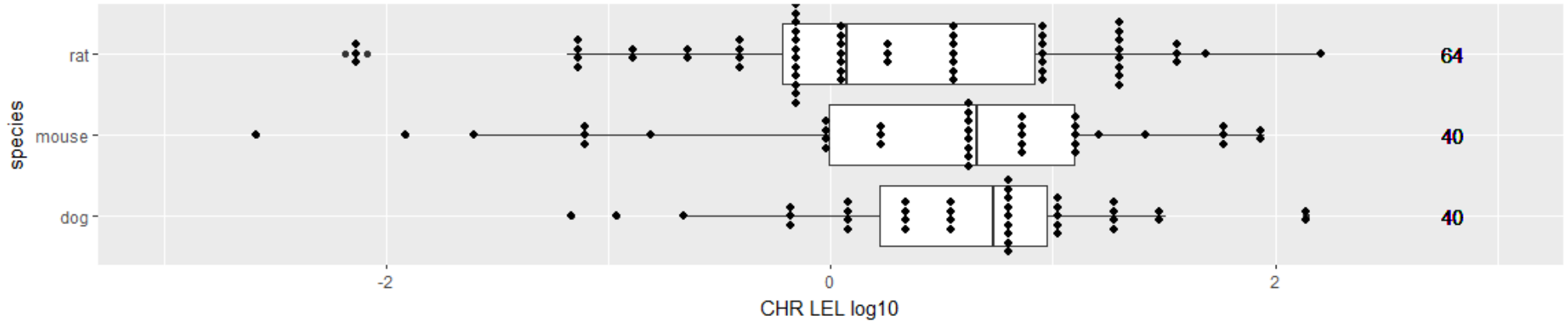
Appendix

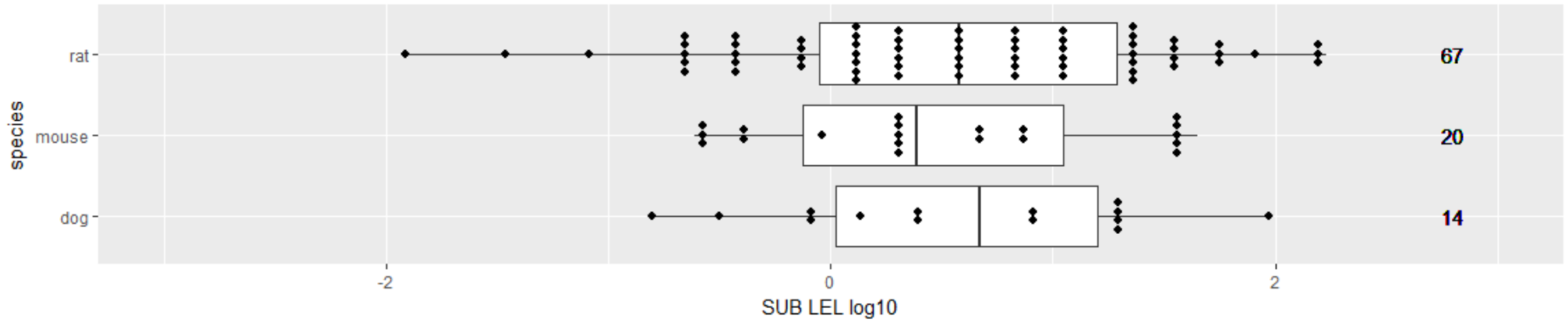

study_source
unpublished_submission
unknown
sanofi_pharma
opp_der
open_lit
ntp
-2
0
2
DEV LEL log10


