Education
Centre for Education
Statistics & Evaluation

Big data, big possibilities, big challenges: Lessons from using quasi-experimental designs in evaluation of educational reforms

Ben Barnes, Director, Evaluation Unit, CESE
Ian Watkins, Principal Data Analyst, Evaluation Unit, CESE
Duncan Rintoul, Manager, Evaluation Capacity Building, Evaluation Unit, CESE

DATE: 19/09/2018

About CESE

Provide data analysis, information and evaluation that improve effectiveness, efficiency and accountability

Collect essential education data and provide a one-stop shop for information needs

Build capacity across the whole education sector so that everyone can make better use of data and evidence

About CESE

Provide data analysis, information and evaluation that improve effectiveness, efficiency and accountability

Collect essential education data and provide a one-stop shop for information needs

Build capacity across the whole education sector so that everyone can make better use of data and evidence

Overview

- Theory of causal attribution
  - Randomised controlled trials just a special case for causal inference
- Planning, conducting and presenting a quantitative evaluation
- Local Schools, Local Decisions (LSLD) evaluation case example
  - Using regression methods to test predictions based on program logic
- Reading Recovery (RR) evaluation case example
  - Using regression and propensity score methods to estimate average treatment effects
- Literacy and Numeracy Action Plan (LNAP) Phase 2 evaluation case example
  - Using propensity score methods to influence research design
- Final thoughts

Overview

- Theory of causal attribution
    - Randomised controlled trials just a special case for causal inference
- Planning, conducting and presenting a quantitative evaluation
- Local Schools, Local Decisions (LSLD) evaluation case example
    - Using regression methods to test predictions based on program logic
- Reading Recovery (RR) evaluation case example
    - Using regression and propensity score methods to estimate average treatment effects
- Literacy and Numeracy Action Plan (LNAP) Phase 2 evaluation case example
    - Using propensity score methods to influence research design
- Final thoughts

Theory of causal attribution

- Rubin Causal Model
  - Unified perspective of causal inference
- Causal effects defined by within-unit differences
  - Potential outcomes defined as outcome value after application of treatment **and** non-application of treatment
  - Fundamental problem of causal inference
- Randomised trials just a special case
  - Assignment mechanism allows within-unit differences to be easily inferred from between-unit observations
  - Often not possible based on ethical and/or political grounds
- Causal effect estimation is possible with observational data, but not necessarily the goal of a quantitative outcome evaluation
  - Regression based methods
  - Propensity score methods

Theory of causal attribution

- Rubin Causal Model
  - Unified perspective of causal inference
- Causal effects defined by within-unit differences
  - Potential outcomes defined as outcome value after application of treatment **and** non-application of treatment
  - Fundamental problem of causal inference
- Randomised trials just a special case
  - Assignment mechanism allows within-unit differences to be easily inferred from between-unit observations
  - Often not possible based on ethical and/or political grounds
- Causal effect estimation is possible with observational data, but not necessarily the goal of a quantitative outcome evaluation
  - Regression based methods
  - Propensity score methods

Planning, conducting and presenting a quantitative evaluation

- Influencing the research design stage
- Scoping and operationalising research questions
  - Features of the policy/initiative
  - Data availability
  - Mapping data to appropriate outcomes
  - Formalising analysis approach
- Conducting the analysis
  - Skills assessment
- Presenting the results
  - Graphical and verbal methods
  - Quantify uncertainty
- Influencing policy and decision-making

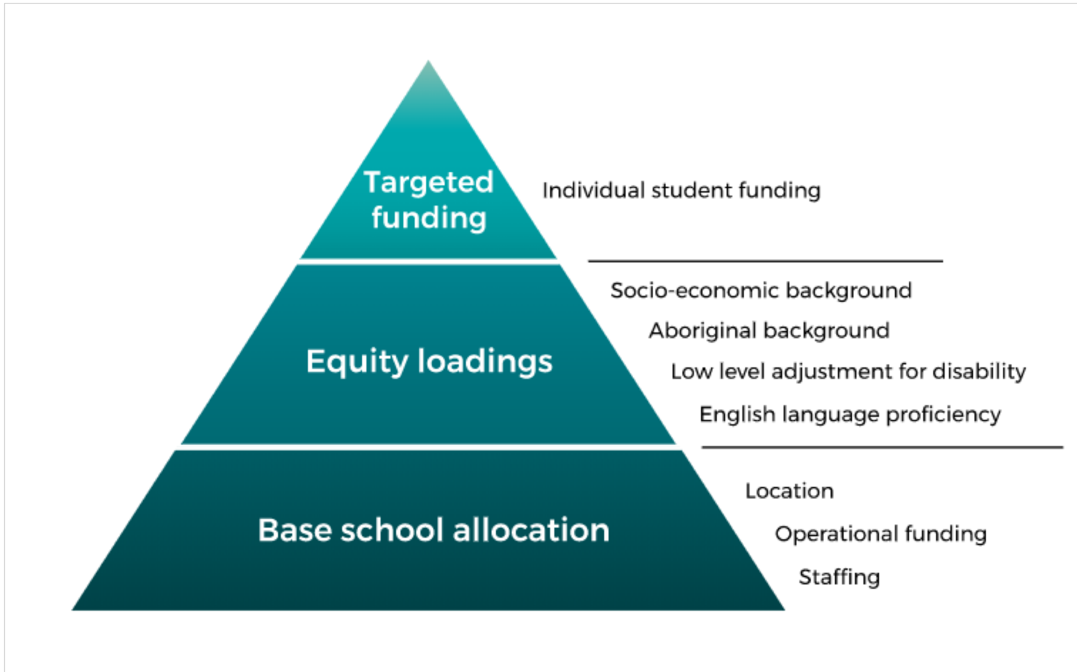
Planning, conducting and presenting a quantitative evaluation

- Influencing the research design stage
- Scoping and operationalising research questions
    - Features of the policy/initiative
    - Data availability
    - Mapping data to appropriate outcomes
    - Formalising analysis approach
- Conducting the analysis
    - Skills assessment
- Presenting the results
    - Graphical and verbal methods
    - Quantify uncertainty
- Influencing policy and decision-making

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

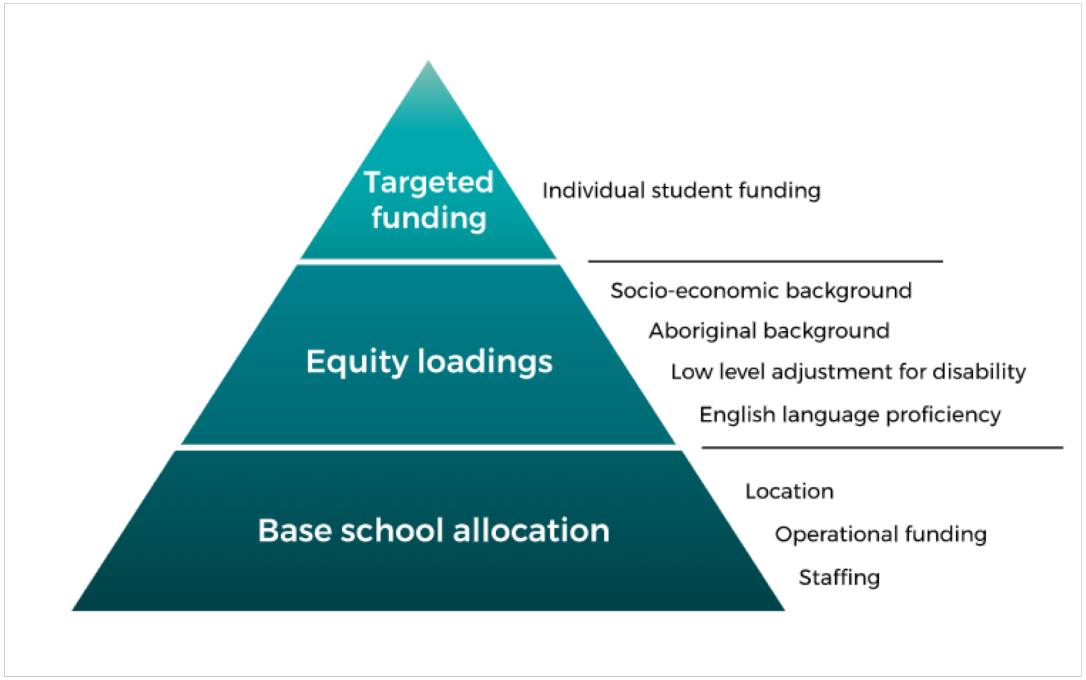
- Progressively implemented since 2012, LSLD includes 37 different initiatives
  - Almost all initiatives in place by the end of 2016
- Includes a phased transition to a needs-based funding model for NSW schools
  - Resource Allocation Model (RAM) - $667M



Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Progressively implemented since 2012, LSLD includes 37 different initiatives
    - Almost all initiatives in place by the end of 2016
- Includes a phased transition to a needs-based funding model for NSW schools
    - Resource Allocation Model (RAM) - $667M



Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Progressively implemented since 2012, LSLD includes 37 different initiatives
  - Almost all initiatives in place by the end of 2016
- Includes a phased transition to a needs-based funding model for NSW schools
  - Resource Allocation Model (RAM) - $667M

What has been the impact of LSLD and RAM funding on school and student outcomes?

- Could not identify counterfactual potential outcomes due to the systemic nature of the policy
- Formulate predictions based on program logic
  - We predicted that schools with higher levels-of-need would show greater changes across time than those with lower levels-of-need

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Progressively implemented since 2012, LSLD includes 37 different initiatives
  - Almost all initiatives in place by the end of 2016
- Includes a phased transition to a needs-based funding model for NSW schools
  - Resource Allocation Model (RAM) - $667M

What has been the impact of LSLD and RAM funding on school and student outcomes?

- Could not identify counterfactual potential outcomes due to the systemic nature of the policy
- Formulate predictions based on program logic
  - We predicted that schools with higher levels-of-need would show greater changes across time than those with lower levels-of-need

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative
  - Operationalise levels-of-need by combining the four RAM equity loadings

- Map data to outcomes
  - Use attendance data from 2011 through 2016
  - Use *Tell Them From Me* (TTFM) data from 2013 and 2016
  - Others…

- Formalise analysis approach
  - Use latent variable growth curve models to estimate school-specific changes in outcomes across time
  - Relate school-specific changes to levels-of-need

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative
  - Operationalise levels-of-need by combining the four RAM equity loadings

- Map data to outcomes
  - Use attendance data from 2011 through 2016
  - Use *Tell Them From Me* (TTFM) data from 2013 and 2016
  - Others…

- Formalise analysis approach
  - Use latent variable growth curve models to estimate school-specific changes in outcomes across time
  - Relate school-specific changes to levels-of-need

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative

$$\text{level of need}_s = \ln\left(\frac{\sum_{i=1}^{4} \text{RAM equity loading}_{si}}{\left(\left(\sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{spc}\right)/4\right)}\right)$$

- Map data to outcomes

$$n_{st} = \text{number of enrolment days}_{st} = \sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{stpc} \cdot \text{school days}_{stpc}$$

$$y_{st} = \text{number of attended days}_{st} = n_{jt} - \sum_{p=1}^{6}\sum_{c=1}^{4} \text{absent days}_{stpc}$$

- Formalise analysis approach

$$y_{st} \sim \text{Binomial}(n_{st}, \pi_{st}),$$

$$g(\pi_{st}) = (\beta_{00} + \beta_{01} \cdot \text{need GMC}_s + \beta_{02} \cdot \text{need GMC}_s^2 + e_{st} + u_{0s}) + (\beta_{10} + \beta_{11} \cdot \text{need GMC}_s + \beta_{12} \cdot \text{need GMC}_s^2 + u_{1s}) \cdot \text{time}_{st} + \beta_{20} \cdot \text{time}_{st}^2 + \beta_{30} \cdot \text{ICSEA CMC}_{st},$$

$$\text{for } t = 0,1,\ldots,T \text{ calendar years and } s = 1,2,\ldots,n_t \text{ schools, and } e_{st} \sim N(0,\sigma_e^2) \text{ and } \mathbf{u}_s \sim MN\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{u0}^2 & \sigma_{01}\\ \sigma_{10} & \sigma_{u1}^2\end{pmatrix}\right].$$

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative

$$\text{level of need}_s = \ln\left(\frac{\sum_{i=1}^{4} \text{RAM equity loading}_{si}}{\left(\left(\sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{spc}\right)/4\right)}\right)$$

- Map data to outcomes

$$n_{st} = \text{number of enrolment days}_{st} = \sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{stpc} \cdot \text{school days}_{stpc}$$

$$y_{st} = \text{number of attended days}_{st} = n_{jt} - \sum_{p=1}^{6}\sum_{c=1}^{4} \text{absent days}_{stpc}$$

- Formalise analysis approach

$$y_{st} \sim \text{Binomial}(n_{st}, \pi_{st}),$$

$$g(\pi_{st}) = (\beta_{00} + \beta_{01} \cdot \text{need GMC}_s + \beta_{02} \cdot \text{need GMC}_s^2 + e_{st} + u_{0s}) + (\beta_{10} + \beta_{11} \cdot \text{need GMC}_s + \beta_{12} \cdot \text{need GMC}_s^2 + u_{1s}) \cdot \text{time}_{st} +$$

$$\beta_{20} \cdot \text{time}_{st}^2 + \beta_{30} \cdot \text{ICSEA CMC}_{st},$$

$$\text{for } t = 0,1,\ldots,T \text{ calendar years and } s = 1,2,\ldots,n_t \text{ schools, and } e_{st} \sim N(0, \sigma_e^2) \text{ and } \mathbf{u}_s \sim MN\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{u0}^2 & \sigma_{01}\\ \sigma_{10} & \sigma_{u1}^2\end{pmatrix}\right].$$

Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

Table G3
Parameter estimates obtained from the final attendance model – primary schools

				Wald based 95% confidence intervals	
Model parameter	Point estimate	Standard error	*p* value	Lower limit	Upper limit
$\boldsymbol{\beta_{00}}$	2.69	0.01	<.005	2.67	2.70
$\boldsymbol{\beta_{01}}$	-0.20	0.01	<.005	-0.22	-0.19
$\boldsymbol{\beta_{02}}$	-0.03	0.01	<.005	-0.04	-0.01
$\boldsymbol{\beta_{10}}$	0.04	0.00	<.005	0.04	0.04
$\boldsymbol{\beta_{11}}$	-0.00	0.00	0.30	-0.00	0.00
$\boldsymbol{\beta_{12}}$	--	--	--	--	--
$\boldsymbol{\beta_{20}}$	-0.01	0.00	<.005	-0.01	-0.01
$\boldsymbol{\beta_{30}}$	0.00	0.00	<.005	0.00	0.00
$\boldsymbol{\sigma_{e}}$	0.11	--	--	--	--
$\boldsymbol{\sigma_{u0}}$	0.21	--	--	--	--
$\boldsymbol{\sigma_{u1}}$	0.03	--	--	--	--
$\mathbf{Corr}(\boldsymbol{\sigma_{u0}}, \boldsymbol{\sigma_{u1}})$	-0.54	--	--	--	--

Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

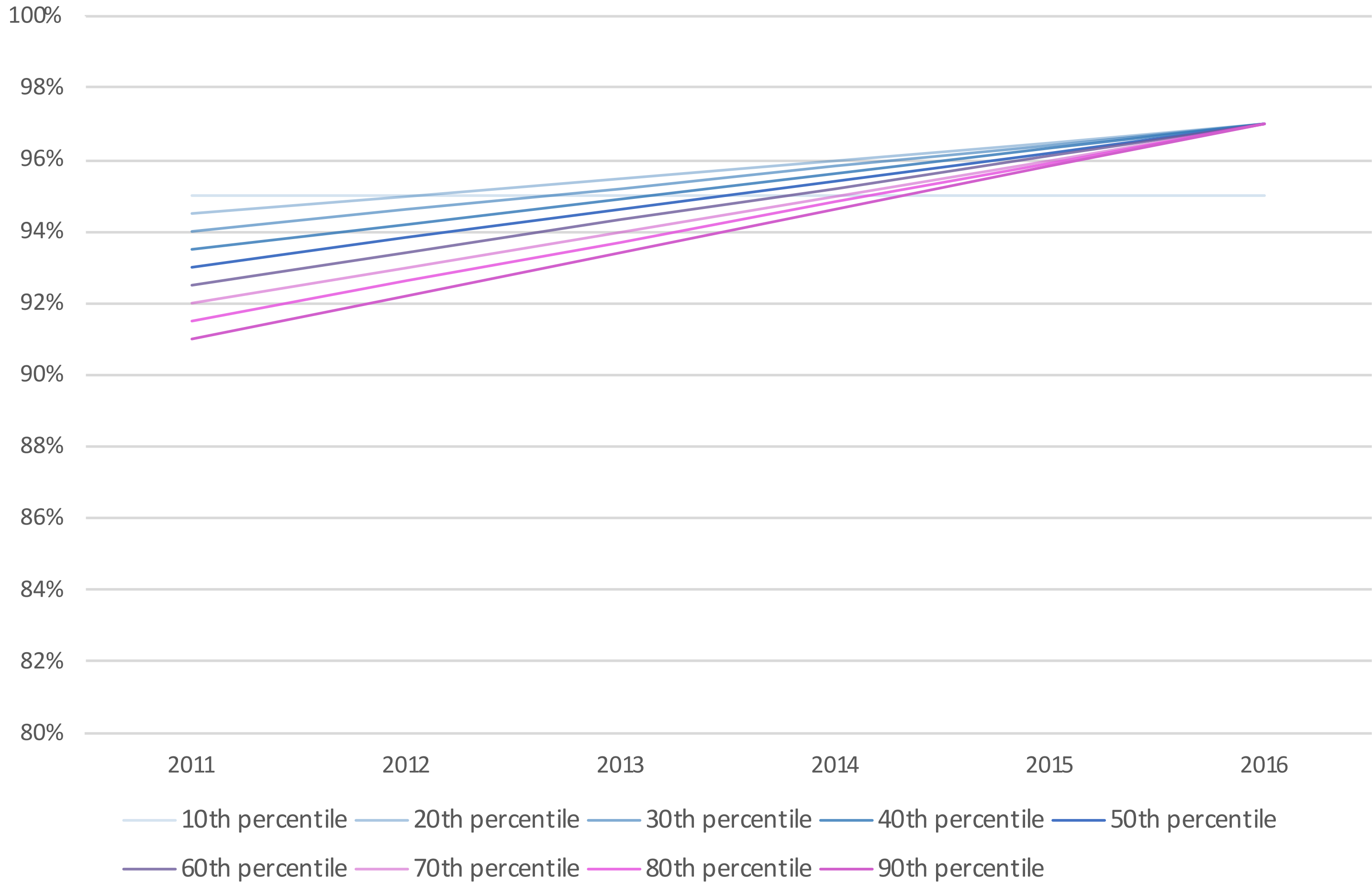
Table G3
Parameter estimates obtained from the final attendance model – primary schools

				Wald based 95% confidence intervals	
Model parameter	Point estimate	Standard error	*p* value	Lower limit	Upper limit
β_{00}	2.69	0.01	<.005	2.67	2.70
β_{01}	-0.20	0.01	<.005	-0.22	-0.19
β_{02}	-0.03	0.01	<.005	-0.04	-0.01
β_{10}	0.04	0.00	<.005	0.04	0.04
β_{11}	-0.00	0.00	0.30	-0.00	0.00
β_{12}	--	--	--	--	--
β_{20}	-0.01	0.00	<.005	-0.01	-0.01
β_{30}	0.00	0.00	<.005	0.00	0.00
σ_{e}	0.11	--	--	--	--
σ_{u0}	0.21	--	--	--	--
σ_{u1}	0.03	--	--	--	--
$\mathrm{Corr}(\sigma_{u0}, \sigma_{u1})$	-0.54	--	--	--	--

Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

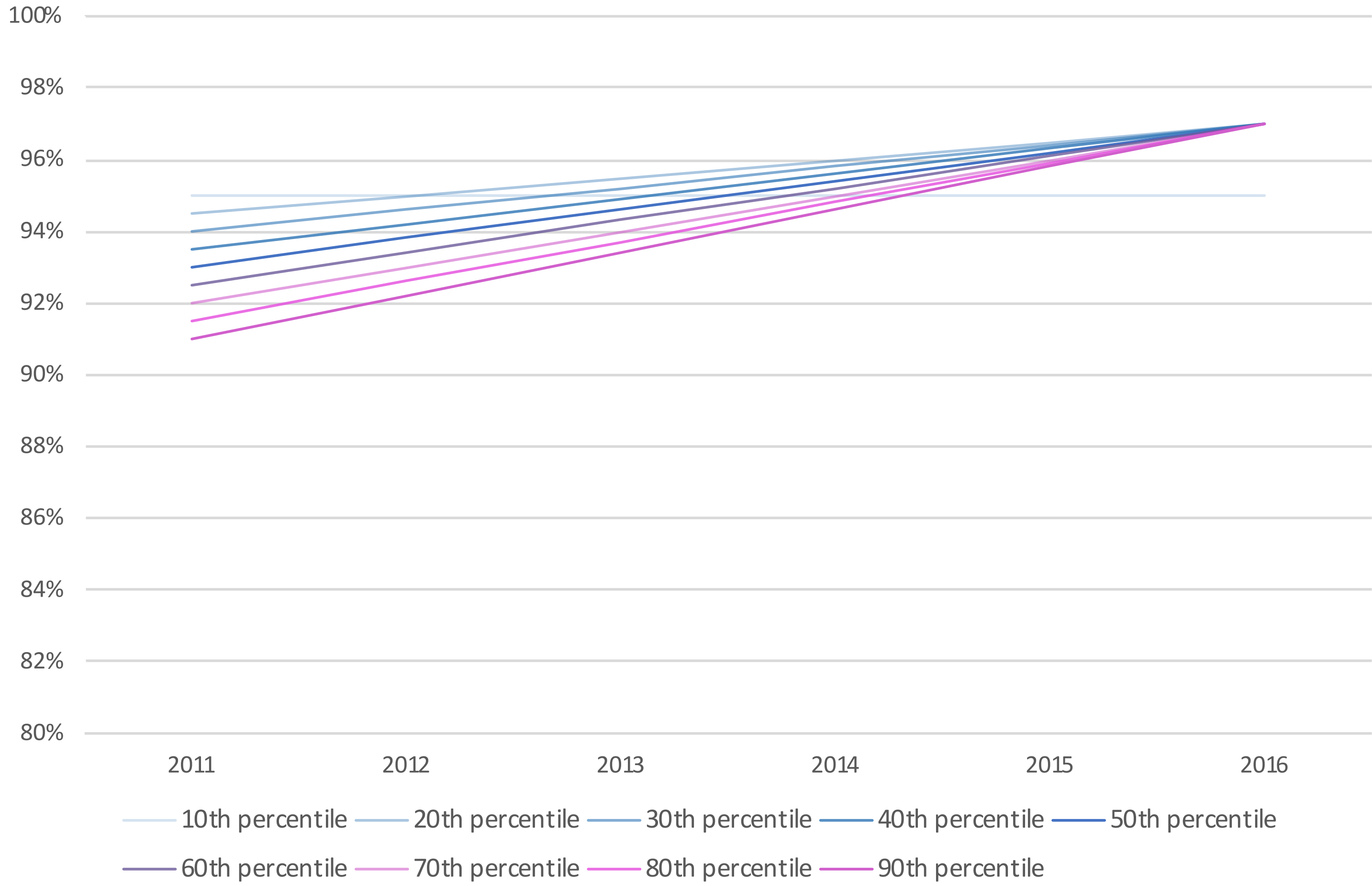
- Hypothesised attendance rates for primary schools with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

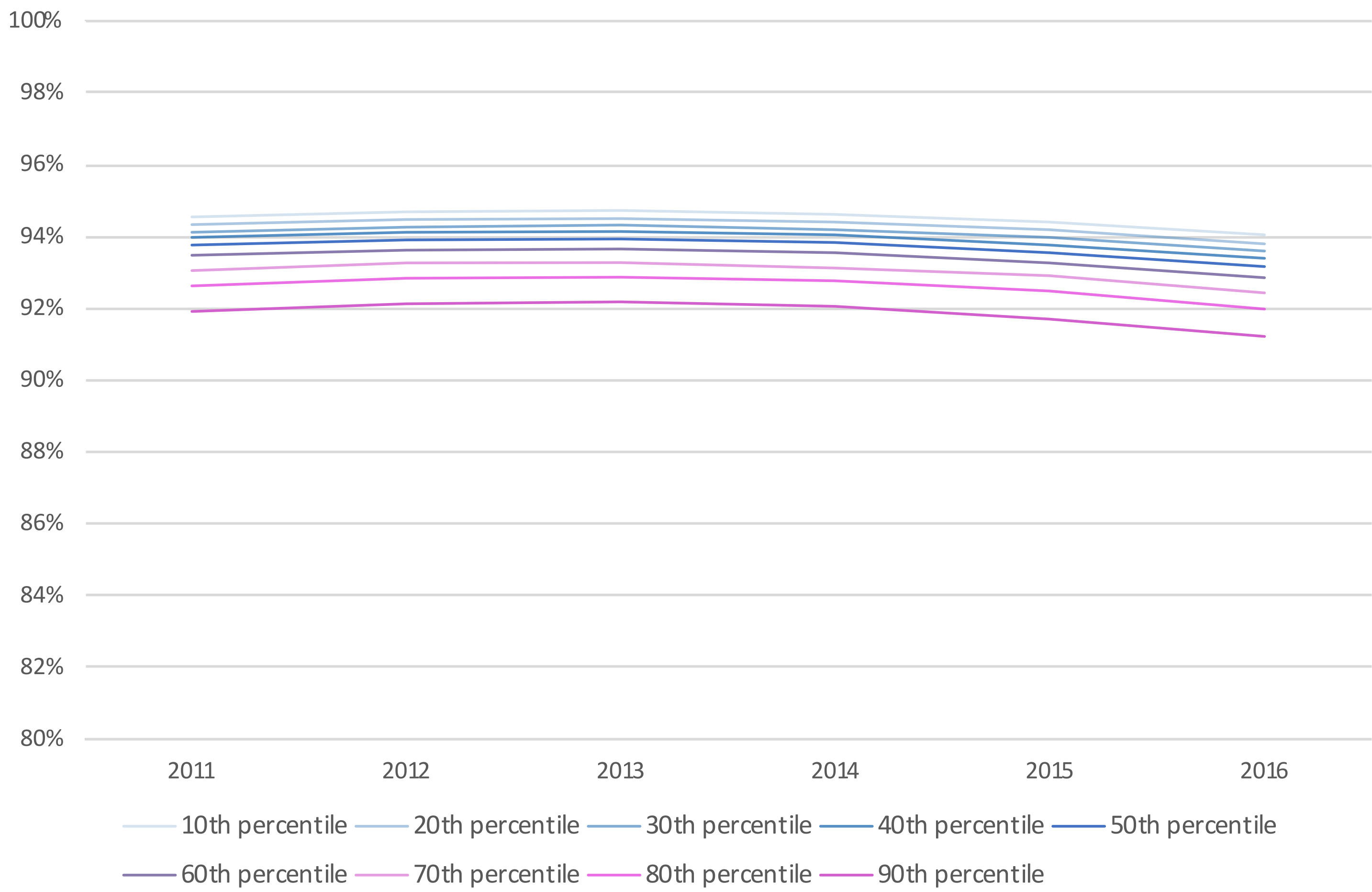
- Hypothesised attendance rates for primary schools with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

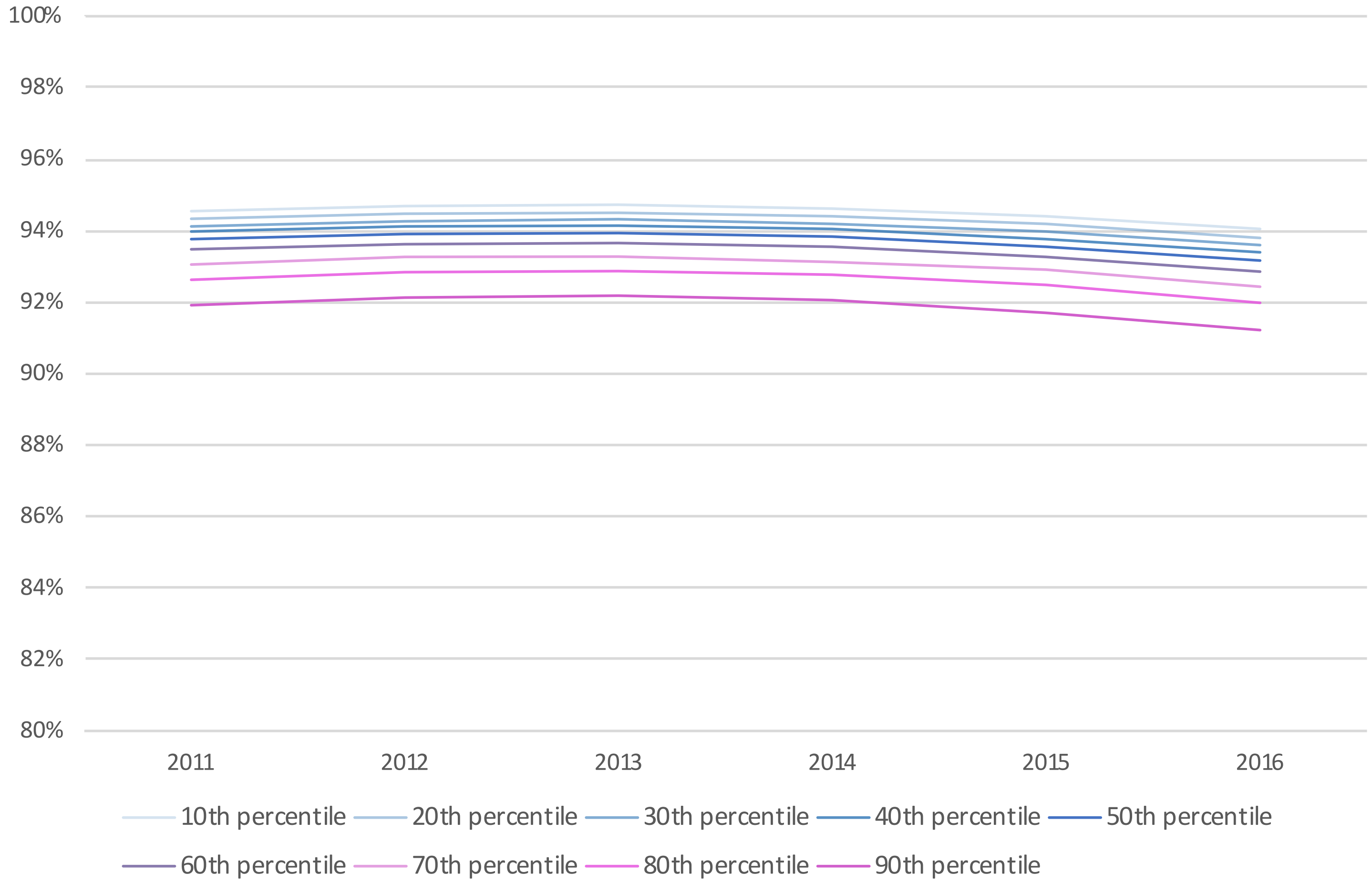
- Predicted attendance rates for **primary schools** with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

