THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Advancing our understanding of cognitive development and motor vehicle crash risk

Citation for published version:

Mirman, JH, Murray, AL, Mirman, D & Adams, S 2021, 'Advancing our understanding of cognitive development and motor vehicle crash risk: A multiverse representation analysis', *Cortex*, vol. 138, pp. 90-100. <https://doi.org/10.1016/j.cortex.2021.01.017>

Digital Object Identifier (DOI):

[10.1016/j.cortex.2021.01.017](https://doi.org/10.1016/j.cortex.2021.01.017)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Cortex

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Advancing our understanding of cognitive development and motor vehicle crash risk: A
multiverse representation analysis

Abstract

Neurobiological and cognitive maturational models are the dominant theoretical account of adolescents' risk-taking behavior. Both the protracted development of working memory (WM) through adolescence, as well as individual differences in WM capacity have been theorized to be related to risk-taking behavior, including reckless driving. In a cohort study of 84 adolescent drivers Walshe et al. (2019) found adolescents who crashed had an attenuated trajectory of WM growth compared to adolescent drivers who never reported being in a crash, but observed no difference in WM capacity at baseline. The objectives of this report were to attempt to replicate these associations and to evaluate their robustness using a hybrid multiverse – specification curve analysis approach, henceforth called multiverse representation analysis (MRA). The authors of the original report provided their data: 84 adolescent drivers with annual evaluations of WM and other risk factors from 2005-2013, and of driving experiences in 2015. The original analysis was implemented as described in the original report. An MRA approach was used to evaluate the robustness of the association between developmental trajectories of WM and adolescents' risk-taking (indexed by motor vehicle crash involvement) to different reasonable methodological choices. We enumerated 6 reasonable choice points in data processing-analysis configurations: (1) model type: latent growth or multi-level regression, (2) treatment of WM data; (3) which waves are included; (4) covariate treatment; (5) how time is coded; and (6) link function/estimation method: weighted least squares means and variance estimation (WLSMV) with a linear link vs logistic regression with maximum likelihood estimation. This multiverse consists of 96 latent growth models and 18 multi-level regression models.

Advancing our understanding of cognitive development and motor vehicle crash risk: A multiverse representation analysis

Maturation theories of adolescent risk-taking stipulate that a comparatively faster rate of development of reward processing areas of the brain (limbic system) relative to the executive control areas of the brain (pre-frontal cortex) is the neurobiological genesis for an increase in risk-taking during adolescence (Casey et al., 2008; Steinberg, 2008). Although leading maturational theories have important differences, they also rest on a few key, shared assumptions. The first claim is that structural and functional changes in the brain are related to changes in adolescent cognition and behavior. Specifically, structural maturation and functional responsivity of the prefrontal cortex are related to improved executive functioning (such as working memory and inhibitory control) and structural maturation and functional responsivity of the striatal and limbic systems are related to the increasing appeal and salience of hedonic rewards and difficulties with emotion regulation. The second claim is that the asynchronous development of these neurobiological systems results in a neurocognitive imbalance between the reward/emotion systems and the executive functioning system, producing a greater predilection for risk-taking behavior during the adolescent developmental period compared to other developmental periods when that imbalance is smaller. Following from this reasoning, it is posited that, within the adolescent developmental period, (1) individual differences in the size of that imbalance are related to greater risk-taking behavior and (2) individual differences in cognitive abilities, such as working memory (Romer, Reyna & Satterthwaite, 2017), may relate to engagement in risk behavior. Support for each of these predictions varies and these models have been both praised and critiqued in detail elsewhere (Mills et al., 2014; Pfeifer & Allen, 2016; Shulman et al., 2016).

Cognitive functions that rely on pre-frontal cortical regions include control, action planning, and memory (Miller & Cohen, 2001), which have demonstrated continued development through adolescence (Luciana et al., 2005). Thus, determining how engagement in risk-taking is predicted by individual differences in these cognitive faculties, as well as differences in how they change over time, is a fertile area of research (Luna, 2010; Luna et al., 2010). Deficits (or individual differences) in working memory have emerged as a candidate predisposing factor for engagement in risk behavior, or compromised task performance, due to the developing ability to hold behavioral or conceptual goals on-line (Kimberg & Farah, 1993) and to theoretical links with inhibitory control and impulsivity (Romer et al., 2009). Luna (2010) argues that working memory becomes more integrative, efficient, and flexible through adolescence, offering greater ability to: (a) perform complex tasks with more accuracy in the face of distraction (Luna, 2010) and (b) leverage prior experience toward behavioral and cognitive performance (Murty et al., 2016).

The extent to which these faculties can reliably predict engagement in different types of risk behaviors (e.g., unprotected sex, alcohol and other drug use, and driving) is unknown. This is important because risk behaviors are not monolithic and vary in (a) their recursive impacts on the developing brain, (b) the contribution of practical experience towards behavioral performance and consequently level of risk, (c) the influence of contextual factors, and d) if the behavior also has a positive or normative function (e.g., substance use vs. driving) (Mirman, 2018).

Notably, a recent cohort study of 84 adolescent drivers (Walshe et al., 2019) reported that crash-involved adolescents had an attenuated trajectory of working memory (WM) development

compared to drivers who never reported being in a crash (Figure 1, left panel). There was no observed effect of individual differences in WM capacity at baseline on future crash likelihood (Walshe et al., 2019). This is an odd finding because it means that, when both were included in the model, the *rate* of WM development - but not baseline WM (i.e., WM in early adolescence) - differentiated adolescent drivers who crashed from those that did not crash. WM capacity measured at the final assessment wave was reported to be inversely associated with crash risk in a simple correlational analysis; information about the association between WM and crash risk was not reported for prior WM assessment waves. This Walshe et al. (2019) study is of unique interest because of the robust measures of working memory, and longitudinal design, as well as the counterintuitive nature of the observed result – an effect of WM change but not baseline WM capacity.

Current Study

To better understand these results, especially in the context of the mixed literature and ongoing preoccupation with maturational models of risk-taking among the scientific and lay community, whether these models involve an “imbalance” or simply implicate deficient executive functioning capacities, we replicated the original analysis and further conducted a Multiverse Representation Analysis, which is a hybridization of a multiverse analysis (Steege et al., 2016) and specification curve analysis (Simonsohn et al., 2015).

