AIR® Advancing Evidence. Improving Lives.

MOSAIC Methods of Synthesis and Integration Center

Exploring Effect Heterogeneity Using Meta-Regression and Machine Learning

Using PreK-12 Mathematics Interventions as a Case Example

Ryan Williams, Martyna Citkowicz, David Miller, Jim Lindsay, and Kirk Walters

AERA SRMA SIG | February 2022

AIR®
Advancing Evidence. Improving Lives.

MOSAIC
Methods of Synthesis and Integration Center

This work was supported by the Institute of Education Sciences

R305A170146

Exploring Heterogeneity in Mathematics Intervention Effects

- Our goal was to examine:
  - How heterogeneous are mathematics intervention effects?
  - What factors best explain heterogeneity?
  - How much heterogeneity can be systematically explained by observable features of studies?

Williams et al. (2022)

Photo by Gayatri Malhotra on Unsplash

Number of Studies at Each Systematic Review Stage

Abstracts screened: 9,384

Full text screened: 2,462

In meta-analysis: 191

Meta-Analysis Using the MUTOS Framework

- **M**ethods, **U**nits, **T**reatments, **O**utcomes, **S**ettings
- Uses model building approach
  - Ran mixed effects models, controlling for methods moderators in all models.
  - Ran separate models for each vector of characteristics in the MUTOS framework.
  - Estimated combined model with moderators that had $p < 0.10$.
- Uses linear meta-regression

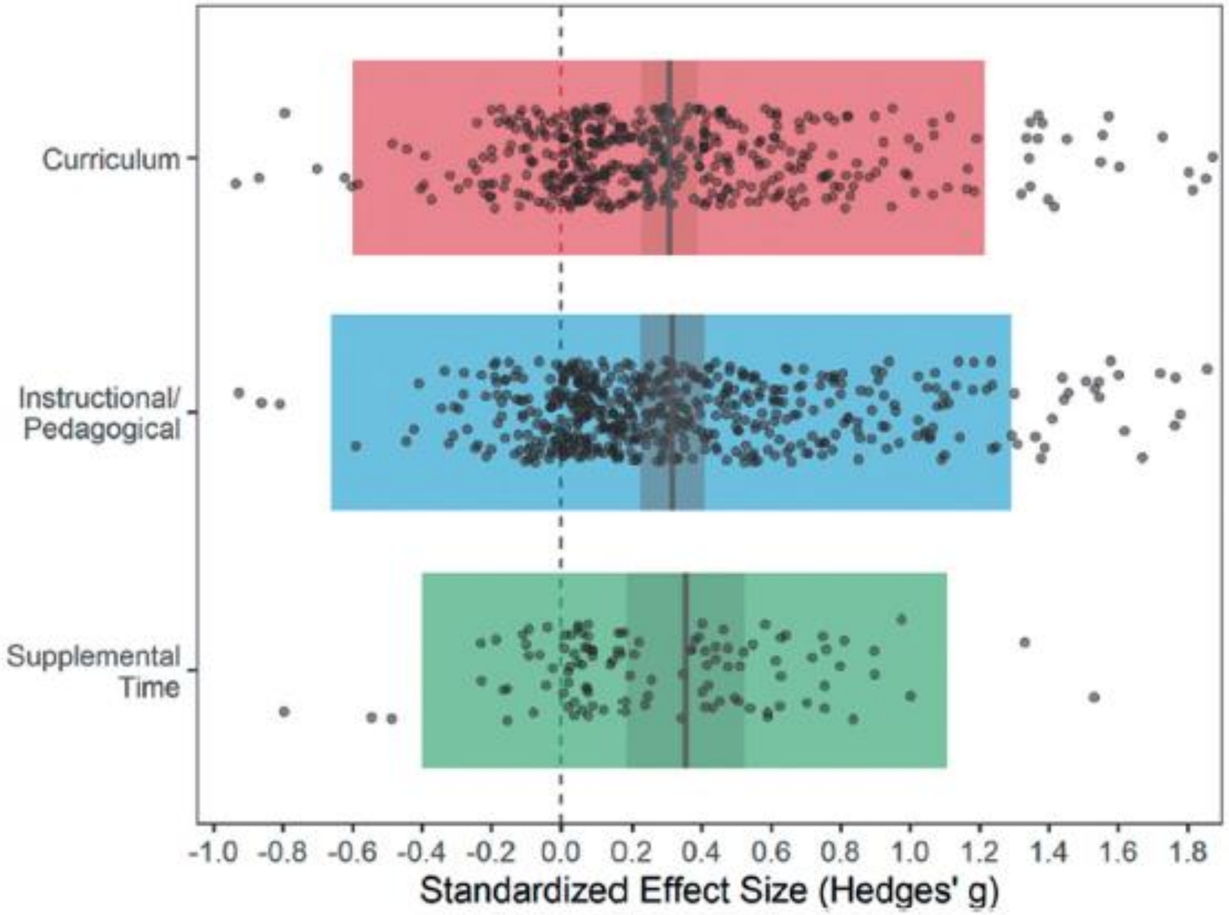
Aloe and Becker (2009); Becker (2017); Cronbach (1982)

AIR® Advancing Evidence. Improving Lives.

MOSAIC Methods of Synthesis and Integration Center

What did we initially find?

How heterogeneous are mathematics intervention effects?



$$\bar{g} = .31, 95\%\, CI[.26, .37];$$
$$\hat{\tau} = .47, 95\%\, CI[.37, .58]$$

Williams et al. (2022)

Which specific factors best explain heterogeneity?

Table 10. Moderator results from mixed-effects meta-regression model.

