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 (*n* = 2,108), we conducted a multiverse analysis to examine the degree to which the finding reported by Christakis et al. was robust to analytic choices. We evaluated 888 analytic models, including linear regression, logistic regression, and two forms of propensity score analysis. Only **166** models (**18.7**%) yielded a statistically significant relationship between early TV exposure and later attention problems, with most models estimating a trivially small effect size. We conclude that the evidence for the harmful effect of early TV use is weak and inconsistent. All data and code necessary to reproduce our analysis is available online via github (<https://github.com/mcbeem/TVAttention>) and as a Docker container (<https://hub.docker.com/repository/docker/mmcbee/rstudio_tvattention>). Preprint [link].

**Keywords**: media, TV, ADHD, attention development, multiverse analysis, specification curve, computational reproducibility, garden of forking paths

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 correct the public understanding of the issue. 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. In spite of the careful language used in the paper, subsequent presentations by the lead author discussing this work blur the lines between associations and definitive causal links. For example, in his TEDxRainier talk, **Christakis (2011)** said:

And we, we tested this some years ago, and what we found was that for the more television children watched before age three, the more likely they were to actually have attentional problems at school age. Specifically, for each hour that they watched before the age of three, their chances of having attentional problems was increased by about ten percent. So a child who watched two hours of TV a day before age three would be twenty percent more likely to have attention problems compared to a child who watched none (7:21 to 7:46).

This statement does not admit any epistemic (or statistical) uncertainty about the nature of this estimate, nor of any limitations to the evidence supporting the claim. It is presented as a scientific fact.

Unsurprisingly, popular media was much less careful in raising alarm about the potential harm of television exposure on children. 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 retractions or contemporary updates 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), or sometimes even a moral panic about the harm they are supposedly wreaking on children (e.g., **Twenge, 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. Recommendations from the American Academy of Pediatrics that children avoid television viewing were based on the notion that TV was inherently harmful for young children (AAP, 1999; Children and Media, 2018); they also reinforced it by lending credibility and weight to the claim.

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, or at least is much more nuanced than initially thought. 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 their relationship is only weak to moderate, but was unable to clarify the direction of causality or the potential that this estimate is contaminated with confounding bias (Nikkelen, Valkenberg, Huizinga, & Bushman, 2014). A recent review came to a similar conclusion (Kostyrka-Allchorne, Cooper, & Simpson, 2017). As is often the case, however, 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, while the more methodologically sound critique (Foster & Watkins, 2010) has 76 citations and the meta-analysis (Nikkelen et al., 2014) has only 54. Not only the general public, but substantial portions of the research community, is still left with the message that TV causes attention problems..

Our in this paper ; see also **Orben, Dienlin, & Przybylski, 2019**data analysis . While many of these decisions can be made in a principled manner in accordance with known best practices, some of them are arbitrary or, at the least, ambiguous (King & Zeng, 2007). For example, should a researcher employ an efficient but assumption-leveraged linear model, or would propensity score analysis be a safer bet? This decision is driven by the researcher’s sense of the risk and reward landscape at play, as well as the desired balance between competing but largely obscured priorities – a preferred position with respect to a bias / efficiency tradeoff when neither bias nor efficiency can be evaluated directly. As such, these decisions are often resolved by the researcher’s idiosyncratic preferences, knowledge and comfort of methods, or what the researcher perceives the audience to expect.

The set of all possible analytic choicesis known as the At the terminus of each decision tree lies a result.; rendering any particular conclusion suspect.results vary across analytic methodsIf most defensible models reach a similar positive conclusion about the presence and magnitude of some effect, then the claim about that effect becomes more credible. Conversely, if most models produce null effects, or if the results vary wildly across models, then the claim made by one particular positive analysis must be viewed as less credible.

In this paper, we present a multiverse analysis of Christakis et al.’s (2004) original claim, using the same NLSY79 dataset, prepared in the same manner as was documented in the 2004 paper. We then subjected this data set to a wide variety of analyses across three general types: propensity score analysis, linear regression, and logistic regression. The models varied across many dimensions which are discussed in detail later. The selection of covariates (adjustment variables) is an important issue, common to all analysis approaches, that we discuss in detail in the next section.

