Analysis of multiple specifications
Online Supplement

Christoph Semken David Rossell

01 April 2021

Contents

This document contains supplementary information for our paper "Analysis of multiple specifications: statistical validity and a Bayesian proposal". In Section 1 we describe how BMS and BMA are used to produce a Bayesian Specification Curve Analysis (BSCA) and how to replicate our main results on teenager well-being using R. Section 2 contains supplementary data analyses that assess the robustness of our main findings for the Youth Risk Behavior Survey (YBRS) and Millenium Cohort Study (MCS) datasets. These robustness analyses include comparing linear to non-linear treatment effects and considering other well-being outcomes than those presented in the main paper. Section 3 describes certain aspects in which our analysis differs from that of Orben and Przybylski (2019). One motivation for these differences was to consider the inclusion of further control covariates, which as discussed in the main paper, is necessary to reduce biases in parameter estimates. We also adjusted the definition of some of the variables – e.g. expressing them on a common scale to facilitate interpretation and comparability, using validated psychometric scales as opposed to (unvalidated) individual outcome variables, fixing minor errors related to variable codes and including unemployed parents in the analysis.

1 Reproducing the BSCA results for the teenager well-being datasets

We explain how to obtain BMA and BSCA results in R by reproducing our findings in the two teenager well-being datasets. Subsection 1 loads R packages and functions needed for the analysis, as well as the Youth Risk Behavior Survey (YBRS) and Millenium Cohort Study (MCS) datasets on adolescent mental well-being and technology use. Subsection 2 applies Bayesian model selection and averaging to both datasets. Finally, Subsections 3 and 4 produce Figs. 1 (single-outcome BSCA) and 2 (multiple-outcome BSCA) presented in the manuscript, respectively.

1.1 Setup

We start by loading the required R packages. For the statistical analysis we use `mombf` and `BAS`, whereas `tidyverse` offers some convenient functions for treating the data. We source file `functions.R`, which contains auxiliary functions to produce the BSCA plots, and load the pre-processed version of the YBRS and MCS datasets in files `yrbs.Rdata` and `mcs.Rdata`, respectively. We cannot provide these pre-processed data due to copyright issues, but you can run the code provided in the replication file `bsca_preanalysis.Rmd` to create the processed data from the raw data and code provided by Orben and Przybylski (2019). See also `bsca_preanalysis.html` for a compiled version displaying the R code and output.

```
library(mombf)
library(BAS)
library(tidyverse)
```

```
source('code/functions.R')
yrbs = new.env(); mcs = new.env()
load('data/export/yrbs.Rdata', yrbs)
load('data/export/mcs.Rdata', mcs)
```

1.2 Bayesian model selection

1.2.1 YRBS data

We begin by analyzing the YRBS data. The data frame `y` stores several outcomes, whereas `x` stores treatment variables and `data` stores other recorded variables. We specify that we wish to use the outcome variable *thought about suicide* (the second column in `y`) by setting `idy=2`. We also specify that we wish to use TV and electronic device use (first and second column in `x`) as treatment variables to be jointly included in all regression models by setting `idx=c(1,2)`. Finally we specify to use race, age, gender, school grade, survey year and body mass index as potential control variables (saved in `cvars` and `cvarsplus`).

```
attach(yrbs)
names(y)
```

```
## [1] "loneliness"     "think suicide"  "plan suicide"   "commit suicide"
## [5] "doctor suicide"
```

```
names(x)
```

```
## [1] "TV Use"                "Electronic Device Use"
```

```
c_names
```

```
## [1] "Race"    "Aged 12" "Aged 13" "Age"     "Sex"     "Grade"   "Year"
## [8] "BMI"
```

```
idy = 2; idx = c(1,2)
datareg = data.frame(y[,idy], x[,idx], data[,c(cvars,cvarsplus)])
names(datareg) = c('y', names(x)[idx], c_names)  # set names
datareg = datareg[rowSums(is.na(datareg))==0, ]  # remove NAs
detach(yrbs)
```

The data.frame `datareg` contains the outcome, treatment and control variables. For illustration, its first few rows are displayed below. These variables have been conveniently coded so they can be entered directly into the usual R regression equation. For instance, `Aged 12` and `Aged 13` are indicators for an individual's age being 12 and 13 years, whereas `Age` contains the age in years, using these 3 columns to code the effect of age allows to capture non-linear effects detected in preliminary exploratory data analyses (see reproduction file `bsca_preanalysis.Rmd`).

```
head(datareg %>% rename('ED Use' = 'Electronic Device Use'))
```

	y	TV Use	ED Use	Race	Aged 12	Aged 13	Age	Sex	Grade	Year	BMI
45	0	medium	0.333	0	1	0	12	1	11	2007	20.85
46	0	low	0.000	0	1	0	12	1	12	2007	23.85
47	0	medium	0.500	0	1	0	12	1	9	2007	18.04
50	1	high	1.000	0	1	0	12	1	12	2007	29.47
54	0	low	0.167	0	1	0	12	1	11	2007	26.36
55	0	high	0.000	0	0	1	13	1	9	2007	21.19

A first step in BSCA is to run Bayesian model selection (BMS), which will assign a score (posterior probability) to each model (possible set of control variables). Next, we use Bayesian model averging (BMA) to combine these estimates. Since the outcome variable is binary we use logistic regression models, setting a uniform prior on the model size (`modelbbprior(1,1)`). The function `mombf:::modelSelectionGLM` computes scores for all 1024 possible models, and may take a while to run.

```
yrbs_ms = mombf:::modelSelectionGLM(
  y ~ ., data=datareg,
  includevars=1, familyglm=binomial(link='logit'),
  priorDelta=modelbbprior(1,1)
)
```

```
## Enumerating 1024 models.........
```

The BMA treatment effect estimates, 95% posterior intervals and marginal posterior probability that the variable has an effect on the outcome are stored in `yrbs_coef`. The output indicates a strong evidence that both treatments have an effect, albeit with opposing signs, and that age, gender, grade and BMI are control covariates that one should include in the model to avoid biases in the treatment effect estimates (driven by under-selection of truly relevant controls).

```
yrbs_coef = coef(yrbs_ms)
options(scipen=999)  # turn off scientific notation
yrbs_coef
```

	estimate	2.5%	97.5%	margpp
(Intercept)	-1.345	-1.764	-1.076	1.000
‘TV Use‘medium	-0.233	-0.280	-0.185	1.000
‘TV Use‘high	-0.047	-0.119	0.027	1.000
‘Electronic Device Use‘	0.629	0.565	0.692	1.000
Race	0.001	0.000	0.026	0.036
‘Aged 12‘	1.153	0.000	2.166	0.790
‘Aged 13‘	-0.001	0.000	0.000	0.009
Age	0.011	0.000	0.072	0.221
Sex	-0.789	-0.833	-0.745	1.000
Grade	-0.086	-0.147	-0.058	1.000
Year2009	0.000	0.000	0.000	0.006
Year2011	0.000	0.000	0.000	0.006
Year2013	0.000	0.000	0.000	0.006
Year2015	0.001	0.000	0.000	0.006
BMI	0.025	0.021	0.029	1.000

Given these coefficients, we use the function `getOR` to obtain odds ratios for increasing the TV use from low to medium/high, and for increasing the electronic device use from 0 to >= 5 hours (coded as EDU=1 in our dataset, leading to setting `treatvals=1` below). The function exponentiates the coefficient estimates and formats the result.

```
getOR(yrbs_coef, treat='TV Use', digits=2)
```

	OR	CI.low	CI.up
‘TV Use‘medium	0.79	0.76	0.83
‘TV Use‘high	0.95	0.89	1.03

```
getOR(yrbs_coef, treat='Electronic Device Use', treatvals=1, digits=2)
```

	OR	CI.low	CI.up
Electronic Device Use 1	1.88	1.76	2

Note that the regression models include simultaneously the two treatment variables, TV and electronic device (ED) usage, which is necessary to avoid biased estimates when treatments are correlated. In these data the correlation is mild, for instance Pearson's correlation between the number of hours of TV use and ED use (columns `q81` and `q82` in `data`) is 0.21.

```
round(cor(na.omit(yrbs$data[c('q81', 'q82')]), method='pearson')[1,2],
      digits=2)
```

```
## [1] 0.21
```

1.2.2 MCS data

Next, we load the MCS dataset and run Bayesian model selection and averaging, analogously to the above analysis of the YRBS data. As described in the main manuscript, for illustration in our analysis we considered 4 outcome variables, 5 treatments and 14 potential control variables (results for other outcome variables are shown in Fig. S1). We display the names of these variables, which have been stored in the `mcs` workspace.

```
attach(mcs)
```

```
names(yvars)
```

```
## [1] "Depressed (adolescent)"
## [2] "Low self-esteem (adolescent)"
## [3] "High total difficulties (parent)"
## [4] "High emotional problems (parent)"
## [5] "High conduct problems (parent)"
## [6] "High hyperactivity/inattention (parent)"
## [7] "High peer problems (parent)"
## [8] "Low pro-sociality (parent)"
```

```
x_names
```

```
## [1] "TV"               "Electronic games" "Social media"     "Other internet"
## [5] "Own computer"
```

```
names(cvars)
```

```
##  [1] "Male"       "Age"        "BMI"        "Motivation" "Ethnicity"
##  [6] "Closeness"  "Father"     "Score"      "Employed"   "Illness"
## [11] "Time"       "Distress"   "Siblings"   "Income"
```