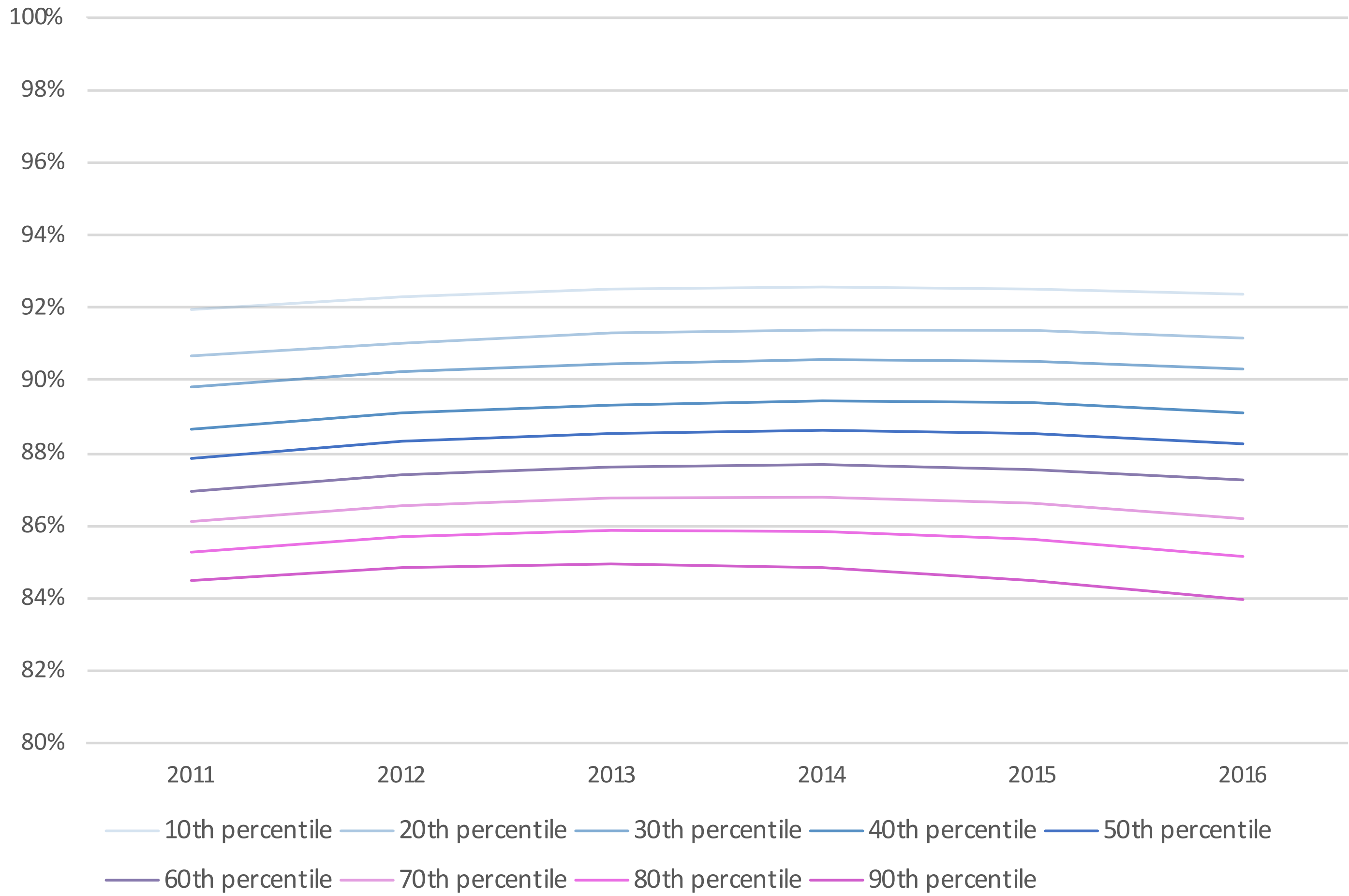
- Predicted attendance rates for **primary schools** with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

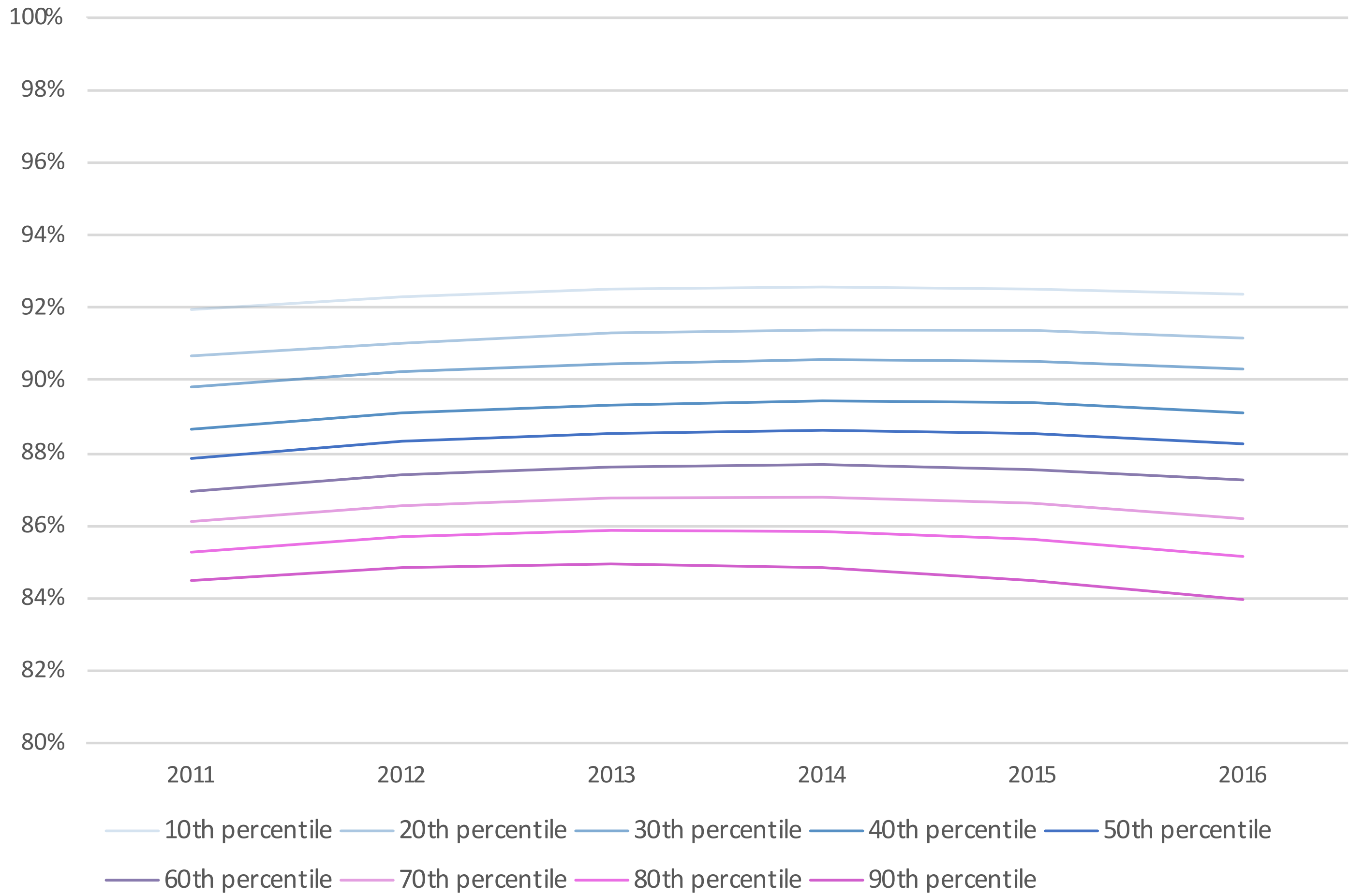
- Predicted attendance rates for **secondary schools** with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

- Predicted attendance rates for **secondary schools** with different levels-of-need



Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative
  - Operationalise levels-of-need by combining the four RAM equity loadings

- Map data to outcomes
  - Use attendance data from 2011 through 2016
  - Use *Tell Them From Me* (TTFM) data from 2013 and 2016
  - Others…

- Formalise analysis approach
  - Use latent variable growth curve models to estimate school-specific changes in outcomes across time
  - Relate the school-specific changes to levels-of-need

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative
  - Operationalise levels-of-need by combining the four RAM equity loadings

- Map data to outcomes
  - Use attendance data from 2011 through 2016
  - Use *Tell Them From Me* (TTFM) data from 2013 and 2016
  - Others…

- Formalise analysis approach
  - Use latent variable growth curve models to estimate school-specific changes in outcomes across time
  - Relate the school-specific changes to levels-of-need

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative

$$\text{level of need}_s = \ln\left(\frac{\sum_{i=1}^{4} \text{RAM equity loading}_{si}}{\left(\left(\sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{spc}\right)/4\right)}\right)$$

- Map data to outcomes
    - During the past 4 weeks how often have you been spoken to by a teacher or principal for any of the following reasons:

	Never	Once a week	2/3 times a week	Almost every day
Being disruptive in class	X			
Making inappropriate comments		X		
Getting into fights	X			
Breaking a school rule		X		
Lying or cheating			X	

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative

$$\text{level of need}_s = \ln\left(\frac{\sum_{i=1}^{4} \text{RAM equity loading}_{si}}{\left(\left(\sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{spc}\right)/4\right)}\right)$$

- Map data to outcomes
  - During the past 4 weeks how often have you been spoken to by a teacher or principal for any of the following reasons:

	Never	Once a week	2/3 times a week	Almost every day
Being disruptive in class	✖			
Making inappropriate comments		✖		
Getting into fights	✖			
Breaking a school rule		✖		
Lying or cheating			✖	

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative

$$\text{level of need}_s = \ln\left(\frac{\sum_{i=1}^{4}\text{RAM equity loading}_{si}}{\left(\left(\sum_{p=1}^{6}\sum_{c=1}^{4}\text{enrolment count}_{spc}\right)/4\right)}\right)$$

- Map data to outcomes

Step 1) $$\Pr(Y_{ij} = k|\alpha_i, \boldsymbol{\beta}_i, \theta_j) = \frac{\exp(k\alpha_i\theta_j + \beta_{ik})}{1 + \sum_{s=1}^{K_i}\exp(s\alpha_i\theta_j + \beta_{is})}$$

Step 2) $$f(\mathbf{y}_j|\widehat{\mathbf{B}}, \theta_j) = \prod_{i=1}^{I}\Pr(Y_{ij} = y_{ij}|\widehat{\mathbf{B}}, \theta_j)$$

Step 3) $$L_j(\widehat{\mathbf{B}}) = \int_{-\infty}^{\infty} f(\mathbf{y}_j|\widehat{\mathbf{B}}, \theta_j)\phi(\theta_j)d\theta_j$$

Step 4) $$\omega(\theta_j|\mathbf{y}_j, \widehat{\mathbf{B}}) = \frac{f(\mathbf{y}_j|\widehat{\mathbf{B}}, \theta_j)\,\phi(\theta_j)}{L_j(\widehat{\mathbf{B}})}$$

Step 5) $$\tilde{\theta}_j = \int_{-\infty}^{\infty}\theta_j\omega(\theta_j|\mathbf{y}_j, \widehat{\mathbf{B}})d\theta_j$$

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Features of the policy/initiative

$$\text{level of need}_s = \ln\left(\frac{\sum_{i=1}^{4} \text{RAM equity loading}_{si}}{\left(\left(\sum_{p=1}^{6}\sum_{c=1}^{4} \text{enrolment count}_{spc}\right)/4\right)}\right)$$

- Map data to outcomes

Step 1) $\Pr(Y_{ij} = k|\alpha_i, \boldsymbol{\beta}_i, \theta_j) = \dfrac{\exp(k\alpha_i\theta_j + \beta_{ik})}{1 + \sum_{s=1}^{K_i} \exp(s\alpha_i\theta_j + \beta_{is})}$

Step 2) $f(\mathbf{y}_j|\widehat{\mathbf{B}}, \theta_j) = \prod_{i=1}^{I} \Pr(Y_{ij} = y_{ij}|\widehat{\mathbf{B}}, \theta_j)$

Step 3) $L_j(\widehat{\mathbf{B}}) = \int_{-\infty}^{\infty} f(\mathbf{y}_j|\widehat{\mathbf{B}}, \theta_j)\phi(\theta_j)d\theta_j$

Step 4) $\omega(\theta_j|\mathbf{y}_j, \widehat{\mathbf{B}}) = \dfrac{f(\mathbf{y}_j|\widehat{\mathbf{B}}, \theta_j)\ \phi(\theta_j)}{L_j(\widehat{\mathbf{B}})}$

Step 5) $\tilde{\theta}_j = \int_{-\infty}^{\infty} \theta_j\omega(\theta_j|\mathbf{y}_j, \widehat{\mathbf{B}})d\theta_j$

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Formalise analysis approach

$$\tilde{\theta}_{js} = (\beta_{00}+\beta_{01} \cdot \text{need GMC}_s + \beta_{02} \cdot \text{need GMC}_s^2 + e_{js} + u_{0s}) + (\beta_{10} + \beta_{11} \cdot \text{need GMC}_s + u_{1s}) \cdot \text{time}_{js} +$$

$$\beta_{20} \cdot \text{ICSEA CMC}_{js} + \beta_{30} \cdot \text{Year 9 CMC}_{js} +, \ldots, + \beta_{60} \cdot \text{Year 12 CMC}_{js},$$

for $s = 1,2, \ldots, 80$ schools, $j = 1,2, \ldots, n_s$ students,

and $e_{js} \sim N(0, \sigma_e^2)$ and $\mathbf{u}_s \sim MN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{01} \\ \sigma_{10} & \sigma_{u1}^2 \end{pmatrix}\right]$.

Scoping and operationalising research questions

Local Schools, Local Decisions (LSLD) evaluation

- Formalise analysis approach

$$\tilde{\theta}_{js} = (\beta_{00}+\beta_{01} \cdot \text{need GMC}_s + \beta_{02} \cdot \text{need GMC}_s^2 + e_{js} + u_{0s}) + (\beta_{10} + \beta_{11} \cdot \text{need GMC}_s + u_{1s}) \cdot \text{time}_{js} +$$

$$\beta_{20} \cdot \text{ICSEA CMC}_{js} + \beta_{30} \cdot \text{Year 9 CMC}_{js} +, \ldots, + \beta_{60} \cdot \text{Year 12 CMC}_{js},$$

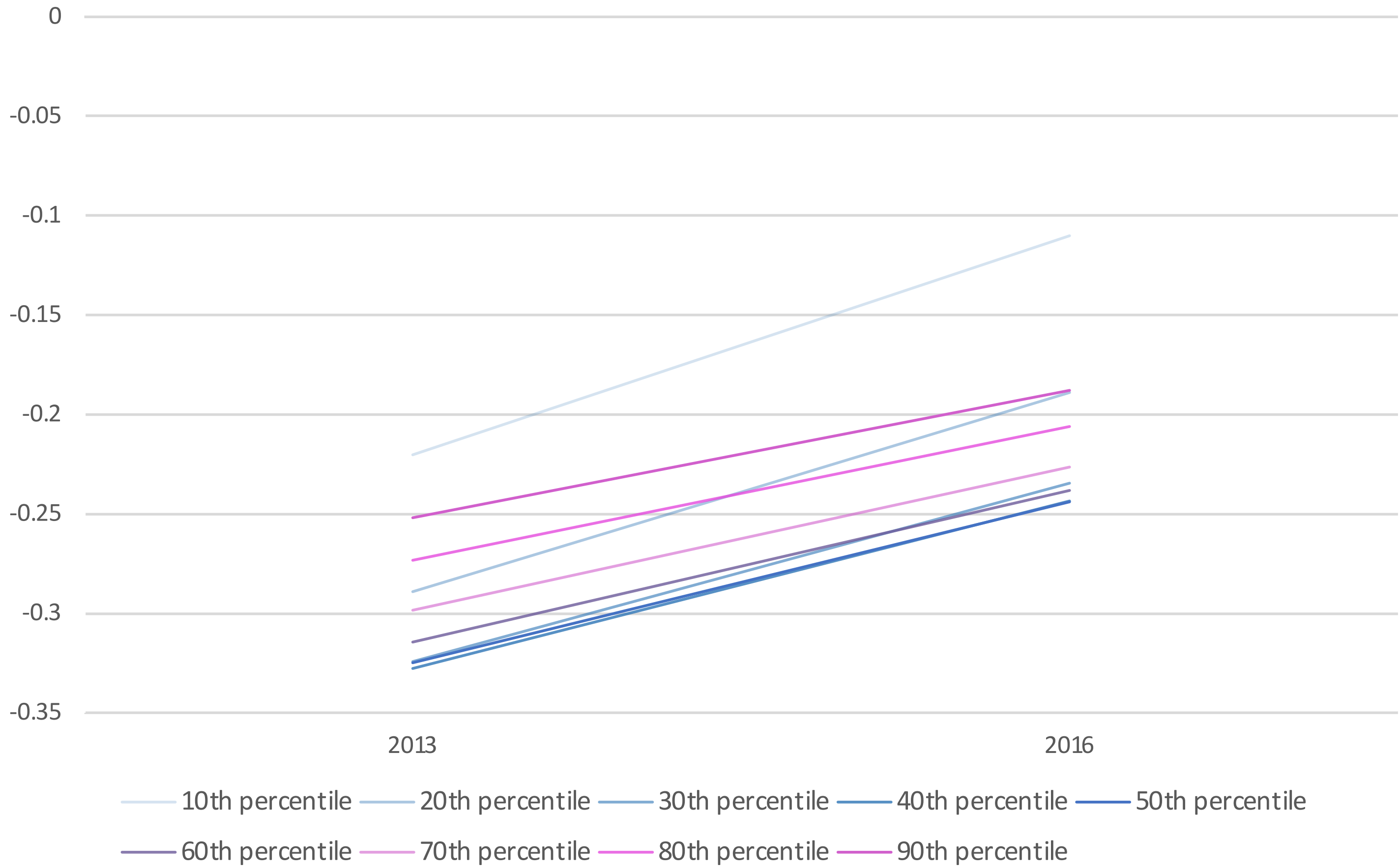
for $s = 1,2, \ldots, 80$ schools, $j = 1,2, \ldots, n_s$ students,

and $e_{js} \sim N(0, \sigma_e^2)$ and $\mathbf{u}_s \sim MN\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{u0}^2 & \sigma_{01}\\ \sigma_{10} & \sigma_{u1}^2\end{pmatrix}\right]$.

Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

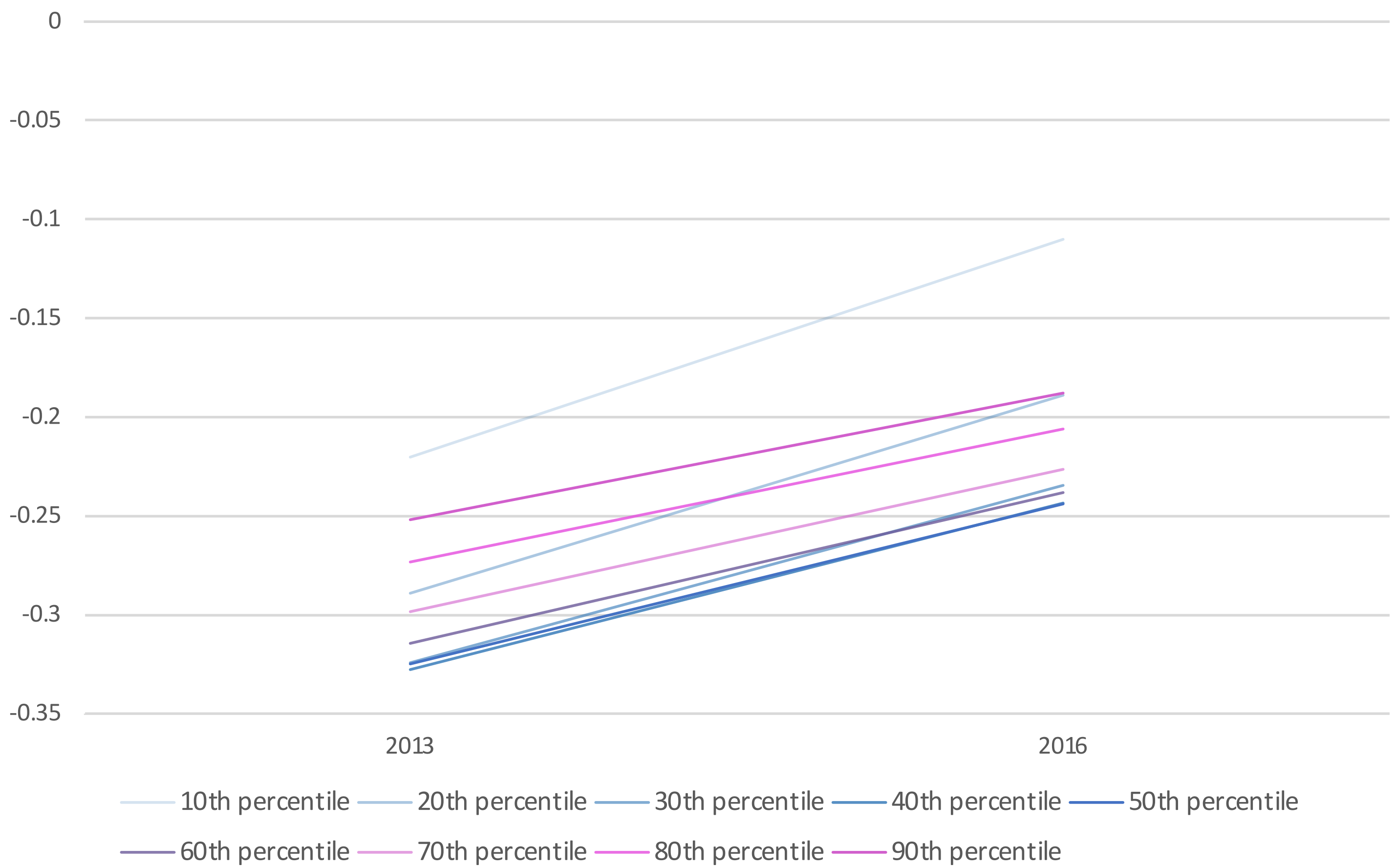
- Predicted **institutional engagement** scores for secondary schools with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

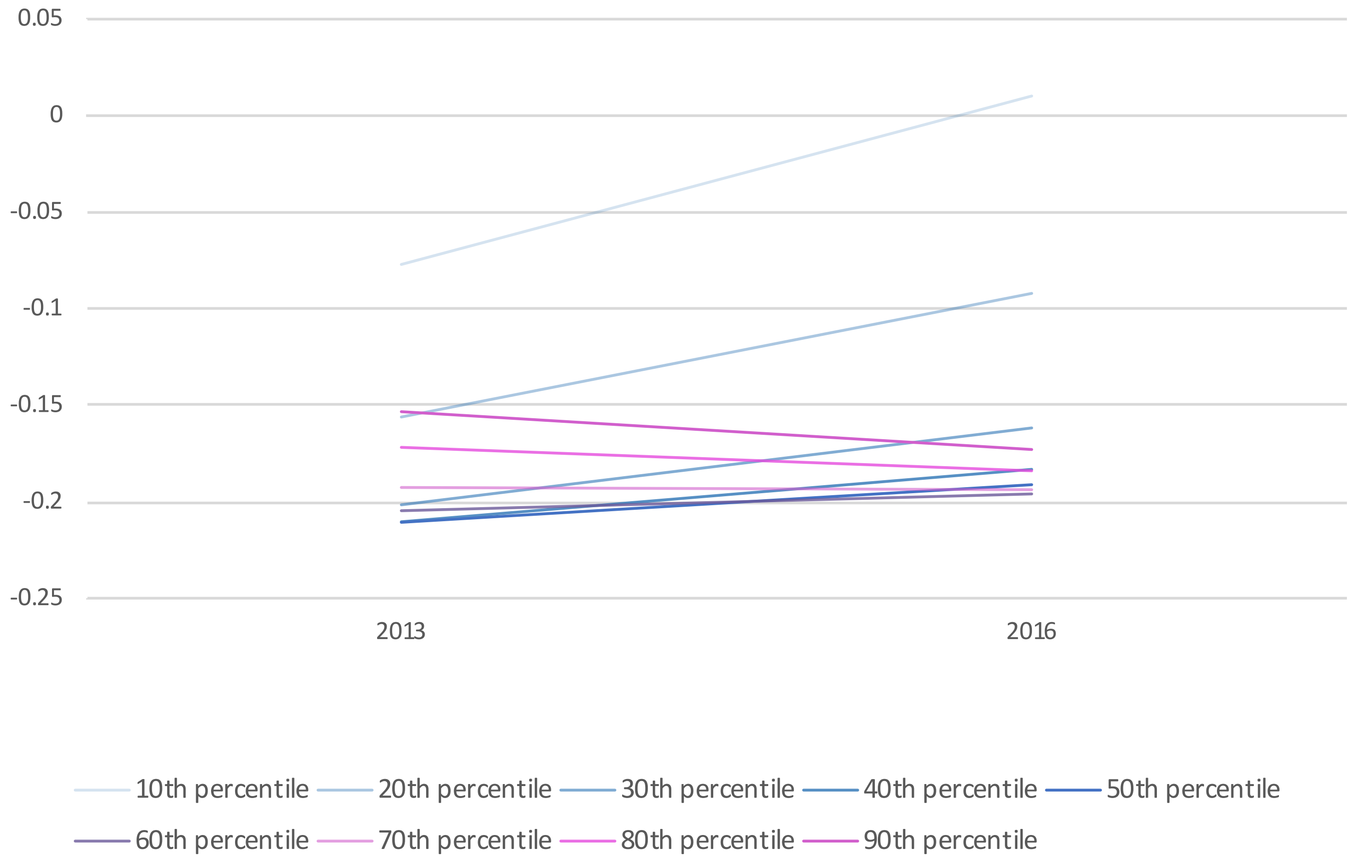
- Predicted **institutional engagement** scores for secondary schools with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

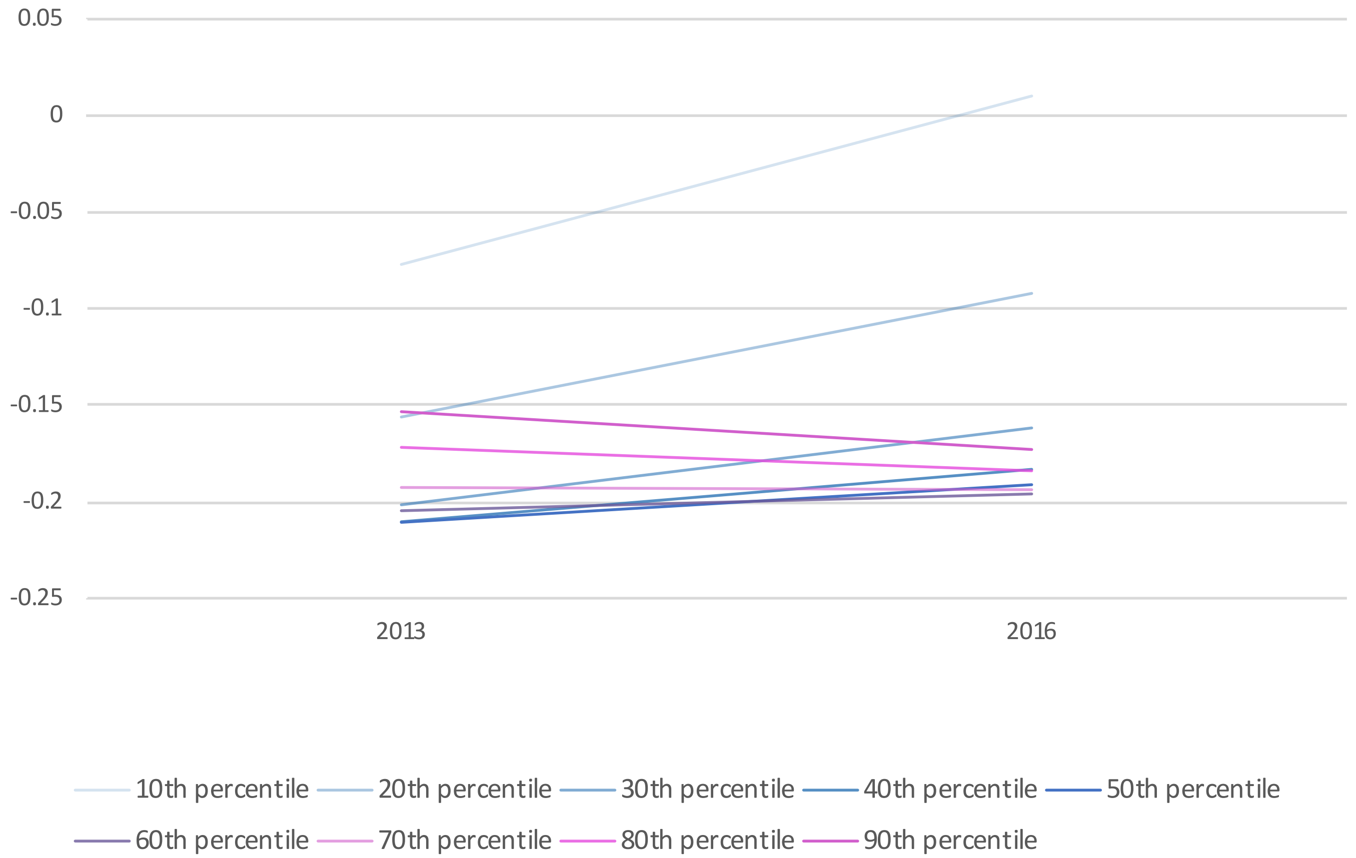
- Predicted **social engagement** scores for secondary schools with different levels-of-need



Presenting the results

Local Schools, Local Decisions (LSLD) evaluation

- Predicted **social engagement** scores for secondary schools with different levels-of-need



Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Literacy intervention targeting students in the bottom 20% of Year 1
- Involves one-to-one literacy tuition over a 12 to 20 week time period
- Offered in over half (approximately 60%) of NSW government primary schools
- In 2018, approximately 14% of all Year 1 students participate in the intervention (costing $50M)

Are literacy outcomes for students who participate in RR greater than those for comparable students who do not participate in RR?

- Possible to identify counterfactual potential outcomes using quasi-experimental methodologies
    - Requires good data and diagnostics
    - Also requires some untestable assumptions

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Literacy intervention targeting students in the bottom 20% of Year 1
- Involves one-to-one literacy tuition over a 12 to 20 week time period
- Offered in over half (approximately 60%) of NSW government primary schools
- In 2018, approximately 14% of all Year 1 students participate in the intervention (costing $50M)

Are literacy outcomes for students who participate in RR greater than those for comparable students who do not participate in RR?

- Possible to identify counterfactual potential outcomes using quasi-experimental methodologies
    - Requires good data and diagnostics
    - Also requires some untestable assumptions

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Features of the policy/initiative
  - Literacy intervention targeting students in the bottom 20% of Year 1

- Map data to outcomes
  - Use Literacy Continuum data from the end of Year 1
  - Also include baseline data from Term 4 Kindergarten and other student- and school-level factors to control for the impact of important confounders

- Formalise analysis approach
  - Use a series of ordered logistic mixed-effects regression models to estimate the effect of RR
  - Model the interaction between treatment and baseline literacy measures to assess treatment effect heterogeneity

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Features of the policy/initiative
    - Literacy intervention targeting students in the bottom 20% of Year 1
- Map data to outcomes
    - Use Literacy Continuum data from the end of Year 1
    - Also include baseline data from Term 4 Kindergarten and other student- and school-level factors to control for the impact of important confounders
- Formalise analysis approach
    - Use a series of ordered logistic mixed-effects regression models to estimate the effect of RR
    - Model the interaction between treatment and baseline literacy measures to assess treatment effect heterogeneity

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Features of the policy/initiative

$$\text{treat}_{\text{js}} \begin{cases} = 1 \text{ if student j in school s was a Reading Recovery student} \\ = 0 \text{ otherwise} \end{cases}$$

- Map data to outcomes

$$y_{\text{jsa}} \begin{cases} = 1 \text{ if student j in school s was placed in the first cluster of the } a^{th} \text{Literacy Continuum aspect at the end of Year 1} \\ = 2 \text{ if student j in school s was placed in the second cluster of the } a^{th} \text{Literacy Continuum aspect at the end of Year 1} \\ \vdots \\ = K \text{ if student j in school s was placed in the } k^{th} \text{ cluster of the } a^{th} \text{Literacy Continuum aspect at the end of Year 1} \end{cases}$$

for $a = 1,2, \ldots, 7$ Literacy Continuum aspects

- Formalise analysis approach

$$\begin{aligned} \Pr(y_{\text{js}} > k) = g(&\beta_1 \cdot \text{treat}_{\text{js}} + \beta_2 \cdot \text{T4 K Reading } 2_{\text{js}} + \beta_3 \cdot \text{T4 K Reading } 3_{\text{js}} + \beta_4 \cdot \text{T4 K Reading } 4_{\text{js}} + \\ &\beta_5 \cdot (\text{treat}_{\text{js}} \cdot \text{T4 K Reading } 2_{\text{js}}) + \beta_6 \cdot (\text{treat}_{\text{js}} \cdot \text{T4 K Reading } 3_{\text{js}}) + \beta_7 \cdot (\text{treat}_{\text{js}} \cdot \text{T4 K Reading } 4_{\text{js}}) + \\ &\textstyle\sum_{c=1}^{34} \gamma_c \cdot x_{\text{cjs}} + u_s - \alpha_k), \end{aligned}$$

for $j = 1,2, \ldots, n_s$ students, $s = 1,2, \ldots, S$ schools, $c = 1,2 \ldots, 34$ control covarites, and $k = 1,2, \ldots, K - 1$ cutpoints

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Features of the policy/initiative

$$\mathrm{treat}_{js}\begin{cases} = 1 \text{ if student j in school s was a Reading Recovery student} \\ = 0 \text{ otherwise} \end{cases}$$