**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 NLSY79 dataset. These datasets were merged via a common ID code variable allowing mother and child data to be linked. We initially downloaded 340 variables from the Child and Young Adult dataset and 40 variables from the NLSY79 dataset(NLSY, 2018). Our project’s Github page (under “Documentation”) presents a spreadsheet mapping our analysis variables to the variable codes and labels from the NLSY dataset. Our raw and processed analysis datasets 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. 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.

**Selection of cases.** We followed the original paper’s criteria for sample selection. For each index year (1996, 1998, and 2000), including those children whose ages at index were between 6 years 9 months and 8 years 9 months.­ Children with severe vision or hearing impairment, as well as those with severe emotional disturbances or orthopedic disabilities were excluded. A total of 2,108 cases were extracted that met these conditions.

**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.

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.

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 project github 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.

**Selection of covariates.** The goal of each of our models was to estimate the causal effect of early TV on mid-childhood attention as accurately as possible. Since this data was collected via an observational longitudinal design, confounding is a near-certainty. Causal inference from observational data, in theory, possible if the proper set of covariates are incorporated into the analysis such that all confounding paths are blocked (Rohrer, 2018). To this end, our models employed two different sets of covariates. 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. We hope that including these covariates reduced confounding bias that would otherwise render the estimates uninterpretable, though we seriously doubt that we have completely eliminated it (**Westfall & Yarkoni, 2016**).

*Original covariates.* The first set of covariates was identical to those employed in the original study. They included the following: cohort (year in which the child’s attention was assessed: 1996, 1998, or 2000), the child’s age when attention was assessed (typically 93 months, but varied between 81 and 105 months), child’s race, child’s sex, the number of children of the mother living in the household, mother’s highest grade completed, the cognitive simulation and emotional support of the home (measured between ages 1 and 3), binary indicators 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, maternal depression as measured by the CES-D in 1992, child’s gestational age at birth (centered at term), and an urbanicity indicator variable in the form of the four levels of the Statistical Metropolitan Sampling Area classification. Where applicable, 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 (in other words, endogenous; Montgomery, Nyhan, & Torres, 2018; Rohrer, 2018).

*Expanded covariates*. The expanded covariate list included all of the original covariate set with the following additions, which we suspected to be plausible confounders for TV use and childhood attention. We added family income, the partner or spouse’s highest level of educational attainment, an indicator variable for low birth weight (less than 2500 grams or 5 lbs 8 oz), and an indicator that the child suffers from a health condition that limited their school and play activities[[1]](#footnote-1). 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), as we suspected this would better capture the relevant information in this variable.

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.

We view temperament as potentially a key confounding variable that could influence both early childhood TV use and mid-childhood attention deficits and doubted that a credible causal effect could 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 temperament to be an important confounder that was not included in Christakis et al.’s analysis.

At the direction of reviewers, who expressed a concern that the behaviors incorporated into our temperament variable might simply be an earlier manifestation of attention deficits, we performed an exploratory factor analysis. A two-factor model with varimax rotation exhibited clean simple structure separating attention from temperament items, and in which the largest absolute standardized cross-loading was 0.133. The correlation between factors was *r* = -0.114. We therefore concluded that attention and temperament were highly distinct variables.

**Analytic approaches.** Our models considered two different outcomes (raw attention vs the within-sex standardized attention scores used in the original analysis), measured TV use at approximately 1.5 and three years of age, and incorporated the two different sets of covariates designated above. Additional features specific to each model are described in the next section.

**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 that were included in the propensity score model. In this manner, propensity score analysis approximates a randomized experiment with respect to the measured covariates included in the analysis (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. One of the most important of these is how to incorporate the propensity scores into the inference. We used two such methods in our multiverse analysis: inverse probability of treatment weighting (IPTW) and stratification. Further, because propensity score analysis requires dichotomizing the predictor variable, we ran analyses using six different percentile cut points to define the high and low TV groups as follows:

* Below 20th percentile / Above 80th percentile
* Below 30th percentile / Above 70th percentile
* Below 40th percentile / Above 60th percentile
* Below 50th percentile / Above 50th percentile
* Below 60th percentile / Above 60th percentile
* Below 70th percentile / Above 70th percentile

In all the propensity score analyses, we used boosted classification trees (as implemented in the *twang* package (v. 1.5, Ridgeway, McCaffrey, Morral, Griffin, and Burgette, 2017) to estimate the propensity scores, using bagging and cross-validation to prevent overfitting. Missing data on covariates is handled automatically by the classification tree approach, in that the missingness is treated as informative and propensity scores can be estimated for cases with missing covariate values.

