Running head: TV AND ATTENTION

Challenging the Link Between Early Childhood Television Exposure and Later Attention Problems: A Multiverse Analysis

Matthew T. McBee1

Rebecca J. Brand2

Wallace Dixon1

1Department of Psychology, East Tennessee State University, Johnson City, TN.

2Department of Psychology, Villanova University, Villanova, PA 19085

**Abstract**

The claim that early childhood television exposure causes later attention problems (Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004) seems to remain strongly held by the popular media as well as by researchers in the field. Using the same NLSY-79 dataset and a superordinate set of covariates, we conducted a multiverse analysis to examine the degree to which the finding reported by Christakis et al. was dependent on analytic choices. We evaluated forty separate models treating attention as a continuous variable and forty-two models treating attention as categorical. Only 6 of these 82 models produced significant results. We conclude that these data offer little evidence for a detrimental causal effect of early childhood use TV on mid-childhood attention problems. Preprint: [link redacted for peer review]. Project: [<https://goo.gl/doxyuG>]

Challenging the Link Between Early Childhood Television Exposure and Later Attention Problems: A Multiverse Analysis

Psychological science is capable of having a broad and deep impact on human lives. In developmental psychology in particular, there is a sense of relevance, indeed urgency, to many of its questions: Is it helpful or harmful to grow up multilingual? Do vaccines cause autism? Does screen time cause attention deficits? Research in this field has the potential to reveal which behaviors, products, and choices are harmful to development and which give children the best chance to grow up happy and healthy. The stakes are high; it is crucial that scientists get it right.

Once an erroneous finding has been disseminated via the media, it is nearly impossible to put Pandora back in the box. Take, for example, the supposed link between vaccines and autism. Even 20 years after Andrew Wakefield’s fraudulent 1998 report (Wakefield et al., retracted) implied a link between the MMR vaccine and autism, and despite numerous findings that such a link does not exist (Committee to Review Adverse Effects of Vaccines, 2012), a substantial proportion of the public still believes that it does (Oliver & Wood, 2014). Cases like this illustrate how important it is that researchers make their best attempt to disseminate accurate findings, and how crucial it is that we engage in and disseminate replication attempts, especially for high-impact findings. We follow Nature Editors (2016) and others in arguing that doing so is a duty to our profession.

Over a decade ago, Christakis and colleagues (Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004) published a paper claiming a relationship between television exposure in toddlerhood and subsequent attention problems at school age. Although longitudinal in nature and including a variety of control variables, the lack of randomized manipulation of TV use made it inappropriate to draw strong causal conclusions from these data. Nevertheless, such causal claims were made, by the lead author himself and subsequently by media outlets the world over. Using Google search in April 2018 for “Does TV cause attention problems,” the first six hits all claim a link between TV and attention problems. One hit (WebMD) uses blatantly causal language in its headline (“Toddler TV Time Can Cause Attention Problems”) and another published by whitedot.org (Lotus, 2018) quotes Christakis as saying “TV ‘rewires’ an infant’s brain,” and says his study shows that “TV watching is a cause [of ADHD].” Although findings described on these pages are wildly overstated and out of date, we can find no expiration dates or retractions associated with them.

Christakis’s story was attractive to the public for many reasons. Since the late 1990s, there has been a growing public interest in how early experiences impact later development (e.g., Nash, 1997). But as far back as the 1950s, new media technologies have been greeted with fear and skepticism (Goode & Ben-Yehuda, 2010; Markey & Ferguson, 2017). Further, there is an intuitive face validity to the claim that TV harms children’s attention. Infants’ attention is undergoing rapid development (Ruff & Rothbart, 1996), and the stimulating pace of screen media rarely resembles the slower pace of real life. Even recommendations from the American Academy of Pediatrics to avoid television viewing seemed to endorse the claim that TV was inherently harmful for young children (AAP, 1999; Children and Media, 2018).

All in all, parents of children born in the last 20 years had ample reason to believe that TV-watching caused attention disorders. Yet recent research indicates that this may not be true at all. A re-analysis of the data set used by Christakis et al. (2004) indicated that the finding was not robust in the face of certain small changes in analytic parameters (Foster & Watkins, 2010). A subsequent meta-analysis on screen media use and attention problems indicated that the relationship between the two is a small to moderate one, but was unable to clarify the direction of causality or the potential involvement of third variables (Nikkelen, Valkenberg, Huizinga, & Bushman, 2014). A recent review came to a similar conclusion (Kostyrka-Allchorne, Cooper, & Simpson, 2017). Tellingly, however, while the paper making a claim of a causal link (Christakis et al., 2004) has 1008 citations in Google Scholar at the time of this writing, the more methodologically sound critique (Foster & Watkins, 2010) has 76 citations and the meta-analysis (Nikkelen et al., 2014) has only 54. The general public is still left with the misleading message that TV causes attention problems, full stop.

Our goals in this paper were two-fold. First, following the suggestion of Nikkelen et al. (see also Valkenburg & Peter, 2013) that the possible link between TV and attention should be examined in light of potential confounders, we wanted to include children’s temperament in the analysis. We view temperament as potentially a key confounding variable that could both result in childhood TV use and manifest later in childhood as attention deficits, and doubt that a credible causal effect can be identified without controlling for it.

Temperament includes the ability to regulate one’s own attention (Posner & Rothbart, 2018; Smith et al., 1997; Thomas, Chess, & Birch, 1968), and as one would predict, certain temperament dimensions predict children’s later attention problems (Auerbach et al., 2008; Gurevitz, Geva, Varon, & Leitner, 2014; Sullivan et al. 2015). In addition, those infants with difficult temperaments may be shown more screen media than other children, as a way to keep them calm and engaged (Brand, Dixon, & Hardesty, 2011). In support of this claim, parents’ perception of infants’ energy level (Nabi & Krcmar, 2016), poor self-regulation (Radesky et al., 2014), and fussiness (Thompson et al., 2013) all predict TV use. In short, we suspected that relationships between early television and later attention problems, to the extent that they exist, might be driven by their shared connection to early attention problems (as measured through temperament).

Our second and more encompassing goal was to further examine the robustness of the original finding through the use of a “multiverse analysis” (Silberzahn et al. 2017; Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). In any research endeavor, a series of analytic decisions must be made, and some of these decisions are necessarily arbitrary (King & Zeng, 2007). These analytic decisions have been described as the “garden of forking paths” (Gelman & Loken, 2013). Trouble occurs when different paths or models lead to different conclusions. The degree to which this is true is called *model dependence*. When most paths through the garden lead to a null finding, an unaware or unscrupulous researcher may tweak analytic decisions until a significant p-value is obtained (termed “p-hacking;” Simmons, Nelson, & Simonsohn, 2011), thus producing a “successful” study but a misleading claim about reality. One way to evaluate the model dependence of a claim, and thus prevent p-hacking, is to subject the data to a wide variety of defensible analyses, to illustrate systematically how sensitive the outcome is to different model specifications. This is what we have done here.

In this paper, we present a multiverse analysis of Christakis et al.’s (2004) original claim, using the same dataset. We incorporated a larger set of control variables than found in either the original paper or the Foster and Watkins re-analysis (2014), including among them children’s temperament, in which we had a particular interest. We then subjected this data set to a wide variety of analyses across three general types: propensity score analysis, linear regression, and logistic regression.

**Propensity Score Analyses.** Given the nature of the data set, we believed that propensity score analysis was the most defensible choice for estimating the causal effect of TV watching on later attention problems. A propensity score is the probability of being in the treatment group (in this case, the group being shown a large amount of TV), conditional on a variety of baseline characteristics (such as mother’s education, household income, and child temperament; cf. Austin, 2011). Once the propensity scores have been estimated, they can be applied via a non-parametric technique such as matching, weighting, or stratification in order to produce a virtual sample that is balanced in expectation on all of the covariates. In this manner, propensity score analysis approximates a randomized experiment with respect to the measured covariates included in the propensity score model (Rosenbaum & Rubin, 1983). Unlike a true experiment, however, propensity score methods do not balance on unobserved or omitted covariates.

Within the propensity score family of analyses, there are still many decisions to be made in the garden of forking paths. In our analysis set, we varied age when TV use was measured (approximately 1.5 versus 3 years) as well as how the outcome was treated (raw vs. standardized attention scores). Because an effect of TV watching might plausibly be different for those who watch a lot of TV versus the average child, we estimated both the average treatment effect (ATE) and the average treatment effect for the treated (ATT). Since propensity score analysis requires dichotomizing the predictor variable, we ran analyses with two different cut points for the high and low TV groups: a median split for one set of analyses and more extreme cut points (greater than 80th percentile and less than 20th percentile) for the other. Finally, we identified a set of covariates with some of the largest residual imbalance statistics and gave those covariates an additional regression adjustment. We ran analyses both with and without this doubly-robust strategy (Guo & Fraser, 2015). Finally, we applied the propensity scores either as weights (e.g., inverse probability of treatment weights) or used them to perform stratification. The result was 36 separate propensity score analyses.

**Linear Regression.** Linear regression models are less robust than propensity score models yet offer substantially enhanced efficiency (in the form of smaller standard errors) and increased statistical power. These models allowed us to detect and more precisely estimate weak effects that could have remained hidden in the noise of the propensity score models, albeit with more risk of exposure to systematic bias due to assumption violations. We ran four linear regression analyses, estimating both raw and standardized attention scores from TV watching at both ~1.5 and ~3yrs.

**Logistic Regression.** Finally, as a close replication attempt of the original study, we analyzed the data set using logistic regression, in spite of our belief that this approach is unjustified given the continuous and apparently linear nature of the response variable. Christakis et al. (2004) divided the continuous attention/behavior problems scale into typical and problematic levels of attention deficit based on a *z =* 1.2 cut point (corresponding to a score on the Behavior Problems Index of 120). The authors argued that using this cut point yielded a rate of problematic attention similar to its incidence in the population. But would their conclusions regarding TV’s effects have been the same if they had used a cut point of 119 or 121? We performed 42 analyses, systematically varying the cut point between problematic and non-problematic attention scores and examining outcomes for TV-watching at both ~1.5 and ~3years. Thus, across the three different sets of analyses, we examined the relationship between early TV exposure and later attention problems in this data set in 82 distinct ways.

**Method**

**Data**

As in Christakis et al. (2004), data for the present investigation were obtained from the National Longitudinal Survey of Youth 1979 (NLSY-79), available via the NLS Investigator web interface (2018). Child data came from the NLSY79 Child and Young Adult dataset. Information on the mothers of these children came from the original NLSY-79 dataset. These datasets were merged via a common ID code variable allowing mother and child data to be linked. We initially downloaded 321 variables from the Child and Young Adult dataset and 41 variables from the NLSY79 dataset(NLSY, 2018). Our project’s Open Science Framework (OSF) page (anonymized link for peer review: <https://goo.gl/doxyuG>) presents a spreadsheet mapping our analysis variables to the variable codes and labels from the NLSY dataset (see “documentation” component). Our downloaded and processed data as well as our analysis code are disclosed on this site, allowing interested readers to replicate or extend our analysis.

Our variable selection process was based on the one reported in the original paper with some additions. As per Christakis et al. (2004), we selected three cohorts of children who were approximately 7 years old during the three “index years” of 1996, 1998, and 2000. Our baseline variable selections matched the original study to the extent possible given the brief description in the original paper, which did not report ID codes for the selected variables. In most cases, we could unambiguously identify variables by searching the NLSY data by question text or question title.