- Map data to outcomes

$$y_{jsa}\begin{cases} = 1 \text{ if student j in school s was placed in the first cluster of the } a^{th}\text{Literacy Continuum aspect at the end of Year 1} \\ = 2 \text{ if student j in school s was placed in the second cluster of the } a^{th}\text{Literacy Continuum aspect at the end of Year 1} \\ \vdots \\ = K \text{ if student j in school s was placed in the } k^{th} \text{ cluster of the } a^{th}\text{Literacy Continuum aspect at the end of Year 1} \end{cases}$$

for $a = 1,2,\dots,7$ Literacy Continuum aspects

- Formalise analysis approach

$$\begin{aligned}\Pr(y_{js} > k) = g(&\beta_1 \cdot \mathrm{treat}_{js} + \beta_2 \cdot \mathrm{T4\ K\ Reading\ 2}_{js} + \beta_3 \cdot \mathrm{T4\ K\ Reading\ 3}_{js} + \beta_4 \cdot \mathrm{T4\ K\ Reading\ 4}_{js} + \\ &\beta_5 \cdot (\mathrm{treat}_{js} \cdot \mathrm{T4\ K\ Reading\ 2}_{js}) + \beta_6 \cdot (\mathrm{treat}_{js} \cdot \mathrm{T4\ K\ Reading\ 3}_{js}) + \beta_7 \cdot (\mathrm{treat}_{js} \cdot \mathrm{T4\ K\ Reading\ 4}_{js}) + \\ &\textstyle\sum_{c=1}^{34} \gamma_c \cdot x_{cjs} + u_s - \alpha_k),\end{aligned}$$

for $j = 1,2,\dots,n_s$ students, $s = 1,2,\dots,S$ schools, $c = 1,2\dots,34$ control covarites, and $k = 1,2,\dots,K-1$ cutpoints

Presenting the results

Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio	Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio
Reading Texts	Level 1 or below	**2.67***	**Phonics**	Level 1 or below	0.96
	Level 2	1.50		Level 2	1.01
	Level 3	**0.42***		Level 3	**0.28***
	Level 4 or above	**0.09***		Level 4 or above	**0.14***
Comprehension	Level 1 or below	1.3	**Phonemic Awareness**	Level 1 or below	0.85
	Level 2	**0.53***		Level 2	**0.52***
	Level 3	**0.30***		Level 3	**0.29***
	Level 4 or above	**0.13***		Level 4 or above	**0.22***
Aspects of Writing	Level 1 or below	0.98	**Concepts about Print**	Level 1 or below	1.11
	Level 2	**0.40***		Level 2	**0.53***
	Level 3	**0.22***		Level 3	**0.33***
	Level 4 or above	**0.11***		Level 4 or above	**0.05***
Aspects of Speaking	Level 1 or below	0.78			
	Level 2	**0.51***			
	Level 3	**0.44***			
	Level 4 or above	**0.34***			

Presenting the results

Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio	Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio
Reading Texts	Level 1 or below	**2.67***	**Phonics**	Level 1 or below	0.96
	Level 2	1.50		Level 2	1.01
	Level 3	**0.42***		Level 3	**0.28***
	Level 4 or above	**0.09***		Level 4 or above	**0.14***
Comprehension	Level 1 or below	1.3	**Phonemic Awareness**	Level 1 or below	0.85
	Level 2	**0.53***		Level 2	**0.52***
	Level 3	**0.30***		Level 3	**0.29***
	Level 4 or above	**0.13***		Level 4 or above	**0.22***
Aspects of Writing	Level 1 or below	0.98	**Concepts about Print**	Level 1 or below	1.11
	Level 2	**0.40***		Level 2	**0.53***
	Level 3	**0.22***		Level 3	**0.33***
	Level 4 or above	**0.11***		Level 4 or above	**0.05***
Aspects of Speaking	Level 1 or below	0.78			
	Level 2	**0.51***			
	Level 3	**0.44***			
	Level 4 or above	**0.34***			

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Features of the policy/initiative
  - Literacy intervention targeting students in the bottom 20% of Year 1

- Map data to outcomes
  - Use Literacy Continuum data from the end of Year 1
  - Also include baseline data from Term 4 Kindergarten and other student- and school-level factors to control for the impact of important confounders

- Formalise analysis approach
  - Use logistic regression model to estimate conditional probabilities of treatment for each student
  - Match students based on the estimated probabilities
  - Use ordinal logistic regression model to estimate the average causal effect of RR in the matched sample

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Features of the policy/initiative
  - Literacy intervention targeting students in the bottom 20% of Year 1

- Map data to outcomes
  - Use Literacy Continuum data from the end of Year 1
  - Also include baseline data from Term 4 Kindergarten and other student- and school-level factors to control for the impact of important confounders

- Formalise analysis approach
  - Use logistic regression model to estimate conditional probabilities of treatment for each student
  - Match students based on the estimated probabilities
  - Use ordinal logistic regression model to estimate the average causal effect of RR in the matched sample

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Formalise analysis approach

  - Step 1 – $\mathrm{Pr}(\mathrm{treat}_j = 1) = \mathrm{g}(\sum_{c=1}^{37} \gamma_c \cdot \mathrm{x}_{cj})$,

    $\text{for j} = 1,2, \ldots, \text{n students, and c} = 1,2 \ldots, 37 \text{ control covarites}$

,

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Formalise analysis approach

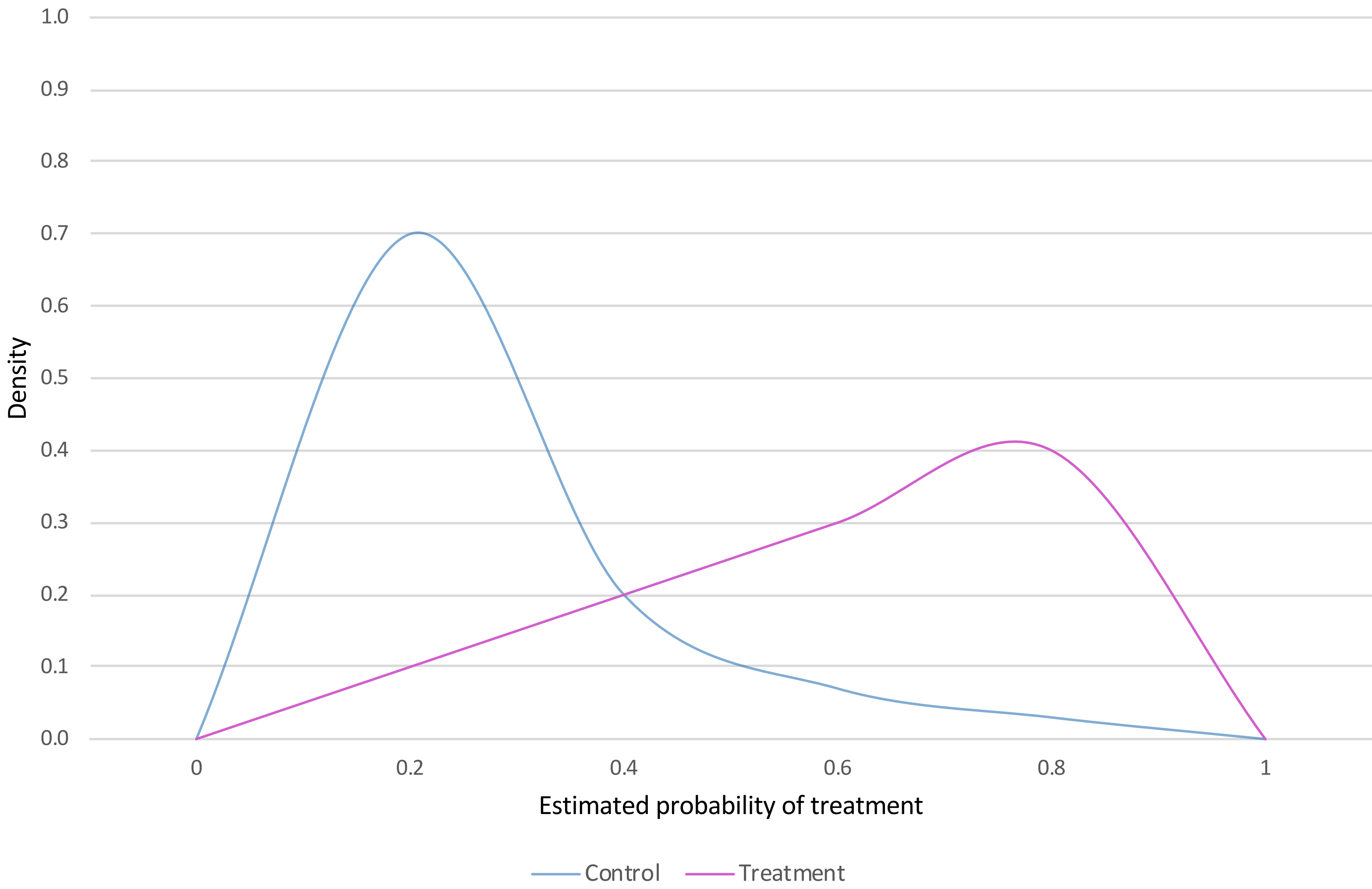
  - Step 1 – $\Pr(\mathrm{treat}_j = 1) = g(\sum_{c=1}^{37} \gamma_c \cdot x_{cj})$,

    $\text{for } j = 1,2,\ldots,n \text{ students, and } c = 1,2\ldots,37 \text{ control covarites}$

,

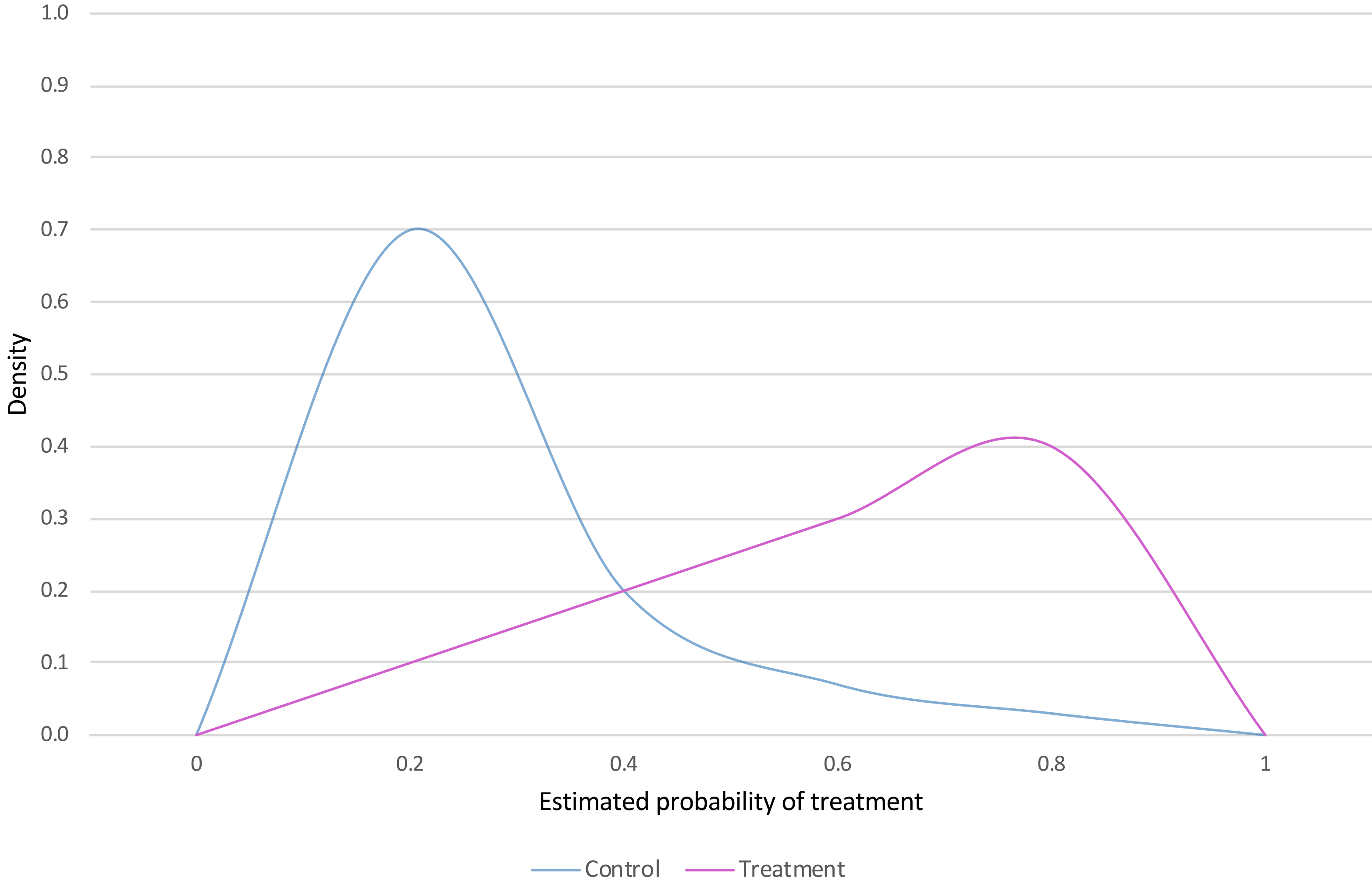
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



Scoping and operationalising research questions

Reading Recovery (RR) evaluation



Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Formalise analysis approach

  - Step 1 – $\mathrm{Pr}(\mathrm{treat}_j = 1) = \mathrm{g}(\sum_{c=1}^{37} \gamma_c \cdot \mathrm{x}_{cj}),$

    $\text{for } j = 1,2, \ldots, n \text{ students, and } c = 1,2 \ldots, 37 \text{ control covarites}$

  - Step 2 – use 1-to-1 nearest neighbour matching with replacement and a caliper of 0.0001

,

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Formalise analysis approach
  - Step 1 – $\Pr(\text{treat}_j = 1) = g(\sum_{c=1}^{37} \gamma_c \cdot x_{cj})$,

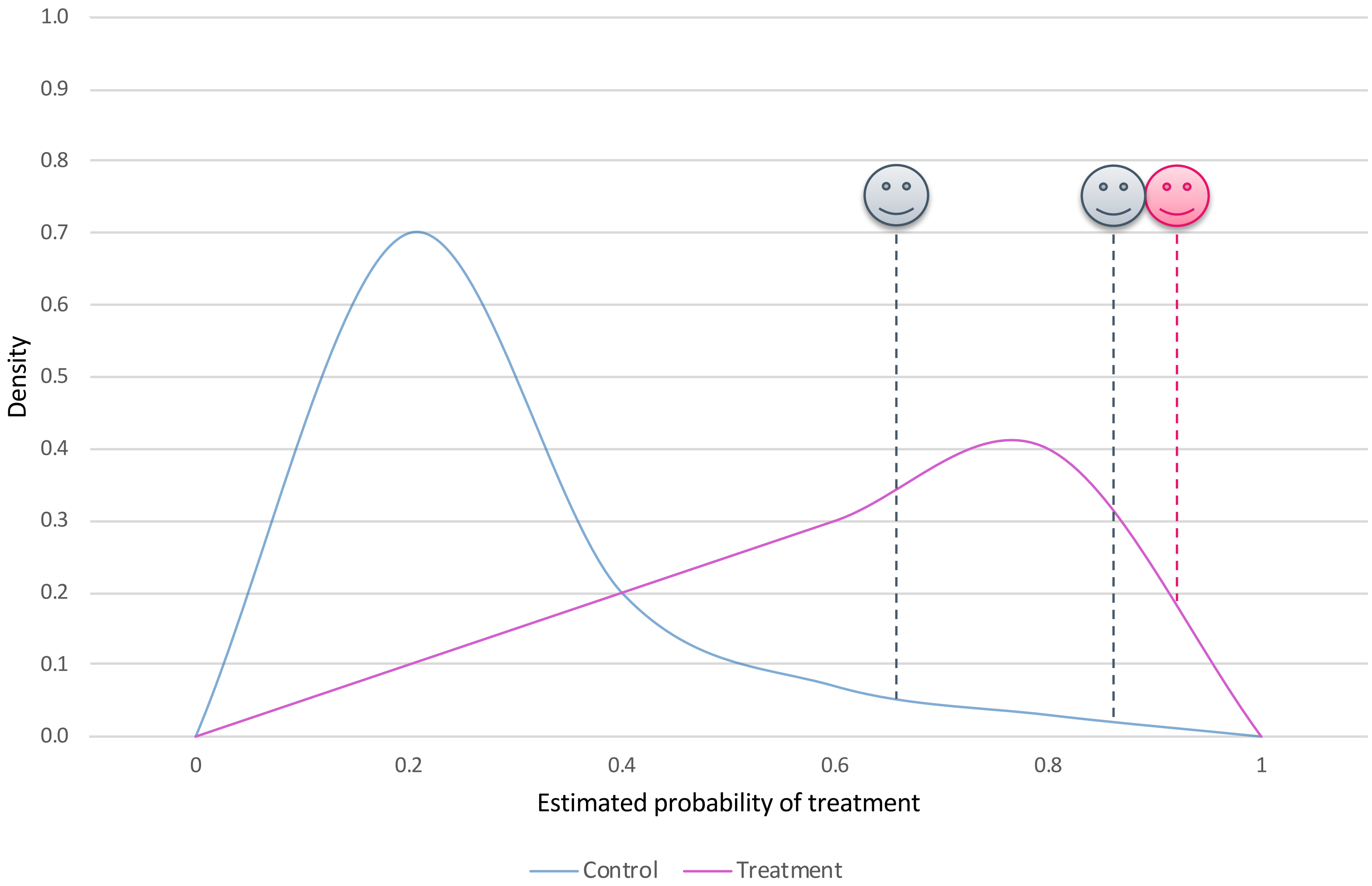
    $\text{for } j = 1,2, \dots, n \text{ students, and } c = 1,2 \dots, 37 \text{ control covarites}$