*Inverse probability of treatment weighting*. In this method, the propensity scores are used to construct IPTW weights which, when applied to the data, equalize the distribution of propensity scores between the treatment and control groups – and by implication, also equalize the distribution of all of the covariates that were included in the propensity score model (Guo & Fraser, 2015). We because the effect of TV exposure might plausibly be different for those who watched a lot of TV versus those who watched an average amountWe ran analyses both with and without the inclusion of sample weights to correct for the design-based oversampling of certain demographic groups. In some analyses, we identified the four covariates with the largest residual imbalance statistics and gave those covariates an additional regression adjustment. These conditions were fully crossed, with 6 (TV cutpoints) x 2 (outcomes) x 2 (TV ages) x 2 (covariate sets) x 2 (treatment effects) x 2 (sample weights) x 2 (doubly-robust) yielding 384 IPTW propensity score models. The *survey* package (v. 3.35-1; **Lumley, 2004, 2019**) was used to estimate the treatment effect after applying IPTW weights.

*Stratification*. An alternative inferential strategy using propensity scores is to stratify on them, calculate a treatment effect specific to each stratum, then combine those stratum-specific estimates to produce the average treatment effect (ATE, Guo & Fraser, 2015). These models were computed for five different numbers of strata (4, 5, 6, 7, or 8), which were fully crossed with 6 (TV cutpoints) x 2 (outcomes) x 2 (TV ages) x 2 (covariate sets) yielding 240 stratification propensity score models. Neither sample weights nor the doubly-robust approach could be implemented in the stratification models, nor could these models estimate the average treatment effect for the treated (ATT). We used the *PSAgraphics* package (v 2.1.1; Helmreich & Pruzek, 2009) to perform the stratified analysis, and calculated *p*-values for the treatment effect estimates using the normal approximation.

**Linear Regression.** Linear regression models are less robust than propensity score models but 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. Linear regression models do not require the TV exposure variable to be dichotomized, but they do require researchers to specify an appropriate functional form for the relationship between TV exposure and attention. We fit regression models of polynomial order one (e.g. linear), two, and three for the TV effect, with linear relations specified for the covariates. The ­*p*­-value for the effect of TV on attention was calculated via a joint *F* test comparing a model containing the TV variables plus all covariates against a baseline model containing only covariates. The point estimate for the TV effect in the poly-2 and -3 models was computed by linearizing the slope (via computation of a partial derivative) at the median TV exposure at age ~1.5 and age ~3. This estimate and its associated standard error were calculated using the *multcomp* package (v. 1.4-10, **Hothon, Bretz, & Westfall, 2008**) via the specification of custom contrast coefficients. We fit models both with and without sample weights, again using the *survey* package to perform the weighted analysis. Finally, we performed our analysis using both listwise deletion of cases with missing values and using multiple imputation (as implemented by the *mice* package, v. 3.6.0; van Buuren & Groothius-Oudshoorn, 2011). However, sample weights could not be combined with multiple imputation, so these conditions were not fully crossed. Using listwise deletion, we fit 2 (outcomes) x 2 (TV ages) x 2 (covariate sets) x 3 (polynomial functional forms) x 2 (sample weights) = 48 models. Using multiple imputation, we fit 2 (outcomes) x 2 (TV ages) x 2 (covariate sets) x 3 (polynomial functional forms) = 24 models. A total of 72 linear regression models were fit to the data.