Moderator	Mean or b*	SE	m	k	df	p
UTOS moderators						
Intervention type					190.00[b]	0.04
Curriculum	0.34	0.04	83	443	66.49	
Pedagogical/Instructional	0.27	0.04	85	553	68.52	
Supplemental	0.53	0.10	24	113	24.07	
Intervention delivery					190.00[b]	0.01
Teacher	0.37	0.05	110	608	64.03	
Technology	0.12	0.08	65	375	30.76	
Interventionist	0.39	0.06	52	380	48.11	
Publication decade*	−0.14	0.06	–	–	36.23	0.04
Methods Moderators						
Outcome type					39.80	<0.01
Researcher-generated measure	0.45	0.05	123	639	101.41	
Standardized achievement measure	0.15	0.05	107	470	76.08	
Publication status					87.69	0.36
Unpublished	0.29	0.05	74	345	53.33	
Published	0.34	0.04	117	764	99.37	
NCEE trial					17.72	0.28
Not an NCEE trial	0.33	0.03	177	1048	124.10	
NCEE trial	0.25	0.07	14	61	15.78	
Assumed correlation					2.84	0.27
Not assumed	0.33	0.03	189	1099	128.34	
Assumed[a]	0.18	0.11	5	10	2.81	
Attrition and baseline equivalence					190.00[b]	0.75
Low-attrition RCT	0.31	0.04	45	255	37.62	
Baseline equivalence satisfied	0.35	0.04	68	292	50.66	
Neither standard satisfied	0.32	0.04	128	562	94.22	
Level of random assignment					190.00[b]	0.74
Student	0.32	0.05	93	547	59.51	
Teacher	0.31	0.05	67	379	62.59	
School	0.38	0.08	33	183	32.05	

Williams et al. (2022)

Which specific factors best explain heterogeneity?

Table 10. Moderator results from mixed-effects meta-regression model.

Moderator	Mean or b^*	SE	m	k	df	p
UTOS moderators						
Intervention type					190.00[b]	0.04
Curriculum	0.34	0.04	83	443	66.49	
Pedagogical/Instructional	0.27	0.04	85	553	68.52	
Supplemental	0.53	0.10	24	113	24.07	
Intervention delivery					190.00[b]	0.01
Teacher	0.37	0.05	110	608	64.03	
Technology	0.12	0.08	65	375	30.76	
Interventionist	0.39	0.06	52	380	48.11	
Publication decade*	−0.14	0.06	–	–	36.23	0.04
Methods Moderators						
Outcome type					39.80	<0.01
Researcher-generated measure	0.45	0.05	123	639	101.41	
Standardized achievement measure	0.15	0.05	107	470	76.08	
Publication status					87.69	0.36
Unpublished	0.29	0.05	74	345	53.33	
Published	0.34	0.04	117	764	99.37	
NCEE trial					17.72	0.28
Not an NCEE trial	0.33	0.03	177	1048	124.10	
NCEE trial	0.25	0.07	14	61	15.78	
Assumed correlation					2.84	0.27
Not assumed	0.33	0.03	189	1099	128.34	
Assumed[a]	0.18	0.11	5	10	2.81	
Attrition and baseline equivalence					190.00[b]	0.75
Low-attrition RCT	0.31	0.04	45	255	37.62	
Baseline equivalence satisfied	0.35	0.04	68	292	50.66	
Neither standard satisfied	0.32	0.04	128	562	94.22	
Level of random assignment					190.00[b]	0.74
Student	0.32	0.05	93	547	59.51	
Teacher	0.31	0.05	67	379	62.59	
School	0.38	0.08	33	183	32.05	

Williams et al. (2022)

AIR® MOSAIC Methods of Synthesis and Integration Center

How heterogeneous are mathematics intervention effects?

Table 10. Moderator results from mixed-effects meta-regression model.

Moderator	Mean or *b**	*SE*	*m*	*k*	*df*	*p*
UTOS moderators						
Intervention type					190.00[b]	0.04
Curriculum	0.34	0.04	83	443	66.49	
Pedagogical/Instructional	0.27	0.04	85	553	68.52	
Supplemental	0.53	0.10	24	113	24.07	
Intervention delivery					190.00[b]	0.01
Teacher	0.37	0.05	110	608	64.03	
Technology	0.12	0.08	65	375	30.76	
Interventionist	0.39	0.06	52	380	48.11	
Publication decade*	−0.14	0.06	–	–	36.23	0.04
Methods Moderators						
Outcome type					39.80	<0.01
Researcher-generated measure	0.45	0.05	123	639	101.41	
Standardized achievement measure	0.15	0.05	107	470	76.08	
Publication status					87.69	0.36
Unpublished	0.29	0.05	74	345	53.33	
Published	0.34	0.04	117	764	99.37	
NCEE trial					17.72	0.28
Not an NCEE trial	0.33	0.03	177	1048	124.10	
NCEE trial	0.25	0.07	14	61	15.78	
Assumed correlation					2.84	0.27
Not assumed	0.33	0.03	189	1099	128.34	
Assumed[a]	0.18	0.11	5	10	2.81	
Attrition and baseline equivalence					190.00[b]	0.75
Low-attrition RCT	0.31	0.04	45	255	37.62	
Baseline equivalence satisfied	0.35	0.04	68	292	50.66	
Neither standard satisfied	0.32	0.04	128	562	94.22	
Level of random assignment					190.00[b]	0.74
Student	0.32	0.05	93	547	59.51	
Teacher	0.31	0.05	67	379	62.59	
School	0.38	0.08	33	183	32.05	

About 10%...

(sad trombone)

Williams et al. (2022)

What are we missing?

Let’s give this job to the machines!

- Data driven approach to examine unexpected or unanticipated relationships
- Recent advances in machine learning algorithms, with specific applications to meta-analysis:
  - classification/regression tree (`metacart`)
  - random forest models (`metaforest`)

Breinman (2001); van Lissa (2017, 2020)

Photo by Andrea De Santis on Unsplash

Let's give this job to the machines!