study_source
sanofi_pharma
opp_der
open_lit
ntp
gsk_pharma
-2
0
2
CHR LEL log10


study_source
sanofi_pharma
opp_der
open_lit
ntp
-2
0
2
SUB LEL log10





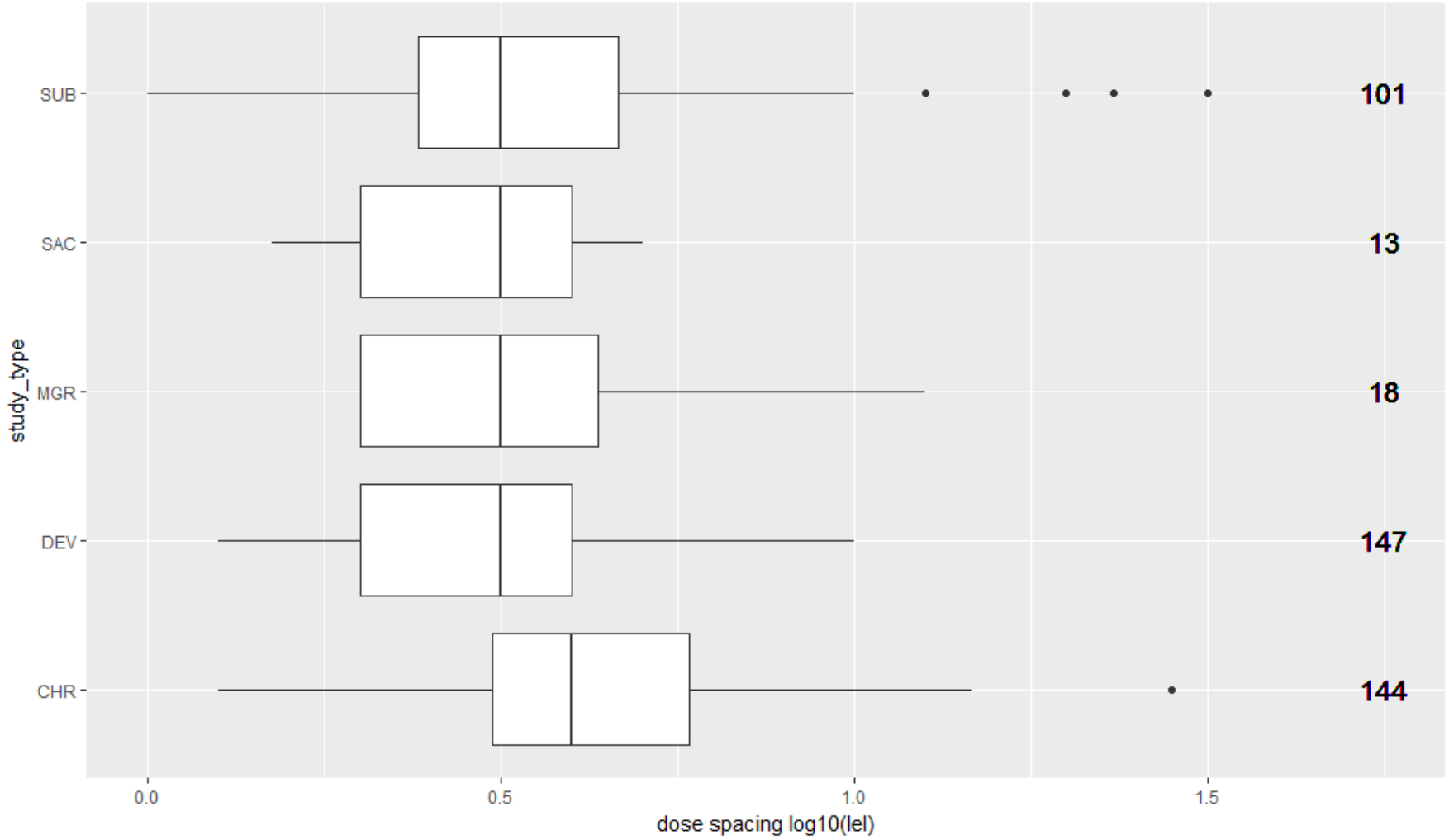


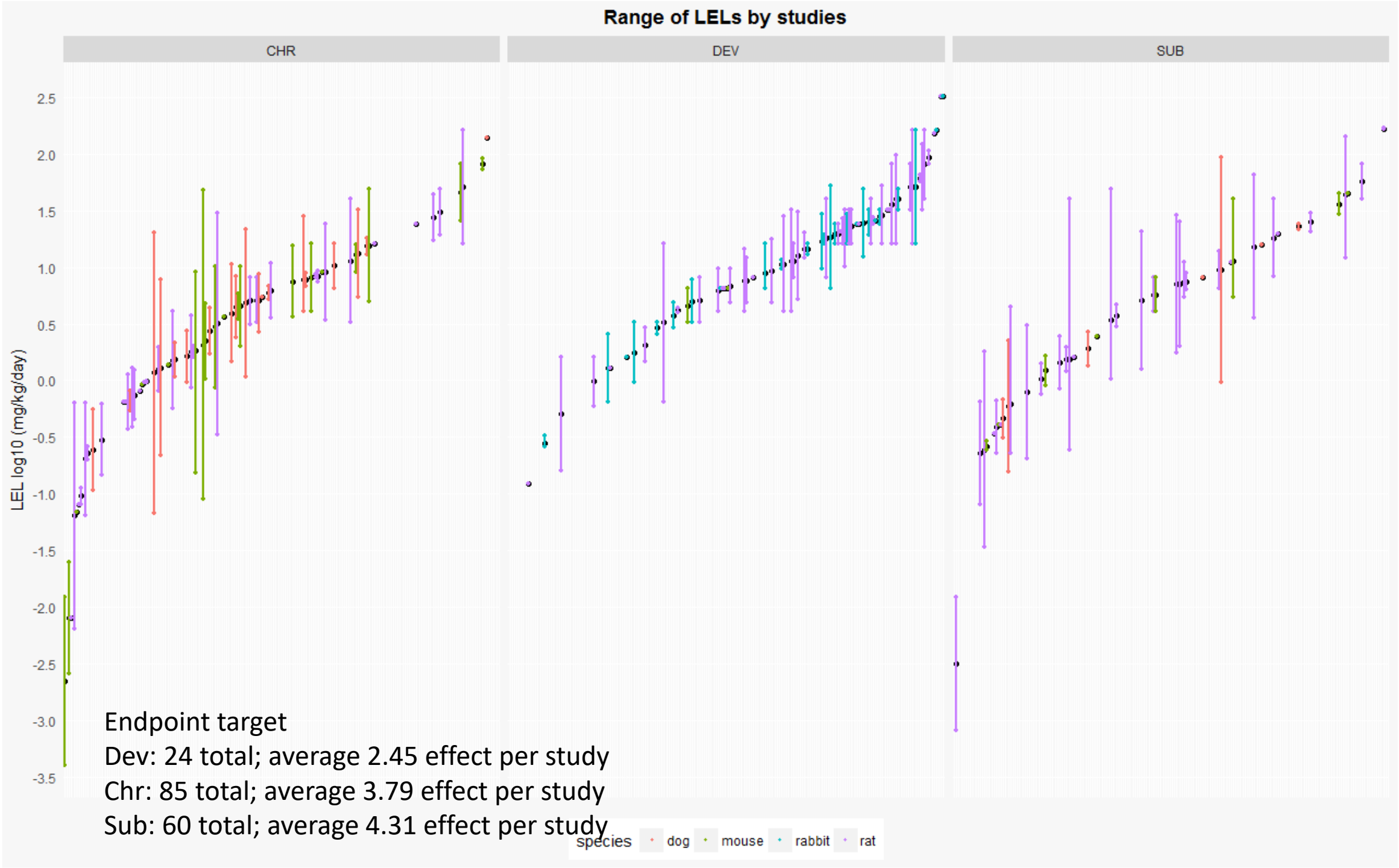
Comparison of LEL across species by study type

- All dose are allometrically scaled
- Number line indicates the number of studies in each group

Comparison of the distribution of dose spacing across studies.

Number line indicates number of studies in each group





Endpoint target

Dev: 24 total; average 2.45 effect per study

Chr: 85 total; average 3.79 effect per study

Sub: 60 total; average 4.31 effect per study