A multiverse analysis acknowledges that a research hypothesis – e.g., rate of WM development is associated with being in a motor vehicle crash – does not fully specify the data *creation* process. Beyond study design choices, from the point of data acquisition, raw data undergo a variety of pre-processing steps that are agnostic to the main question of interest. These steps can include how to derive variables (e.g., discretizing, transforming, standardizing) and how to construct a longitudinal analytic data set (e.g., binning data, including or excluding observations). Importantly, these decisions about how to construct the analytic data set can lead to different statistical approaches and consequently the type of conclusions drawn from the data (Patel et al., 2015). A specification curve analysis (SCA) extends this idea such that a research hypothesis does not fully specify the data *analysis* process (Simonsohn et al., 2015). Its proponents recommend a three-step process: (1) identify reasonable choice points, (2) display results graphically allowing for the identification for inflection points in decision-making that lead to different results, and (3) statistical analysis of the full set of permutations identified in step 2.

In the current report, we hybridized these two similar methods to span the data creation to analysis pipeline and to allow for a more intuitive and accessible visualization and analysis approach in comparison to steps 2 and 3 as outlined in SCA. Clarifying the robustness of the Walshe et al. (2019) results is especially important because although maturational models are the dominant theoretical account of adolescents’ risk-taking behavior, excitement about these models rests on their claim to predict real-world risk-taking behavior. Both the protracted development of working memory through adolescence, as well as individual differences in working memory capacity have been theorized to be related to risk-taking behavior. However, tests of these claims have largely been restricted to cross-sectional studies and lab-based assessments, and analyses of these data have not been conducted utilizing open science principles, making it difficult to ascertain the basic and applied scientific value of maturational or imbalance models of adolescent risk-taking.

METHOD

Data

This verification report presents the results of a re-analysis and re-interpretation of the data presented in Walshe et al. (2019). The protocol was pre-registered and can be found here: <https://osf.io/r9fgm>.

Briefly, sampling from a community cohort study of teenagers, 84 drivers were identified based on their responses to self-report questions about driving. The analysis used assessments of working memory (WM) at 6 time points beginning at age 11 to 13 years with the final WM assessment occurring in 2013 and 2014, when participants were 18 to 20 years old. In 2015, prior crash involvement as a driver was determined through self-report. Participants were 19-20 years-old at the point of this assessment. Assessments were completed annually except for the final assessment (wave 6) that occurred two years after wave 5. The WM battery consisted of four assessments: the 2-back, Corsi Blocks, spatial working memory, and Digit Span. The 2-back was not administered at wave 1. Additional methodological details are available elsewhere (Khurana et al., 2013a).

We received the original de-identified dataset via an email from the authors subsequent to the execution of a data use agreement. Using these data, we conducted a replication of the authors' original analysis to determine if we obtained the same results and then conducted a multiverse analysis to determine the robustness of the result.

Exact replication: Computational verification

The replication analysis followed the description provided in the original report (Walshe et al., 2019) as closely as possible; the authors answered a few clarification questions through email concerning the analytic code they supplied with their data and variables in the dataset. The dataset for the computational verification was thus created by (1) omitting all of wave 1 data because the 2-back was not administered in wave 1, (2) excluding the digit span test, (3) standardizing each WM measure (Corsi, 2-back, and spatial working memory) within each wave, and (4) averaging them together to make a composite WM score. These data were then analyzed using a latent growth model with an intercept and linear slope factor. A binarized crash variable was regressed on these factors using weighted least squares means and variance estimation (WLSMV) thus providing an estimate of the effect of a latent continuous crash risk variable on WM initial status and growth. In the original report, the linear slope but not intercept factor showed a significant effect on crash risk.

Multiverse representative analysis (MRA)

We identified six key choice points across the data creation and data analysis pipeline where a reasonable researcher aiming to test the same hypothesis could have made a different choice. The choice points are summarized in Table 1 and described in more detailed below.

Table 1. Summary of Multiverse Representational Analysis

Choice point	Original analysis	Multiverse considered
Model	Latent growth model	<ul style="list-style-type: none"> Latent growth model Multi-level regression
Working memory composite scores	Within-wave standardization	<ul style="list-style-type: none"> Within-wave standardization Common-baseline standardization Longitudinal latent variable model

		<ul style="list-style-type: none"> • Factor analysis scores
Wave inclusion	2-6	<ul style="list-style-type: none"> • 2-6 • 1-6 • 1-5
Covariate inclusion	Yes	<ul style="list-style-type: none"> • Yes (+ sex) • No
Coding of time	First wave = 0	<ul style="list-style-type: none"> • First wave = 0 • Last wave = 0
Link function	Linear	<ul style="list-style-type: none"> • Linear • Logistic

Model

In the original analysis the authors used a latent growth model to model working memory trajectories and regressed crashes on the intercept and linear slope factors. We identified a multi-level regression (MLR) model with working memory scores as the outcome and group (crash or no crash), time, and time by crash group interaction as predictors as a reasonable alternative. This modeling approach also assesses crash group differences in WM scores, but inverts their position in the regression: in the MLR model, WM scores are the outcome variable and crash group is the predictor variable.

Working memory scores

In the original analysis, a composite score was derived by standardizing Corsi block, 2-back and spatial working memory scores within each wave and averaging the resultant z-scores (Figure 1, left panel). This method is likely to distort the between-individual differences in growth trends that are the subject of the original authors' hypotheses, but it corresponds to the analyses of the original authors.

We identified three reasonable alternatives: (1) deriving a composite score using a standardization of each task that utilized scores across all waves before averaging the resultant z-scores (we call this the "common baseline" approach; Figure 1, middle panel). (2) Using a one-step longitudinal latent variable model with the working memory tasks acting as indicators of a latent working memory factor (we call this the "latent" approach). This method is advantageous because it does not assume that each task is equally well related to working memory. Further, we imposed only the minimal constraints necessary for identification in order to allow for the possibility that the measurement of working memory does not show developmental invariance (Murray et al., 2019). (3) Using a two-step process where Confirmatory Factor Analysis (CFA) is used to derive factor scores in a first step (Figure 1, right panel), which are then used in the model fitting in a second step.