**Logistic Regression.** Finally, to replicate the analysis used in 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 within-sex standardized attention score of 120). They argued that using this cut point yielded a rate of problematic attention similar to its incidence in the population, but would their conclusions have been the same if they had used a different cut point? We defined multiple dichotomous outcome variables by varying the standardized attention cutpoint from 110 to 130. We produced percentile-equivalent cutpoints on the raw attention outcome. Like the linear regression models, we fit models with and without sample weights, and also with and without multiple imputation of missing data (though these could not be combined). With listwise deletion, this yielded 21 (attention cutpoints) x 2 (outcomes) x 2 (TV ages) x 2 (covariate sets) x 2 (sample weights) = 336 models. Using multiple imputation, we fit 21 (attention cutpoints) x 2 (outcomes) x 2 (TV ages) x 2 (covariate sets) = 168 models, for a total of 504 logistic regression models. However, because of sparseness on the attention outcome (particularly the raw version), frequently the imposition of two adjacent cutoffs (e.g, 121 and 122) would produce identical categorizations of the outcome and therefore redundant results. After purging these redundancies, we were left with 192 unique logistic regression models. We used the *survey* and *mice* packages to incorporate sample weights and to perform multiple imputation.

In total, we fit 888 non-redundant models to the data, including 384 IPTW propensity score, 192 stratified propensity score, 72 linear regression, and 192 logistic regression models.

**Results**

Space limitations prevented us from displaying detailed results from any specific model in this paper. However, the github repository “Results” directory contains a subdirectory for every analysis conducted, which includes descriptive statistics, diagnostic tables and plots, and formatted model results, which were produced using the *stargazer* package (v. 5.2.1; Hlavac, 2015). Figures may be browsed on the github site. However, tables will appear as unrendered HTML code unless the link to the table is prefixed with <http://htmlpreview.github.com/?>. The easiest way to examine the results in detail is to download our Github repository to your local computer; once downloaded, html tables will render in your web browser upon opening.

**Descriptive statistics**

Tables 1 and 2 provide descriptive statistics for the continuous and categorical variables. Figure 1 displays a set of scatterplots displaying the relationship between TV consumption and within-sex standardized attention measured at age 7. The plots in the second column portray the relationship between TV measured at age ~3 and attention, both without controls (top row) and with controls (bottom row). Missing data on covariates dramatically reduced the sample size for these residualized points, so the figure also displays imputed data taken from the first (of 10) multiple imputations. The figure contains two different non-parametric smoothed regression lines: the solid blue line fits to complete data only, while the dashed red line fits to all the data, including the imputed portion. These scatterplots are characterized by a complete lack of apparent gross relationship between TV and attention, but also by the presence of a non-linear “wiggle” in the smoothed trajectory. This non-linearity is diminished but not eliminated by the controlling for covariates, especially when imputation was used. We will argue that this feature of the data has important implications for the some of the model results.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tables 1 & 2 and Figure 1 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**IPTW propensity score analysis results**. Figure 2 summarizes the IPTW propensity score model results. The top panel displays the point estimates and 95% confidence intervals of the treatment effect estimate (rescaled to a Cohen’s *d* metric) for the within-sex standardized attention outcome when TV is measured at age ~1.5 (left subpanel) and at age ~3 (right subpanel). The middle panel displays the same for the raw attention outcome. Both panels are arranged such that a higher score means worse (or more impaired) attention. Statistically significant estimates are designated by colored points and confidence limits that do not include zero. The median effect size across all IPTW models was *d* = 0.068, with empirical 2.5th and 97.5th percentiles of [-0.082, 0.203]. The bottom panel shows the distribution of ­p-values for the TV effect from the models, presented in order from largest to smallest. Overall, 99 out of 384 models (25.8%) produced a statistically significant TV effect on attention, and direction of all of these was TV is harmful.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Figure 2 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Stratification propensity score analysis results.** Stratification model results are presented in Figure 3, which has the same structure as Figure X. The median effect size across all IPTW models was *d* = -0.016, with empirical 2.5th and 97.5th percentiles of [-0.141, 0.079], where higher values indicate worse attention. Only one of the 240 models (0.4%) produced statistically significant results, and its point estimate indicated a beneficial effect of TV exposure.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Figure 3 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Linear regression results.** Linear regression model results are presented in Figure 4. The top and center rows depict the TV slope point estimate and confidence intervals in terms of raw units (e.g., the expected change in attention given a one-hour per day change in TV exposure). The estimates for models incorporating squared or cubed functional forms were calculated by linearizing the slope at the median level of TV use. Converting these point estimates to standardized regression coefficients (betas) allows the estimates from the within-sex standardized attention and raw attention outcomes to be placed on a common scale. The median beta coefficient for the regression models was = 0.028, with empirical 2.5th and 97.5th percentiles of [-0.040, 0.163]. Thirteen of the 72 models (18.1%) produced statistically significant estimates for the effect of TV; all of these were in the direction of harm.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Figure 4 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Logistic regression results**. Logistic regression model results are presented in Figure 5. Effect size point estimates and confidence bounds are given in odds ratio (OR) units, where ORs greater than one indicate a higher risk of being classified into the ‘problematic attention’ category as defined by the attention cutpoint (which varied across models). The median OR was 1.035, with empirical 2.5th and 97.5th percentiles of [0.988, 1.175]. Sixty of the 192 models (31.3%) produced significant estimates; all in the direction of harm.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Figure 5 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Overall summary.** Figure 6 summarizes the distribution of ­*p*-values across models. Overall, 175 of 888 models (19.7%) produced statistical significance. The right margin of the figure is a histogram of the distribution of *p*-values. The dashed horizonal line at *p* = .05 shows the threshold for significance. This figure can be interpreted as a uniform quantile-quantile plot. The *p*-values would fall along the diagonal reference line if they were uniformly distributed, as would be expected under the null. However, we caution readers against interpreting this *p*-value distribution as a *p*-curve offering evidence of a substantive harmful effect of TV exposure for reasons we will elaborate in the next section. The null can be false without a particular favored alternative being true.