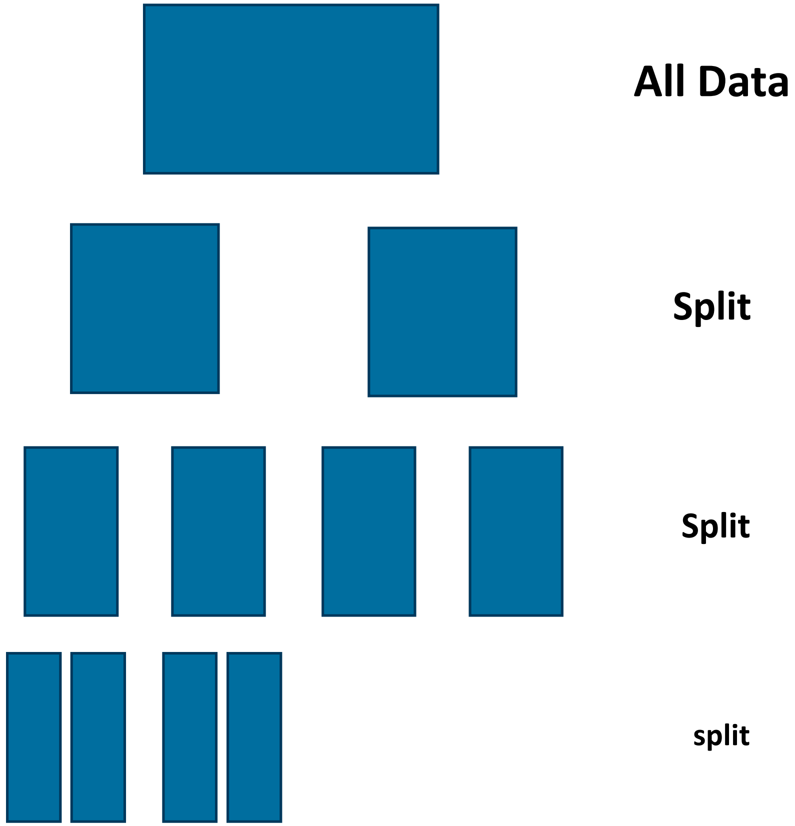
- Data driven approach to examine unexpected or unanticipated relationships
- Recent advances in machine learning algorithms, with specific applications to meta-analysis:
  - classification/regression tree (`metacart`)
  - random forest models (`metaforest`)

Breinman (2001); van Lissa (2017, 2020)

Photo by Andrea De Santis on Unsplash

Let’s give this job to the machines!

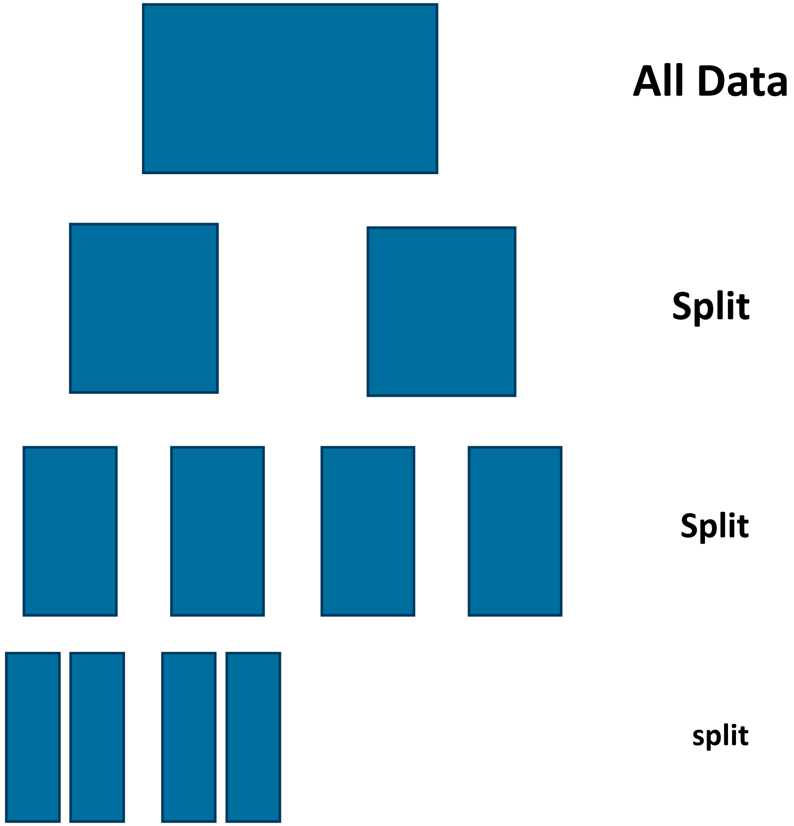
- Random forests for meta-analysis
  1. Identify moderators (from candidates) to split the data into maximally homogeneous groups
  2. Continue recursively until a stopping criterion is satisfied (e.g., number of cases in each final split – terminal nodes)
  3. Repeat steps 1 & 2 many, many times over bootstrapped samples of the original data --> making use of substantive and idiosyncratic variation
  4. The result is a set of pooled effect size predictions that that capture complex interactions, linear, and nonlinear relationships



AIR® MOSAIC Methods of Synthesis and Integration Center

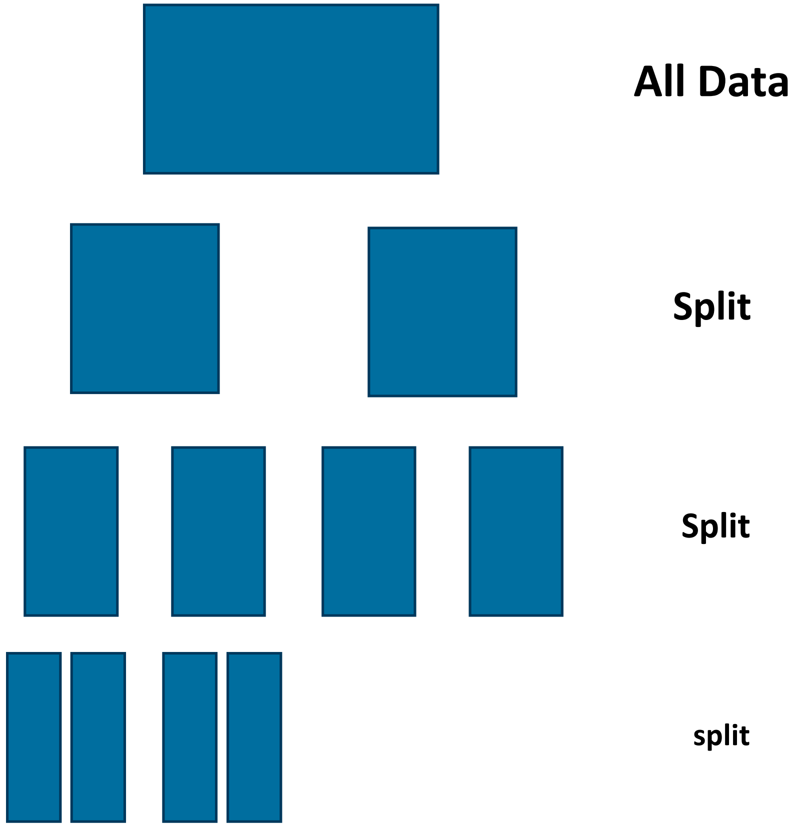
Let’s give this job to the machines!