Inclusion of waves

In the original analysis the authors included data from wave 2-6 though wave 1 data were available. Their rationale was that the 2-back task was not administered at wave 1. We identified two reasonable alternatives: 1) Include data from all 6 waves, basing the wave 1 composite on the measures that were available and 2) Include data from waves 1-5 because the outcome - involvement in a crash as a driver - likely occurred *prior* to the wave 6 working memory assessment for most participants. We do not know exactly when the crashes occurred

in this sample as this was not assessed by the original authors, but newly licensed adolescent drivers are most likely to crash in the first 6 months of licensure, with (population-level) risk rapidly decreasing thereafter (Curry et al., 2015a). Twenty-five participants (30%) reported having been in at least 1 crash. Of these, only 1 started driving less than 6 months before the wave 6 WM evaluation; 21 (84%) were driving for over a year prior to the wave 6 WM evaluation and 18 (72%) were driving for over 2 years. Based on this timing of assessments, it is unlikely that a meaningful proportion of those 25 crashes occurred between the wave 6 WM assessment in 2013/14 and when they self-reported on their crash involvement as a driver at the follow-up assessment in 2015. Rather, it is more likely that wave 6 WM was assessed post-crash. Time of crash was not captured in the original study.

Inclusion of covariates

In the original analysis the authors included three covariates derived from a self-reported survey: driving errors, reckless driving, and risk behavior. As these covariates were only weakly and (mostly) non-significantly correlated with the outcome, we identified excluding the covariates as a reasonable alternative. We also added driver sex as a covariate based on prior literature that male adolescents differ from female adolescents with respect to crash risk. However, in this sample of 84 participants, 25 reported having been in a crash, of which 11 were male and 14 were female (of the non-crashers, 28 were male and 31 were female; $\chi^2(1)=0.003$, $p=0.96$).

Link function

In the original analysis, the authors used weighted least squares means and variance estimation (WLSMV) and regressed a binarized crash variable on working memory intercept and slope factors and covariates. In doing so they were estimating a linear regression of the observed binary crash variable on the predictors. As a reasonable alternative we identified a logistic regression of a binary crash variable on the predictors. Note that this applies only to the latent growth curve models because the multi-level models used a continuous outcome (working memory score) where a logistic regression would not have been appropriate.

Coding of time

In their original analysis, the authors set time=0 at the first wave, meaning that the intercept factor captured individual differences in WM capacity at baseline. As a reasonable alternative, we identified setting time=0 at the last wave in the LGCs because this was the most proximate wave to the crashes where individual differences in WM capacity should matter the most. When wave 6 was included in the analyses we thus set wave 6=0 and when wave 6 was omitted we set wave 5=0.

Interpretation

The computational reproduction directly informed our assessment of the internal integrity of the original data, code, and reporting pipeline. Second, the pattern of multiverse results was inspected for key decision points that lead to different types of results (e.g., does the choice of a latent growth curve model versus multi-level model lead to different results?), and thus lead to different inferences. We also determined how representative the original result was of the specified multiverse by calculating a multiverse representation quotient (MRQ) and by using new data visualization approaches we developed (e.g., a stoplight graph). This novel MRQ

indicator was calculated as the number of analytical configurations that yield the same qualitative result (i.e., a significant effect of WM slope) out of the total number of analytical configurations in the multiverse. So, if 5 out of 114 configurations matched the results reported in Walshe et al 2019 that would yield an MRQ of 4.39% for the original reported result. The stoplight graphing system we developed is a simple visual depiction of each choice point using a hierarchical taxonomic diagram with green circles representing replication, orange circles representing a marginal trend in the same direction, and red circles representing failed replication. Finally, the results were evaluated with respect to their alignment with prevailing maturational models of risk-taking. According to the extant literature on WM and risk-taking from this cohort study (see Romer et al., 2011) we would have expected an effect of WM capacity to be prospectively and concurrently predictive of engagement in crash risk among adolescent drivers such that better WM would be associated with less crash involvement; of note, this is not what was reported in Walshe et al 2019 but it would have been the effect most consistent with prior data and with theoretical frameworks that implicate deficient, or less mature, executive functioning and an increased predilection for risk behavior. Specifically, we would have expected to observe this pattern especially in the models with zero at wave 6 because this time-period was likely concurrent with crashes. Relatedly, to our knowledge, there is no prior evidence to suggest an independent effect of slope, or rate, of WM change on crash likelihood independent of the intercept term, as was reported by Walshe et al 2019.

Results

WM composite score trajectories for each group are shown in Figure 1. The common baseline standardized and CFA scores (middle and right panels, respectively) show an upward trajectory reflecting WM performance improvement in all three groups (non-drivers, drivers with a crash history, drivers without a crash history) across the 6 waves of data collection. The original report used within-wave standardized scores (left panel), which look approximately flat because these scores are computed separately relative to within-wave performance. The critical question is whether the crashed and never crashed groups reliably differ in terms of the WM trajectory (slope), as claimed by Walshe et al. (2019). Full results are provided in Tables S1 (latent growth curve models) and S2 (multi-level regressions) of Supplementary Materials and visualized in Figures 2 and 3. Factor loadings are provided in Table S3 -- the factor scores were dominated by spatial working memory, which is why their trajectories in Figure 1 are somewhat different from the composite scores that weight each WM task equally.

Computational verification

We first reviewed the code and data set provided by the authors and identified some discrepancies with the original report. For example, for the Corsi Blocks Tapping task the paper reports percent correct, but the data set has whole numbers that only go up to 15 except for wave 6, which has values very slightly less than whole numbers (e.g., 9.996, 7.980); these values cannot produce the central tendency statistics shown in Table 2 of the original report. Also, there were 18 observations with Corsi = 0, which is a possible value but would mean severe failure on the task. Similarly, in reviewing the model code it was clear it specified a different model from the one reported in the paper (e.g., in the code provided no covariance between the covariates was included and the slope factor loading for the wave 6 working memory composite was freely estimated). The authors however confirmed that both the code and the data were correct. Given this, we a) ran the model specified by the authors' code on the data provided to determine if we could produce the results reported in the paper and b)

implemented the methods reported in the paper (which is the official research record) and ran that on the data provided by the authors.