**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? In this case, most modeling approaches lead to non-significance; achieving statistical significance is highly model dependent. The finding of significant harm of TV use on attention development is not a robust one.

The most straightforward method of visualizing the relationship -- the simple scatterplots presented in Figure 1 – suggests a lack of compelling evidence for this purported relationship. And yet, 175 of our models produced statistically significant results. We believe that significance in these models is largely triggered by models responding to curvilinearity of the TV-attention relationship, with the remainder mostly consisting of Type-I errors - under a simple binomial model of independent events, 44 false positives (95% range of 32 to 58)[[2]](#footnote-2) would be expected under the null.

**IPTW propensity score analysis post-mortem**. Table 3 describes how the significance of these models varied across the TV cutpoints used to define the low- and high-TV group. The highest rate of significance was associated with the 60th percentile cutoff. Figure 8 displays a zoomed-in version of the residualized scatterplot of TV versus attention (e.g., panel D in Figure 1). The blue loess line is the same as displayed in Figure 1 and is identical across panels. Each panel shows the low- and high-TV cutoffs, and also displays (via the horizonal lines and their shaded confidence bands) the conditional mean of attention in the low- and high-TV groups. These conditional means are similar to the quantities being detected by the propensity score analysis. The figures reveal that the 60th percentile cutoff places the dividing line between low- and high-TV use almost precisely in the center of the non-linear “hump” of the fitted curve, thus mazimizing the raw effect size of the difference in conditional means – while at the same time retaining sufficient data in both the low- and high-TV groups to produce the relatively narrow confidence intervals needed to make that difference statistically significant. The figure reveals that the 50th percentile cutoff and the < 40th/ > 60th percentile cutoffs also accomplish this reasonably well. We believe that it is no coincidence that these models had the highest significance rates.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Table 3 / Figure 8 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Logistic regression analysis post-mortem.** Table 4 shows significance by the cutoff used to define the “normal” and “impaired” attention categories under listwise deletion and multiple imputation. Significance is strongly dependent on the cutoff. It also occurs at a far higher rate under listwise deletion than under multiple imputation, which is suggestive of missing data bias because the sample sizes of the models under listwise deletion were about half that of those with multiple imputations. The two loess fitted lines in Figure 1, panel D show that the nonlinear hump in the data is far less pronounced under multiple imputation than under listwise deletion.