  - Step 2 – use 1-to-1 nearest neighbour matching with replacement and a caliper of 0.0001

,

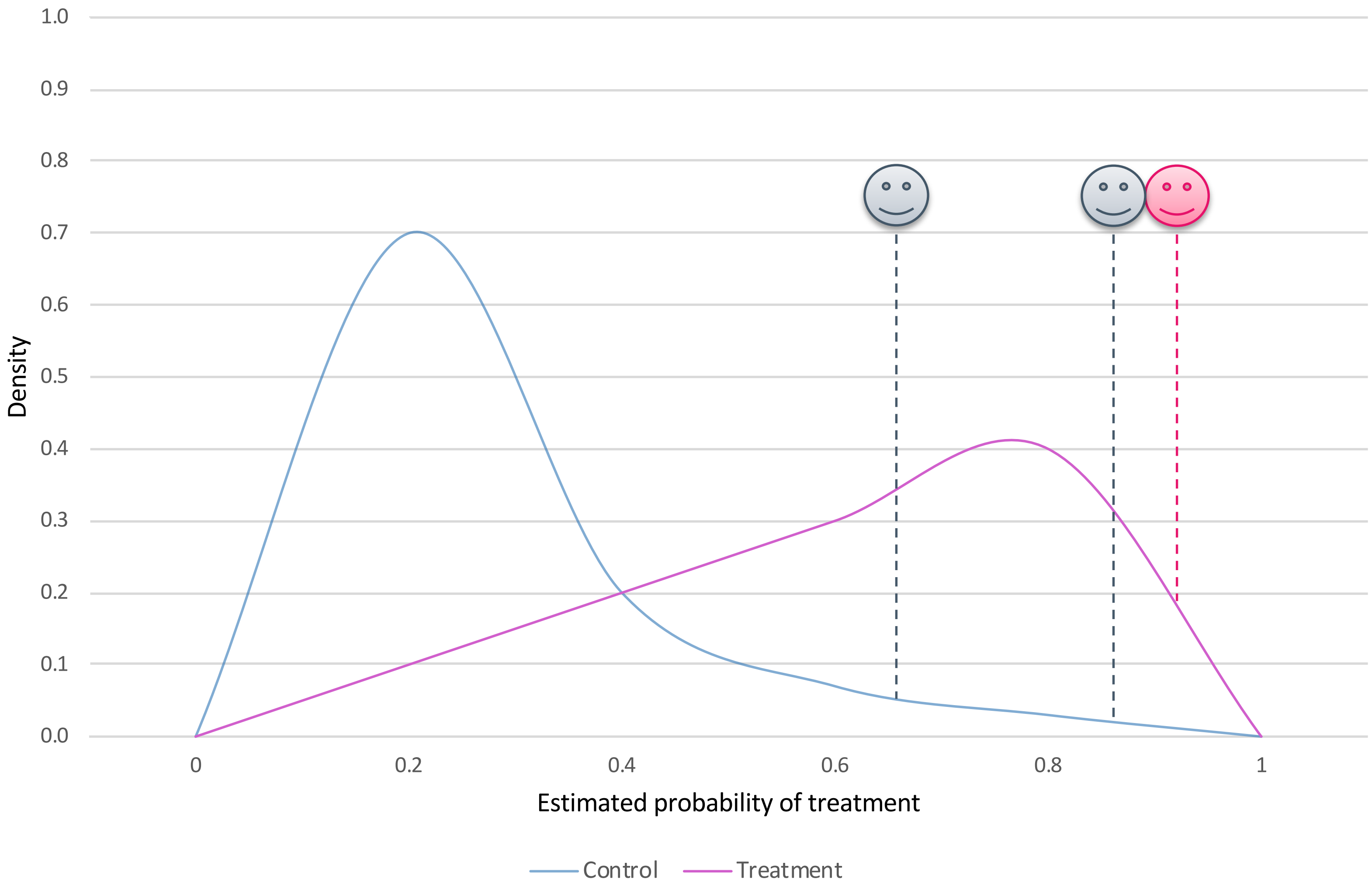
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



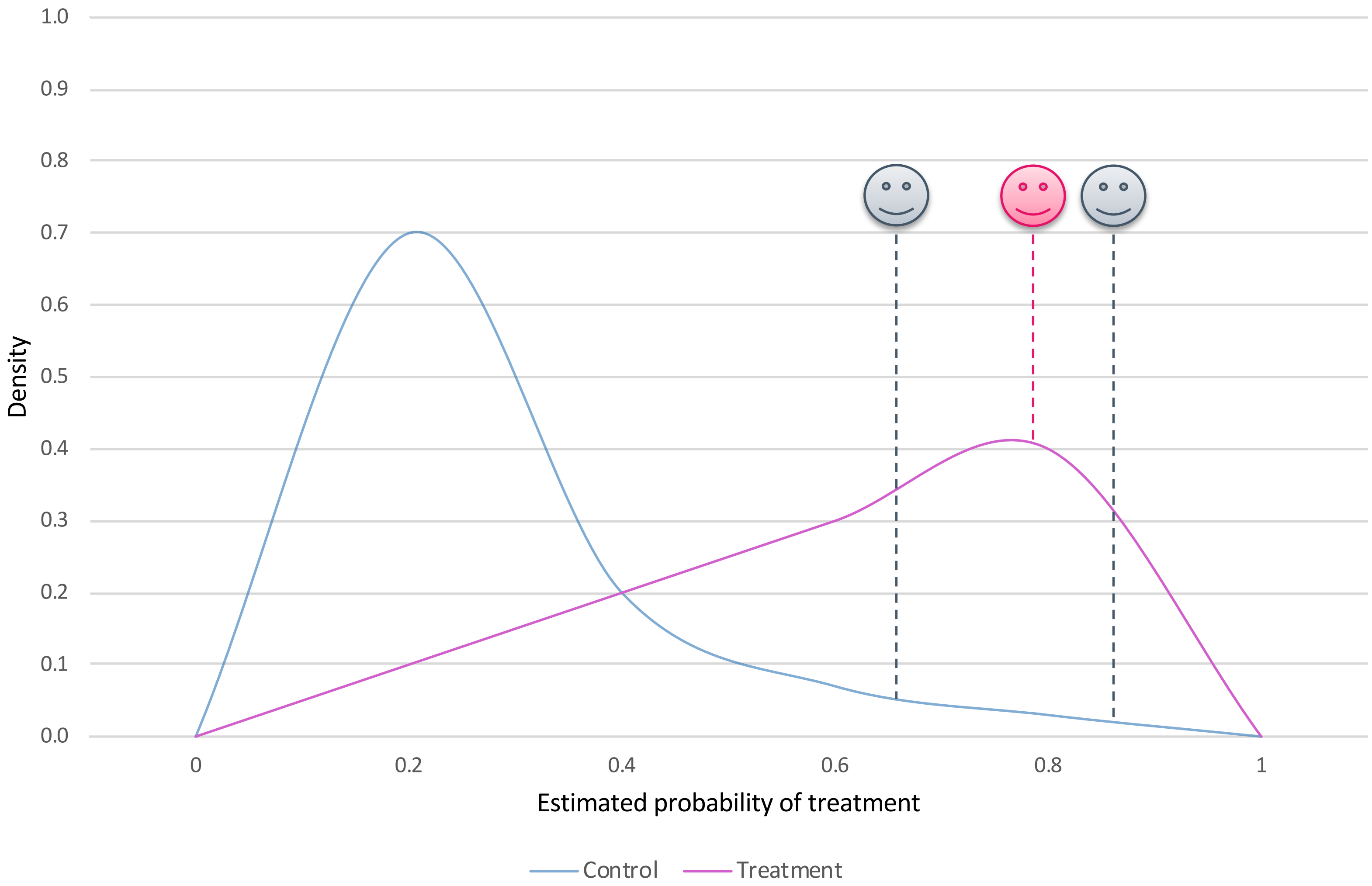
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



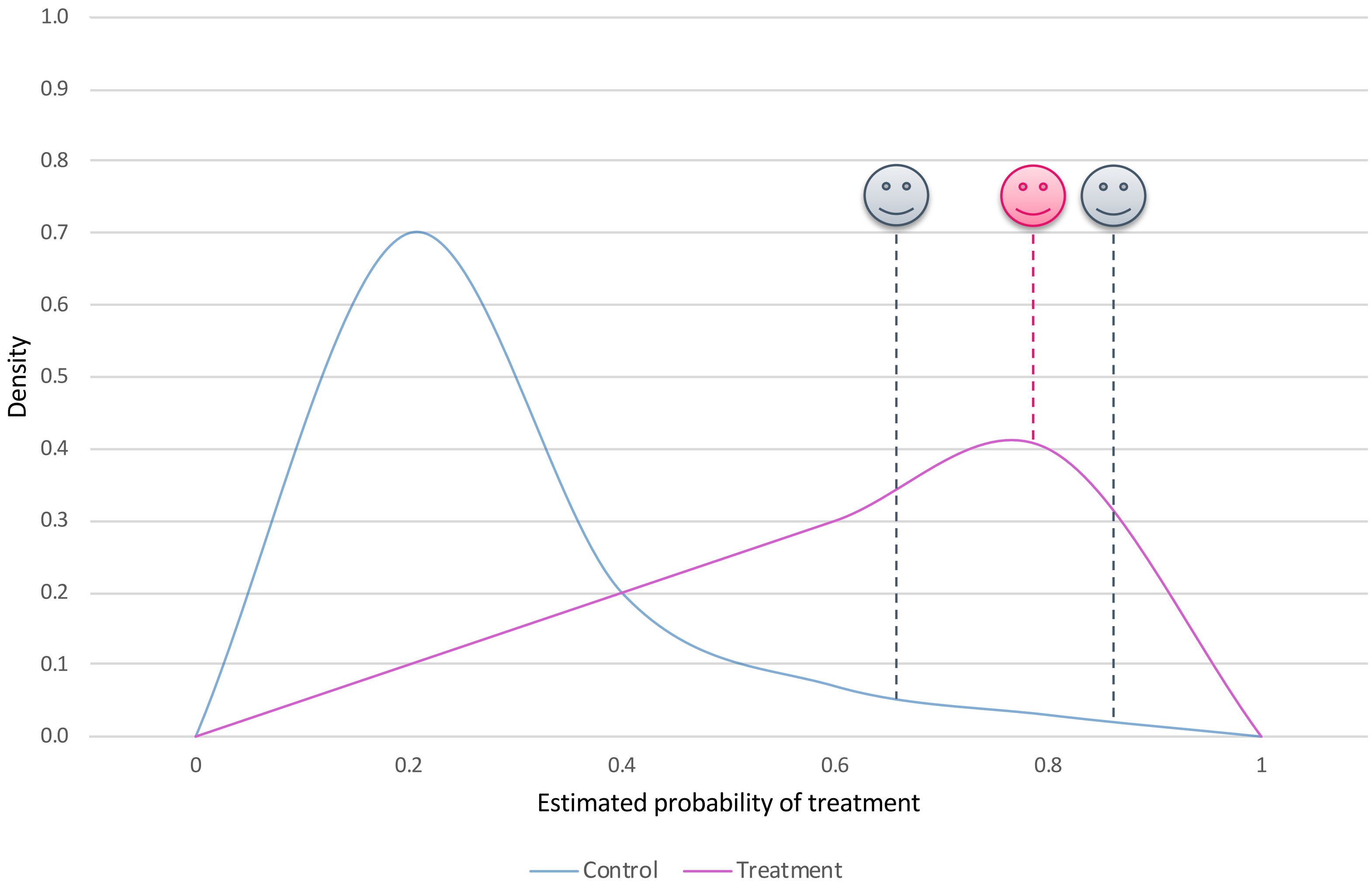
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



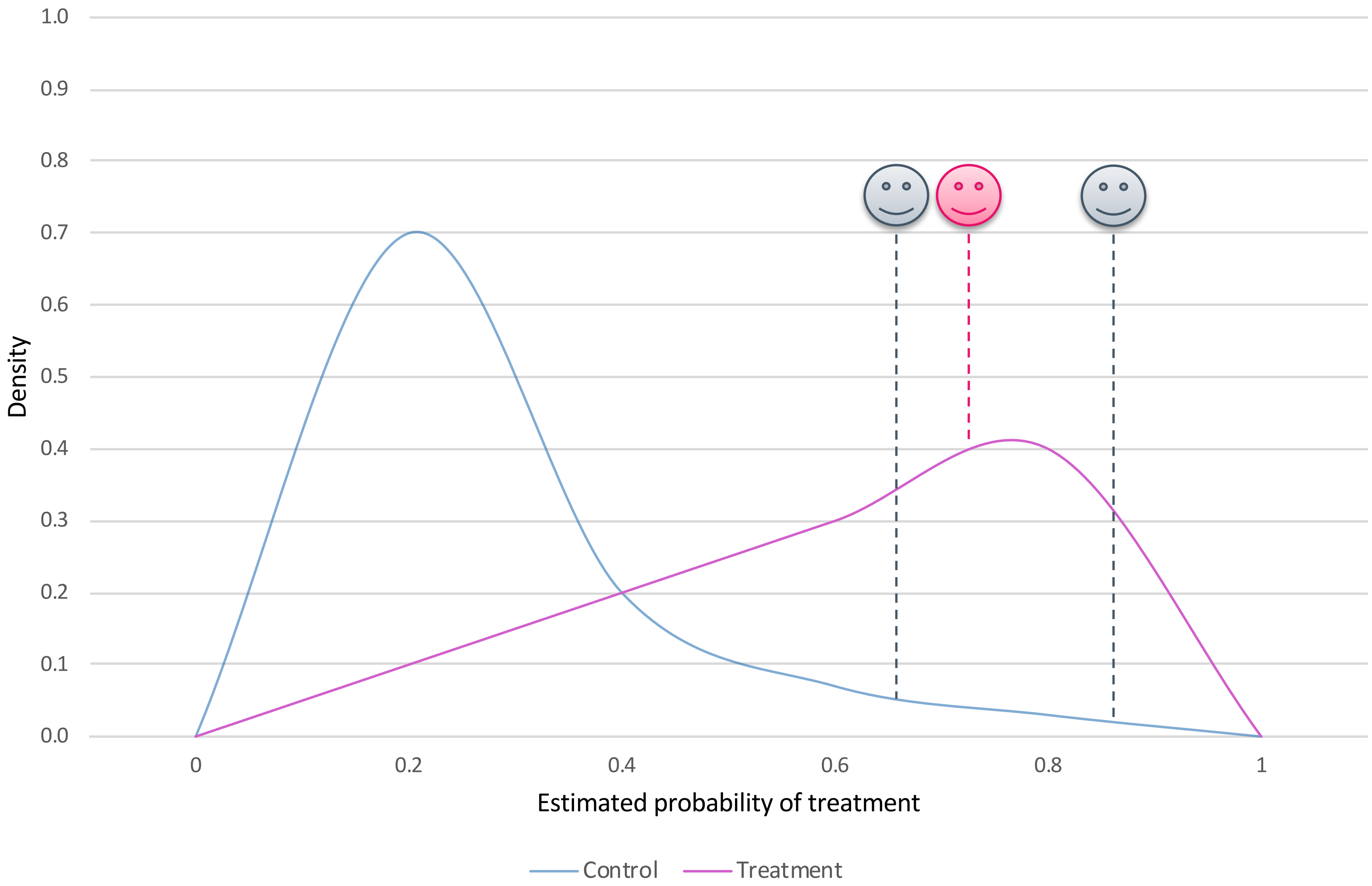
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