Most of these variables are based on survey questions that were repeatedly administered on a biennial basis. We were interested in these variables as potential confounders (common causes) of both early childhood TV exposure and mid-childhood attention deficits or hyperactivity. Thus, nearly all of these variables were selected from survey administrations contemporaneous with the TV exposure observation. Two exceptions were maternal self-esteem, which was asked only in 1987, and maternal depression (CES-D), which was assessed only in 1992. Depending on the cohort, depression could have been assessed up to four years before birth or the same year the child was born; and self-esteem from one to five years before birth. In spite of this problem of timing, we included these two variables because the original paper did. In any case, we expected a moderate degree of stability over time in these constructs (Lovibond, 1998; Trzesniewski, Donnellan, & Robins, 2003), which may ameliorate some concern about the timing of their measurement.

**Selection of cases.** We followed the original paper’s criteria for sample selection. For each index year (1996, 1998, and 2000), we included those children whose ages at index were between 6 years 9 months and 8 years 9 months.­ A total of 2,145 cases were extracted.

**Variables.** As in the original study, our measure of attention was the standardized score on the hyperactivity subscale of the five-item Behavior Problems Index (BPI), which was standardized to an IQ-like metric (M = 100, SD = 15) within sex, as per the original study. However, we also retained the raw attention scores which were unadjusted for sex. The five items addressed children’s ability to concentrate and pay attention, as well as their confusion, impulsivity, obsessions, and restlessness or inability to sit still. We created a dichotomized variable for our logistic regression models based on the following description from the original paper:

We created a binary classification representing attentional problems as either present or absent, using a cut point of 120 on the same-gender standardized BPI subscale score. That is, children with scores 1.2 standard deviations (SDs) above the mean were classified as having attentional problems (Christakis et al., 2004, p. 709)

Television use was calculated as in the original study. Items measuring hours per day of television watched by the child on both weekdays and weekends days were converted to average hours of TV by multiplying weekday hours per day by five, adding to this weekend hours per day multiplied by two, and dividing by seven. We took this measurement from three and two waves prior to the index year, such that TV was measured at approximately age 1.5 and age 3, though the exact age of each child during these waves could vary to some extent. Following the original study, other variables included the mother’s race, the child’s gender, the number of children of the mother living in the household, mother’s highest grade completed, a binary indicator of maternal alcohol use and cigarette smoking during pregnancy, a binary indicator of whether the child’s father lived in the household, maternal self-esteem as assessed by the Rosenberg Self-Esteem Scale in 1987, and maternal depression as measured by the CES-D in 1992. Rather than a continuous gestational age at birth variable, we created a binary indicator of pre-term delivery (child born before 37 weeks of gestation). Instead of the original study’s urban/rural indicator variable, we incorporated the four levels of the Statistical Metropolitan Sampling Area classification. We did not include calendar year at index as a covariate. Finally, we did not incorporate sample weights into our analyses, as it is unclear how they should be incorporated into stratified propensity score models or combined with propensity score weights.

We added the following covariates, which we suspected to be plausible confounders for TV use and childhood attention: family income, spouse or partner’s educational attainment, child BMI, a binary indicator of poor health (e.g., does the child have a medical condition that limits ordinary childhood activities?), and a binary indicator of low birth weight (weight less than 2,500 grams or 5 lbs, 8 oz). All of these were extracted from the first wave of data availability to avoid conditioning on post-treatment variables, since they could have potentially biased our estimates if they were mediators or colliders (Montgomery, Nyhan, & Torres, 2018; Rohrer, 2018).

Finally, we created a variable assessing infants’ temperament. According to the NLSY website (NLSY Temperament, 2018), the temperament scale included items taken from work by Mary Rothbart, Joseph Campos, and Jerome Kagan. We chose the six available items that represented aspects of difficult temperament, as defined by Rothbart and Bates (2006), which included irritability, high-intensity affect, and negative mood. These items included assessments of how often the child cries when seeing a stranger, how often she is afraid of dogs or cats, how often she cries with doctors or nurses, how often the caregiver has trouble calming the child, and how often the child cries compared to others. Our temperament variable was the mean of these items, each of which was represented on a 5-point scale.

It was necessary to correct some out-of-range values prior to analysis. We set any BMI value outside the range of 13-22 (based on CDC norms) to missing, and truncated the following variables to the top of their ranges: TV use in average hours per day exceeding 16 (following the original study), highest grade completed exceeding 24 (as this would imply more than eight years of post-graduate education), and annual income of $839,078 (the NLSY description of this variable includes a comment that this value is probably untrustworthy).

The file “variable name propagation spreadsheet.xlsx” on the OSF page (under “Documentation”) provides a crosswalk from our substantive, conceptual variable names to NLSY alphanumeric variable names. The analysis code is the canonical description of how the variables were constructed and should resolve any vagueness or ambiguity in the preceding description.

**Results**

**Descriptive statistics**

Tables 1 and 2 provide descriptive statistics for the continuous and categorical variables. These were formatted using the *stargazer* package (Hlavac, 2015). Figure 1 displays histograms of TV use measured at ages 1.5 and 3. Recall that propensity score analysis requires that the treatment variable be categorical, and that we used two different sets of cut points (50th percentile and below-20th vs. above-80% percentile) to define the low- and high-TV groups. Table 3 provides descriptive statistics for TV use for dual sets of low- and high-TV categories based on the age 1.5 and age 3 data. Space limitations prevented us from including the descriptive statistics broken down by TV category by age of measurement and cut point. They can be found on our project’s OSF page ([goo.gl/93uWt4](file:///C:\\Users\\rbrand\\AppData\\Local\\Microsoft\\Windows\\Temporary%20Internet%20Files\\Content.Outlook\\SFRHOTIT\\goo.gl\\93uWt4)) under Tables → Descriptives.

Figure 2 displays a set of scatterplots displaying the relationship between TV consumption and within-sex standardized attention measured at age 7. The lack of any obvious systematic relationship between TV and attention is apparent. These observations hold regardless of whether TV use is measured at age 1.5 or age 3, whether the response variable is adjusted for covariates or not, and whether the standardized or raw attention measure is considered. We note that a similar figure was not presented by Christakis et al (2004). Doing so would likely have dramatically reduced the credibility of a claimed link between TV and attention deficits.

**Propensity Score Models**

We fit a total of 36 different propensity score analysis models to the data. The models varied on the following dimensions:

* **Outcome**: raw versus standardized attention
* **Use of propensity scores**: weights versus stratified analysis
* **Age when TV use was measured:** 1.5 versus 3 years
* **Additional covariate adjustment:** yes (“doubly robust”) versus no
* **Causal effect estimand**: average treatment effect (ATE) versus average treatment effect for the treated (ATT)
* **Cut points for defining high and low TV groups**: < 20th / > 80th percentiles versus median split

The stratified analyses could not estimate the ATT, could not incorporate additional post-stratification covariate adjustment, and could not incorporate sample weights. Thus, there are a total of four stratification models, 2 (outcome) x 2 (TV age), and 32 models that use the propensity scores as inverse probability of treatment weights (IPTW; Guo & Fraser, 2015): 2 (outcome) x 2 (TV age) x 2 (doubly robust) x 2 (estimand) x 2 (cut points). The *R* packages *twang* v1.5 (*Toolkit for Weighting and Analysis of Non-Equivalent Groups;* Ridgeway, McCaffrey, Morral, Griffin, & Burgette, 2017) and *survey* v3.33 (Lumley, 2017) were used to fit the weighting models, while the *PSAgraphics* package v2.1.1 (Helmreich & Pruzek, 2009) was used to fit the stratification models. The *twang* package uses boosted classification trees to estimate the propensity scores such that the covariate balance is optimized, therefore slightly different sets of propensity scores were generated for our models estimating ATE versus ATT.

We examined three pieces of diagnostic information to detect possible problems with the propensity score models that could have resulted in erroneous results. We computed the model’s correct classification rate (“hit rate”) in order to verify that the covariates were adequately related to TV use. Our models generated hit rates in the 77-82% range. We also examined the distribution of propensity scores for the low- and high-TV groups in order to determine whether adequate common support existed to justify proceeding with the analysis. Figure 3 provides a typical example plot; the rest are presented on our project’s OSF page (Figures → Diagnostic: Propensity Score Distribution by Groups). Common support was evinced for all the models. We also examined balance statistics after applying the propensity scores to determine whether covariates were adequately balanced across groups. These are presented on the project site in both textual and graphical form.

Figures 4 and 5 display the relative influence of the covariates in the propensity score models; these display the strength of relationship between each variable and the model’s classification decisions. Child age was a strong predictor of high TV use, especially in the ~1.5 age group. Recall that due to the timing of data collection waves, children’s age at the ~1.5 time point could actually range quite a bit. Finding that TV use increases sharply from early infancy to toddlerhood matches previous findings (Anand & Krosnick, 2005; Duch, Fisher, Ensari, & Harrington, 2013). Other important predictors of TV use in both age groups were maternal depression, the cognitive stimulation of the home, income, the emotional support of the home, and maternal self-esteem, which also matches previous findings (Anand & Krosnick, 2005; Certain & Kahn, 2002; Vaala & Hornik, 2014; Vandewater et al., 2007). Confirming our suspicions, temperament was also predictive of TV use in this data set, falling in the moderate range, comparable to parent education level and child BMI. A follow-up simple linear regression of TV at age 3 on temperament revealed a statistically significant relationship, b = 0.484 (0.103), *p* < .001. However, the model R2 was only .011.

We also examined plots of the covariate imbalance before and after applying the propensity scores in order to ensure that the models produced adequate balance (available on the OSF page). These plots illustrate that nearly all covariate imbalance was reduced to *d* = 0.2 or smaller when TV was measured at age ~1.5, and to *d* = 0.1 or smaller when TV was measured at age ~3.

In the stratified propensity score models, we created five strata based on the quantiles of the marginal distribution of propensity scores. Therefore, each stratum encompassed equal sample sizes. The stratum boundaries for TV at age ~1.5 were located at the 26th, 38th, 50th, and 62nd percentiles. The boundaries for TV at age ~3 were located at the 40th, 46th, 51st, and 58th percentiles. Figures 6 and 7 display covariate balance for continuous and categorical variables for the stratified propensity score models. In all cases, we deemed the achieved covariate balance to be acceptable. In all cases, we deemed the achieved covariate balance to be acceptable.

**Linear regression models**. Next, we ran a series of linear regression models predicting attention outcomes from early TV watching. Due to missing data on many variables, listwise deletion would have resulted in an analysis based on only about 40% of the sample. We therefore employed multiple imputation via chained equations (MICE) via the *mice* package (van Buuren & Groothuis-Oudshoorn, 2011) to construct multiply-imputed datasets. The advantage of MICE over classical implementations of multiple imputation is that MICE allows each variable to follow its own distribution rather than assuming that the entire data matrix is multivariate normal. Our continuous variables were imputed using predictive mean matching, binary variables were imputed via logistic regression imputation, nominal variables with greater than two levels by polytomous regression imputation, and ordered categorical variables with more than two levels by proportional odds regression imputation. We created ten imputed datasets based on fifty iterations.

Linear regression models were fit to the data including the covariates as control variables. Categorical variables were dummy coded. The focal predictor in each model was the continuous TV use variable (measured at age ~1.5 or ~3), which was rescaled such that a one-unit change resembled the magnitude of moving from the low- to high-TV group in the propensity score models using the 20th percentile / 80th percentile cut points. Complete regression tables may be found on our OSF page.

Figures 8 and 9 summarize the results of these propensity score and linear regression analyses for the within-sex standardized attention and the raw attention scores, respectively. The 95% confidence intervals for 36 out of 40 models contain zero. Therefore, only four of 40 versions of the analysis reject the null hypothesis that early childhood TV use is associated with attention at the conventional level of significance. Further, it is clear from the figures that even the estimated effect sizes that are statistically significant are miniscule.