The code below runs BMA for all 5 outcome variables. One option is to use the `BAS` package, which implements MCMC sampling to explore the model space (sets of control covariates to be potentially included in the regression). A faster alternative analysis is to set `fast = TRUE`, which limits the analysis to the top 100 models (i.e. those shown in the BSCA plot) according to a prior screening (stored in `pp_lin`, see `bsca_preanalysis.Rmd` for details). The posterior probabilities of any model after the 100 first ones is vanishingly small, and the final BSCA results are virtually identical to those of the analysis using the `BAS` package.

```
fast = TRUE

mcs_coef = list(); mcs_ms = list()
for (idy in 1:length(yvars)) {
  # select data
  yvar = yvars[idy]; yname = names(yvars)[idy]
  datareg = na.omit(data[c(yvar, x_vars, cvars)])
  names(datareg) = c('y', x_names, names(cvars))

  # BMA
  if (fast) {
    models = as.matrix.pp(
      pp_lin[[idy]], nummodels=100, numvars=length(datareg)
    )
    mcs_ms[[yname]] = mombf:::modelSelectionGLM(
      y ~ ., data=datareg, models=models,
      familyglm= binomial(link='logit'),
      priorDelta=modelbbprior(1,1)
    ); cat('\n')
  } else {
    mcs_ms[[yname]] = BAS:::bas.glm(
      y~., data=datareg, family=binomial(link='logit'),
      betaprior=bic.prior(), modelprior=beta.binomial(),
      method='MCMC', n.models=150
    )
  }
  mcs_coef[[yname]] = coef(mcs_ms[[yname]])
}
```

```
## Enumerating 100 models.........
## Enumerating 100 models.........
## Enumerating 100 models.........
## Enumerating 100 models.........
## Enumerating 100 models.........
## Enumerating 100 models.........
## Enumerating 100 models.........
## Enumerating 100 models.........
```

```
detach(mcs)
```

We can inspect the BMA results for the four outcomes. For brevity here we focus on *parent-assessed* high total difficulties and *adolescent-assessed* depression (see Subsection 3 below for a plot summarizing the BMA results for the two other outcomes). The analysis provides strong evidence that, according to parents, social media decrease the odds of total

difficulties (marginal posterior probability=1, up to rounding), whereas electronic games increase those odds. We also find strong evidence that BMI, educational motivation, closeness to parents, the primary caregiver's word ability score and psychological distress, presence of a longstanding illness and the household income are necessary control covariates, as well as strong evidence that other covariates are not needed.

```
mcs_coef[["High total difficulties (parent)"]]
```

	estimate	2.5%	97.5%	margpp
(Intercept)	1.888	1.046	2.758	1.000
TV	-0.005	0.000	0.000	0.017
'Electronic games'	0.571	0.349	0.789	1.000
'Social media'	-0.717	-0.954	-0.471	1.000
'Other internet'	-0.001	0.000	0.000	0.004
'Own computer'	0.000	0.000	0.000	0.003
Male	0.000	0.000	0.000	0.005
Age	-0.003	0.000	0.000	0.028
BMI	0.043	0.027	0.058	1.000
Motivation	-0.790	-0.939	-0.643	1.000
Ethnicity	0.000	0.000	0.000	0.003
Closeness	-0.537	-0.651	-0.427	1.000
Father	0.049	0.000	0.302	0.244
Score	-0.055	-0.074	-0.037	1.000
Employed	-0.004	-0.040	0.000	0.032
Illness	1.038	0.880	1.197	1.000
Time	-0.011	-0.120	0.000	0.128
Distress	0.113	0.098	0.128	1.000
Siblings	-0.001	-0.012	0.000	0.031
Income	-0.001	-0.002	-0.001	1.000

The output for adolescent-assessed depression can be interpreted analogously. Briefly, here there is strong evidence that social media and other internet use increase the odds of depression, in contrast with the results of the earlier parent-assessed outcome. Some of the needed control covariates are also different, for instance males self-report lower odds of depression than females, whereas gender did not play a role in the parental assessment.

```
mcs_coef[["Depressed (adolescent)"]]
```

	estimate	2.5%	97.5%	margpp
(Intercept)	4.100	3.195	4.971	1.000
TV	0.000	0.000	0.000	0.005
'Electronic games'	0.006	0.000	0.050	0.028
'Social media'	0.999	0.648	1.348	1.000
'Other internet'	1.036	0.667	1.408	1.000
'Own computer'	-0.001	0.000	0.000	0.008
Male	-1.113	-1.272	-0.955	1.000
Age	0.001	0.000	0.000	0.014
BMI	0.045	0.028	0.062	1.000
Motivation	-1.607	-1.760	-1.452	1.000
Ethnicity	-0.003	0.000	0.000	0.018
Closeness	-0.783	-0.892	-0.671	1.000
Father	0.000	0.000	0.000	0.008
Score	0.014	0.000	0.040	0.501
Employed	0.000	0.000	0.000	0.005
Illness	0.513	0.340	0.687	1.000
Time	0.000	0.000	0.000	0.009
Distress	0.002	0.000	0.025	0.109
Siblings	-0.009	-0.110	0.000	0.110
Income	0.000	0.000	0.001	0.086

To summarize the treatment effects of interest we again use the auxiliary function `getOR`. These correspond to odds ratios for increasing the use of social media from 0 (no usage) to >7 hours (coded as 1 in our dataset, hence we set `treatvals=1`). We first obtain odds-ratios and 95% posterior intervals for social media and electronic games on parent-assessed total difficulties.

```
getOR(mcs_coef$`High total difficulties (parent)`, treat='Social media',
      treatvals=1, digits=2)
```

	OR	CI.low	CI.up
Social media 1	0.49	0.39	0.62

Next we report the odds ratios for adolescent self-assessed depression and low self-esteem.

```
getOR(mcs_coef$`Depressed (adolescent)`, treat='Social media',
      treatvals=1, digits=2)
```

	OR	CI.low	CI.up
Social media 1	2.72	1.91	3.85

```
getOR(mcs_coef$`Low self-esteem (adolescent)`, treat='Social media',
      treatvals=1, digits=2)
```

	OR	CI.low	CI.up
Social media 1	1.52	1	2.76

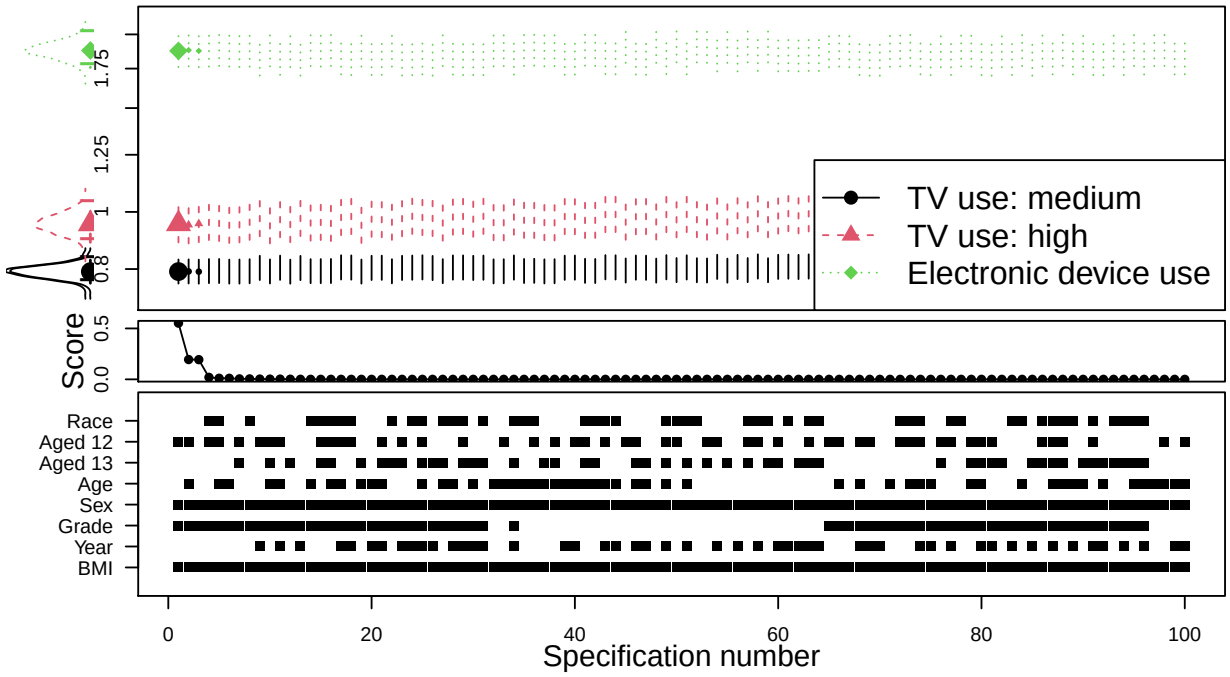
1.3 Reproducing Figure 1 (single-outcome BSCA)