- Why random forests?
  1. They tend to be robust to issues of overfitting
  2. Nonparametric technique that are especially useful for capturing nonlinear relationships and complex interactions
  3. They provide diagnostic information on moderator importance



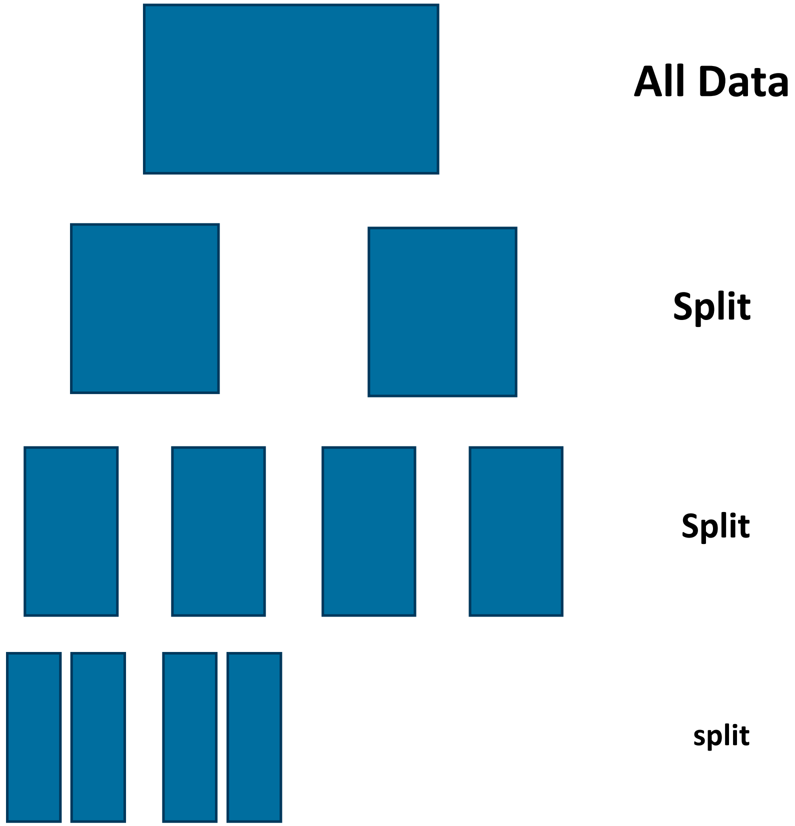
So, what did we do?

- **Step 1**
  - Identify an initial set of candidate moderators to inform the random forest model
    - These were the initial 43 MUTOS-aligned moderators we used in our planned meta-regression



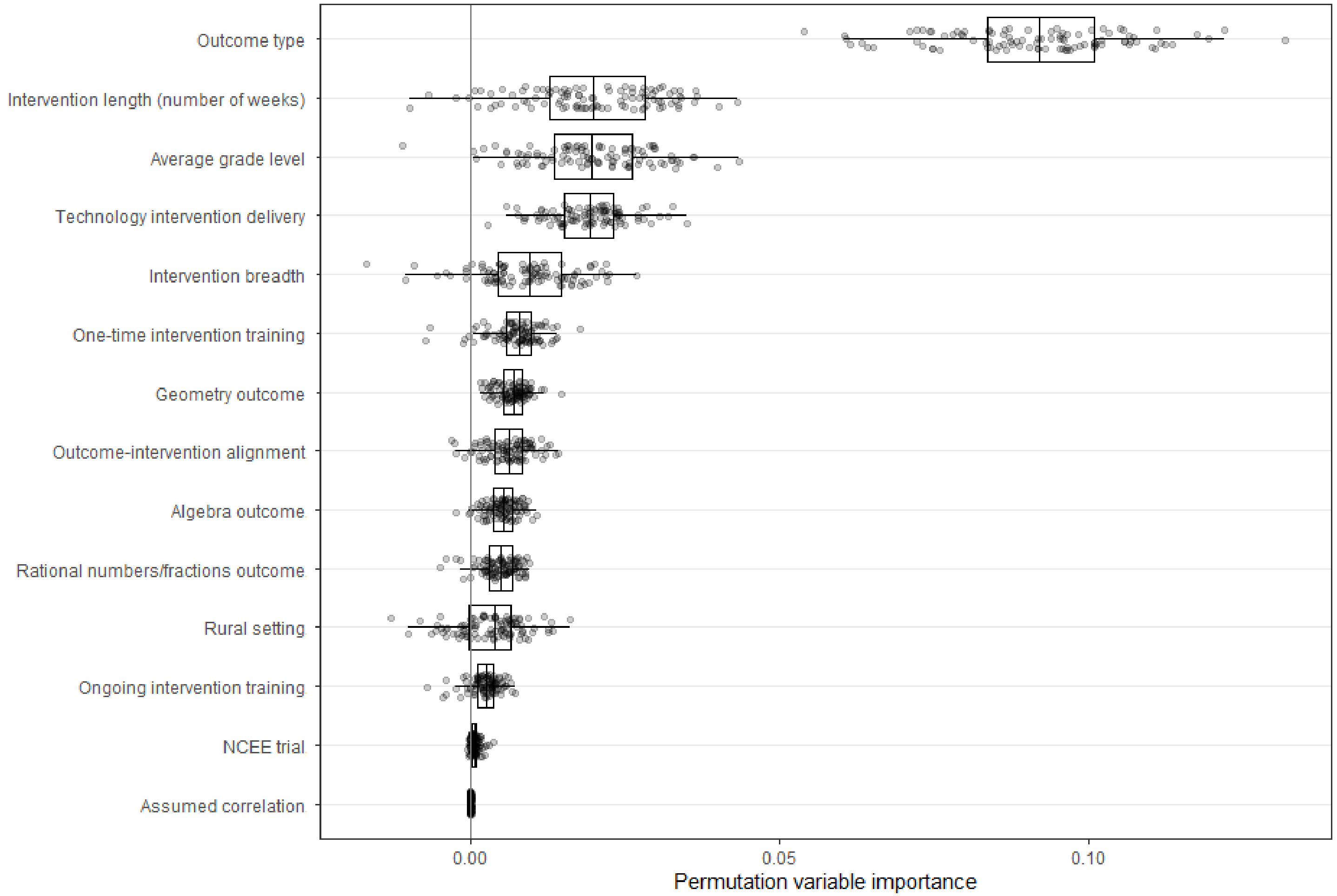
So, what did we do?