**Logistic regression.** As previously discussed, it is inappropriate to use logistic regression to analyze these data because the outcome variable is continuous. However, we report results from logistic regression models because they comport with the results from the original paper. Doing so allowed us to explore the consequences of varying the 120 threshold that Christakis et al. employed for defining a problematic level of attention deficit. We used the same multiply-imputed dataset that we used for the linear regression models. For each analysis, we report the odds ratio of the relationship between TV consumption at ages ~1.5 and ~3 and the probability of being in the high “problematic” category of attention based on setting the thresholds from 110 to 130 after controlling for the covariate set. Results are summarized in Figure 10.

As shown in the figure, the results of the logistic regression analysis are highly sensitive to the choice of cut point. Statistically significant (though quite imprecise) estimates of the relationship between TV (at age ~3) and attention emerge for cut points of 123 or 124, with the remainder non-significant. All of the confidence intervals for TV measured at age ~1.5 correspond with non-significant hypothesis test results as they include an odds ratio of 1.0.

**Discussion**

The broad goal of this paper was to determine the model dependency of the claim that early TV watching causes attentional problems (Christakis et al., 2004). In other words, to what degree does finding a relationship depend on the analysis model used? As only 6 out of 82 analyses suggested a causal relationship, all with tiny effect sizes, we assert that the claim is highly model-dependent.

Indeed, starting with the most straightforward method of visualizing the relationship -- a simple scatterplot with TV watching and attention problems as shown in Figure 1 – suggests that the purported relationship is illusory. This plot reveals essentially a flat line when TV use is measured at age 1.5 and a mildly s-shaped curvilinear relationship between TV and attention when TV use is measured at age 3 (top right panel). This curvilinear relationship, such as it is, is nearly dampened out of existence by the introduction of covariates (bottom right panel).

Other methods of estimating the relationship only further strengthen our opinion that no relationship exists in these data between TV and attention. Specifically, none of the propensity score analysis variants using the within-sex standardized outcome, nor any analyses using linear regression, provided any evidence for the claim. Only four of the propensity score models using the raw attention score, and two of 42 logistic regressions corresponding closely to the original choice of cutpoint, produced statistically significant results. Given the tiny minority of analytic paths that produced a significant result, we think the most reasonable conclusion is that there is *no causal effect* of hours of TV watching on attention problems, at least as defined by scores in the Behavior Problems Index.

We also examined the role of temperament in connection with TV watching and later attention problems. Our hunch at the outset of this project was that any relationship between early TV-watching and later attention problems might be the result of the third variable of temperament. In fact, however, there was little sign of a meaningful relationship between TV and attention to be explained. We did find the predicted link between children’s temperament and their early TV-viewing to some degree. Specifically, temperament emerged as a moderately important predictor in the propensity score models for TV use at age 3 but not at age 1.5. However, as our follow-up simple regression shows, in practical terms this is very small effect.

In some ways, the field has already moved beyond the broad-brush claims from the original paper: that hours of TV (in general) cause attention problems (for everyone). As Kostyrka-Allchorne et al. put it, it seems foolish to expect that screen time “as an undifferentiated activity,” (2017, p. 52) predicts much of anything. Recent research about screen media use in children has gotten more precise – investigating the specific effects of violent content, fantastical content, pace of scene-change, and the viewer’s voluntary control of the action, among other factors (Huber et al., 2018).

In one such line of work, Lillard and Peterson (2011) found that certain cartoons appear to temporarily attenuate children’s executive functioning, including planning and delay of gratification. The culprit in these cartoons was first thought to be the fast pace of the scene changes, but subsequent work suggested that the fantastical content seemed to be the cause (Lillard, Drell, Richey, Boguszewski, & Smith, 2015). Two things are notable about this line of research, however. The first is that it (and similar lines of research) was founded on the desire to locate a *mechanism* for the purported negative effect of TV – an effect that our multiverse analysis suggests is probably nonexistent. Second, while the experimental approach taken by Lillard and colleagues is thoughtful and well-controlled, the negative effects on executive functions are short-term in nature, and the authors note that it is unclear whether this might lead to long-term deficits. It seems equally plausible that fantastical content exercises the executive functions in much the same way as running exercises the leg muscles. Leg strength may be drained after a sprint, but improved in the long term. There is even some evidence that fantastical components to stories and problems lead to *improved* learning in preschoolers (Weisberg, Hirsh-Pasek, Golinkoff, & McCandliss, 2014). In any event, it would seem premature to suggest that children be shielded from fantastical content, which is such a rich part of childhood.

In summary, the multiverse analysis presented in this paper used a large, nationally representative dataset to ask the same question in 82 different ways: Is there any reason to believe that TV watching in early childhood causes attention problems in later childhood? In only six of these 82 analyses was the answer “yes.” Thus, we think the true answer is likely “no.” In fact, screen media may not be all that special. It may be just one more part of life that has the power to entertain, teach, confuse, distract, or inspire.

**References**

American Academy of Pediatrics, Committee on Public Education. (1999). Media education.

*Pediatrics*, *104*, 341-343. doi:10.1542/peds.2011-1753

Anand, S., & Krosnick, J. A. (2005). Demographic predictors of media use among infants, toddlers, and preschoolers. *American Behavioral Scientist*, *48*(5), 539-561. doi: 10.1177/0002764204271512

Auerbach, J. G., Berger, A., Atzaba‐Poria, N., Arbelle, S., Cypin, N., Friedman, A., & Landau, R. (2008). Temperament at 7, 12, and 25 months in children at familial risk for ADHD. *Infant and Child development*, *17*(4), 321-338. doi: 10.1002/icd.579

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of

confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399-424.

Brand, R.J., Hardesty, S., & Dixon, Jr., W.E. (2010, March). *Toddlers’ difficult temperament*

*predicts television use*. Poster presented at the Biennial Meetings of the Society for

Research in Child Development, Montreal, Quebec, Canada.

Certain, L. K., & Kahn, R. S. (2002). Prevalence, correlates, and trajectory of television viewing among infants and toddlers. *Pediatrics*, *109*(4), 634-642. doi: 10.1542/peds.109.4.634

Children and Media – Tips for Parents (2018). Retreived from: https://www.aap.org/en-

us/about-the-aap/aap-press-room/pages/Children-And-Media-Tips-For-Parents.aspx

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L, & McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics, 113*(4), 708-713. doi: 10.1542/peds.113.4.708

Committee to Review Adverse Effects of Vaccines. (2012). *Adverse effects of vaccines:*

*Evidence and causality*. Washington, DC: National Academies Press.

Duch, H., Fisher, E. M., Ensari, I., & Harrington, A. (2013). Screen time use in children under 3 years old: a systematic review of correlates. *International Journal of Behavioral Nutrition and Physical Activity*, *10*(1), 102.

Foster, E. M. & Watkins, S. (2010). The value of reanalysis: TV viewing and attention problems. *Child Development, 81*(1), 368-375. doi: 10.1111/j.1467-8624.2009.01400.x

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be

a problem, even when there is no “fishing expedition” or “p-hacking” and the research

hypothesis was posited ahead of time. Unpublished manuscript. *Department of Statistics,*

*Columbia University*. Retrieved from <http://www.stat.columbia.edu/~gelman/research/>

unpublished/p\_hacking.pdf

Goode, E., & Ben-Yehuda, N. (2010). *Moral panics: The social construction of deviance*. West

Sussex, UK:Wiley.

Guo, S. Y. & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and application* (2nd ed.). Thousand Oaks, CA: SAGE.

Gurevitz, M., Geva, R., Varon, M., & Leitner, Y. (2014). Early markers in infants and toddlers for development of ADHD. *Journal of Attention Disorders*, *18*(1), 14-22. doi: 10.1177/1087054712447858

Helmreich, J. E. & Pruzek, R. M. (2009). PSAgraphics: An R Package to support propensity score analysis. *Journal of Statistical Software, 29*(6), 1-23. doi: 10.18637/jss.v029.i06

Hlavac, M. (2015). *stargazer: Well-formatted regression and summary statistics tables.* [R package]. Version 5.2. <http://CRAN.R-project.org/package=stargazer>

Huber, B., Yeates, M., Meyer, D., Fleckhammer, L., & Kaufman, J. (2018). The effects of screen media content on young children’s executive functioning. *Journal of experimental child psychology*, *170*, 72-85.

King, G. & Zeng, L. (2007). Detecting model dependence in statistical inference: A response. *International Studies Quarterly, 51.* 231-241.

Kostyrka-Allchorne, K., Cooper, N.R., & Simpson, A. (2017). The relationship between

television exposure and children’s cognition and behaviour: A systematic

review. *Developmental Review*, *44*, 19-58. <https://doi.org/10.1016/j.dr.2016.12.002>

Lillard, A. S., Drell, M. B., Richey, E. M., Boguszewski, K., & Smith, E. D. (2015). Further examination of the immediate impact of television on children’s executive function. *Developmental Psychology*, *51*(6), 792-805. doi: 10.1037/a0039097

Lillard, A. S., & Peterson, J. (2011). The immediate impact of different types of television on young children's executive function. *Pediatrics*, e1-e6. doi: 10.1542/peds.2010-1919

Lovibond, P. F. (1998). Long-term stability of depression, anxiety, and stress syndromes. *Journal of Abnormal Psychology*, *107*, 520-526. doi: 10.1037/0021-843X.107.3.520

Lumley, T. (2017). *Survey: Analysis of complex survey samples* [R package]. version 3.32.

Markey, P.M., & Ferguson, C.J. (2017). *Moral combat: Why the war on violent video games is wrong*. Dallas, TX: BenBella Books.

Montgomery, J. M., Nyhan, B. & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*. Advance online publication. https://doi.org /[10.1111/ajps.12357](https://doi.org/10.1111/ajps.12357)

Nabi, R. L., & Krcmar, M. (2016). It takes two: the effect of child characteristics on US parents’ motivations for allowing electronic media use. *Journal of Children and Media*, *10*, 285-303. doi: 10.1080/17482798.2016.1162185

Nash, J. M. (1997, February). Fertile minds. *Time*, 149. No. 5.

NLS Investigator (2018, August 6). Retrieved from https://www.nlsinfo.org/investigator/pages/search.jsp?s=NLSY79

National Longitudinal Survey of Youth - Temperament (How My Child Usually Acts) (2018, July 11). Retrieved from <https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/temperament-how-my-child-usually-acts>

Nature Editorial Staff. (2016). Go forth and replicate! *Nature, 536*, 373. doi:10.1038/536373a

Nikkelen, S. W., Valkenburg, P. M., Huizinga, M., & Bushman, B. J. (2014). Media use and

ADHD-related behaviors in children and adolescents: A meta-analysis. *Developmental*

*Psychology*, *50*(9), 2228. http://dx.doi.org/10.1037/a0037318

Oliver, J. E., & Wood, T. (2014). Medical conspiracy theories and health behaviors in the United States. *JAMA Internal Medicine*, *174*(5), 817-818. doi: 10.1001/jamainternmed.2014.190

Plomin, R., Corley, R., DeFries, J. C., & Fulker, D. W. (1990). Individual differences in

television viewing in early childhood: Nature as well as nurture. *Psychological*

*Science*, *1*, 371-377.

Posner, M. I., & Rothbart, M. K. (2018). Temperament and brain networks of

attention. *Philosophical Transactions of the Royal Society B*, *373*, 20170254.

DOI: 10.1098/rstb.2017.0254

Radesky, J. S., Silverstein, M., Zuckerman, B., & Christakis, D. A. (2014). Infant self-regulation and early childhood media exposure. *Pediatrics*, *133*(5), e1172-e1178. doi: 10.1542/peds.2013-2367

Ridgeway, G., McCaffrey, D., Morral, A. Griffin, B., & Burgettey, L. (2017). *Toolkit for weighting and analysis of nonequivalent groups (TWANG)* [R package]. Santa Monica, CA: RAND Corporation.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science,* *1* (1), 27-42. <https://doi.org/10.1177/2515245917745629>

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in

observational studies for causal effects. *Biometrika*, *70*, 41-55. [https://doi.org/10.1093/ biomet/70.1.41](https://doi.org/10.1093/%20biomet/70.1.41)

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M.

Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality*

*development (6th ed.), 3,* 99-166. John Wiley & Sons Inc.

Ruff, H.A., & Rothbart, M.K. (1996). *Attention in early development: Themes and variations*. New York: Oxford University Press.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F.,... Nosek, B. A. (2017, September 21). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. <http://doi.org/10.17605/OSF.IO/QKWST>

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366. doi: 10.1177/0956797611417632

Smith, P. H., Dixon Jr, W. E., Jankowski, J. J., Sanscrainte, M. M., Davidson, B. K., &

Loboschefski, T. (1997). Longitudinal relationships between habituation and temperament in infancy. *Merrill-Palmer Quarterly*, *46*, 291-304.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702-712. <https://doi.org/10.1177/1745691616658637>

Sullivan, E. L., Holton, K. F., Nousen, E. K., Barling, A. N., Sullivan, C. A., Propper, C. B., & Nigg, J. T. (2015). Early identification of ADHD risk via infant temperament and emotion regulation: a pilot study. *Journal of Child Psychology and Psychiatry*, *56*(9), 949-957. doi: 10.1111/jcpp.12426

Thomas, A., Chess, S., & Birch, H. G. (1968). *Temperament and behaviour disorders in children*. New York: University Press.

Thompson, A. L., Adair, L. S., & Bentley, M. E. (2013). Maternal characteristics and perception of temperament associated with infant TV exposure. *Pediatrics*, *131*(2), e390-e397. doi 10.1542/peds.2012-1224

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology*, *84*, 205-220. doi: 10.1037/0022-3514.84.1.205

Vaala, S. E., & Hornik, R. C. (2014). Predicting US infants' and toddlers' TV/video viewing rates: Mothers' cognitions and structural life circumstances. *Journal of Children and Media*, *8*(2), 163-182. doi: 10.1080/17482798.2013.824494

Valkenburg, P. M., & Peter, J. (2013). The differential susceptibility to media effects

model. *Journal of Communication*, *63*, 221-243. doi:10.1111/jcom.12024

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1-67. doi: 10.18637/jss.v045.i03

Vandewater, E. A., Rideout, V.J., Wartella, E.A., Huang, X., Lee, J.H., & Shim, M.S. (2007). Digital childhood: electronic media and technology use among infants, toddlers, and preschoolers. *Pediatrics*, *119*(5), e1006-e1015. doi: 10.1542/peds.2006-1804

Wakefield, A.J., Murch, S.H., Anthony, A., Linnell, J., Casson, D.M., Malik, M., ... & Walker-

Smith, J.A. (1998). RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific

colitis, and pervasive developmental disorder in children. *The Lancet*.

Weisberg, D.S., Hirsh-Pasek, K., Golinkoff, R.M., & McCandliss, B.D. (2014). Mise en place: Setting the stage for thought and action. *Trends in Cognitive Sciences*, *18*(6), 276-278. doi: 10.1016/j.tics.2014.02.012

Table 1

Marginal descriptive statistics for continuous variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
| Variable | Valid n | Mean | Std Dev | Min | Max |
|  | | | | | |
| Age (yrs) when attention was measured | 2145 | 7.75 | 0.61 | 6.75 | 8.75 |
| Annual family income (thousands) | 1994 | 33.20 | 24.50 | 0.00 | 189.92 |
| Attention | 2145 | 2.63 | 0.40 | 1.00 | 3.00 |
| Attention within-sex SS | 2110 | 101.40 | 13.81 | 83.00 | 136.00 |
| BMI | 1491 | 17.23 | 2.22 | 13.02 | 21.97 |
| CES-D Depression score (1992) | 2126 | 47.07 | 7.94 | 32.30 | 79.90 |
| Cognitive stimulation of home age 1-3 | 1940 | 97.49 | 16.21 | 11.10 | 148.20 |
| Emotional support of home age 1-3 | 1796 | 97.88 | 16.54 | 31.60 | 124.70 |
| Mother's age at birth | 2145 | 28.47 | 2.63 | 22.00 | 36.00 |
| Mother's years of schooling | 2132 | 12.94 | 2.49 | 0.00 | 20.00 |
| Number of children in household | 2134 | 1.65 | 1.20 | 0.00 | 7.00 |
| Partner's years of schooling | 1785 | 13.26 | 2.72 | 0.00 | 20.00 |
| Rosenberg self-esteem score (1987) | 2077 | 45.04 | 8.44 | 23.50 | 59.70 |
| Temperament | 1998 | 2.01 | 0.69 | 1.00 | 5.00 |
| TV hours per day age 1.5 | 2029 | 2.22 | 3.07 | 0.00 | 16.00 |
| TV hours per day age 3 | 2060 | 3.69 | 3.14 | 0.00 | 16.00 |
|  | | | | | |
|  | | | | | |

Table 2

Marginal descriptive statistics for categorical variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| Variable | Value | n | Percent |
|  | | | |
| alcohol | No | 1070 | 49.88% |
|  | Yes | 946 | 44.10% |
|  | . | 129 | 6.01% |
| fatherAbsent | No | 1708 | 79.63% |
|  | Yes | 408 | 19.02% |
|  | . | 29 | 1.35% |
| female | Female | 1046 | 48.76% |
|  | Male | 1099 | 51.24% |
| lowBirthWt | No | 1846 | 86.06% |
|  | Yes | 140 | 6.53% |
|  | . | 159 | 7.41% |
| poorHealth | No | 1944 | 90.63% |
|  | Yes | 132 | 6.15% |
|  | . | 69 | 3.22% |
| preterm | No | 1773 | 82.66% |
|  | Yes | 221 | 10.30% |
|  | . | 151 | 7.04% |
| race | Black | 578 | 26.95% |
|  | Hispanic | 406 | 18.93% |
|  | White | 1161 | 54.13% |
| smoking | No | 1472 | 68.62% |
|  | Yes | 538 | 25.08% |
|  | . | 135 | 6.29% |
| SMSA | Not in SMSA | 388 | 18.09% |
|  | SMSA; central city unknown | 693 | 32.31% |
|  | SMSA; in central city | 307 | 14.31% |
|  | SMSA; not central city | 651 | 30.35% |
|  |  | 106 | 4.94% |
|  | | | |
|  | | | |

Note: Period denote missing values. *alcohol* = indicator of any maternal alcohol use in pregnancy. *FatherAbsent* = child’s father does not live in household. *female* = child gender is female. *poorHealth* = child has medical condition limiting usual childhood activities. *preterm* = child was born < 37 weeks gestation. *race* = child race category. *smoking* = indicator of any maternal smoking in pregnancy. *SMSA* = statistical metropolitan sampling area classification

Table 3

Descriptive statistics for TV use and sample sizes for defining the low- and high-TV groups at ages 1.5 and 3 for the 20th/80th and 50th/50th percentile cut points.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Age 1.5 | |  | Age 3 | |
|  |  |  |  |  |  |  |
| Cutpoint | Statistic | Low-TV | High-TV |  | Low-TV | High-TV |
|  |  |  |  |  |  |  |
| 20/80 | n | 717 | 434 |  | 439 | 422 |
|  | Mean | 0.00 | 6.76 |  | 0.83 | 8.55 |
|  | Median | 0.00 | 5.29 |  | 1.00 | 7.29 |
|  | Min | 0.00 | 3.57 |  | 0.00 | 5.14 |
|  | Max | 0.00 | 16.00 |  | 1.57 | 16.00 |
|  |  |  |  |  |  |  |
| 50/50 | n | 1043 | 986 |  | 1127 | 933 |
|  | Mean | 0.27 | 4.28 |  | 1.75 | 6.03 |
|  | Median | 0.00 | 3.14 |  | 2.00 | 4.86 |
|  | Min | 0.00 | 1.43 |  | 0.00 | 3.14 |
|  | Max | 1.29 | 16.00 |  | 3.00 | 16.00 |
|  |  |  |  |  |  |  |