1.3.1 YRBS data

We use function `single_bsca` to plot the single-outcome BSCA for the YRBS data and the outcome *thought about suicide*. The argument `coefidx` specifies the names of the treatment variables that should be plotted. The function also allows specifying optional arguments such as the treatment names to be displayed in the legend (argument `x.labels`), variable names to be displayed in the bottom panel displaying variable configurations (argument `var.labels`), and omitting variables from that panel (argument `omitvars`, useful when there are many variables or when several columns code for the non-linear effect of a single variable and are always included together, such as year in the YRBS data). The labels on the y axis are stored in an array whose names (optionally) are the original values (argument `y.labels`). Here, we turn the estimated coefficient into the odds ratio by exponentiating it.

```
idx_fit = c(2:4)
id_years = c(12:14)
y_labels = c(0.8, 1, 1.25, 1.5, 1.75, 2)
names(y_labels) = log(y_labels)  # y scale as odds ratio

single_bsca(
  yrbs_ms, coefidx=idx_fit, omitvars=c(1, idx_fit, id_years),
  x.labels=yrbs$x_labels, var.labels=yrbs$c_names, y.labels=y_labels
)
```

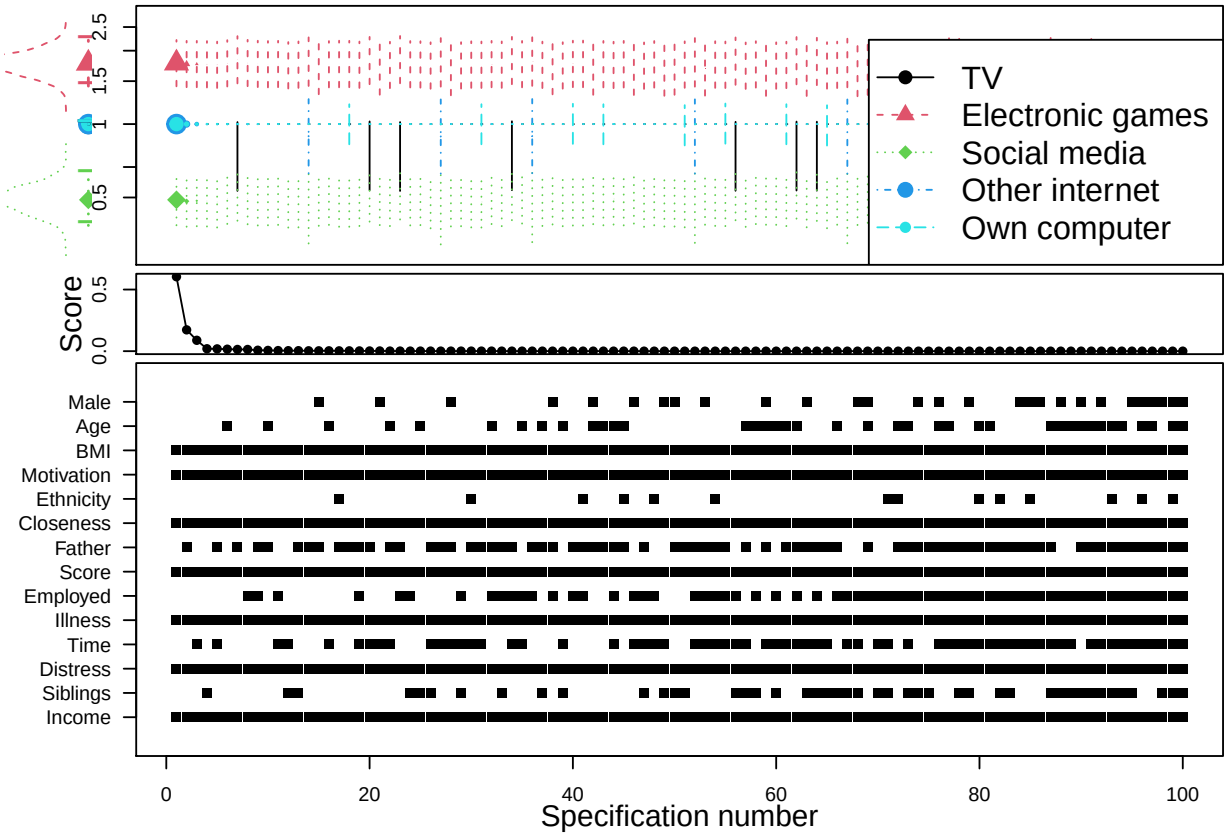


1.3.2 MCS data

Alternatively, we can plot the single-outcome BSCA for the MCS data and the parent-assessed outcome *high total difficulties*.

```
y_labels = c(1/2, 1/1.5, 1, 1.5, 2, 2.5)
names(y_labels) = log(y_labels)  # y scale as odds ratio

single_bsca(
  mcs_ms$`High total difficulties (parent)`, coefidx=2:6,
  x.labels=mcs$x_names, var.labels=names(mcs$cvars), y.labels=y_labels,
  height.vars=0.55
)
```



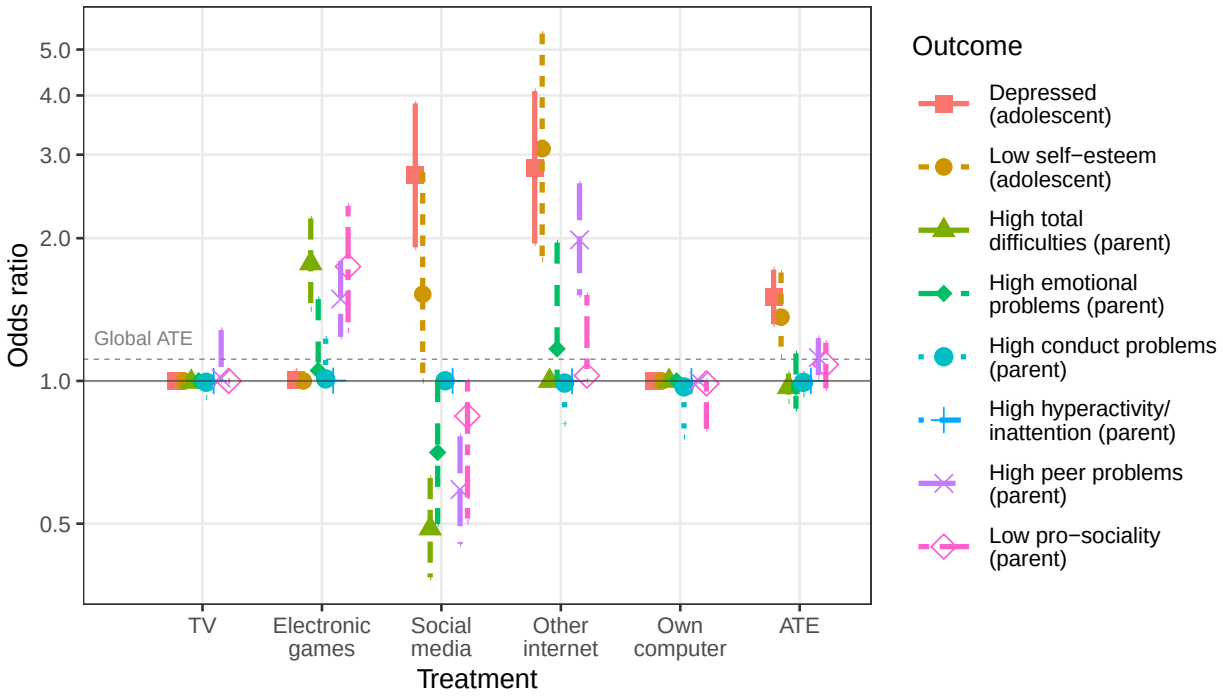
1.4 Reproducing Figure 2 (multiple-outcome BSCA)

Our MCS data analysis included eight outcomes and five treatments of interest, for a total of 40 treatment-outcome combinations. To reduce the burden associated with producing a single BSCA plot for each outcome, the function `multi_bsca` summarizes the BMA results in a single plot. This plot allows one to easily evaluate and compare effects of several treatments on several outcomes. For instance, below all treatments have a similar effect on adolescent-assessed depression and on self-stem. However, these effects are non-comparable to those on parent-assessed total difficulties and emotional problems.[1] We also add the ATE across treatments (`add.ate=TRUE`) and the simple average across outcomes (`add.avg=TRUE`). The average of the ATE estimates is the global ATE.

```
g = multi_bsca(mcs_coef, ms=mcs_ms, conversion=exp,  y.scale='Odds ratio',
           treatments=c('TV', 'Electronic games', 'Social media',
                        'Other internet',  'Own computer'),
```

[1]Using the `BAS` package (if `fast=FALSE`), currently shows normal approximations for the 95% interval in the summary plot. Thus, it looks different from the `mombf` version, which estimates the 95% interval using posterior sampling. Since the latter are more exact, we show it in the main paper. However, running the above analysis using `BAS` shows that the main results remain unchanged (up to rounding) if one samples the entire model space. Adding the ATE is not supported for the `BAS` package.

```
            add.ate=TRUE, add.global.ate=TRUE) +
  scale_y_continuous(trans='log', breaks=c(1/2,1:5), minor_breaks=NULL) +
  theme(legend.key.size=unit(0.9, 'cm'))
g + geom_hline(yintercept=1, lwd=0.2, colour=g$theme$panel.border$colour)
```



2 Robustness checks

In this section, we provide robustness checks for the results presented in the main paper.