- **Step 2**
  - Figure out which moderators are most important



AIR® MOSAIC Methods of Synthesis and Integration Center

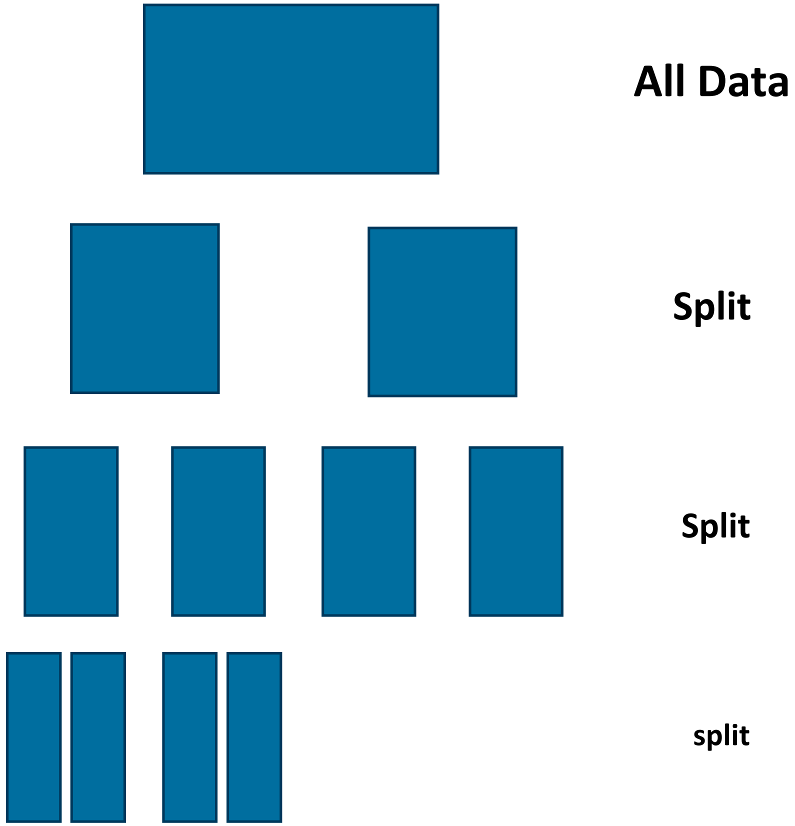
What Moderators Were Most Important?



AIR MOSAIC Methods of Synthesis and Integration Center

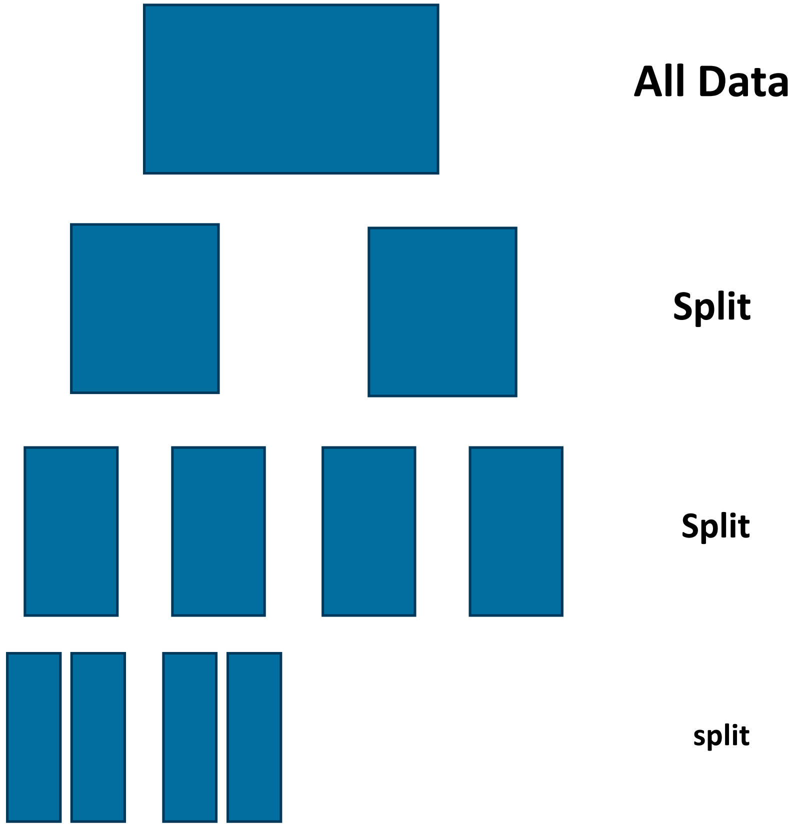
So, what did we do?

- **Step 3**
  - ”Fine tune” the predictions from the moderators that advance.
    - » 13 of the original 43 satisfied our criteria for variable importance



So, what did we do?

- **Step 4**
  - Cross-validate the predictions from the random forest models using leave-1-out
  - Use the cross-validated predictions as “super moderator” variables in a typical meta-regression framework to see how much heterogeneity is explained by the random forest-generated predictions
  - Do the same for the the MUTOS meta-regression to create a fair comparison



AIR® MOSAIC Methods of Synthesis and Integration Center

How Much Heterogeneity Can Be Explained?

Approach	R^2
MUTOS	8%
Random forest	13%

- Random forest model overall explains more heterogeneity.
- Though both approaches leave a lot of heterogeneity unexplained (common finding in large-scale meta-analyses).
- *Note:* These R^2 values were based on leave-1-out cross-validation. The predictions for a study's effect sizes were based on models that did not include the study's effect sizes in the model fitting.

What Moderators Were Most Important?