Running the code provided by the authors on the data provided by the authors did not replicate the key result reported in the manuscript. Instead, we found that the variation in linear slope of WM growth was not related to reporting a crash ($P = .183$). Variation in the WM intercept was not associated with crashes ($P = .624$). The model fit was acceptable but did not reach conventional cut-offs for good fit (CFI=.92, TLI=.90, RMSEA=.06, SRMR=.17)

Following the methods exactly as reported, the variation in linear slope of WM growth was not related with reporting a crash ($P = .425$). Variation in the WM intercept was not associated with crashes ($P = .198$) and model fit was poor (CFI=.71, TLI=.60, RMSEA=.13, SRMR=.24)

Multiverse representative analysis (MRA)

We specified a multiverse of 114 possible models consisting of 96 (84%) latent growth curve models and 18 (16%) multi-level regressions. Of the 96 latent growth curve models fit in our multiverse analysis, the effect of WM slope on crash risk was $p > .05$ in all models (Figure 2). One model produced a statistically marginal effect ($p = 0.066$; shown in orange in Figure 2), all others had clearly non-significant effects of WM trajectories ($p > 0.2$). Convergence failures were observed in 23 models. Estimation issues that rendered the test of the WM slope untrustworthy occurred in 7 models. Thus, this branch of the multiverse that contained 84% of the total number of models yielded a MRQ of 0%.

Of the 18 multi-level regressions fit in the multiverse analysis (Figure 3), the key effect was $p < .05$ in 6 out of 18 (33%) of the models; p -values ranged from $p = .031$ to $p = .049$ (shown in green in Figure 3). Two models produced statistically marginal effects (both $p = 0.062$, shown in orange in Figure 3) and the remaining 10 out of 18 models (56%) produced clearly non-significant effects of WM trajectory ($p > .4$). Convergence warnings were observed in 8 models. Thus, this branch of the multiverse that contained 16% of the total number of models yielded a MRQ of 33%.

Considering both branches together, a statistically significant effect of slope ($p < .05$) was found in 6 out of 114 (5.3%) of the models, and 3 additional models (2.6%) showed statistically marginal effects. In other words, 105 out of 114 (92%) of the models yielded clearly non-significant associations between crash risk and WM trajectories.

The pattern of multiverse results revealed the following consequential decision points. The first was statistical framework: latent growth curve model versus multi-level model which may reflect differences in the distributional properties of the outcome variable (continuous WM score for MLR, binary crash outcome for LGC), the models' assumptions about distributional properties, and the extent to which those assumptions were violated. The second critical decision point was the inclusion of wave 6 data: none of the multi-level regression models that excluded wave 6 data produced the originally reported result, thus identifying this decision point as critical to the association between WM trajectory and crash group observed in this branch of the multiverse. Finally, equal weighting of WM measures in a WM composite score (within-wave and common baseline) was more likely to produce a statistically significant effect. The CFA method, which is based on covariance between the WM measures, showed no difference between groups, possibly because those scores were dominated by the spatial WM test.

Post-hoc analysis: Longitudinal measurement invariance

To gain further insights into the source of differences across branches with different working memory scoring methods we tested longitudinal invariance (configural, metric, and scalar), using a longitudinal CFA approach (Murray et al., 2019). First, a configural model with the minimal identification constraints was fit, where one loading and one intercept was fixed equal over time and the mean and variance of the wave 1 working memory factor were fixed to 0 and 1 respectively. Residual covariances between the same items measured over time were also freely estimated. Configural invariance was judged to hold if this baseline model showed good fit based on $TLI > .95$, $CFI > .95$, $RMSEA < .05$ and $SRMR < .08$ (Hu & Bentler, 1999; Schermelleh-Engel et al., 2003). Metric invariance was tested by imposing loading equality constraints for each item over time. Metric invariance was judged to hold if the decrease in CFI was $< .050$, the increase in RMSEA $< .010$, and the increase in SRMR $< .025$ (Chen, 2007). If metric invariance held, scalar invariance was tested by imposing equality constraints for all intercepts over time. It was judged to hold if with the addition of scalar invariance constraints CFI decreased by no more than .05, RMSEA increased by $< .010$ and SRMR increased by $< .005$. If at the metric stage invariance did not hold, modification indices and expected parameter changes were used to guide the iterative release of constraints to aim to achieve partial metric invariance. In these cases, scalar constraints were not added for any items that did not show metric invariance.

The configural model showed acceptable fit ($CFI = .950$, $TLI = .921$, $RMSEA = .049$, $SRMR = .075$). The addition of metric invariance constraints resulted in a substantial deterioration in fit that exceeded recommended thresholds ($CFI = .894$, $TLI = .845$, $RMSEA = .068$, $SRMR = .112$). Partial metric invariance was, however, achieved according to CFI and RMSEA by removal of constraints on spatial working memory at wave 2, then at wave 1. Though the SRMR difference remained above the recommended threshold ($CFI = .918$, $TLI = .888$, $RMSEA = .060$, $SRMR = .108$), modification indices and expected parameter changes did not point to the removal of any further constraints, therefore, this model was used as the baseline for testing scalar invariance. The addition of scalar constraints to this model led to an improvement in fit ($CFI = .938$, $TLI = .915$, $RMSEA = .050$, $SRMR = .101$) and suggested that partial scalar invariance held.

These results suggest that there was some non-invariance in the spatial working memory test that may have contributed to the difference between using unweighted composites and weighted composites (factor scores and latent models). In particular, the spatial working memory threshold loadings at wave 1 and 2 were much larger than at later waves, suggesting that they were more strongly related to overall working memory performance at earlier ages. Further, the intercept of spatial working memory was higher at wave 2 than at other waves, which could affect estimates of working memory trajectories.

DISCUSSION

The computational verification was not successful, and the robustness evaluation indicated the originally reported results are not representative of the multiverse. The original effect was only observable under narrow data processing and analytic parameters. Therefore, we found no compelling evidence that the effect reported in Walshe et al 2019 - that crash risk is associated with an attenuated trajectory of WM, but not baseline WM - exists in either the sample or the population.