2.1 YRBS: alternative outcomes

In their analysis of the YRBS data, besides 'thought about suicide', Orben and Przybylski (2019) used several well-being measures to study the effect of technology use: loneliness, planned suicide, attempted to commit suicide and saw a doctor about suicide. We have reproduced Fig. 1 for these outcomes in Fig. S1. All associations are qualitatively similar, although the magnitudes vary.

2.2 YRBS: linear regression

All outcome variables in the YRBS are binary (e.g. 0 = did not think about suicide, 1 = thought about suicide). Orben and Przybylski (2019) used linear regression, which is unsuitable for binary outcomes. Instead, we used logistic regression to model the probability of the outcome being one. The logistic regression coefficient is easily interpretable; it is the log odds-ratio associated of said probability relative to the reference group, for example the

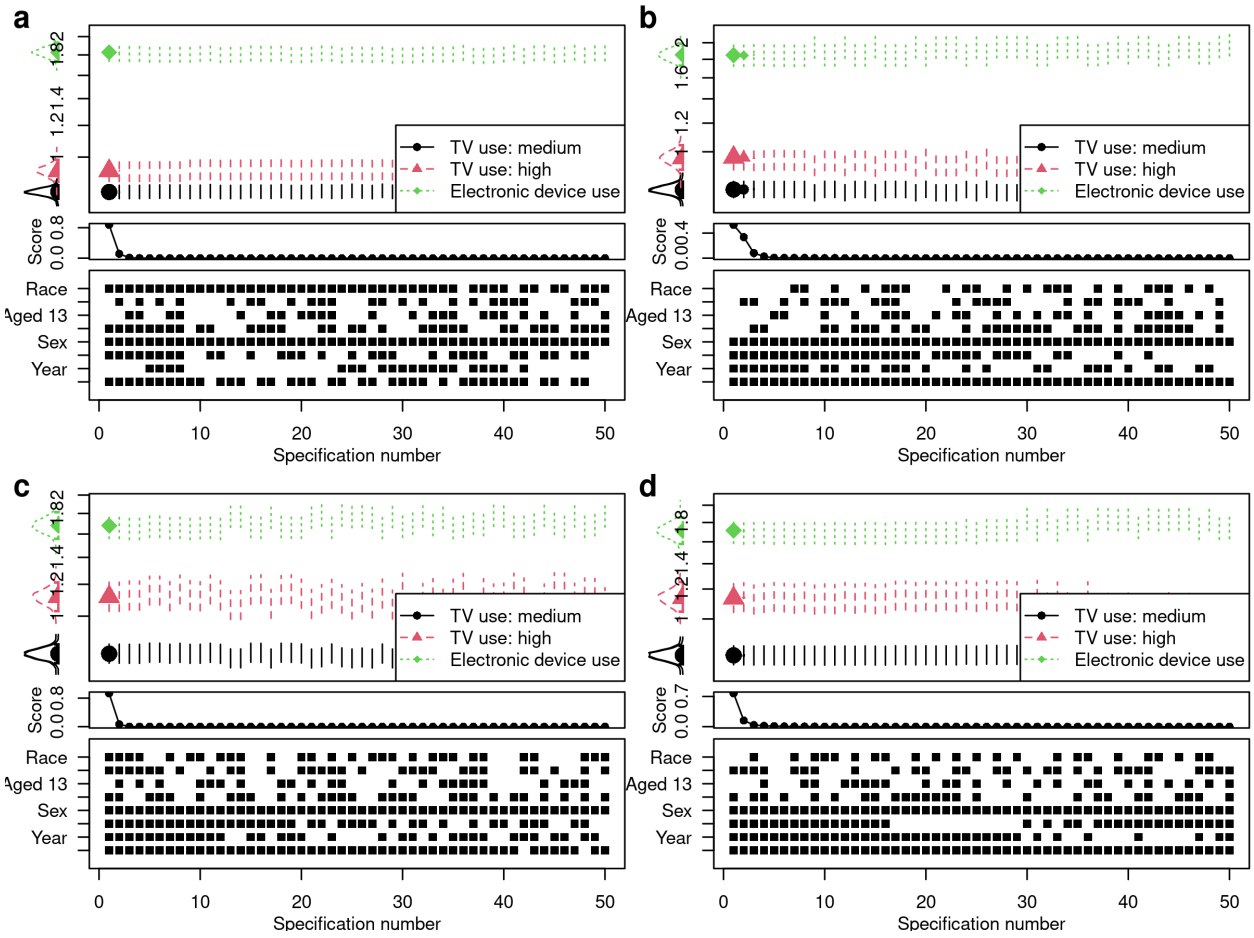


Figure S1: BSCA for other outcomes using YRBS data: (a) loneliness, (b) planned suicide, (c) attempted to commit suicide and (d) saw a doctor about suicide. Otherwise, everything – including the order of control variables in the bottom panel – is as in Fig. 1 of the main text.

log odds-ratio of the probability that a teenager who watches a moderate amount of TV thinking about suicide relative to a teenager who does not watch TV.

For robustness, Fig. S2 shows the single-outcome BSCA for a linear regression model. Our main results remain qualitatively unaltered: electronic device use is associated with an increase in the probability of thinking about suicide, whereas moderate TV use is associated with a decrease.

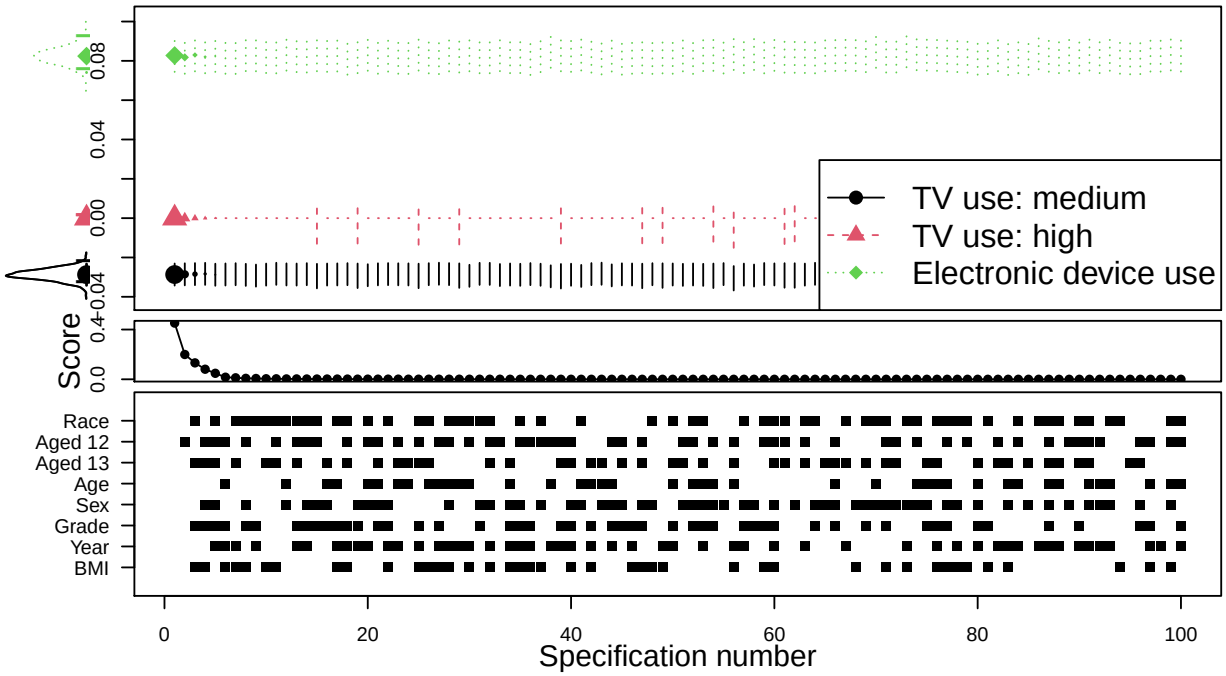


Figure S2: YRBS data. BSCA for the effect technology uses on thinking about suicide. Estimates are obtained using the linear probability model, by iterating over all possible models – including with only one TV coefficient removed. Otherwise, everything is specified as in Fig. 1 of the main text.

2.3 YRBS: lineary of association with TV use

Preliminary exploratory data analyses (see the supplementary file `bsca_preanalysis.html`) revealed that in both the YRBRS and MCS datasets almost all treatments had a monotone, near-linear association. The only exception was TV usage in the YRBS data, which displayed a U-shaped association with adolescent well-being, see Fig. S3.