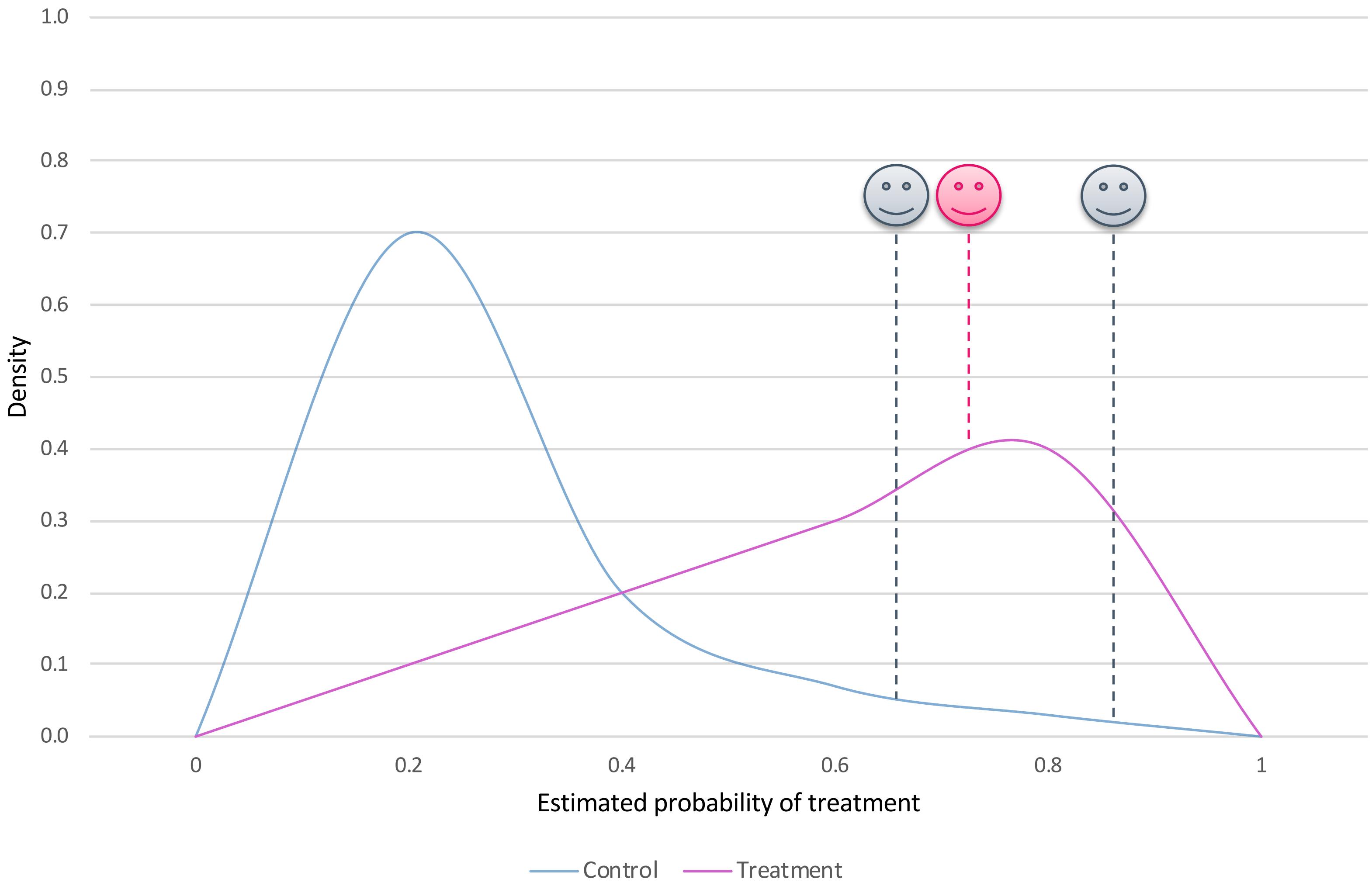
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



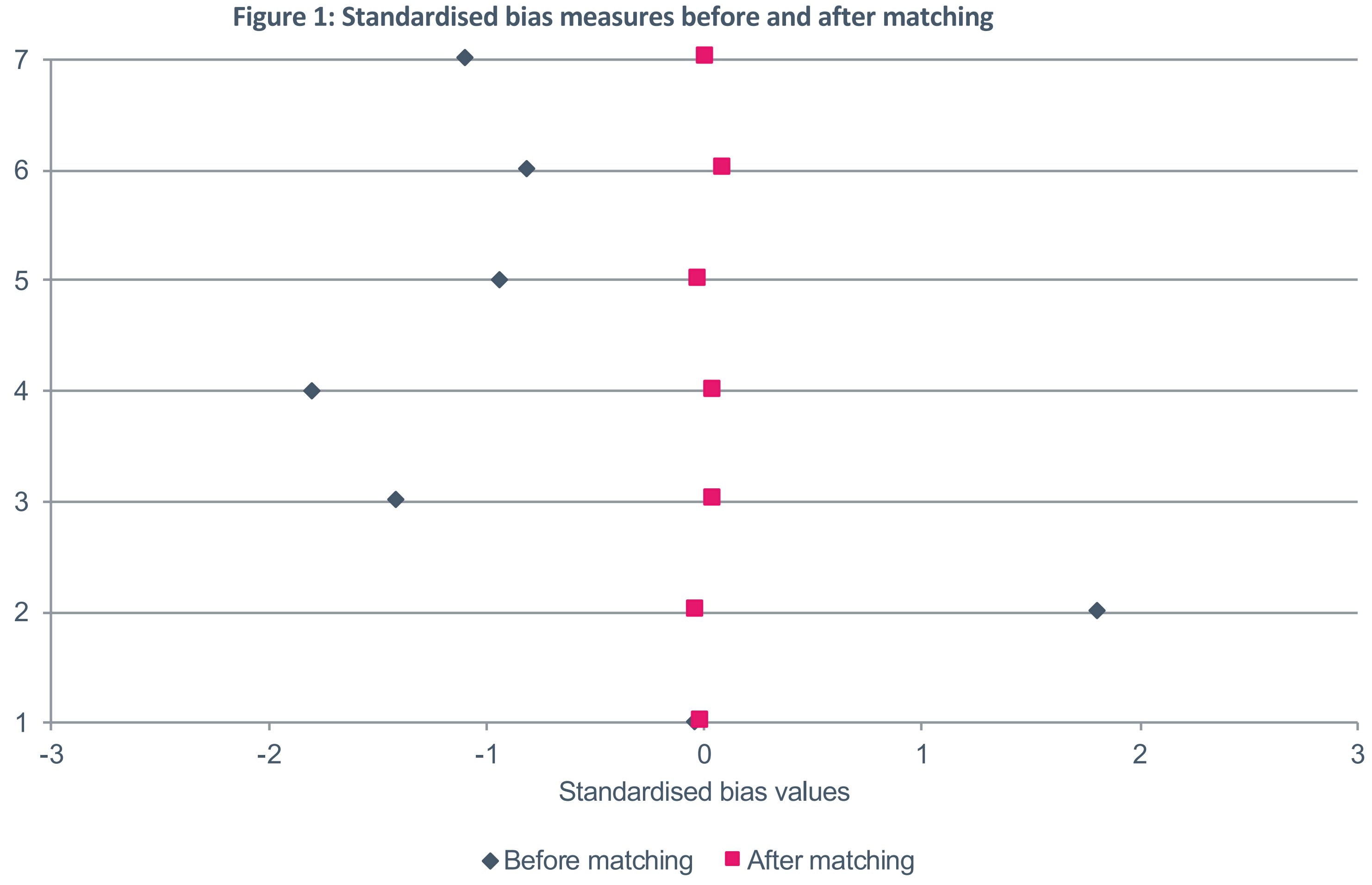
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



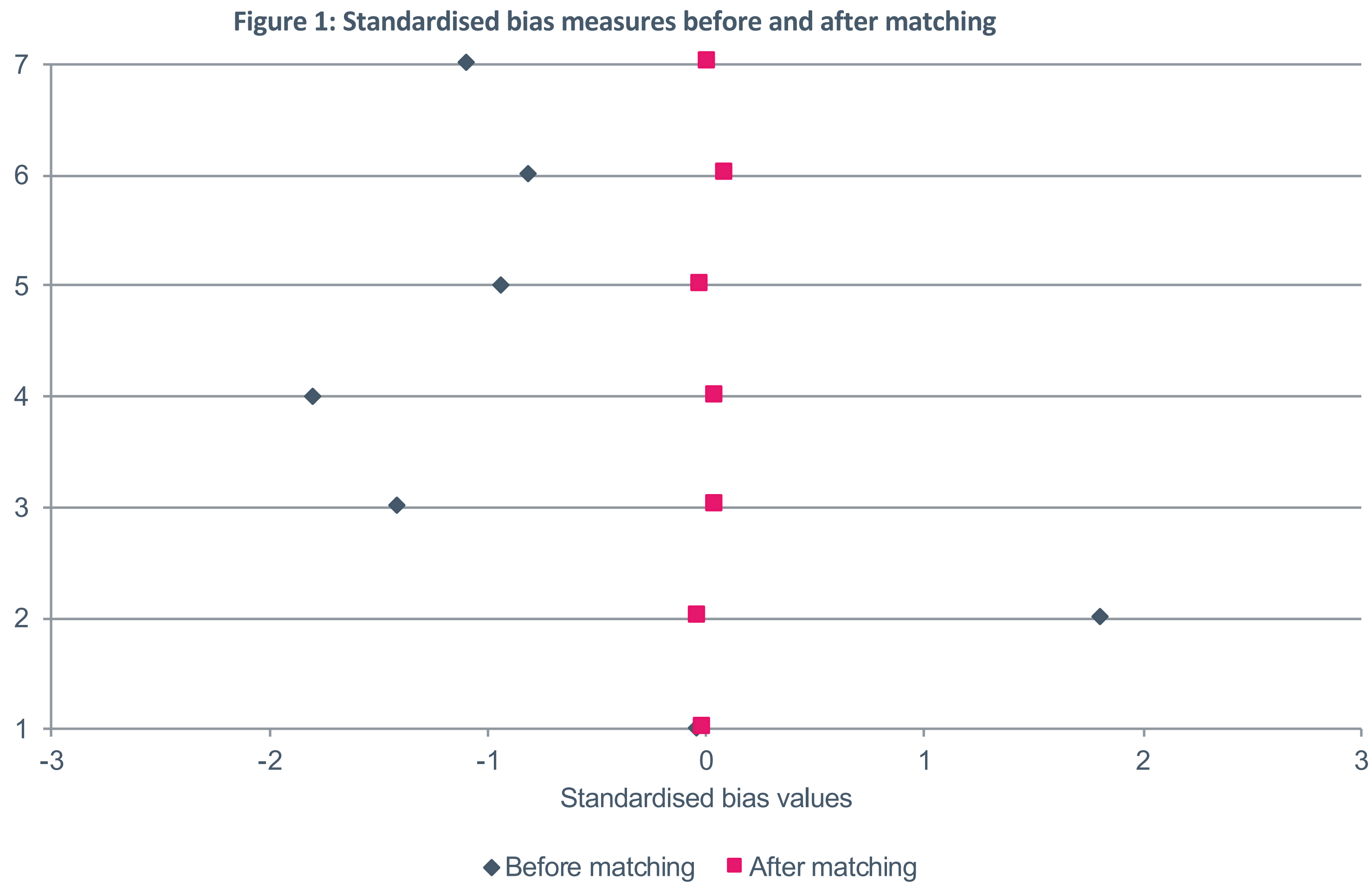
Scoping and operationalising research questions

Reading Recovery (RR) evaluation



Scoping and operationalising research questions

Reading Recovery (RR) evaluation



Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Formalise analysis approach

  - Step 1 – $\mathrm{Pr}(\mathrm{treat}_j = 1) = g(\sum_{c=1}^{37} \gamma_c \cdot x_{cj}),$

    $\text{for } j = 1,2, \ldots, n \text{ students, and } c = 1,2 \ldots, 37 \text{ control covarites}$

  - Step 2 – use 1-to-1 nearest neighbour matching with replacement and a caliper of 0.0001

  - Step 3 –

$$\mathrm{Pr}(y_j > k) = g(\beta_1 \cdot \mathrm{treat}_j + \beta_2 \cdot \mathrm{T4\ K\ Reading\ 2}_j + \beta_3 \cdot \mathrm{T4\ K\ Reading\ 3}_j + \beta_4 \cdot \mathrm{T4\ K\ Reading\ 4}_j + \beta_5 \cdot (\mathrm{treat}_j \cdot \mathrm{T4\ K\ Reading\ 2}_{js}) + \beta_6 \cdot (\mathrm{treat}_j \cdot \mathrm{T4\ K\ Reading\ 3}_j) + \beta_7 \cdot (\mathrm{treat}_j \cdot \mathrm{T4\ K\ Reading\ 4}_j) - \alpha_k),$$

$\text{for } j = 1,2, \ldots, n \text{ students, and } k = 1,2, \ldots, K-1 \text{ cutpoints}$

Scoping and operationalising research questions

Reading Recovery (RR) evaluation

- Formalise analysis approach
    - Step 1 – $\Pr(\text{treat}_j = 1) = g(\sum_{c=1}^{37} \gamma_c \cdot x_{cj})$,

      $\text{for } j = 1,2, \dots, n \text{ students, and } c = 1,2 \dots, 37 \text{ control covarites}$

    - Step 2 – use 1-to-1 nearest neighbour matching with replacement and a caliper of 0.0001

    - Step 3 –

$$\Pr(y_j > k) = g(\beta_1 \cdot \text{treat}_j + \beta_2 \cdot \text{T4 K Reading } 2_j + \beta_3 \cdot \text{T4 K Reading } 3_j + \beta_4 \cdot \text{T4 K Reading } 4_j + \beta_5 \cdot (\text{treat}_j \cdot \text{T4 K Reading } 2_{js}) + \beta_6 \cdot (\text{treat}_j \cdot \text{T4 K Reading } 3_j) + \beta_7 \cdot (\text{treat}_j \cdot \text{T4 K Reading } 4_j) - \alpha_k),$$

$$\text{for } j = 1,2, \dots, n \text{ students, and } k = 1,2, \dots, K-1 \text{ cutpoints}$$

Influencing the research design phase

Literacy and Numeracy Action Plan (LNAP) Phase 2 evaluation

- Features of the policy/initiative ($85M)
    - Instructional leadership
    - Diagnostic assessment
        - Transition from NSW Literacy and Numeracy Continua to new Learning Progressions
    - Differentiated teaching
    - Tiered interventions

Does the Strategy result in improved K-2 student outcomes?

- Transition to new Learning Progressions means we will not have the same outcome measures across treatment arms
- Use historical data to select a sample of similar schools for early transition to Learning Progressions

Influencing the research design phase

Literacy and Numeracy Action Plan (LNAP) Phase 2 evaluation

- Features of the policy/initiative ($85M)
    - Instructional leadership
    - Diagnostic assessment
        - Transition from NSW Literacy and Numeracy Continua to new Learning Progressions
    - Differentiated teaching
    - Tiered interventions

Does the Strategy result in improved K-2 student outcomes?