Alternative interpretations of the original result

On balance, there was very limited evidence to support any association between WM trajectory and crash involvement and it is difficult to generate implications that concern only an attenuated WM trajectory without implicating WM capacity. Monitoring, screening, or training WM as advised by the original authors will not be useful in predicting or preventing crash risk because WM capacity was never prospectively associated with crashing in any of the models. Indeed, reliable longitudinal WM measurement is itself a substantial challenge, as suggested by our measurement invariance analyses and other work (Taylor et al., 2020). Even if we assume the unlikely possibility that the association exists, because crash timing was not assessed we must logically allow the alternative interpretation that crashing led to reduced WM performance (via a crash-related injury) rather than slow WM development predisposed adolescents to crash. As described in the methods section, 21 (84%) of adolescents had over a year of driving experience prior to the wave 6 WM evaluation and 18 (72%) were driving for over 2 years prior to the wave 6 WM evaluation. Thus, it is likely that wave 6 WM was assessed post-crash for the majority of participants, rather than pre-crash. Concussions or mild TBI in young people have been shown to negatively affect WM performance (Keightley et al., 2014). Of note, we cannot determine from the original study data whether the crash caused an injury. Unfortunately, time of crash was not captured in the original study, nor was whether the young person was at-fault for the crash.

Implications for Research on Adolescent Risk Behavior

Other papers arising from this same cohort study have reported an inverse association between WM measured at an earlier wave with different types of health risk-behavior measured at a future wave that has primarily been mediated through impulsivity operationalized as “acting without thinking” and/or a behavioral delayed discounting task. This general pattern appears to have been reported for a composite risk behavior index (e.g., alcohol use, gambling, cigarettes, fighting) (Romer et al., 2009, 2011), substance use (Khurana et al., 2015b), and risky sex (Khurana et al., 2012, 2015). Interestingly (Khurana et al., 2013) found the opposite pattern as Walshe et al., 2019 with baseline WM, but not growth trajectories predicting alcohol use. It is hard to pinpoint the source of this variation theoretically because maturational models of risk-taking stipulate no differences by behavior type, and empirically one would expect consistency within the same dataset.

Across the papers arising from this dataset there is variation with respect to: waves used (also how waves are referenced changes across papers with different waves being called baseline or time one for example), which variables are included (e.g., other predictor variables and covariates), how variables were derived (e.g., which measures and items were included, arithmetic and statistical methods employed for derivation), whether the analysis focused on WM capacity vs. trajectory influences, and whether this association would be direct or mediated, and the overall analytical frameworks employed (SEM vs LGCM). Terminology also changes across papers (e.g., the same monetary reasoning task was called Temporal Discounting, Delay Discounting, and Inability to Delay Gratification). Variation like this could be due to a number of non-mutually exclusive reasons such as researchers’ knowledge and expertise with the data growing over the course of the study, vagaries of the peer review process including idiosyncratic reviewers across papers (Smith, 2006), the substantial amount of researcher degrees of freedom inherent in a cohort study with multiple measures of multiple constructs (Steege et al., 2016), as well as questionable research practices (Chambers, 2013; John et al., 2012; Klapwijk et al., 2021). Irrespective of the reason, as consumers of the literature it is difficult to be

confident about the strength and nature of the relationship between working memory (capacity or trajectory), potential mediators such as impulsivity, and health risk behaviors in this sample or in the population. Measurement and scientific reporting issues such as these (Dang et al., 2020; Flake & Fried, 2019) are not limited to this particular area but rather are broad-reaching in the psychological sciences.

Although it is not currently a component of maturational models of adolescent risk-taking, another source of variation is the type of risk behavior. Risk behaviors vary in the influence of contextual factors and biobehavioral underpinnings, so the contribution of cognitive factors may vary by behavior type. For example, automobile crashes are often cited as a prototypical example of adolescent risk-taking behavior. Importantly, the majority of adolescent drivers' crashes are caused by task inexperience (e.g., not slowing down entering curve, then over-correcting coming out it) (Curry et al., 2011; McCartt et al., 2009). Although younger age at licensure is associated with increased crash risk, population-level crash rates peak at licensure, then decrease rapidly (Curry et al., 2015b) – much too fast to be attributed to person-level brain-based maturation. A multi-level (individual- and population-level) computational cognitive modeling analysis involving two randomized controlled trials of effective crash prevention behavioral interventions for young drivers (Mirman et al., 2018; Thomas et al., 2016) and administrative crash databases from the United States, Australia, and Canada indicated strong evidence for skilled learning processes consistent with spontaneous discoveries of superior cognitive-behavioral strategies (e.g., how to manage scenarios with impeded sight lines) as underpinning the near universal patterns of changes in population-level crash rate reductions following licensure (Mirman et al., 2019).

Moreover, the kinds of crashes and behaviors that are considered to be overtly risky, such as speeding, reckless driving, and driving while impaired, are both uncommon and have a large gender disparity. Male adolescents have twice the motor vehicle fatality rate than female adolescents (Swedler et al., 2012). Compared with females, the higher death rate among male adolescents is due to driving less crashworthy vehicles, greater likelihood of not wearing a seatbelt, more frequent reckless driving/speeding (i.e., higher velocity crashes that confer greater risk for serious injury), and driving with multiple male adolescent passengers (McCartt & Teoh, 2015; Simons-Morton et al., 2011; Swedler et al., 2012; Williams et al., 2003).

In the real-world negative effects of peer passenger carriage are largely (but not exclusively) caused by a small subset of adolescent male drivers, driving with their male friends who share similar interests, risk-accepting normative beliefs, enjoyment of sensation-seeking activities, and concern fatal crash scenarios (not all crash scenarios) (Chen et al., 2000; Curry et al., 2012; Mirman et al., 2012; Mirman, 2018; Ouimet et al., 2015; Simons-Morton et al., 2011). Further, other studies demonstrate positive effects of adolescents being pro-social passengers (e.g., helping the driver manage their cell phone, wayfinding) (McDonald & Sommers, 2015). As such, the real-world data from field trials, on-road behavioral studies, computational modeling studies, and actuarial data point to important roles of gender and sex differences, context, and practical experience. These findings are illustrative of the importance of calls to re-tool models of risk-taking to better consider how risk-taking behaviors are constrained socio-culturally (Willoughby et al., 2014). Currently, maturational models of risk-taking make no claim about why these effects might exist.

Among young drivers, the extant laboratory-based research has found small and inconsistent associations on between individual differences in various aspects of executive functioning and driver behavior (Walshe et al., 2017). For example, some studies report that measures of executive functioning are positively associated with riskier driving (e.g., speeding) (Graefe & Schultheis, 2013), including working memory specifically (Starkey & Isler, 2016). Inverse

associations between executive functioning abilities and risky driving have also been observed (Mantyla et al., 2009), while other studies show positive associations between working memory and driver performance (e.g., lane changing) (Ross et al., 2014). Clearer, is the work with atypically developing youth. For example, atypically developing youth with deficits in executive functioning, such as those with attention deficit/ hyperactivity disorder (ADHD), are known to be at greater risk for crashing (Curry et al., 2017) and for aberrant driver behaviors compared to drivers without ADHD (Bishop et al., 2019; Fabiano et al., 2011). However, among the laboratory and experimental studies, methodological problems include heterogeneity in how executive functioning and driver behavior were assessed, reliance on cross-sectional designs that introduce temporal ambiguity, and small sample sizes that can produce unstable effect sizes.