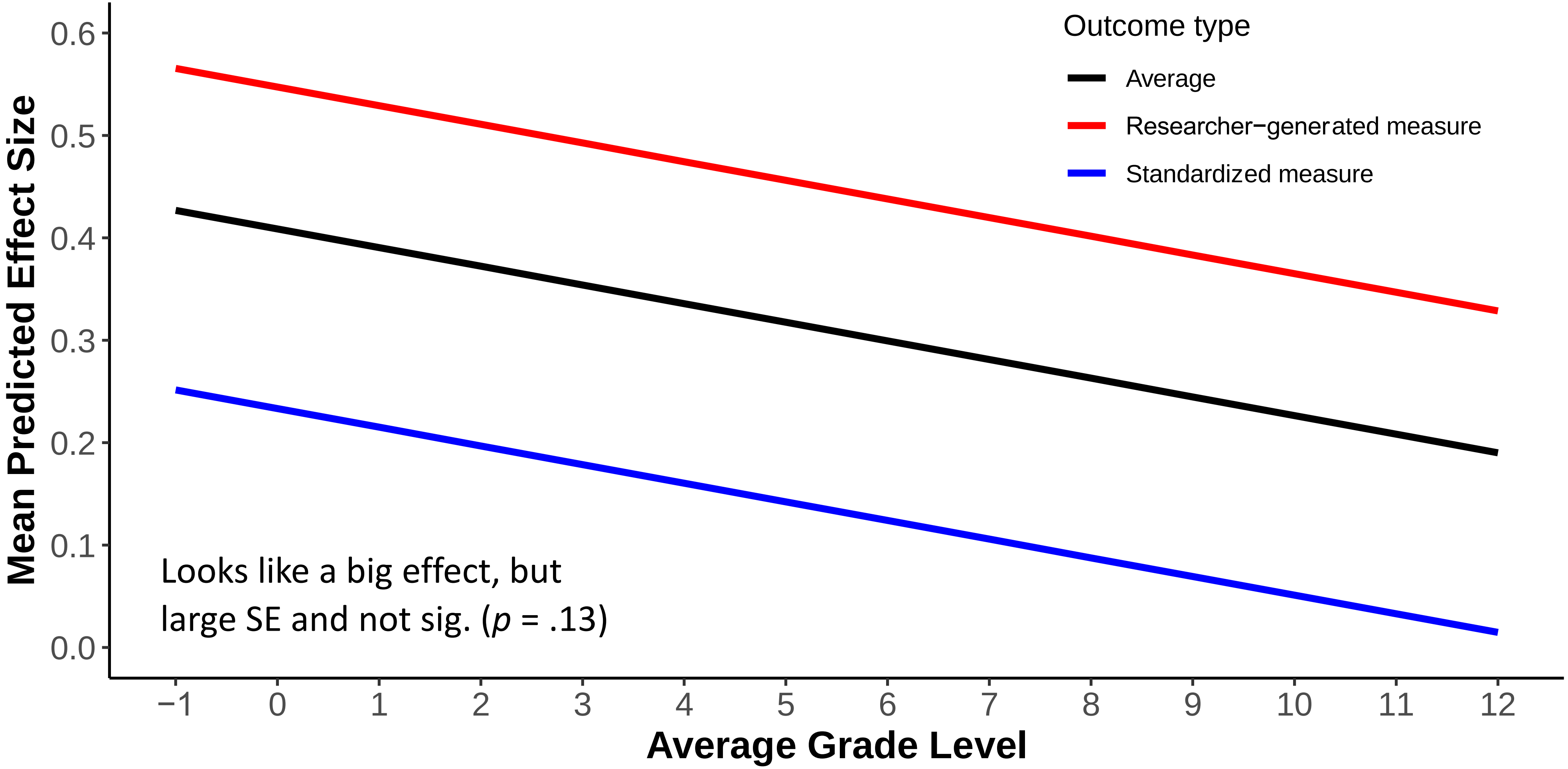
Moderator	MUTOS	Random Forest
Researcher-generated vs. standardized measure	✓	✓
Tech. delivery vs. other intervention delivery	✓	✓
Average student grade level	✗	✓
Intervention length (number of weeks)	✗	✓
Supplemental time vs. other intervention types	✓	✗
Publication year	✓	✗

✓ Statistically significant/ranked as important for improving effect size predictions

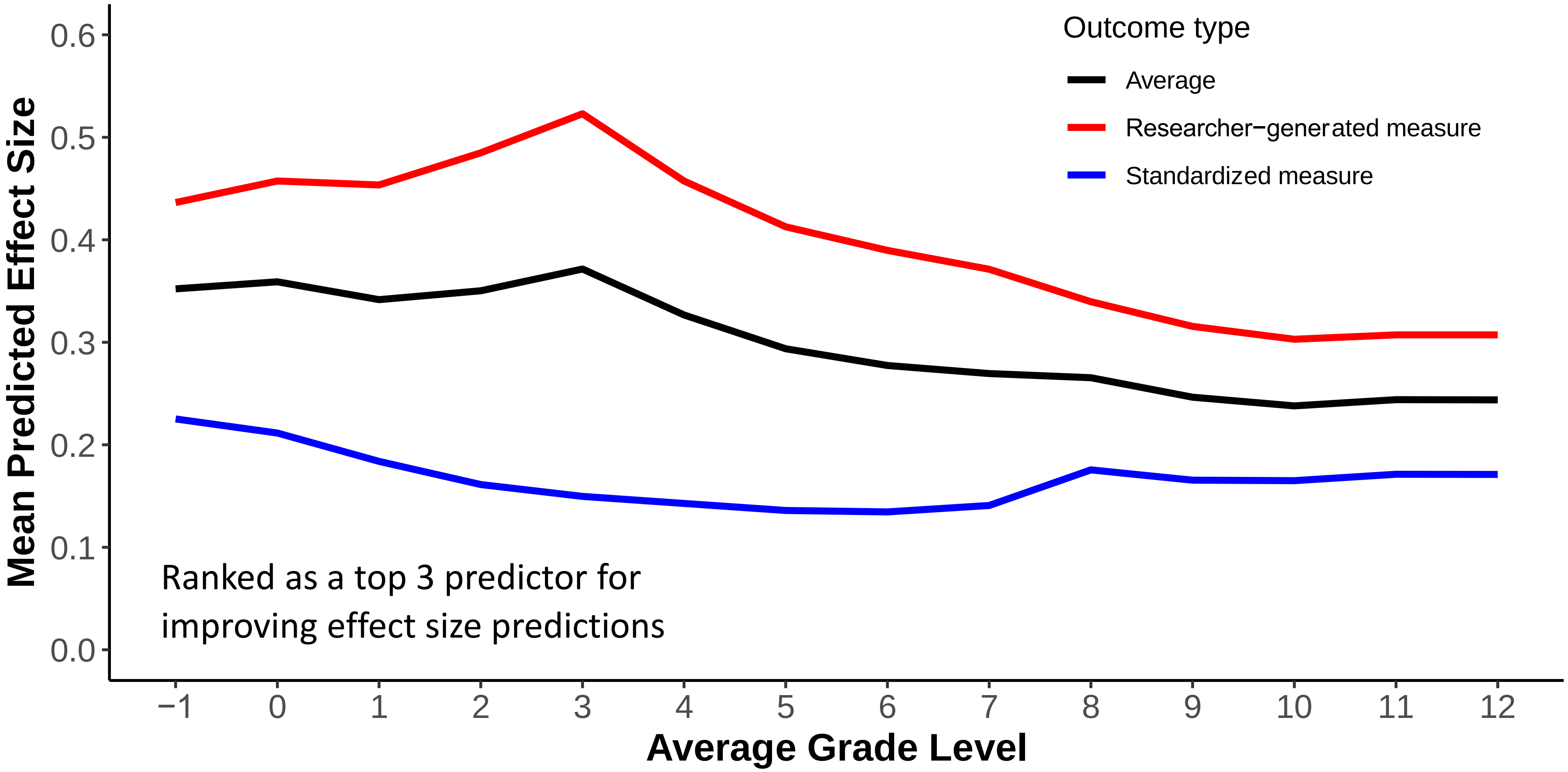
✗ Not significant/not retained in model building process