Therefore, we estimated separate coefficients for medium and high TV use based on coefficient similarity in Fig. S3 (see the main text for the exact coding). For completeness, we also performed a BSCA with an assumed linear TV effect. As expected, the association was in between that for medium and high usage, as shown in Fig. S4. Note that the estimate still suggests that higher TV use is associated with lower odds of thinking about suicide, as in our main analysis.

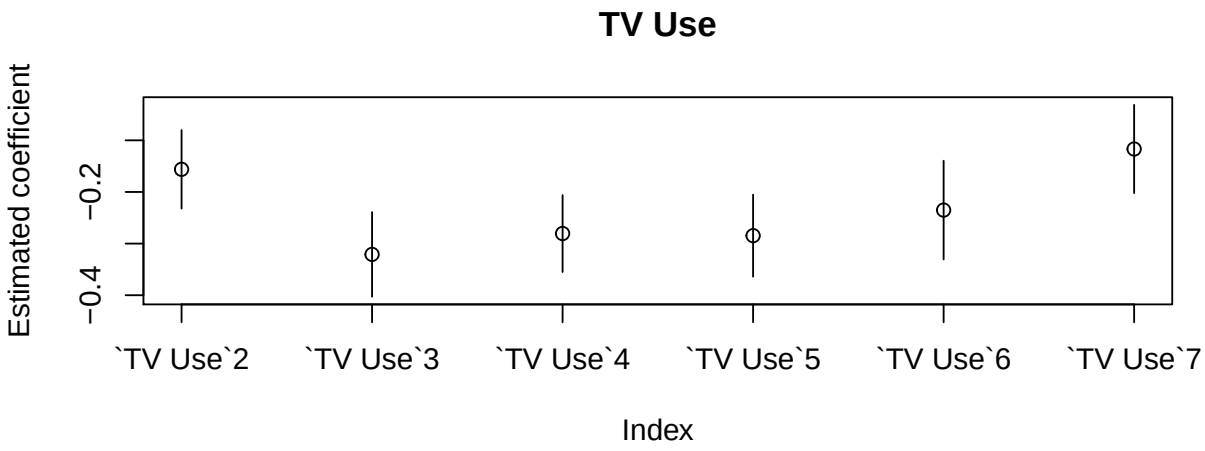


Figure S3: YRBS data. Estimated coefficient for different values of TV use on thinking about suicide. Coefficients were estimated by MLE for a model including all treatment and control variables. Estimated coefficients are qualitatively similar for other outcomes.

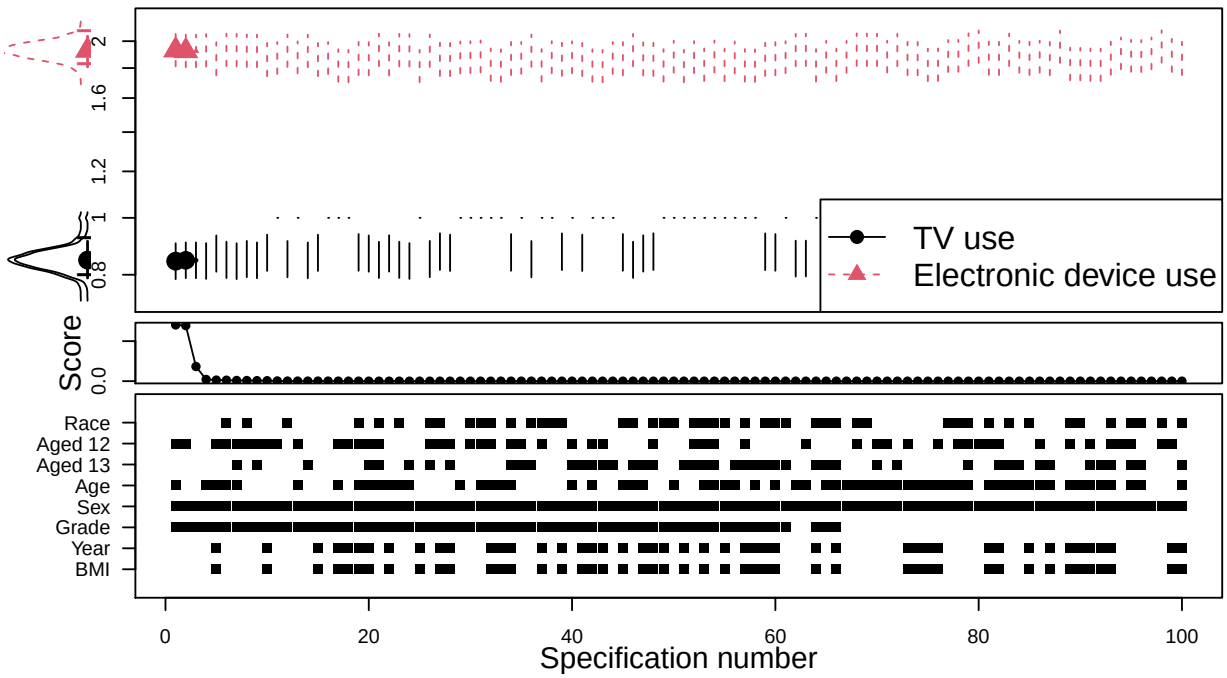


Figure S4: YRBS data. Bayesian SCA for the effect of technology uses on thinking about suicide. TV use is numeric, with 0 = no usage and 1 = 5 hours or more. Otherwise, everything is specified as in Fig. 1 of the main text.

2.4 MCS: full BSCAs

Fig. 2 of the main paper shows the summary of coefficients for different (parent- and self-assessed) outcomes. We showed the multiple-outcome BSCA for brevity. Fig. S5 shows the single-outcome BSCAs for those outcomes discussed in the text (self-assessed depression and low self-esteem, parent-assessed high total difficulties and emotional problems), whereas Fig. S6 shows those for the remaining parent-assessed outcomes.

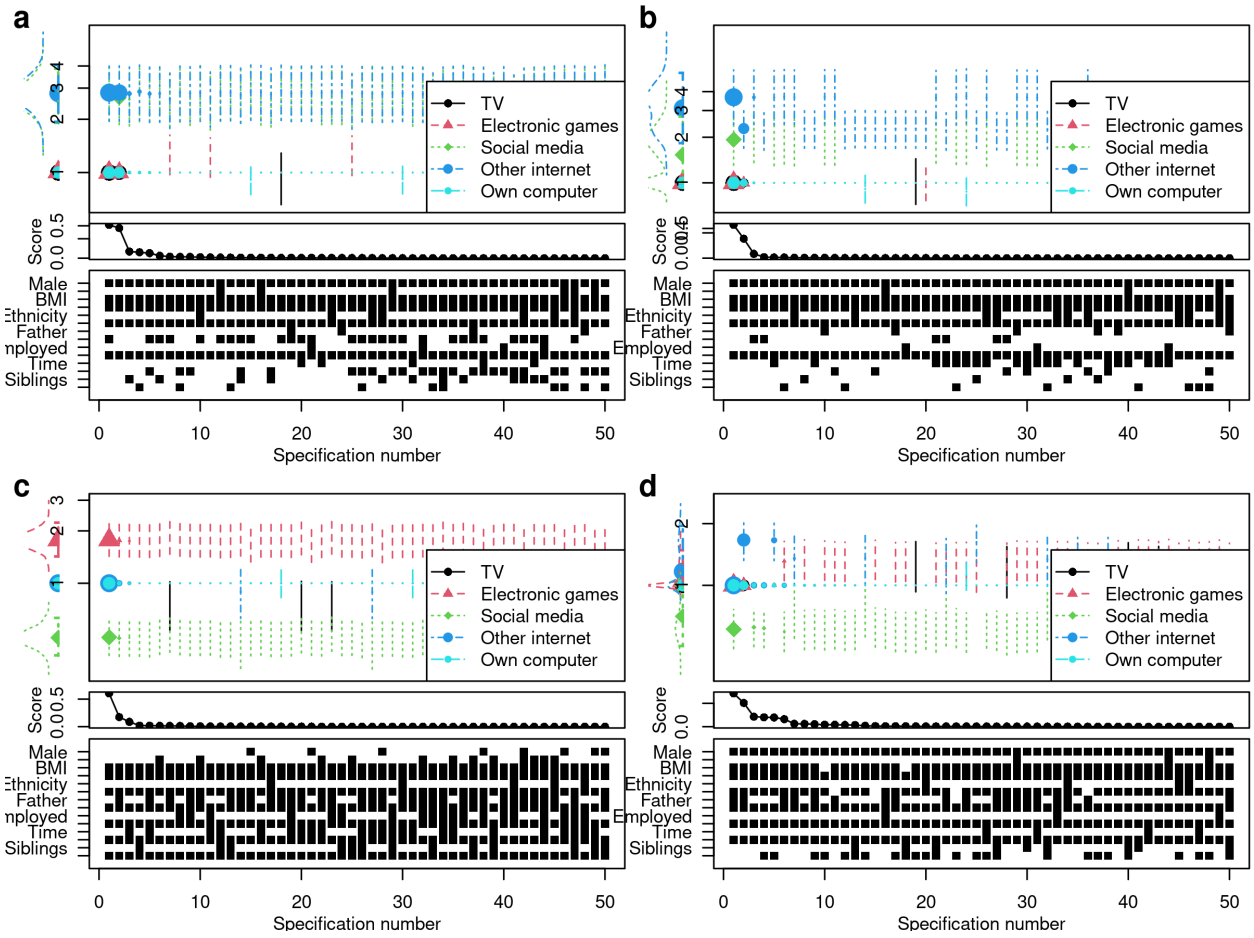


Figure S5: MCS data. BSCAs for outcomes: (a) adolescent-assessed depression, (b) adolescent-assessed low self-esteem, (c) parent-assessed high total difficulties and (d) parent-assessed high emotional problems. For variable codings and confounders, see Fig. 2 of the main text.