- Transition to new Learning Progressions means we will not have the same outcome measures across treatment arms
- Use historical data to select a sample of similar schools for early transition to Learning Progressions

Influencing the research design phase

Literacy and Numeracy Action Plan (LNAP) Phase 2 evaluation

- Our goal was to estimate the probability that each school would be involved in the LNAP in 2017 using historical student data
    - Schools with similar probabilities of treatment are expected to have similar levels of the observed covariates
    - Probability distributions can then be matched; approximating the conditions of an RCT

- We selected all the Kindergarten students who were enrolled in a public school in the first term of 2013 (n = 71,633), tracked their school movements and academic performance up to Year 3, and then used a school-level logistic regression model to estimate the conditional probabilities

Influencing the research design phase

Literacy and Numeracy Action Plan (LNAP) Phase 2 evaluation

- Our goal was to estimate the probability that each school would be involved in the LNAP in 2017 using historical student data
  - Schools with similar probabilities of treatment are expected to have similar levels of the observed covariates
  - Probability distributions can then be matched; approximating the conditions of an RCT

- We selected all the Kindergarten students who were enrolled in a public school in the first term of 2013 (n = 71,633), tracked their school movements and academic performance up to Year 3, and then used a school-level logistic regression model to estimate the conditional probabilities

Influencing the research design phase

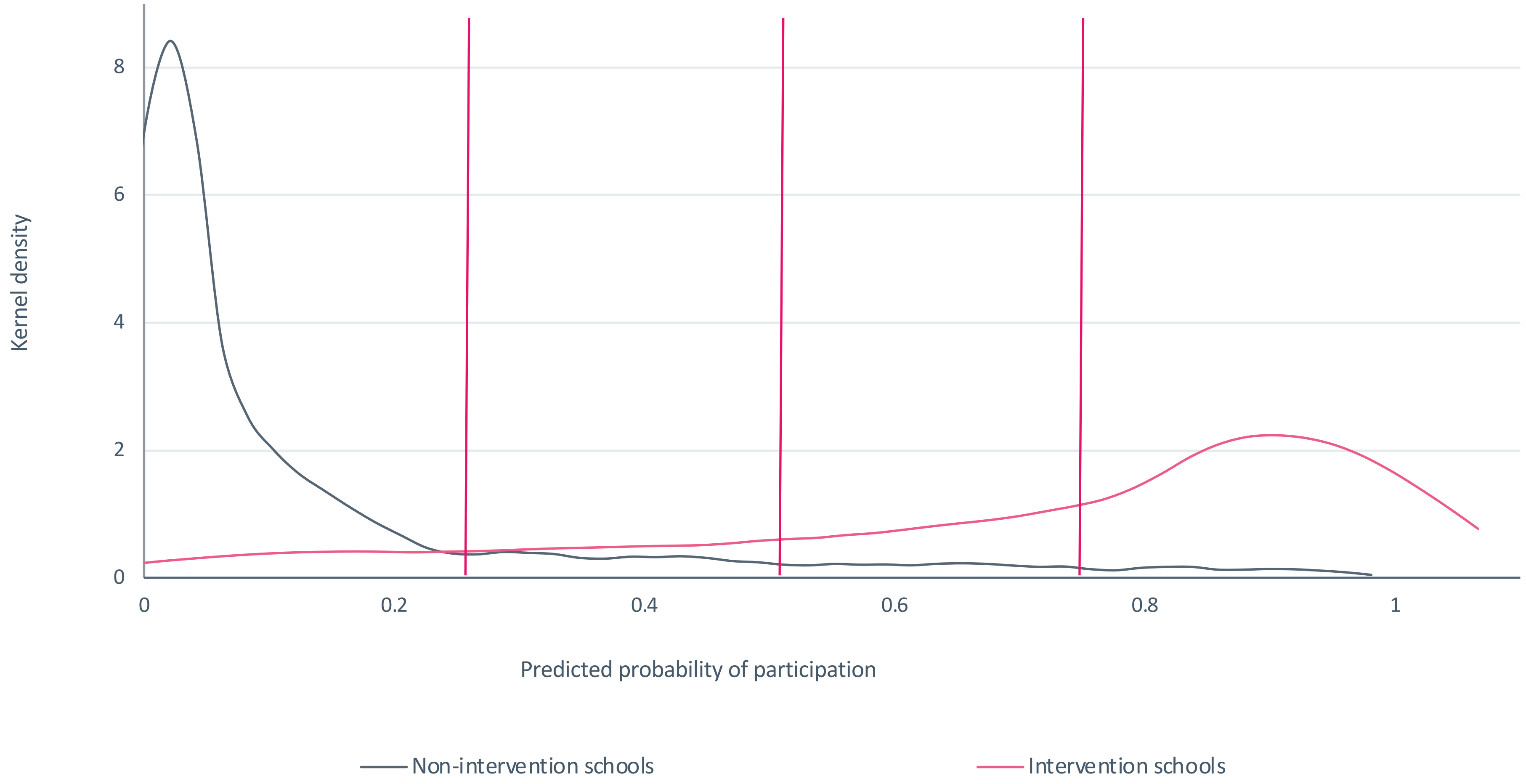


Table 2: Distributions of the intervention and non-intervention schools across the four strata					
	Strata 1	Strata 2	Strata 3	Strata 4	Total
Non-intervention schools	829 (86.26%)	73 (7.60%)	37 (3.85%)	22 (2.29%)	961
Intervention schools	50 (10.20%)	58 (11.84%)	96 (19.59%)	286 (58.37%)	490

Influencing the research design phase

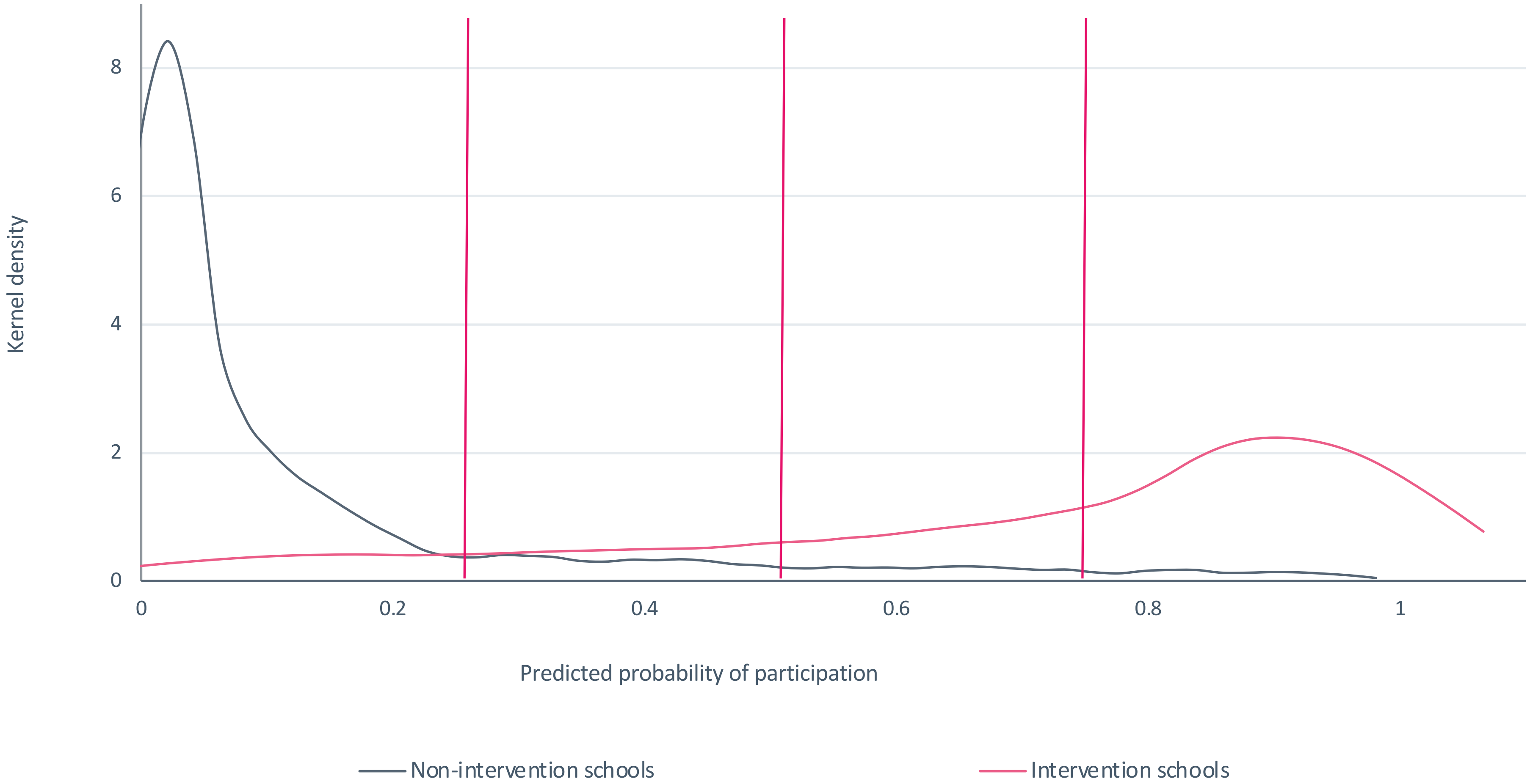
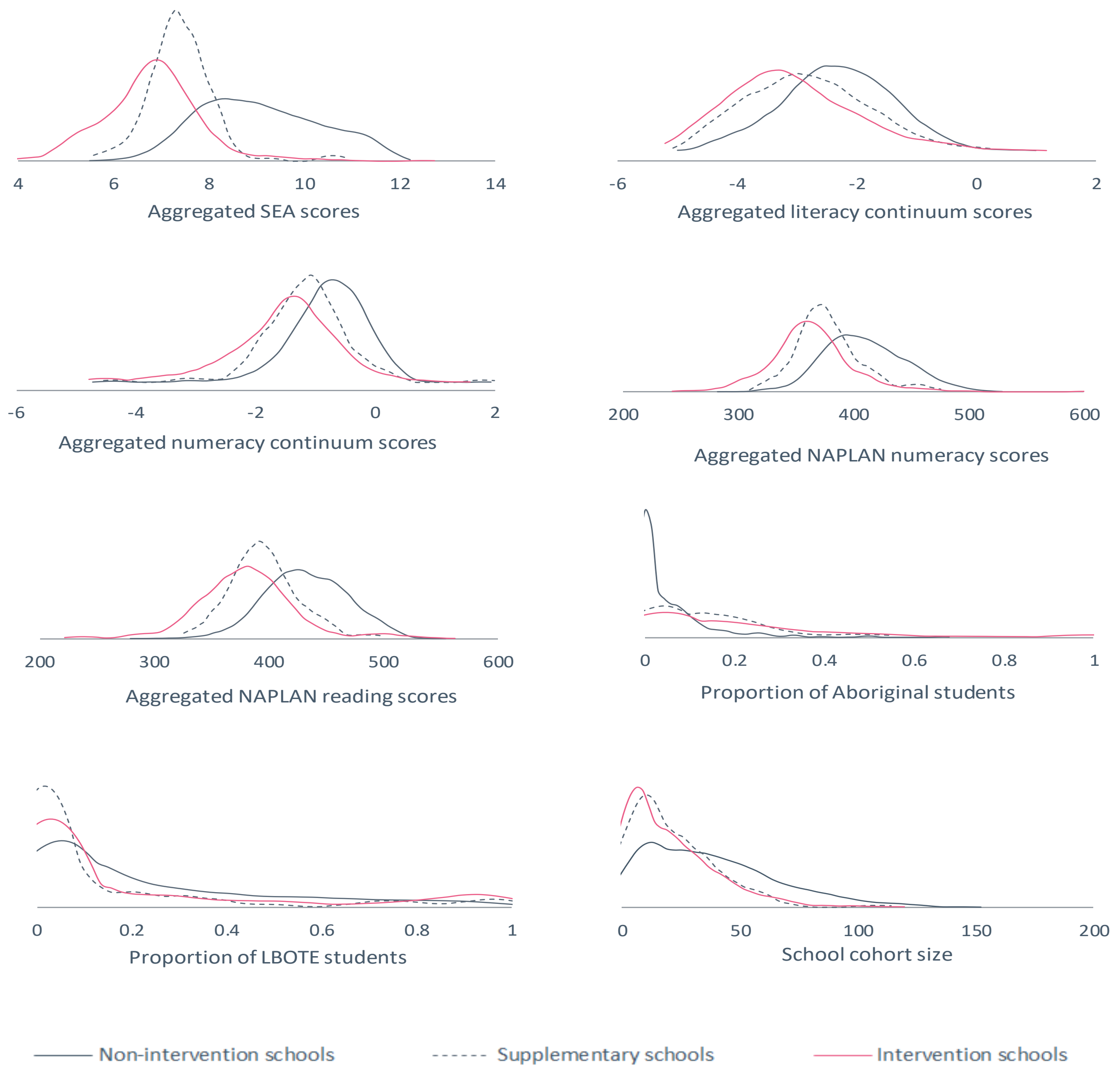
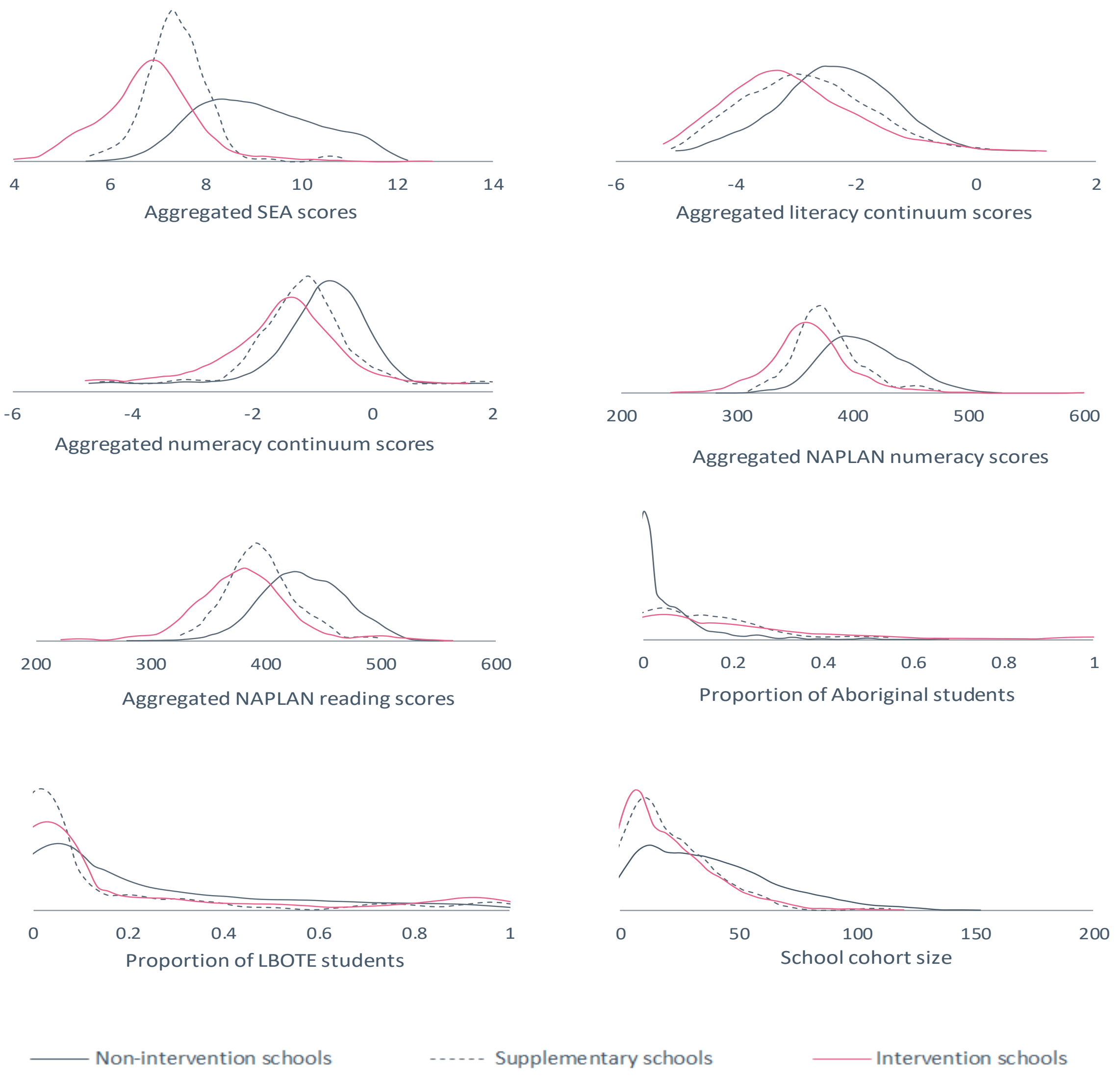


Table 2: Distributions of the intervention and non-intervention schools across the four strata					
	Strata 1	Strata 2	Strata 3	Strata 4	Total
Non-intervention schools	829 (86.26%)	73 (7.60%)	37 (3.85%)	22 (2.29%)	961
Intervention schools	50 (10.20%)	58 (11.84%)	96 (19.59%)	286 (58.37%)	490

Influencing the research design phase



Influencing the research design phase



Final thoughts

- Where possible, consider an RCT
  - But…

- Without randomisation, advanced statistical methods can help answer your evaluation questions: e.g.
  - Use program logic to look for differential effects (dose/response)
  - Create a pseudo-control group **after** the fact
  - Create a comparison group **before** data collection (without randomisation)

- It is possible to convey complex statistical findings in a easy to understand manner

Get in touch with CESE

- Send us an email – **info@cese.nsw.gov.au**
- Follow us on Twitter - **@nswcese**
- Sign up to our newsletter – **www.cese.nsw.gov.au/contact-us**