Figure 9 illustrates how this nonlinearity differentially affected the logistic models using various cutpoints to define normal and impaired attention. In each panel of this figure, TV use has been ‘binned up’ into eight categories. This allowed us to plot the proportion of cases classified as exhibiting impaired attention by TV category. The size of the square points is proportional to the number of cases in that category of TV use. A weighted linear regression line fitted to those points represents the logistic model’s behavior; the *p*-value on each panel is that for each’s line’s slope. When the attention cutpoint is low, the nonlinear “hump” manifests itself in the dichotomized data as well. The backside of that hump both pulls the point estimate of the slope toward zero and also results in a high degree of scatter of points about the line, increasing the standard error of the slope (as displayed by the shaded confidence region). As the cutpoint rises, the base rate of impaired attention goes down, pushing all of the points downward. This tends compress the nonlinearity, so that there is no longer backside to pull the regression line toward zero slope, nor is there much variation of points around the line to add uncertainty to the estimate. In short, the logistic regression approach both accentuates the nonlinearity in the data and, as the attention cutpoints rise, quashes one section of it, leaving an apparently linear trend for the models to respond to.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Table 4 / Figure 9 about here

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

That the significance rate for each attention cutpoint was far lower when multiple imputation was used (and thus, the nonlinear feature is diminished, see Table 4) supports our contention that this set of results, like the IPTW model results, was driven by this nonlinearity.

**Conclusion**. 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 may be 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 888 models that we fit to the data did not find evidence that the TV-attention link claimed by Christakis et al. (2004) is robust to model specification. The significance exhibited by a minority (175, 19.7%) of the models appears to be related to overfitting a small feature of the data, and one that we suspect to have been driven by chance (and therefore not replicable). 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 | 2108 | 7.75 | 0.61 | 6.75 | 8.75 |
| Annual family income (thousands) | 1958 | 33.42 | 24.53 | 0 | 189.92 |
| Attention (raw) | 2108 | 2.64 | 0.39 | 1 | 3 |
| Attention within-sex SS | 2075 | 101.25 | 13.79 | 83 | 136 |
| CES-D Depression score (1992) | 2089 | 46.97 | 7.87 | 32.3 | 79.9 |
| Cognitive stimulation of home age 1-3 | 1907 | 97.61 | 16.15 | 11.1 | 148.2 |
| Emotional support of home age 1-3 | 1765 | 97.99 | 16.58 | 31.6 | 124.7 |
| Gestational age at birth | 1960 | -1.41 | 1.96 | -14 | 7 |
| Mother's age at birth | 2108 | 28.48 | 2.62 | 23 | 36 |
| Mother's years of schooling | 2095 | 12.95 | 2.48 | 0 | 20 |
| Number of children in household | 2097 | 1.64 | 1.2 | 0 | 7 |
| Partner's years of schooling | 1757 | 13.28 | 2.7 | 1 | 20 |
| Rosenberg self-esteem score (1987) | 2040 | 45.07 | 8.4 | 23.5 | 59.7 |
| Temperament | 1961 | 2.01 | 0.69 | 1 | 5 |
| TV hours per day age 1.5 | 1993 | 2.23 | 3.07 | 0 | 16 |
| TV hours per day age 3 | 2023 | 3.68 | 3.12 | 0 | 16 |
|  |  |  |  |  |  |



Table 2

Marginal descriptive statistics for categorical variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Value | n | Percent |
|  |  |  |  |
| Maternal alcohol use in pregnancy | No | 1050 | 49.81% |
|  | Yes | 932 | 44.21% |
|  | (missing) | 126 | 5.98% |
|  |  |  |  |
| Cohort (interview wave when attention was assessed) | 1996 | 829 | 39.33% |
| 1998 | 796 | 37.76% |
|  | 2000 | 483 | 22.91% |
|  |  |  |  |
| Father absent from household | No | 1681 | 79.74% |
|  | Yes | 399 | 18.93% |
|  | (missing) | 28 | 1.33% |
|  |  |  |  |
| Child sex | Female | 1034 | 49.05% |
|  | Male | 1074 | 50.95% |
|  |  |  |  |
| Low birth weight (< 2500g) | No | 1812 | 85.96% |
|  | Yes | 138 | 6.55% |
|  | (missing) | 158 | 7.50% |
|  |  |  |  |
| Health condition that limits school or play | No | 1917 | 90.94% |
| Yes | 122 | 5.79% |
|  | (missing) | 69 | 3.27% |
|  |  |  |  |
| Premature birth | No | 1744 | 82.73% |
|  | Yes | 216 | 10.25% |
|  | (missing) | 148 | 7.02% |
|  |  |  |  |
| Child race | Black | 572 | 27.13% |
|  | Hispanic | 397 | 18.83% |
|  | White | 1139 | 54.03% |
|  |  |  |  |
| Maternal smoking in pregnancy | No | 1447 | 68.64% |
|  | Yes | 528 | 25.05% |
|  | (missing) | 133 | 6.31% |
|  |  |  |  |
| Standard metropolitan statistical area (urbanicty) | Not in SMSA | 382 | 18.12% |
| SMSA; central city unknown | 680 | 32.26% |
|  | SMSA; in central city | 302 | 14.33% |
|  | SMSA; not central city | 639 | 30.31% |
|  | (missing) | 105 | 4.98% |
|  |  |  |  |