2.5 MCS: linear regression

Unlike in the YRBS dataset, which has binary outcomes, the outcomes in the MCS dataset are numerical. Parent assessments use the Strengths and Difficulties Questionnaire (SDQ), which provides a score for several well-being categories. Adolescent self-assessments come in three different forms: depressive symptoms according to the short version of the Mood

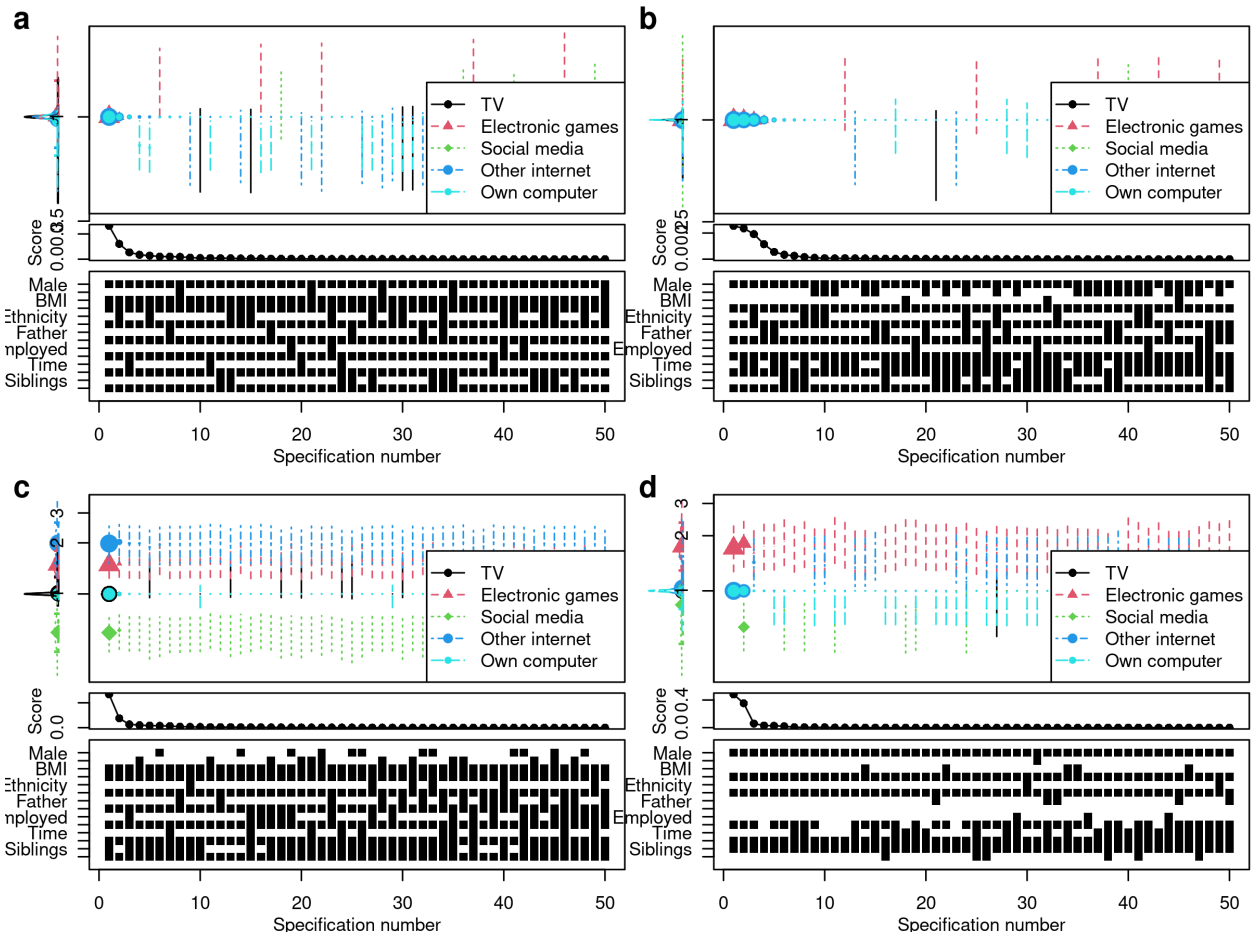


Figure S6: MCS data. BSCAs for parent-assessed outcomes (SDQ): (a) conduct problems (b) hyperactivity/inattention, (c) peer problems and (d) prosociality. For variable codings and confounders, see Fig. 2 of the main text.

and Feelings Questionnaire (SMFQ), self-esteem according to the Rosenberg scale, and a non-standardized "wellbeing grid" (with the prompt "On a scale of 1 to 7 where '1' means completely happy and '7' means not at all happy, how do you feel about the following parts of your life?" and various outcomes, e.g. 'Your life as a whole'). Individual questions use 3 item (SDQ and SMFQ), 4 item (Rosenberg) or 7 item (wellbeing grid) Likert scales and the aggregate scores (where available) also have different supports.

In our main analysis we converted the outcomes into a binary outcome where 1 codes for an abnormal outcome and 0 for a normal outcome. This conversion was done for two reasons. First, we could use logistic regression for both the YRBS and MCS data, making the results (i.e. odds ratios between low and high usage) comparable. Second, because the scales are mostly standardized, there exist recommended cutoffs derived from population data to define our binary bad (abnormal) / good (normal) outcome. The cutoffs (given in Fig. 2 of the main text) come from Goodman (1997) and Goodman, Meltzer, and Bailey (1998) for the SDQ, Rosenberg (1965) and Nguyen et al. (2019) for the Rosenberg scale, and Thabrew et al. (2018) for the SMFQ. The 'well-being grid' outcomes are not included in Fig. 2, since they do not come from standardized scales and no population-based abnormality cutoffs are available, see next section.

For robustness, we also obtained results using the linear regression model on the original numerical outcomes. The linear regression results for the outcomes discussed in the main paper are shown in Fig. S7. One important difference is that social media is not included in the results for the self-esteem outcome. This is likely due to the high co-linearity with other internet (Pearson's correlation = -0.63). Otherwise, our main results retain their estimated signs.

In addition, Figs. S8 and S9 show the linear results for the remaining parent-assessed and adolescent-assessed outcomes, respectively. Notably, unlike in the logistical model, in Fig. S8 other (non-social media) internet usage and owning a computer are negatively and positively associated with parent-assessed prosociality. As shown in Fig. S9, technology use has very different estimated effects on different types of self-assessed happiness. Social media, for example, is positively associated with being happy with ones friends, but negatively with being happy with ones looks. In contrast, none of the technology uses considered are associated with happiness regarding school, school work or family.

3 Further differences with Orben and Przybylski (2019)

Our analysis is based on the YRBS and MCS datasets also used by Orben and Przybylski (2019) and largely followed their treatment of the data, which was facilitated by their commendable sharing of the R code used for the analysis. However, we deviated from their data treatment choices (in the MCS data) when we felt these were potentially problematic. We also enlarged the set of possible control variables in both datasets. In this section, we explain these differences.