Collectively, these issues have made it difficult to integrate the evidence base. Further, as shown by our analysis, these variations can also be simply due to researcher degrees of freedom and have nothing to do with the underlying constructs of interest. Cohort studies involving many self-report measures across measurement periods are particularly susceptible to having high researcher degrees of freedom across the data construction to analysis pipeline and should be subjected to intense scrutiny.

Implications

A more accurate and comprehensive evidence base is essential for the development of theories of young people's risk-taking and for realigning and refocusing prevention efforts on the modifiable psychological, behavioral, and contextual factors that truly have the most potential to advance public health for young people.

The importance of this sea change is critical. For example, motor vehicle crashes are the leading cause of death and disability for young people around the world (Global Status Report on Road Safety 2018). Graduated driver licensing provisions remain the single most effective population-level intervention available to reducing the risk for young drivers (Masten et al., 2011; Williams, 2017) but the identification of individual-level behavioral interventions that can be deployed at scale has severely lagged, with calls for methodological rigor and for innovation increasing (Mirman, 2019; Mirman et al., 2019; Peek-Asa et al., 2014; Scott-Parker & Senserrick, 2017).

It is difficult to construct evidence-based behavioral interventions if the evidence base on individual-level risk and protective factors is substantially weaker than it appears. Our results highlight the importance of evaluating published literature with extreme skepticism. This skepticism should not be restricted to interrogating counterintuitive results. Intuitive results require a similar level of interrogation and transparent evidentiary support.

Figure Titles and Captions

Figure 1. WM composite scores by group. The left panel shows within-wave standardized composite scores published in Walshe et al 2019; middle panel shows the same data with common baseline standardized scores; left panel shows the CFA-based composite scores. Points with vertical lines show the observed group means and SE in each wave. Lines are approximate linear trendlines to illustrate the trends.

Figure 2. Multiverse of LGC models. Decision points are indicated by grey rectangles and row labels on the left side. The filled circles represent the model outcomes, color-coded to indicate statistical significance of the WM slope term: $p > .10$ are red, $.10 > p > .05$ is orange, and gray indicates model convergence failure.

Figure 3. Multiverse of MLR models. Decision points are indicated by grey rectangles and the filled circles represent the model outcomes (top row is with covariates included, bottom row is without covariates). The model outcomes are color-coded to indicate statistical significance of the WM slope term: $p > .10$ are red, $.10 > p > .05$ is orange, $p < 0.05$ is green.

References

- Bishop, H. J., Curry, A. E., Stavrinou, D., & Mirman, J. H. (2019). Characterizing the Learning-to-Drive Period for Teens with Attention Deficits. *Journal of Developmental & Behavioral Pediatrics*, 40(8), 581–588. <https://doi.org/10.1097/DBP.0000000000000706>
- Casey, B. J., Getz, S., & Galvan, A. (2008). The adolescent brain. *Developmental Review*, 28(1), 62–77. <https://doi.org/10.1016/j.dr.2007.08.003>
- Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49(3), 609–610. <https://doi.org/10.1016/j.cortex.2012.12.016>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chen, L. H., Baker, S. P., Braver, E. R., & Li, G. (2000). Carrying passengers as a risk factor for crashes fatal to 16- and 17- year-old drivers. *Journal of the American Medical Association*, 283(12), 1578–1582. <https://doi.org/10.1001/jama.283.12.1578>
- Curry, A. E., Hafetz, J., Kallan, M. J., Winston, F. K., & Durbin, D. R. (2011). Prevalence of teen driver errors leading to serious motor vehicle crashes. *Accident Analysis and Prevention*, 43(4), 1285–1290. <https://doi.org/10.1016/j.aap.2010.10.019>
- Curry, A. E., Metzger, K., Pfeiffer, M. R., Elliott, M. R., Winston, F. K., & Power, T. J. (2017). Motor vehicle crash risk among adolescents and young adults with attention-deficit/hyperactivity disorder. *JAMA Pediatrics*, 171(8), 756–763. <https://doi.org/10.1001/jamapediatrics.2017.0910>
- Curry, A. E., Mirman, J. H., Kallan, M. J., Winston, F. K., & Durbin, D. R. (2012). Peer passengers: How do they affect teen crashes? *Journal of Adolescent Health*, 50(6). <https://doi.org/10.1016/j.jadohealth.2011.10.016>

Curry, A. E., Pfeiffer, M. R., Durbin, D. R., & Elliott, M. R. (2015a). Young driver crash rates by licensing age, driving experience, and license phase. *Accident Analysis & Prevention*, 80, 243–250.

<https://doi.org/10.1016/j.aap.2015.04.019>

Curry, A. E., Pfeiffer, M. R., Durbin, D. R., & Elliott, M. R. (2015b). Young driver crash rates by licensing age, driving experience, and license phase. *Accident Analysis & Prevention*, 80, 243–250.

<https://doi.org/10.1016/j.aap.2015.04.019>

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences*, 24(4), 267–269.

<https://doi.org/10.1016/j.tics.2020.01.007>

Fabiano, G. A., Hulme, K., Linke, S., Nelson-Tuttle, C., Pariseau, M., Gangloff, B., Lewis, K., Pelham, W. E., Waschbusch, D. A., Waxmonsky, J. G., Gormley, M., Gera, S., & Buck, M. (2011). The Supporting a Teen's Effective Entry to the Roadway (STEER) Program: Feasibility and Preliminary Support for a Psychosocial Intervention for Teenage Drivers With ADHD. *Cognitive and Behavioral Practice*, 18(2), 267–280. <https://doi.org/10.1016/j.cbpra.2010.04.002>

Flake, J. K., & Fried, E. I. (2019). *Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them*. PsyArXiv. <https://doi.org/10.31234/osf.io/hs7wm>

Global status report on road safety 2018. (n.d.). Retrieved December 1, 2020, from

<https://www.who.int/publications-detail-redirect/9789241565684>

Graefe, A. C. ;, & Schultheis, M. T. (2013). Examining neurocognitive correlates of risky driving behavior in young adults using a simulated driving environment. *International Conference on Virtual Rehabilitation (ICVR)*, 235–241.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:

Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5), 524–532.