Figure 1

Density plots for TV consumption at age 1.5 and age 3

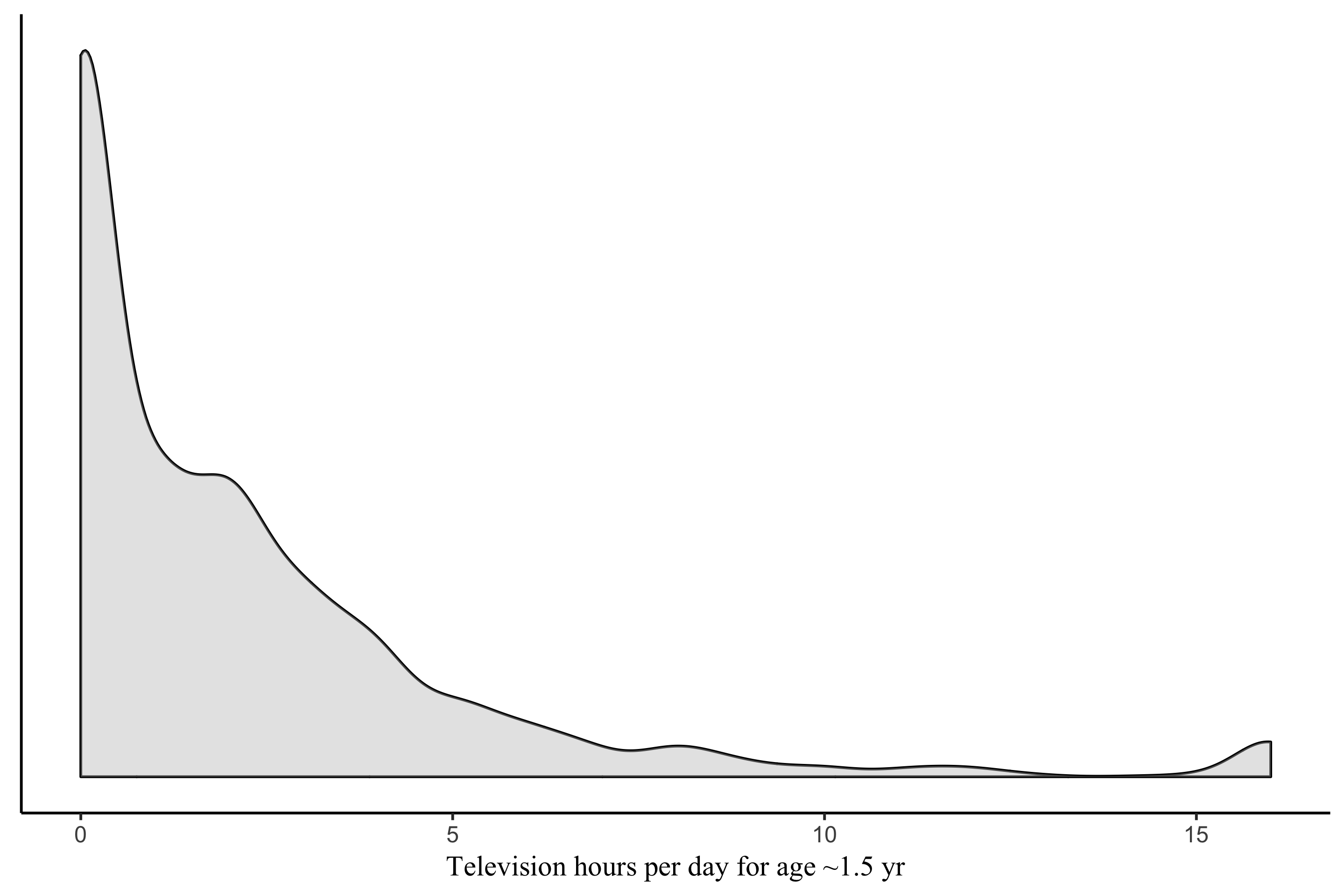
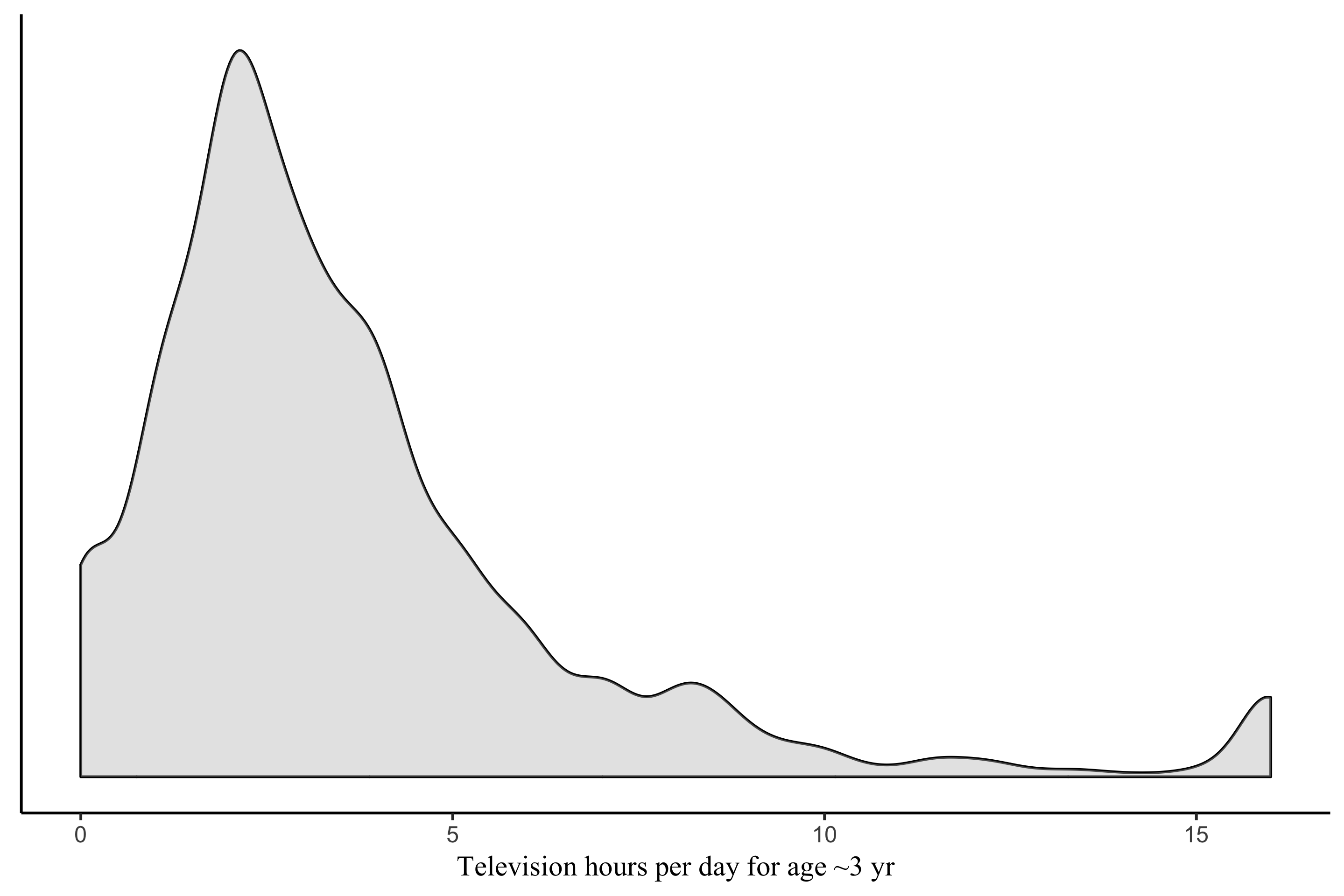


Figure 2

Scatterplots between early childhood TV use (left column age ~1.5, right column age ~3) and standardized within-sex attention score at age 7. *Top row*: raw data. *Bottom row*: adjusted attention score with effect of covariates removed.

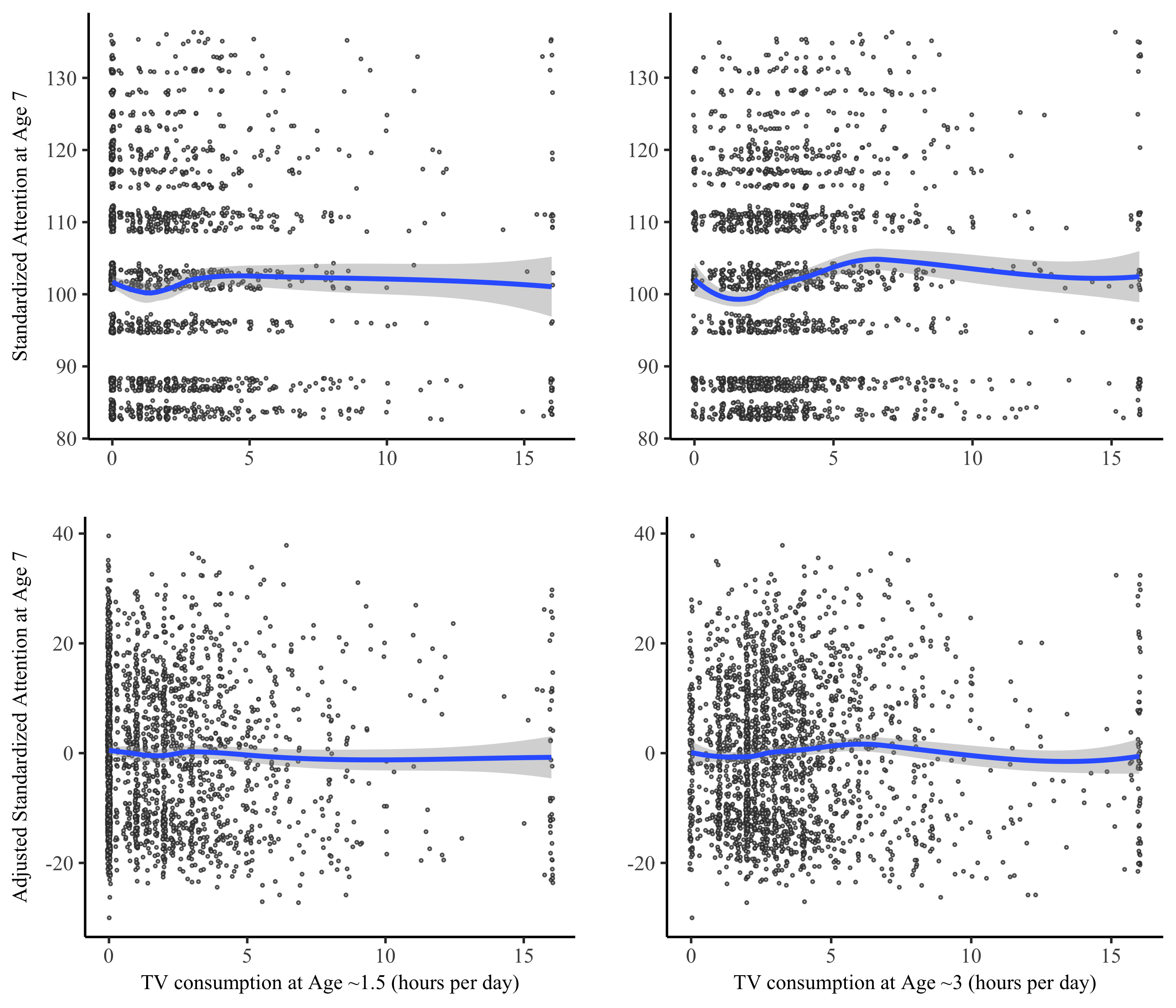
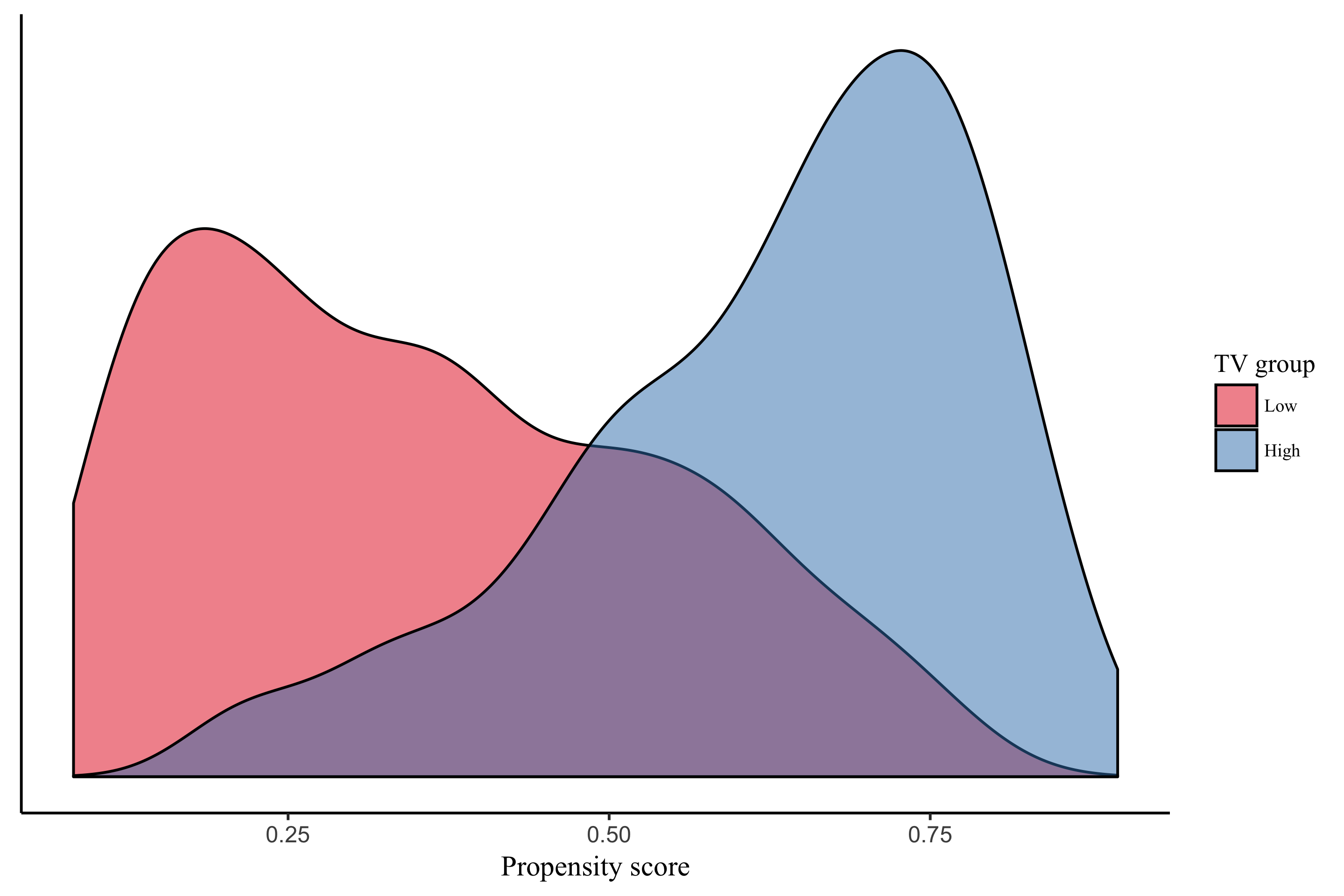
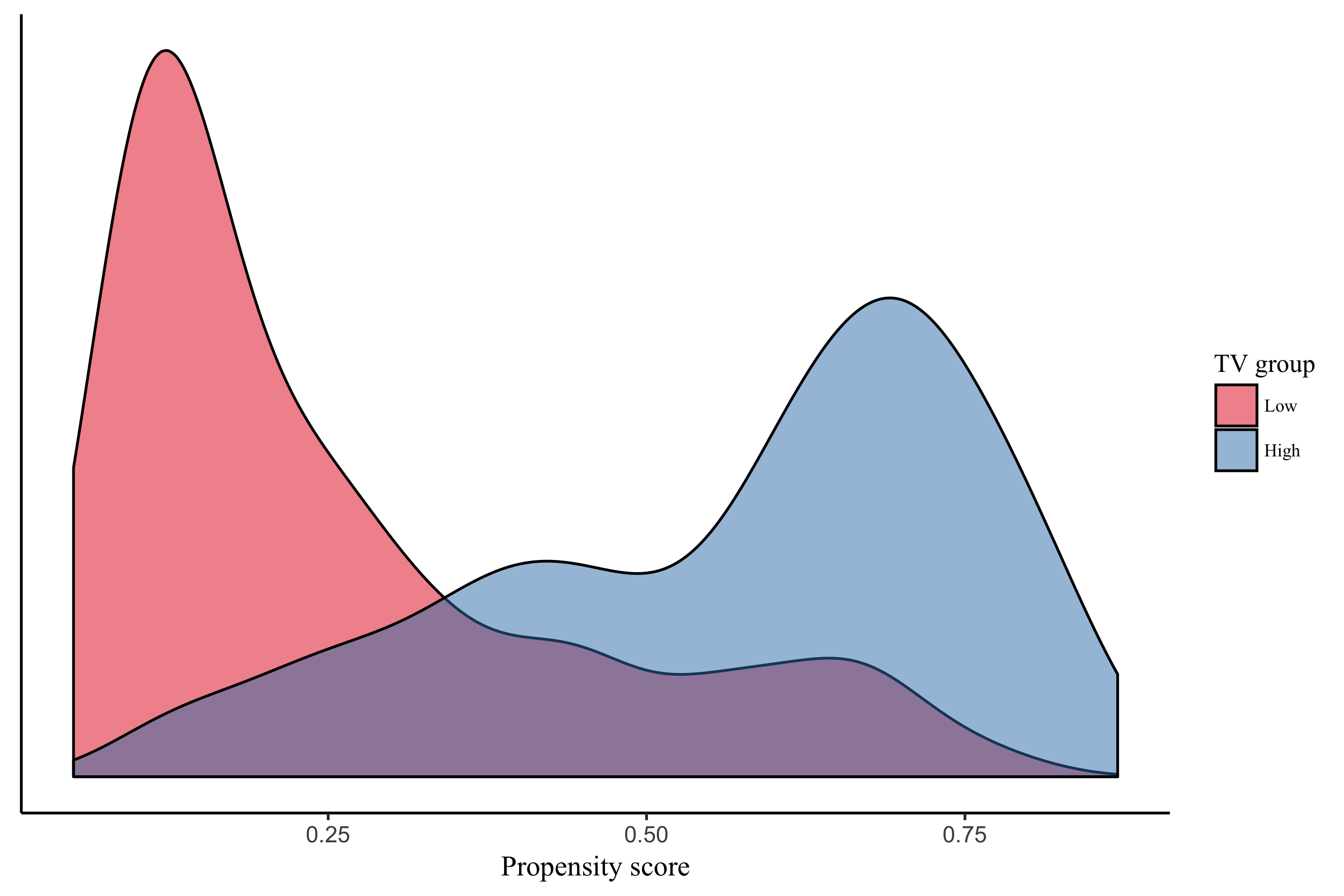