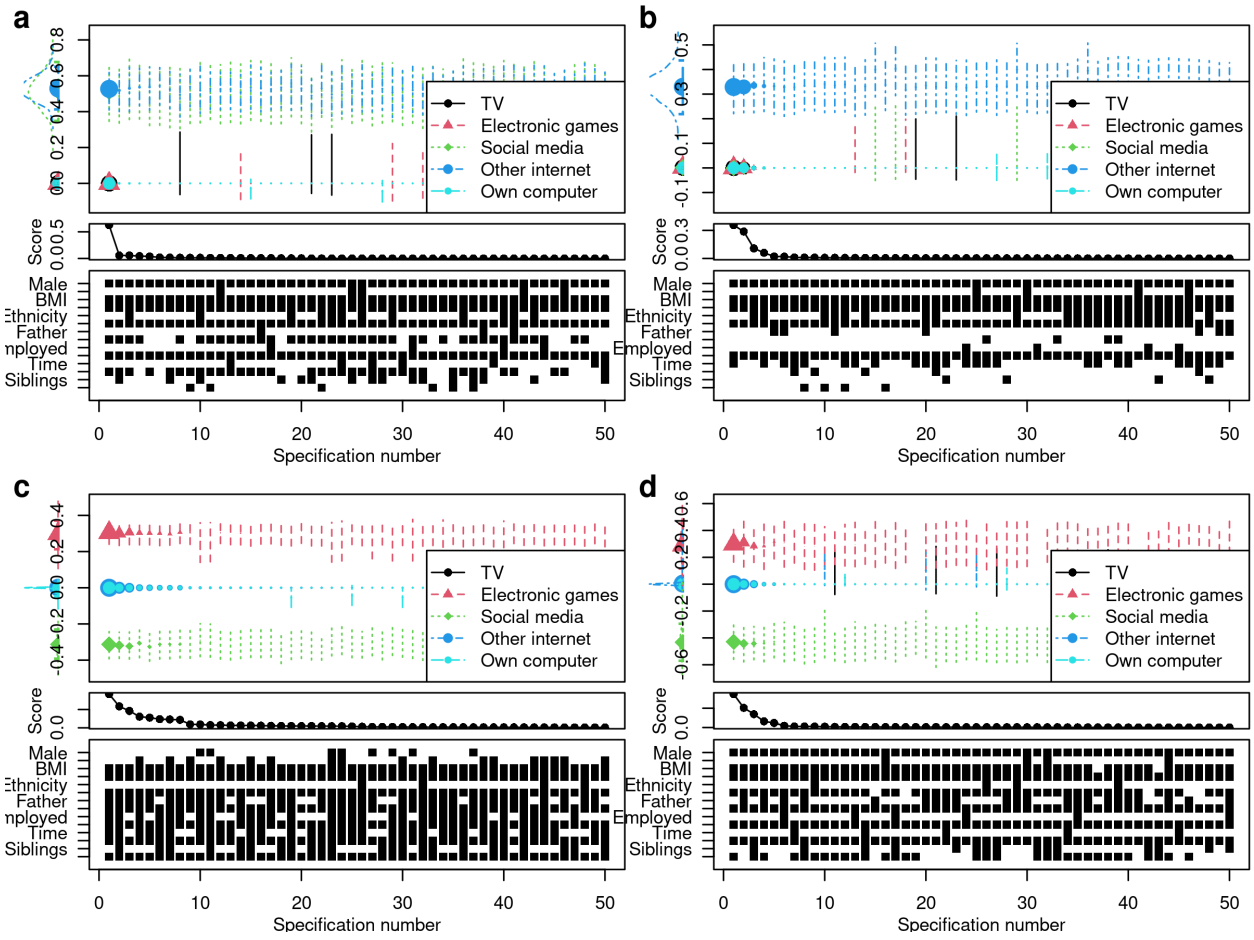


Figure S7: MCS data. Linear regression BSCA for main outcomes: (a) adolescent-assessed depression, (b) adolescent-assessed self-esteem (inverted), (c) parent-assessed total difficulties and (d) parent-assessed emotional problems.

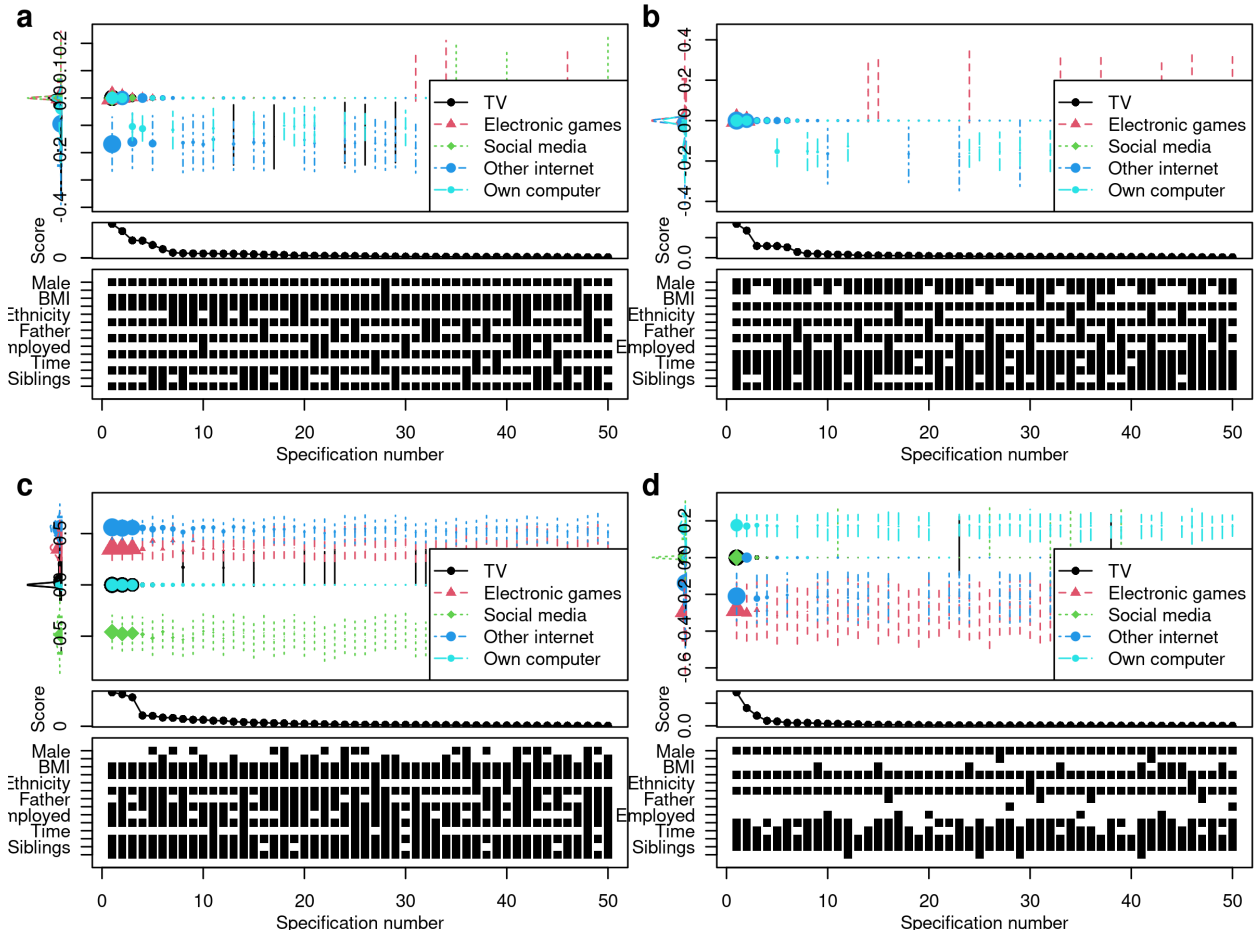


Figure S8: MCS data. Linear regression BSCA for parent-assessed outcomes (SDQ): (a) conduct problems (b) hyperactivity/inattention, (c) peer problems and (d) prosociality.

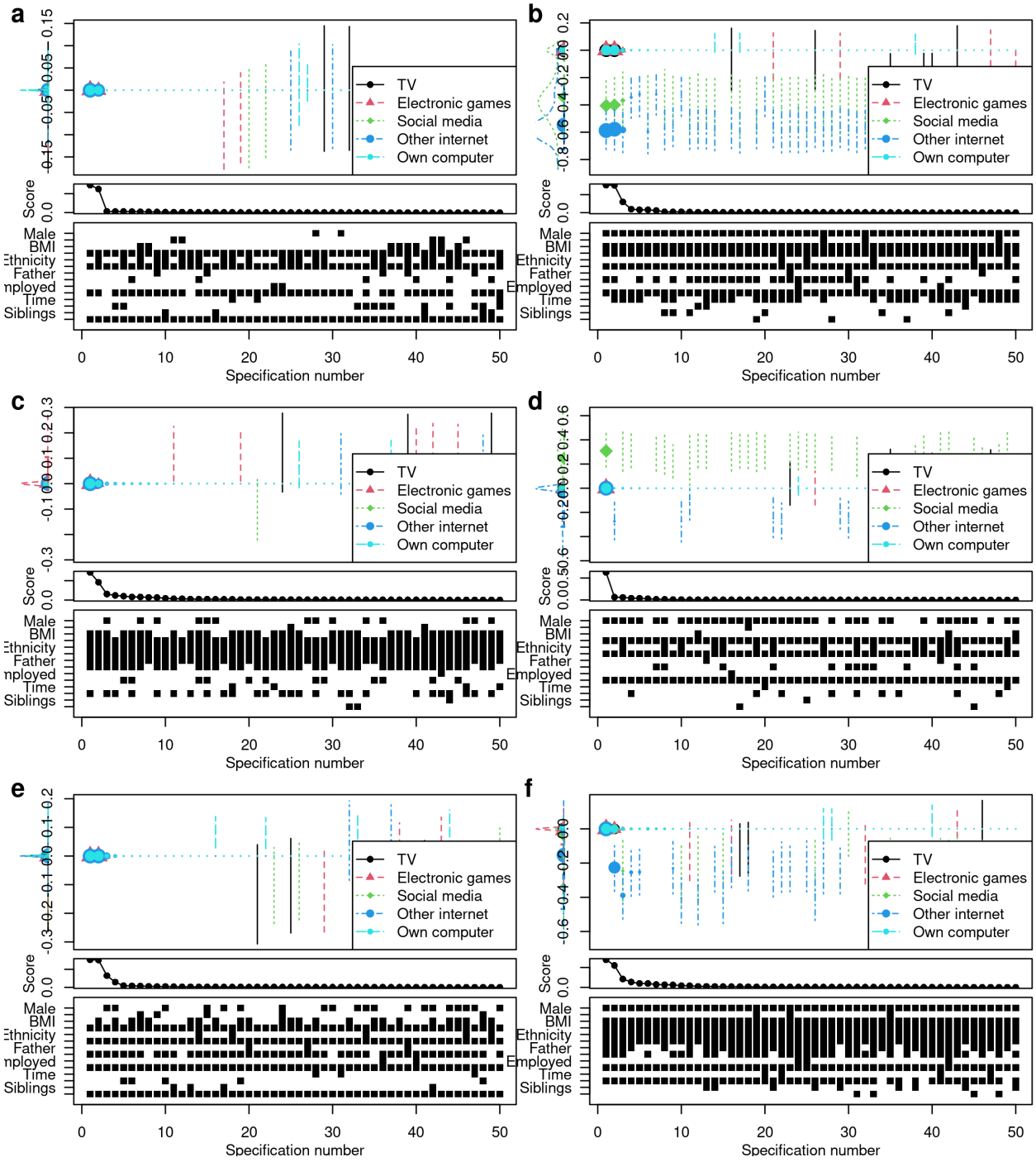


Figure S9: MCS data. Linear regression BSCA for adolescent-assessed outcomes (well being grid): happy with (a) school work (b) looks, (c) family, (d) friends, (e) school and (f) life as a whole.