<https://doi.org/10.1177/0956797611430953>

Keightley, M. L., Singh Saluja, R., Chen, J.-K., Gagnon, I., Leonard, G., Petrides, M., & Ptito, A. (2014). A Functional Magnetic Resonance Imaging Study of Working Memory in Youth after Sports-Related Concussion: Is It Still Working? *Journal of Neurotrauma*, 31(5), 437–451.

<https://doi.org/10.1089/neu.2013.3052>

Khurana, A., Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., & Hurt, H. (2012). Early adolescent sexual debut: The mediating role of working memory ability, sensation seeking, and impulsivity. *Developmental Psychology*, 48(5), 1416. <https://doi.org/10.1037/a0027491>

Khurana, A., Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., & Hurt, H. (2013a). Working memory ability predicts trajectories of early alcohol use in adolescents: The mediational role of impulsivity. *Addiction*, 108(3), 506–515. <https://doi.org/10.1111/add.12001>

Khurana, A., Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., & Hurt, H. (2013b). Working memory ability predicts trajectories of early alcohol use in adolescents: The mediational role of impulsivity. *Addiction*, 108(3), 506–515. <https://doi.org/10.1111/add.12001>

Khurana, A., Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., & Hurt, H. (2015a). Stronger Working Memory Reduces Sexual Risk Taking in Adolescents, Even After Controlling for Parental Influences. *Child Development*, 86(4), 1125–1141. <https://doi.org/10.1111/cdev.12383>

Khurana, A., Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., & Hurt, H. (2015b). Experimentation vs. Progression in Adolescent Drug Use: A Test of an Emerging Neurobehavioral Imbalance Model. *Development and Psychopathology*, 27(3), 901–913. <https://doi.org/10.1017/S0954579414000765>

- Kimberg, D. Y., & Farah, M. J. (1993). A Unified Account of Cognitive Impairments Following Frontal Lobe Damage: The Role of Working Memory in Complex, Organized Behavior. *Journal of Experimental Psychology: General*, 122(4), 411–428. <https://doi.org/10.1037/0096-3445.122.4.411>
- Klapwijk, E. T., van den Bos, W., Tamnes, C. K., Raschle, N. M., & Mills, K. L. (2021). Opportunities for increased reproducibility and replicability of developmental neuroimaging. *Developmental Cognitive Neuroscience*, 47, 100902. <https://doi.org/10.1016/j.dcn.2020.100902>
- Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development*, 76(3), 697–712. <https://doi.org/10.1111/j.1467-8624.2005.00872.x>
- Luna, B. (2010). DEVELOPMENTAL CHANGES IN COGNITIVE CONTROL THROUGH ADOLESCENCE. *Adv Child Dev Behav*, 37, 233–278. <https://doi.org/10.1038/jid.2014.371>
- Luna, B., Padmanabhan, A., & O’Hearn, K. (2010). What has fMRI told us about the Development of Cognitive Control through Adolescence? *Brain and Cognition*, 72(1), 101–113. <https://doi.org/10.1016/j.bandc.2009.08.005>
- Mantyla, T., Karlsson, M. J., & Marklund, M. (2009). Executive control functions in simulated driving. *Applied Neuropsychology*, 16, 11–18.
- Masten, S. V., Foss, R. D., & Marshall, S. W. (2011). Graduated Driver Licensing and Fatal Crashes Involving 16- to 19-Year-Old Drivers. *JAMA*, 306(10), 1098. <https://doi.org/10.1001/jama.2011.1277>
- McCartt, A. T., Mayhew, D. R., Braitman, K. A., Ferguson, S. A., & Simpson, H. M. (2009). Effects of age and experience on young driver crashes: Review of recent literature. *Traffic Injury Prevention*, 10(3), 209–219. <https://doi.org/10.1080/15389580802677807>

- McCartt, A. T., & Teoh, E. R. (2015). Type, size and age of vehicles driven by teenage drivers killed in crashes during 2008-2012. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention*, 21(2), 133–136. <https://doi.org/10.1136/injuryprev-2014-041401>
- McDonald, C. C., & Sommers, M. S. (2015). Teen Drivers' Perceptions of Inattention and Cell Phone Use While Driving. *Traffic Injury Prevention*, 16 Suppl 2, S52-8. <https://doi.org/10.1080/15389588.2015.1062886>
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Mills, K. L., Goddings, A. L., Clasen, L. S., Giedd, J. N., & Blakemore, S. J. (2014). The developmental mismatch in structural brain maturation during adolescence. *Developmental Neuroscience*, 36(3–4), 147–160. <https://doi.org/10.1159/000362328>
- Mirman, Jessica Hafetz. (2019). A dynamical systems perspective on driver behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, 63, 193–203. <https://doi.org/10.1016/j.trf.2019.04.010>
- Mirman, J.H. (2018). Agency, Adolescence, and Motor Vehicle Crash Risk. *Journal of Adolescent Health*, 62(5), 509–510. <https://doi.org/10.1016/j.jadohealth.2018.02.009>
- Mirman, J.H., Albert, D., Jacobsohn, L. J., & Winston, F. K. (2012). Factors associated with adolescents' propensity to drive with multiple passengers and to engage in risky driving behaviors. *The Journal of Adolescent Health*, 50(6), 634–640. <https://doi.org/10.1016/j.jadohealth.2011.10.256>
- Mirman, J.H., Curry, A. E., Elliott, M. R., Long, L., & Pfeiffer, M. R. (2018). Can Adolescent Drivers' Motor Vehicle Crash Risk Be Reduced by Pre-Licensure Intervention? *Journal of Adolescent Health*, 62(3), 341–348. <https://doi.org/10.1016/j.jadohealth.2017.09.015>

Mirman, J.H., Curry, A. E., & Mirman, D. (2019). Learning to drive: A reconceptualization. *Transportation Research Part F: Traffic Psychology and Behaviour*, 62, 316–326.