AIR® MOSAIC Methods of Synthesis and Integration Center

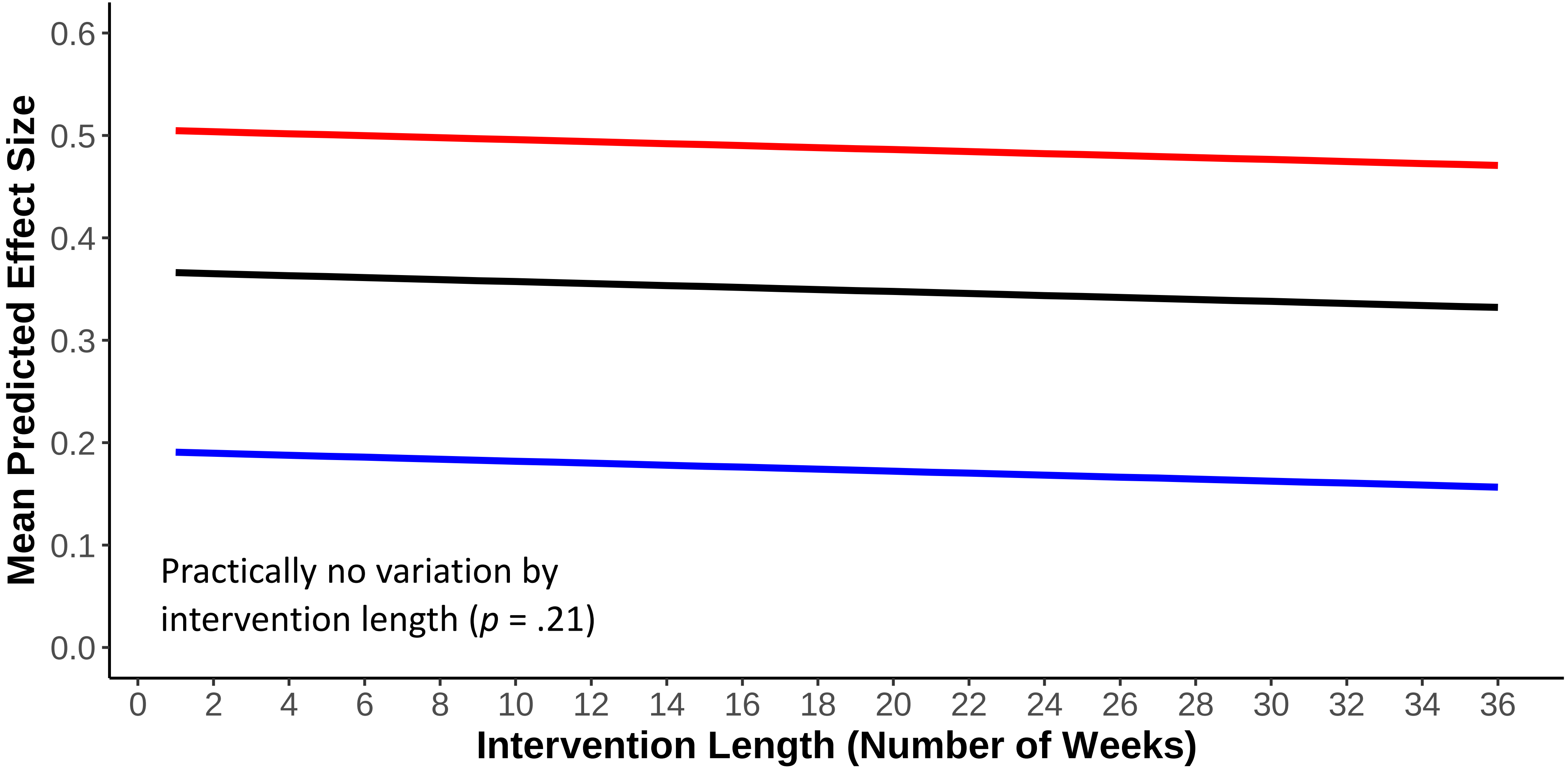
Grade Level: MUTOS Meta-Regression



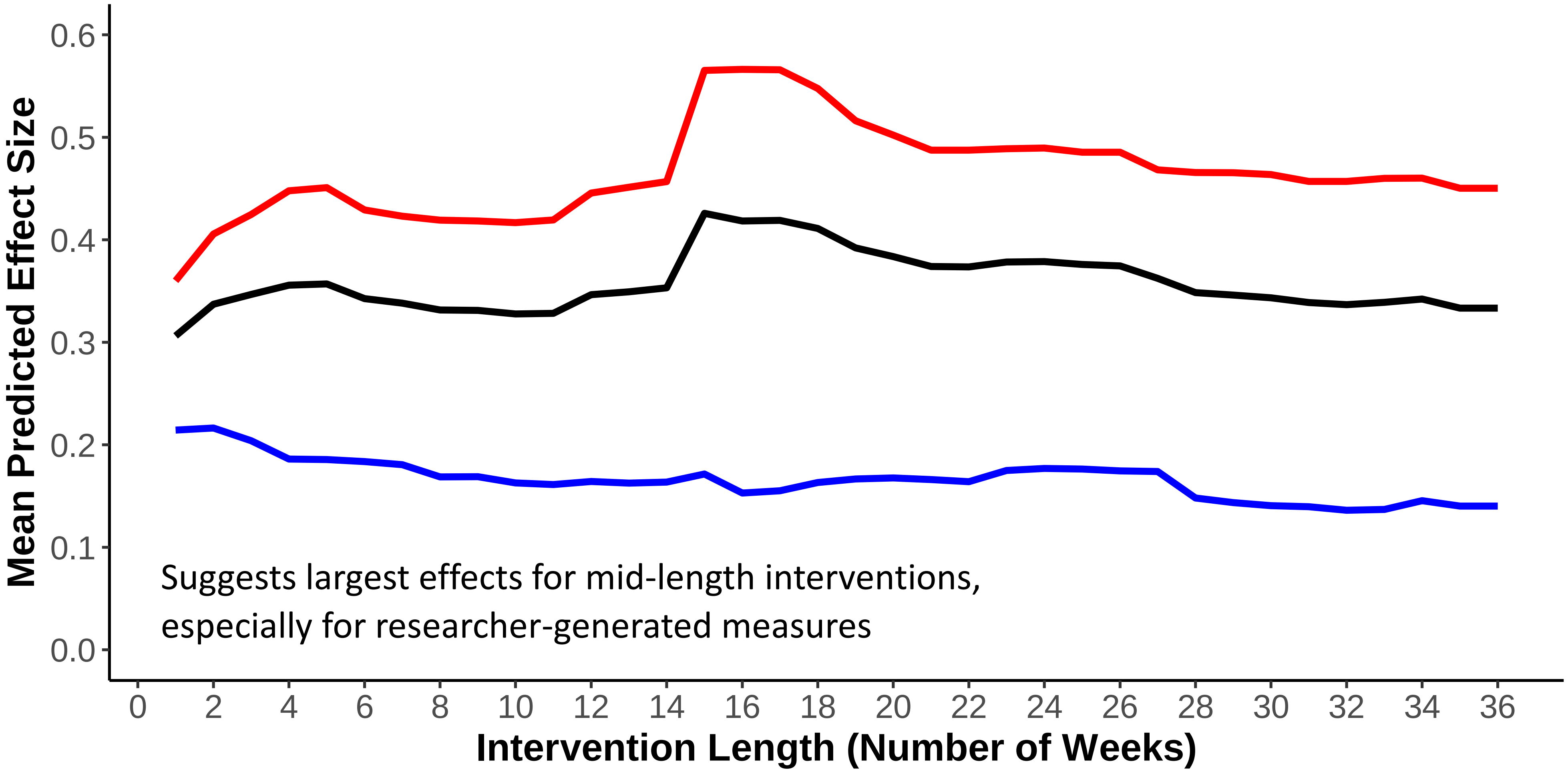
Grade Level: Random Forest



Intervention Length: MUTOS Meta-Regression



Intervention Length: Random Forest



Strategies and considerations…

- I’m still a pretty firm believer in using theory, *as much as possible*, to guide model building.
  - BUT, meta-regression is still pretty much a small sample technique and moderators can quickly overwhelm the number of studies and effect sizes in hand.

Strategies and considerations…

- I’m still a pretty firm believer in using theory, *as much as possible*, to guide model building.
  - BUT, meta-regression is still pretty much a small sample technique and moderators can quickly overwhelm the number of studies and effect sizes in hand
- Data-driven approaches like random-forest models can help us better understand the relative explanatory value of our best theory-driven approaches

Strategies and considerations…

- I’m still a pretty firm believer in using theory, *as much as possible*, to guide model building.
  - BUT, meta-regression is still pretty much a small sample technique and moderators can quickly overwhelm the number of studies and effect sizes in hand
- Data-driven approaches like random-forest models can help us better understand the relative explanatory value of our best theory-driven approaches
- Data-driven approaches come with their own set of critiques, which shouldn’t be ignored

Strategies and considerations…