Figure 3

Propensity score distributions for the low- and high-TV groups. *Top row*: 20th/80th percentile cut points for defining the groups. *Bottom row*: median split cut point. *Left column*: TV use measured at age 1.5. *Right column*: TV use measured at age 3.



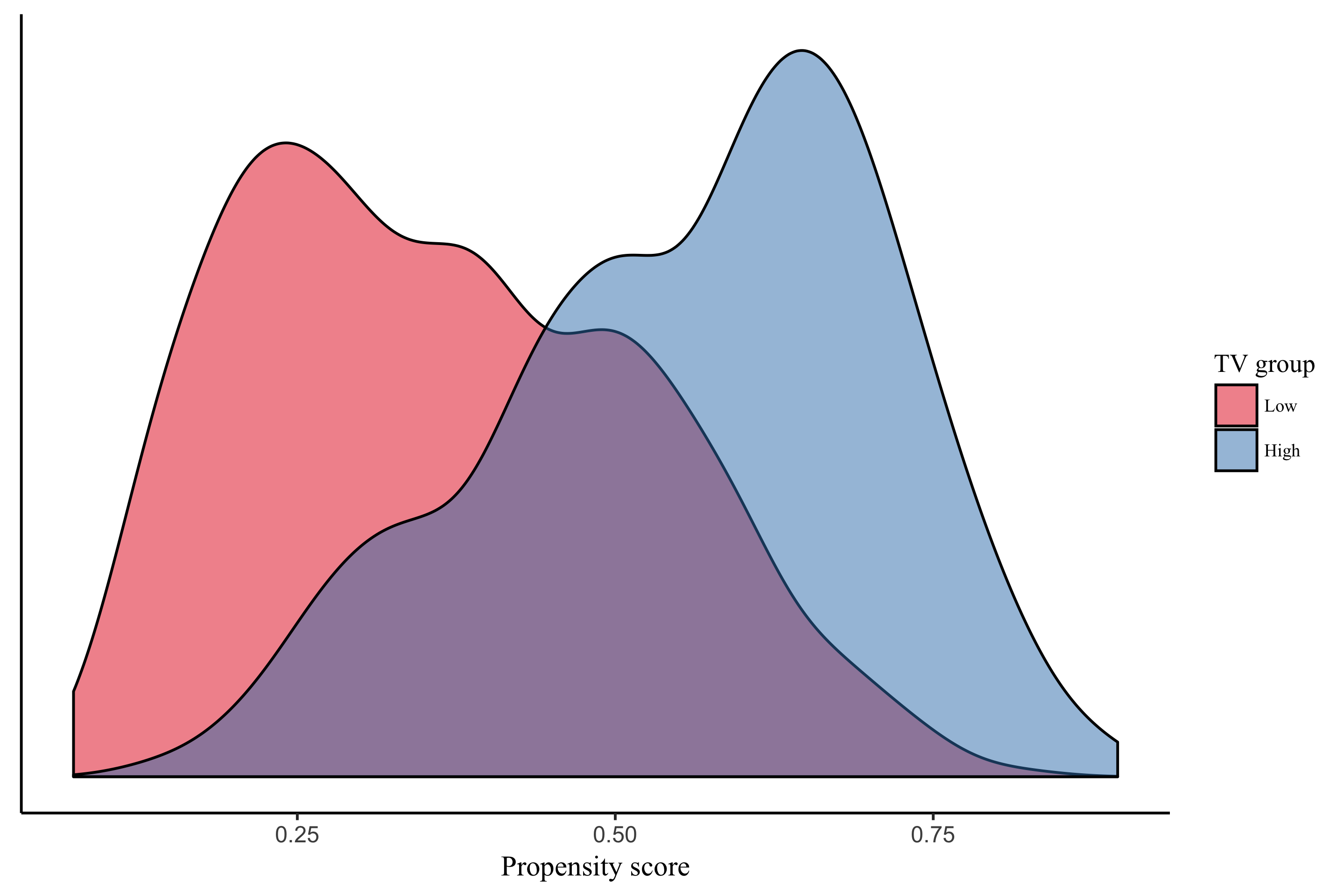
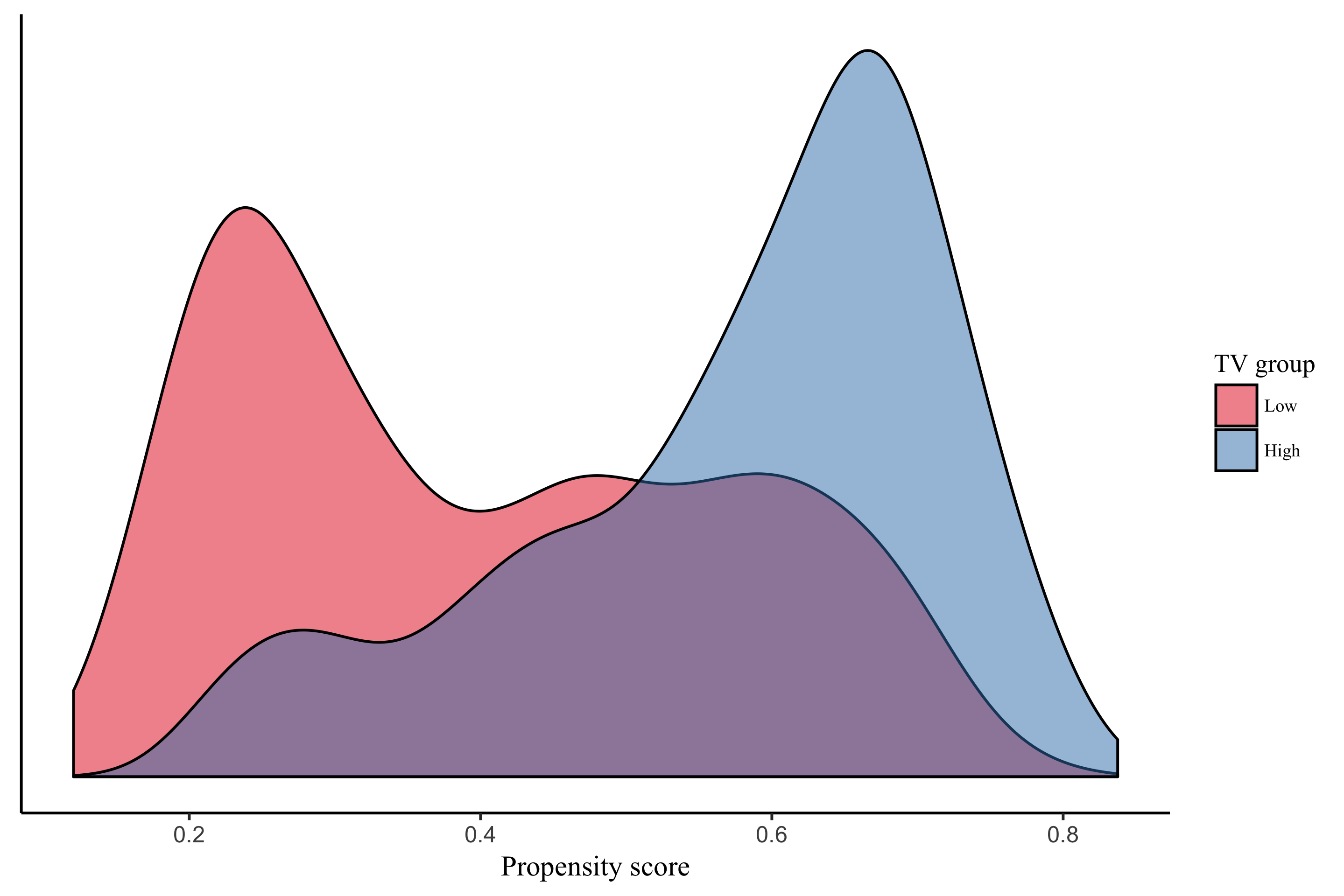
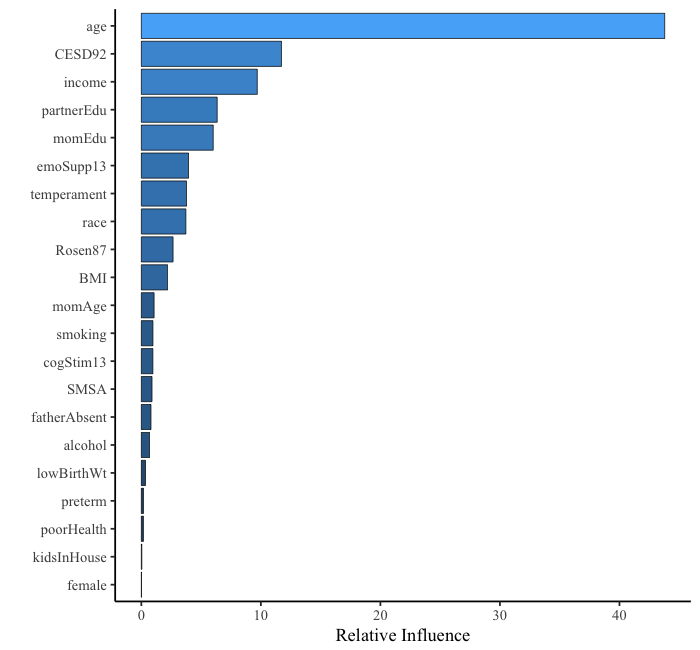