Table 3

IPTW propensity score model results by TV cutpoint

|  |  |  |  |
| --- | --- | --- | --- |
| TV cutpoint percentile | Non-sig | Sig | Proportion sig |
|  |  |  |  |
| 20/80 | 53 | 11 | 0.172 |
| 30/70 | 54 | 10 | 0.156 |
| 40/60 | 47 | 17 | 0.266 |
| 50 | 44 | 20 | 0.312 |
| 60 | 37 | 27 | 0.422 |
| 70 | 50 | 14 | 0.219 |
|  |  |  |  |



*Note:* When two numbers are given for the cutpoint percentile, this implies that TV use between those percentiles were dropped. So 20/80 means that the low-TV group was defined as the 20th percentile or lower and the high-TV group as the 80th percentile or higher. When a single value is given, it means TV use below that percentile was categorized low-TV, and TV use above it was categorized as high-TV. *Non-sig*: the number of models using the specified attention cutpoint and missing data treatment that did not yield statistical significance. *Sig*: the number of models that yielded statistical significance.

Table 3

Logistic regression results by attention cutpoint and missing data treatment

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Listwise | | |  | Multiple imputation | | |
| Attention cutpoint |  | Non-sig | Sig | Proportion sig |  | Non-sig | Sig | Proportion sig |
|  |  |  |  |  |  |  |  |  |
| 110 |  | 8 | 0 | 0.000 |  | 4 | 0 | 0.000 |
| 111 |  | 8 | 0 | 0.000 |  | 4 | 0 | 0.000 |
| 112 |  | 8 | 0 | 0.000 |  | 4 | 0 | 0.000 |
| 113 |  | 8 | 0 | 0.000 |  | 4 | 0 | 0.000 |
| 114 |  | 8 | 0 | 0.000 |  | 4 | 0 | 0.000 |
| 115 |  | 7 | 1 | 0.125 |  | 4 | 0 | 0.000 |
| 116 |  | 7 | 1 | 0.125 |  | 4 | 0 | 0.000 |
| 117 |  | 6 | 2 | 0.250 |  | 4 | 0 | 0.000 |
| 118 |  | 6 | 2 | 0.250 |  | 4 | 0 | 0.000 |
| 119 |  | 5 | 3 | 0.375 |  | 3 | 1 | 0.250 |
| 120 |  | 3 | 5 | 0.625 |  | 2 | 2 | 0.500 |
| 121 |  | 3 | 5 | 0.625 |  | 2 | 2 | 0.500 |
| 122 |  | 3 | 5 | 0.625 |  | 2 | 2 | 0.500 |
| 123 |  | 1 | 7 | 0.875 |  | 2 | 2 | 0.500 |
| 124 |  | 1 | 7 | 0.875 |  | 2 | 2 | 0.500 |
| 125 |  | 0 | 8 | 1.000 |  | 2 | 2 | 0.500 |
| 126 |  | 0 | 8 | 1.000 |  | 2 | 2 | 0.500 |
| 127 |  | 0 | 8 | 1.000 |  | 2 | 2 | 0.500 |
| 128 |  | 0 | 8 | 1.000 |  | 3 | 1 | 0.250 |
| 129 |  | 0 | 8 | 1.000 |  | 3 | 1 | 0.250 |
| 130 |  | 0 | 8 | 1.000 |  | 3 | 1 | 0.250 |
|  |  |  |  |  |  |  |  |  |

Note: *Non-sig*: the number of models using the specified attention cutpoint and missing data treatment that did not yield statistical significance. *Sig*: the number of models that yielded statistical significance.