<https://doi.org/10.1016/j.trf.2019.01.010>

Murray, A. L., Obsuth, I., Eisner, M., & Ribeaud, D. (2019). Evaluating Longitudinal Invariance in Dimensions of Mental Health Across Adolescence: An Analysis of the Social Behavior Questionnaire. *Assessment*, 26(7), 1234–1245. <https://doi.org/10.1177/1073191117721741>

Murty, V. P., Calabro, F., & Luna, B. (2016). The role of experience in adolescent cognitive development: Integration of executive, memory, and mesolimbic systems. *Neuroscience and Biobehavioral Reviews*, 70, 46–58. <https://doi.org/10.1016/j.neubiorev.2016.07.034>

Ouimet, M. C., Pradhan, A. K., Brooks-Russell, A., Ehsani, J. P., Berbiche, D., & Simons-Morton, B. G. (2015). Young drivers and their passengers: A systematic review of epidemiological studies on crash risk. *Journal of Adolescent Health*, 57(1), s24-S35.e6.

<https://doi.org/10.1016/j.jadohealth.2015.03.010>

Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, 68(9), 1046–1058. <https://doi.org/10.1016/j.jclinepi.2015.05.029>

Peek-Asa, C., McGehee, D. V., & Ebel, B. E. (2014). Increasing safe teenaged driving: Time to integrate the growing evidence base. *JAMA Pediatrics*, 168(8), 703–704.

<https://doi.org/10.1001/jamapediatrics.2014.582>

Pfeifer, J. H., & Allen, N. B. (2016). The audacity of specificity: Moving adolescent developmental neuroscience towards more powerful scientific paradigms and translatable models. *Developmental Cognitive Neuroscience*, 17, 131–137. <https://doi.org/10.1016/j.dcn.2015.12.012>

Romer, D., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Farah, M., & Hurt, H. (2009). Executive cognitive functions and impulsivity as correlates of risk taking and problem behavior in

preadolescents. *Neuropsychologia*, 47(13), 2916–2926.

<https://doi.org/10.1016/j.neuropsychologia.2009.06.019>

Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., Yang, W., & Hurt, H. (2011). Does adolescent risk taking imply weak executive function? A prospective study of relations between working memory performance, impulsivity, and risk taking in early adolescence. *Developmental Science*, 14(5), 1119–1133. <https://doi.org/10.1111/j.1467-7687.2011.01061.x>

Ross, V., Jongen, E. M. M., Wang, W., Brijs, T., Brijs, K., Ruiter, R. A. C., & Wets, G. (2014). Investigating the influence of working memory capacity when driving behavior is combined with cognitive load: An LCT study of young novice drivers. *Accident Analysis and Prevention*, 62, 377–387. <https://doi.org/10.1016/j.aap.2013.06.032>

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, 8(2), 23–74.

Scott-Parker, B., & Senserrick, T. (2017). A call to improve sampling methodology and reporting in young novice driver research. *Injury Prevention*, 23(1), 8–9. <https://doi.org/10.1136/injuryprev-2016-042105>

Shulman, E. P., Smith, A. R., Silva, K., Icenogle, G., Duell, N., Chein, J., & Steinberg, L. (2016). The dual systems model: Review, reappraisal, and reaffirmation. *Developmental Cognitive Neuroscience*, 17, 103–117. <https://doi.org/10.1016/j.dcn.2015.12.010>

Simons-Morton, B. G., Ouimet, M. C., Zhang, Z., Klauer, S. G., Lee, S. E., Wang, J., Chen, R., Albert, P., & Dingus, T. A. (2011). The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers. *Journal of Adolescent Health*, 49(6), 587–593. <https://doi.org/10.1016/j.jadohealth.2011.02.009>

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. *SSRN Electronic Journal*.

<https://doi.org/10.2139/ssrn.2694998>

Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. *Journal of the Royal Society of Medicine*, 99(4), 178–182. <https://doi.org/10.1177/014107680609900414>

Starkey, N. J., & Isler, R. B. (2016). The role of executive function, personality and attitudes to risks in explaining self-reported driving behaviour in adolescent and adult male drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 38, 127–136.

<https://doi.org/10.1016/j.trf.2016.01.013>

Steege, 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>

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, 28(1), 78–106. <https://doi.org/10.1016/j.dr.2007.08.002>

Swedler, D. I., Bowman, S. M., & Baker, S. P. (2012). Gender and age differences among teen drivers in fatal crashes. *Annals of Advances in Automotive Medicine*, 56, 97–106.

Taylor, B. K., Frenzel, M. R., Eastman, J. A., Wiesman, A. I., Wang, Y.-P., Calhoun, V. D., Stephen, J. M., & Wilson, T. W. (2020). Reliability of the NIH toolbox cognitive battery in children and adolescents: A 3-year longitudinal examination. *Psychological Medicine*, 1–10.

<https://doi.org/10.1017/S0033291720003487>

Thomas, F. D., Rilea, S. L., Blomberg, R. D., C., Peck, R., & Korbela, K. T. (2016). Evaluation of the safety benefits of the risk awareness and perception training program for novice teen drivers. In *Report No. DOT HS 812 235* (Issue January).

Walshe, E. A., Ward McIntosh, C., Romer, D., & Winston, F. (2017). Executive Function Capacities, Negative Driving Behavior and Crashes in Young Drivers. *International Journal of Environmental Research and Public Health*, 14(11), 1314. <https://doi.org/10.3390/ijerph14111314>

Walshe, E. A., Winston, F. K., Betancourt, L. M., Khurana, A., Arena, K., & Romer, D. (2019). Working Memory Development and Motor Vehicle Crashes in Young Drivers. *JAMA Network Open*, 2(9), e1911421. <https://doi.org/10.1001/jamanetworkopen.2019.11421>

Williams, A.F., McCartt, A. T., & Geary, L. (2003). Seatbelt use by high school students. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention*, 9(1), 25–28.

Williams, Allan F. (2017). Graduated driver licensing (GDL) in the United States in 2016: A literature review and commentary. *Journal of Safety Research*, 63, 29–41.
<https://doi.org/10.1016/j.jsr.2017.08.010>

Willoughby, T., Good, M., Adachi, P. J. C., Hamza, C., & Tavernier, R. (2014). Examining the link between adolescent brain development and risk taking from a social–developmental perspective (reprinted). *Brain and Cognition*, 89, 70–78. <https://doi.org/10.1016/j.bandc.2014.07.006>