- I'm still a pretty firm believer in using theory, *as much as possible*, to guide model building.
  - BUT, meta-regression is still pretty much a small sample technique and moderators can quickly overwhelm the number of studies and effect sizes in hand
- Data-driven approaches like random-forest models can help us better understand the relative explanatory value of our best theory-driven approaches
- Data-driven approaches come with their own set of critiques, which shouldn't be ignored
  - BUT, the bar is pretty low right now for advances in applied meta-regression
- This is a VERY active area of research in research synthesis methods and it's much needed!
  - And we'll probably look back at what we did here in 5 years and wonder, "what in the world were they thinking…"

DISCUSS!

- Some resources:
  - OSF code, data, etc: https://osf.io/f9gud/files/?view_only=c97ba1316ff44606b8954d686e4d2d8b
  - Shiny app: https://airshinyapps.shinyapps.io/math_meta_database/
  - Paper: https://www.tandfonline.com/doi/abs/10.1080/19345747.2021.2009072
- Suggested readings:

van Lissa, C. J. (2017). MetaForest: Exploring heterogeneity in meta-analysis using random forests. https://doi.org/10.31234/osf.io/myg6s

van Lissa, C. J. (2020). Small sample meta-analyses: Exploring heterogeneity using MetaForest. In R. Van De Schoot & M. Miočević (Eds.), Small sample size solutions (open access): A guide for applied researchers and practitioners. CRC Press. https://www.crcpress.com/Small-Sample-Size-Solutions-Open-Access-A-Guide-for-Applied-Researchers/Schoot-Miocevic/p/book/9780367222222

Ryan Williams

Principal Researcher
rwilliams@air.org