Figure 4

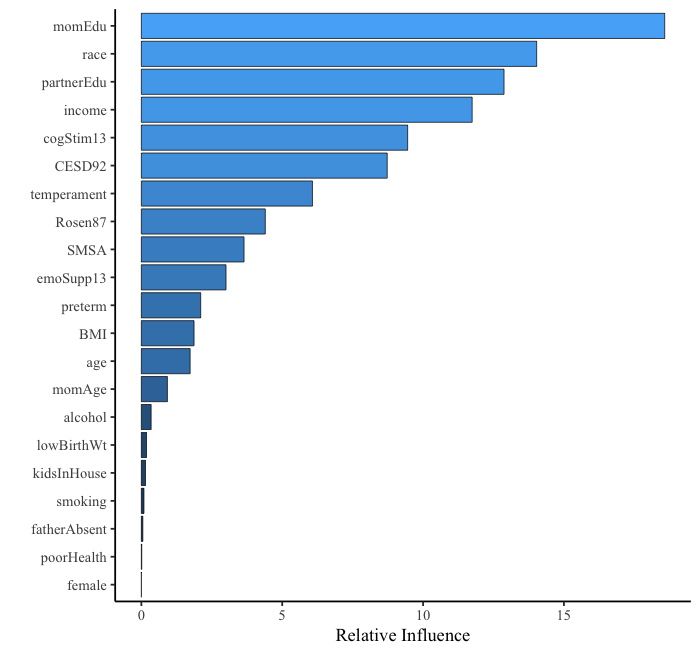
Relative influence of variables in the propensity score model predicting TV use at age 1.5.



Note: Based on TV use groups based on the 20th / 80th percentile cut points and propensity scores optimized to estimate the ATE. *Age* = child age at index year. *CESD9*2 = maternal CES-D depression score measured in 1992. *Income* = family income. *Rosen87* = maternal Rosenberg self-esteem score measured in 1987. *CogStim13* = cognitive stimulation of home environment. *EmoSupp13* = emotional support of home environment. *PartnerEdu* = partner’s educational attainment. *temperament* = child’s temperament score. *BMI* = child’s body mass index. *momEdu* = mother’s educational attainment. *SMSA* = statistical metropolitan sampling area category for the home. *poorHealth* = binary indicator of child health problems limiting usual activities. *momAge* = mother’s age when child was born. *smoking* = binary indicator of maternal smoking during pregnancy. *lowBirthWt* = binary indicator of low birth weight. *race* = child’s race category. *kidsInHouse* = number of children of the mother in the household. *preterm* = child was born at < 37 weeks gestation. *alcohol* = binary indicator of maternal alcohol use during pregnancy. *fatherabsent* = child’s father does not live in the household. *female* = binary indicator that child is female.

Figure 5

Relative influence of variables in the propensity score model predicting TV use at age 3.



Note: Based on TV use groups defined by the 20th / 80th percentile cut points and propensity scores optimized to estimate the ATE. Note for Figure 4 describes variable labels.

Figure 6

Covariate balance for stratification model continuous variables. TV use measured at age ~1.5.

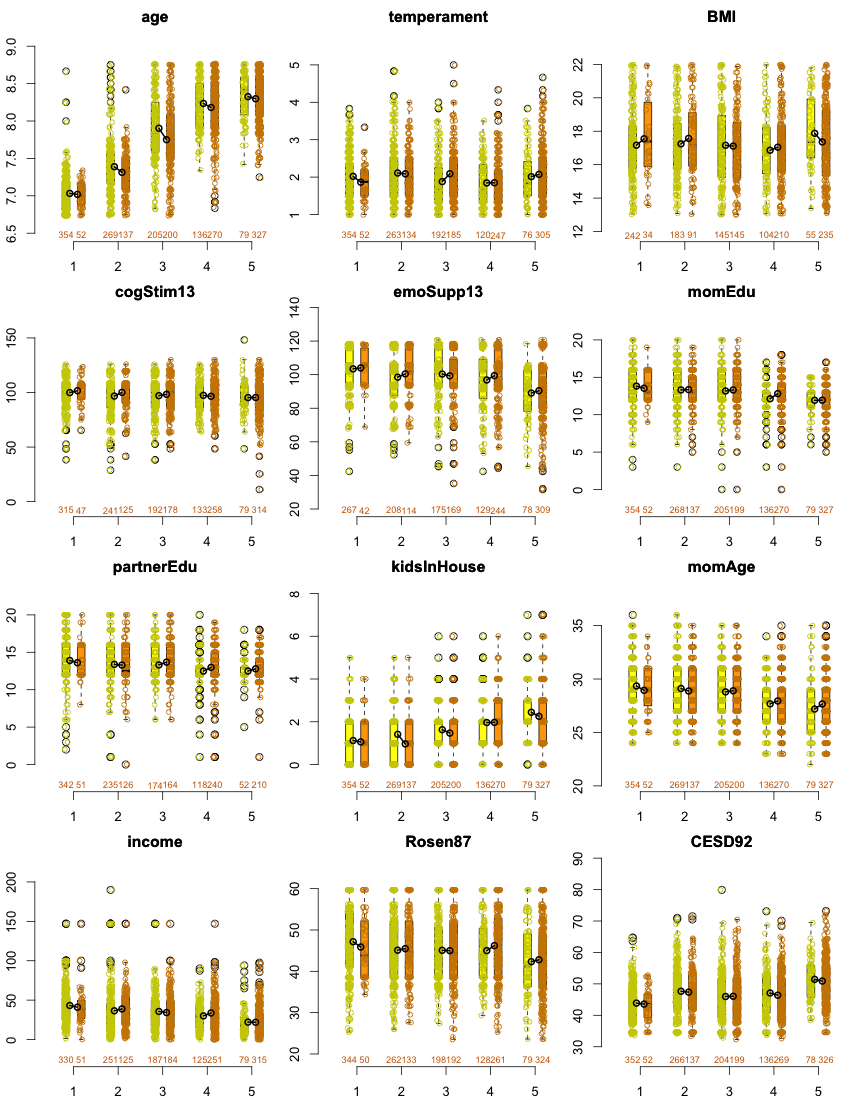


Figure 7

Covariate balance for stratification model categorical variables. TV use measured at age ~1.5.

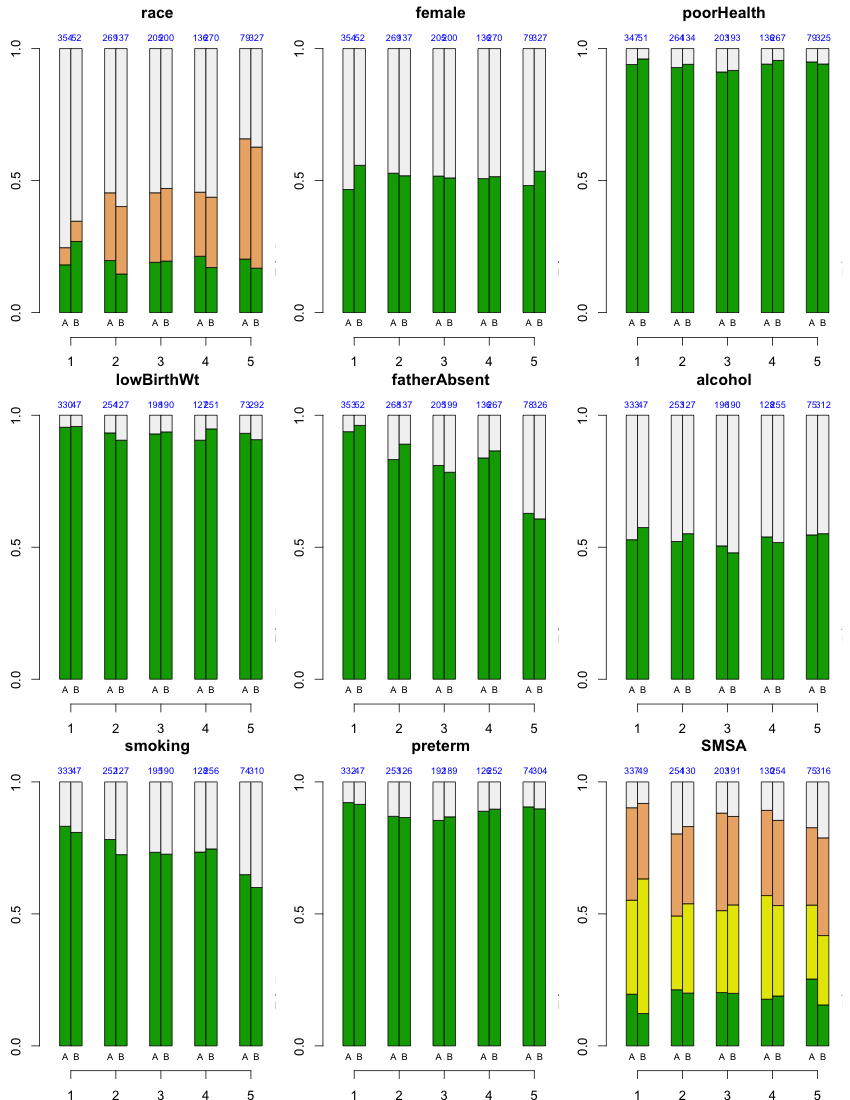
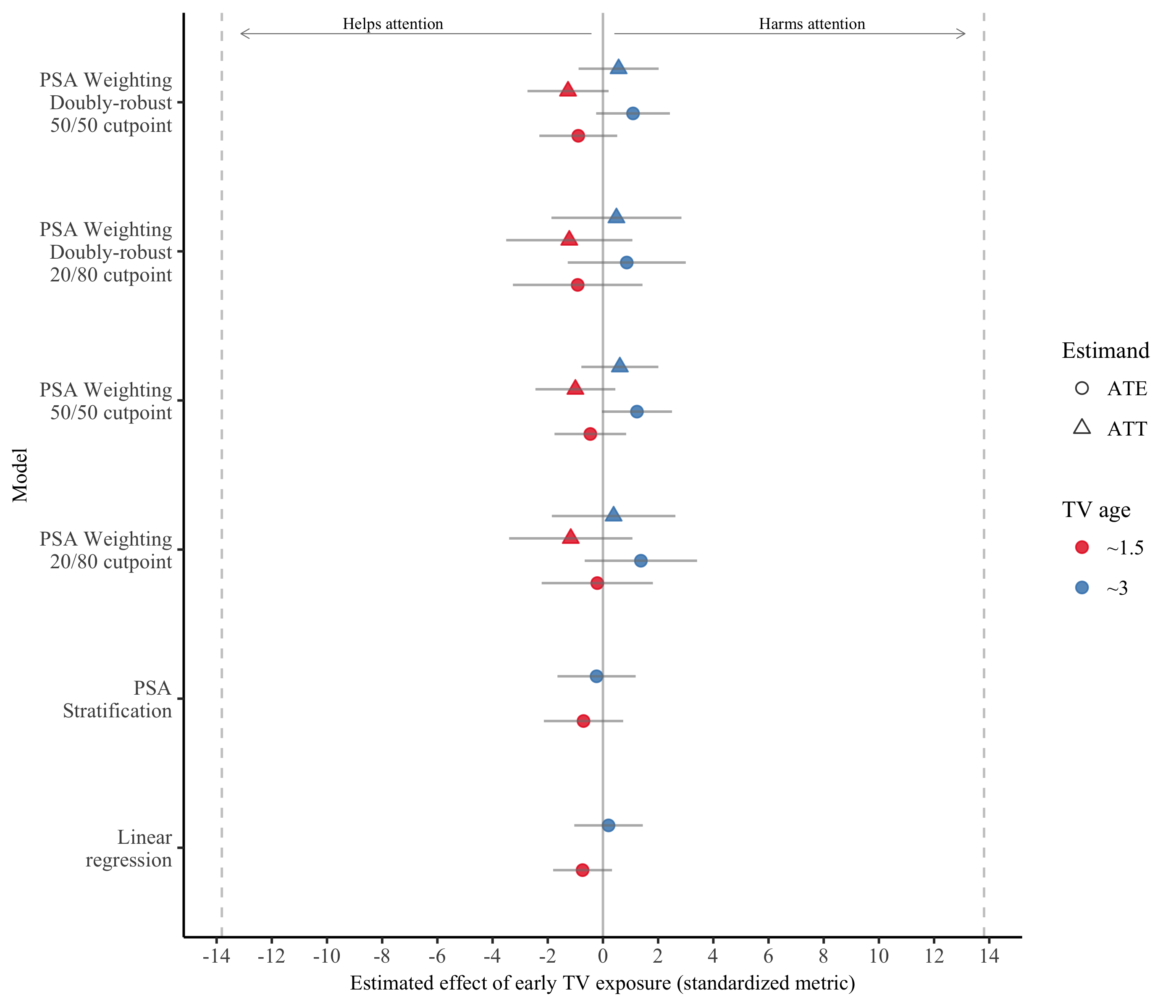