Table S1: Summary statistics for parent-assessed MCS outcome variables used in Orben and Przybylski (2019)

	Obs.	Mean	S.D.	Min	25%	Median	75%	Max
fpsdro00	10265	0.53	0.68	-1	0	1	1	1
fpsdhs00	10214	0.54	0.65	-1	0	1	1	1
fpsdtt00	10242	0.42	0.71	-1	0	1	1	1
fpsdsp00	10234	0.48	0.68	-1	0	1	1	1
fpsdmw00	10242	0.50	0.66	-1	0	1	1	1
fpsdfs00	10198	0.67	0.60	-1	0	1	1	1
fpsdfb00	10255	0.92	0.32	-1	1	1	1	1
fpsdud00	10221	0.74	0.54	-1	1	1	1	1
fpsddc00	10228	0.32	0.72	-1	0	0	1	1
fpsdnc00	10240	0.52	0.66	-1	0	1	1	1
fpsdoa00	10240	0.80	0.46	-1	1	1	1	1
fpsdpb00	10220	0.71	0.57	-1	1	1	1	1
fpsdcs00	10249	0.94	0.29	-1	1	1	1	1
fpsdgb00	10249	0.46	0.67	-1	0	1	1	1
fpsdfe00	10271	0.68	0.57	-1	0	1	1	1
fconduct	11488	9.58	1.63	1	9	10	11	11
fhyper	11481	8.01	2.40	1	7	8	10	11
fprosoc	11489	8.31	1.85	0	7	9	10	10
fpeer	11491	9.26	1.82	1	8	10	11	11
femotion	11486	8.95	2.14	1	8	10	11	11
febdtot	11471	32.81	5.99	3	30	34	37	41

3.1 Re-scaling of variables

Orben and Przybylski (2019) transformed all outcome variables into a common 1-10 point scale prior to their analysis, so that it is easier to compare outcome values, estimated treatment effects, and combining them across outcomes. Unfortunately, this pre-processing step was not coherently applied to all MCS outcomes (see `1_3_prep_mcs.R` in their replication files), which led to several outcomes not being in the 1-10 scale (some actually have negative values, whereas one outcome takes on values as large as 41, see Table S1). As a consequence, the estimated treatment effects for these outcomes are not really comparable to the outcomes that were in the 1-10 scale, leading to difficulties in interpreting their SCA plot.

3.2 Questionary outcomes: individual questions versus validated scales

The analysis by Orben and Przybylski (2019) used as outcomes individual questions that make up common scales (e.g. all the values ending in 00 in Table S1). This is done for the

adolescent-assessed Mood and Feelings Questionnaire – short version (SMFQ), the adolescent-assessed Rosenberg scale and the parent-assessed Strengths and Difficulties Questionnaire (SDQ). While we think it is generally useful to look at different outcomes, we do not recommend breaking up questions that make up established scales, unless there is a strong reason for doing so. This is because the combined scores have well-established psychometric properties, such as internal consistency, test-retest reliability and validity (see Stone et al. 2010; Sinclair et al. 2010; Thabrew et al. 2018 for the SDQ, Rosenberg scale and SMFQ, respectively). Hence, our analyses focused on the combined scores, rather than on individual questions. This point was also addressed by Orben and Przybylski (2020) themselves.

3.3 Multiple treatment variables

As described in Section 1, in situations where there are multiple treatment variables it is statistically preferable to include them jointly in the model, to ameliorate the confounding between their estimated effects. The YRBS data has two treatments: TV and electronic devise use. The MCS data has five treatments: TV, electronic games, social media, other internet and own computer. In our analyses we always jointly included all treatments.

3.4 Control variables

In our analysis, we included more control variables than Orben and Przybylski (2019). For the YRBS analysis, the only control variable they sometimes included is race. We added age, sex, grade, year of the survey and body-mass index (BMI), several of which were statistically significant (see Fig. 1 and Fig. S1). For the MCS analysis, Orben and Przybylski (2019) included a larger set of control variables (see Fig. 2), to which we added sex, age and BMI, which again turned out to be important. We also departed from their analysis in how we treated two MCS control covariates: primary caretaker's employment and education status.

Regarding employment, Orben and Przybylski (2019) used the NS-SEC 5 category for the current job (1="Manager" through 5="Routine"), meaning that all kids with unemployed primary caretakers (for which the variable is coded `NA`) are excluded from the analysis. This is actually a non-negligible proportion of the dataset, see Table S2. We believe that such exclusion may introduce biases, by restricting the scope of the inference to kids with employed parents caretakers. Instead, in our analysis we included a binary variable to control for employment status (1=employed or self-employed; 0 otherwise), which allows including kids with unemployed parents into the analysis.

Regarding education status, the covariate is also coded on a 1-5 scale, but there are two special values (95= "Overseas qualification" and 96="None of these"), see Table S3. Orben and Przybylski (2019) included the 95-96 values in their linear regression analysis, which is inappropriate: one cannot interpret the estimated coefficients as capturing the association between the outcome and the caretaker's education status. Possible alternative analyses are to either exclude individuals with these codes (excluding a non-negligible proportion of individuals), or use a non-linear coding for the covariate's effect. For simplicity, and given that we already included numerous other control covariates, in our analysis we excluded the education variable.

Table S2: Primary caretaker employment status (1-5). Number of individuals with each value

1	2	3	4	5	NA
3276	1832	696	273	2012	3795

Table S3: Primary caretaker education status (1-5). Number of individuals with each value

1	2	3	4	5	95	96	NA
660	2598	1496	3784	1170	331	1104	741

References

Goodman, R, H Meltzer, and V Bailey. 1998. "The Strengths and Difficulties Questionnaire: A Pilot Study on the Validity of the Self-Report Version." *Adolescent Psychiatry* 7 (3): 6.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry* 38 (5): 581–86. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x.

Nguyen, Dat Tan, E. Pamela Wright, Christine Dedding, Tam Thi Pham, and Joske Bunders. 2019. "Low Self-Esteem and Its Association with Anxiety, Depression, and Suicidal Ideation in Vietnamese Secondary School Students: A Cross-Sectional Study." *Frontiers in Psychiatry* 10. https://doi.org/10.3389/fpsyt.2019.00698.

Orben, Amy, and Andrew K. Przybylski. 2019. "The Association Between Adolescent Well-Being and Digital Technology Use." *Nature Human Behaviour* 3 (2): 173–82. https://doi.org/10.1038/s41562-018-0506-1.

———. 2020. "Reply to: Underestimating Digital Media Harm." *Nature Human Behaviour* 4 (4): 349–51. https://doi.org/10.1038/s41562-020-0840-y.

Rosenberg, Morris. 1965. *Society and the Adolescent Self-Image*. Princeton University Press.

Sinclair, Samuel J., Mark A. Blais, David A. Gansler, Elisabeth Sandberg, Kimberly Bistis, and Alice LoCicero. 2010. "Psychometric Properties of the Rosenberg Self-Esteem Scale: Overall and Across Demographic Groups Living Within the United States." *Evaluation & the Health Professions* 33 (1): 56–80. https://doi.org/10.1177/0163278709356187.

Stone, Lisanne L., Roy Otten, Rutger C. M. E. Engels, Ad A. Vermulst, and Jan M. A. M. Janssens. 2010. "Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review." *Clinical Child and Family Psychology Review* 13 (3): 254–74. https://doi.org/10.1007/s10567-010-0071-2.

Thabrew, Hiran, Karolina Stasiak, Lynda-Maree Bavin, Chris Frampton, and Sally Merry. 2018. "Validation of the Mood and Feelings Questionnaire (MFQ) and Short Mood and Feelings Questionnaire (SMFQ) in New Zealand Help-Seeking Adolescents." *International Journal of Methods in Psychiatric Research* 27 (3): e1610. https://doi.org/10.1002/mpr.

1610.