Figure 1

Scatterplots of early childhood TV use versus standardized within-sex attention score at age 7. *Left column*: TV measured at age ~1.5. *Right column*: TV measured at age ~3. *Top row*: raw data. *Bottom row*: adjusted (residualized) attention score with effect of covariates removed. Bottom panel: red ‘x’ points are adjusted based on imputed covariate values. Solid blue smoothing line fits to non-missing data only; red dashed smoothing line fit all data (including imputed values).

A close up of a map

Description automatically generated

Note: point locations are slightly jittered to reduce overplotting

Figure 2

A picture containing screenshot

Description automatically generatedIPTW propensity score model results summary

Note: Left panels represent the effect of TV measured at age ~1.5. Right panels represent the effect of TV measured at age ~3. Top and middle rows plot effect sizes and their 95% Cis in Cohen’s *d* units. Bottom row display the *p*-values; the horizontal dotted reference line displays the .05 threshold for significance. Statistically significant points are shaded.

Figure 3

A close up of a map

Description automatically generatedStratification propensity score model results summary

Note: Left panels represent the effect of TV measured at age ~1.5. Right panels represent the effect of TV measured at age ~3. Top and middle rows plot effect sizes and their 95% Cis in Cohen’s *d* units. Bottom row display the *p*-values; the horizontal dotted reference line displays the .05 threshold for significance. Statistically significant points are shaded.

Figure 4

Regression model results

A screenshot of a cell phone

Description automatically generated

Note: Left panels represent the effect of TV measured at age ~1.5. Right panels represent the effect of TV measured at age ~3. Top and middle rows plot effect sizes and their 95% Cis in raw units; the expected change in the outcome given a one-unit (one hour) change in TV watching. Bottom row display the *p*-values; the horizontal dotted reference line displays the .05 threshold for significance. Statistically significant points are shaded.

Figure 5

Logistic regression results summary

A screenshot of a cell phone

Description automatically generated

Note: Left panels represent the effect of TV measured at age ~1.5. Right panels represent the effect of TV measured at age ~3. Top and middle rows plot effect sizes and their 95% Cis in odds ratio units; the expected change in odds of being classified into the impaired attention category given a one-unit (one hour) change in TV watching. Bottom row display the *p*-values; the horizontal dotted reference line displays the .05 threshold for significance. Statistically significant points are shaded.

Figure 6

*p­­*-value summary for all models

A close up of a map

Description automatically generated

Note: The histogram on the right margin shows the distribution of *p*-values across analyses. The horizontal dotted line represents the .05 threshold for significance. The dashed diagonal reference line illustrates the expected location on the p-values under the null hypotheses, when they would follow a uniform distribution.

Figure 8

IPTW propensity score model post-mortem

A close up of a map

Description automatically generated

Note: Panels display the zoomed-in residualized attention outcome (e.g., controlling for covariates versus TV measured at age ~3. Each panel depicts a different set of TV percentile cutpoints for defining the low- and high-TV groups; which are displayed as dark vertical lines. The blue curve on each plot is the loess smoother, illustrating the location and scale of the nonlinear feature. The red horizontal lines and their shaded uncertainty regions display the conditional mean attention in the low- and high-TV categories.

Figure 9

Logistic regression post-mortem

A screenshot of a computer

Description automatically generated

Note: Panels display the probability of impaired attention (*y*-axis), as defined by the attention cutpoint displayed on each panel, versus TV measured at age ~3 (*x*-axis). The size of each point is proportional to the number of cases in that level of TV consumption. A fitted regression line (weighted by the sample size within each TV grouping) and its slope *p*-value are depicted on each plot; these depict the performance of the logistic regression model in each situation. As the cutoffs rise, the association between TV and the probability of impaired attention seems to increase.

1. A prior version of this analysis also included the child’s body mass index (BMI), but we removed that variable at the direction of a reviewer, who was concerned about potential endogeneity. [↑](#footnote-ref-1)
2. In R, this can be calculated as follows: qbinom(p=c(.025, .5, .975), size=888, prob=.05) [↑](#footnote-ref-2)