Figure 8

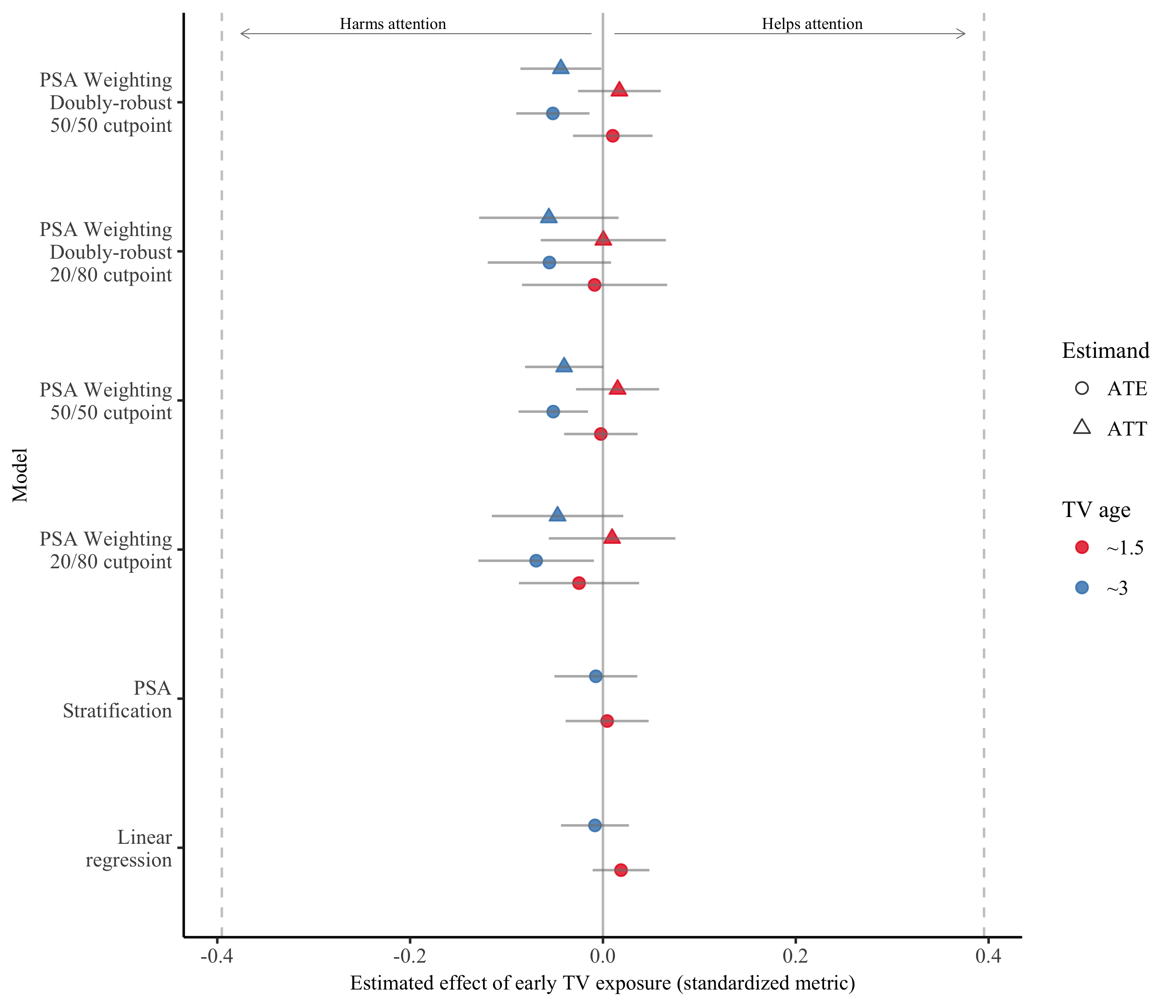
Summary of propensity score and regression model results for the standardized within-sex attention outcome.



Note: Figure displays point estimates and 95% confidence intervals for each model. The vertical reference line in the center indicates no effect of TV on attention. Confidence intervals including zero are non-significant at the level. The dotted vertical reference lines indicate 1 effect sizes. The regression coefficients represent the effect of a one-unit change of TV use on attention, where the unit is defined by the distance in median TV use from the low-TV to the high-TV categories. Thus, the regression estimate is on a similar scale as the propensity score estimates. The linear regression and stratified propensity score models did not employ cut points.

Figure 9

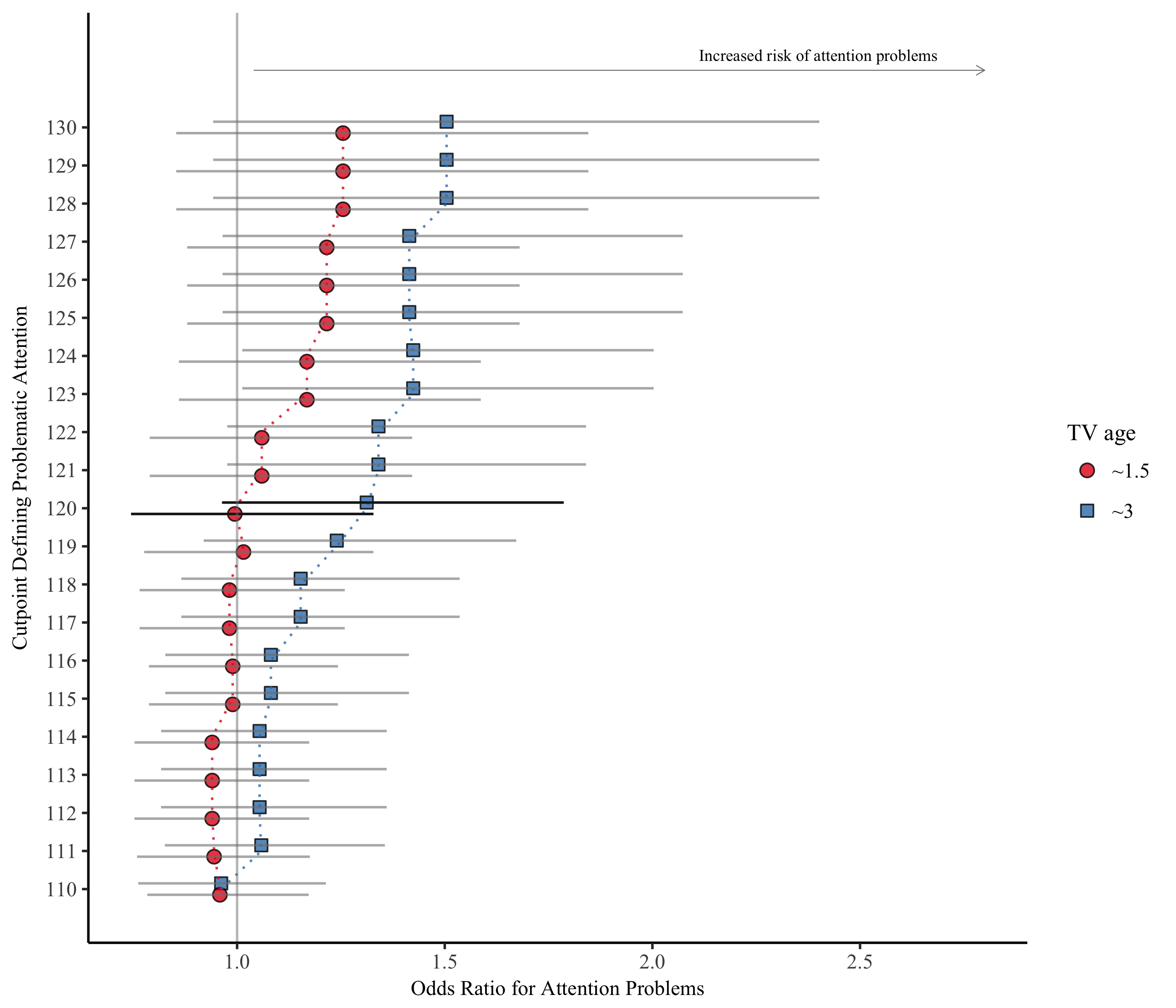
Summary of propensity score and regression model results for the raw attention outcome.



Note: Figure displays point estimates and 95% confidence intervals for each model. The vertical reference line in the center indicates no effect of TV on attention. Confidence intervals including zero are non-significant at the level. The dotted vertical reference lines indicate 1 effect sizes. The regression coefficients represent the effect of a one-unit change of TV use on attention, where the unit is defined by the distance in median TV use from the low-TV to the high-TV categories. Thus, the regression estimate is on a similar scale as the propensity score estimates.

Figure 10

Logistic regression model results by age when TV use was measured and cut point defining problematic attention.



Note: Lines indicate the width of the 95% confidence intervals. The vertical reference line at an odds ratio of 1.0 denotes no effect of TV on the probability of attention problems. Confidence intervals including zero are non-significant at the level. The bolded results denoted the 120 cut point used by Christakis et al. (2004).