
Is (poly-) substance use associated with impaired inhibitory control? A mega-analysis controlling for confounders.

http://researchonline.ljmu.ac.uk/id/eprint/11125/

Citation (please note it is advisable to refer to the publisher’s version if you intend to cite from this work)

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/
Is (poly-) substance use associated with impaired inhibitory control?

A mega-analysis controlling for confounders

8Laboratoire de Psychologie Médicale et d'Addictologie, ULB Neuroscience Institute (UNI), CHU Brugmann-Université Libre de Bruxelles (U.L.B.), Brussels, Belgium;

9University of Cyprus, Nicosia, Cyprus;

10The Mind Research Network and Lovelace Biomedical and Environmental Research Institute, Albuquerque, New Mexico;

11Leiden University, Cognitive Psychology Unit & Leiden Institute for Brain and Cognition, Leiden, the Netherlands;

12The Mind Research Network, The University of Texas at Dallas, Texas, USA;

13University of Rochester Medical Center, School of Medicine and Dentistry, Rochester, USA;

14Department of Psychiatry, University of Vermont, Burlington, USA;

15Centre for Addiction and Mental Health, Campbell Family Mental Health Research Institute and Institute for Mental Health Policy Research, Toronto, Canada;

16School of Psychological Sciences, University of Melbourne, Melbourne, Australia;

17Department of Psychiatry, University of Michigan, Michigan, USA;

18Institute of Cognitive Science, University of Colorado Boulder, Colorado, USA;

19Work Group Addictive Behaviours, Risk Analyses and Risk Management, Faculty of Psychologie, Technische Universität Dresden, Germany;

20Department of Psychology, University of Liège, Belgium;

21Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands;

22University of Bamberg, Department of Clinical Psychology and Psychotherapy, Bamberg, Germany;

23Department of Psychiatry and Biobehavioral Sciences at the University of California, Los Angeles, USA;

24Psychological Neuroscience Lab, Research Center in Psychology (CIPsi), School of Psychology, University of Minho, Braga, Portugal;

25Turning Point, Eastern Health and Eastern Health Clinical School, Monash University, Melbourne, Australia;

26Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands;

27Northern Kentucky University, Highland Heights, USA;

28Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, USA;

29School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK;

30Maastricht University, Faculty of Psychology and Neuroscience, The Netherlands;

31Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea;

32Department of Clinical Medical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea;

33Department of Psychology, Charles Schmidt College of Science, Florida Atlantic University, USA;

34Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston SC, USA;

35Experimental and Clinical Pharmacopsychology, Department of Psychiatry, Psychotherapy, and Psychosomatics, Psychiatric Hospital, University of Zürich, Zürich, Switzerland;
University of California Los Angeles, Department of Psychology, Los Angeles, CA, USA;
School of Psychiatry, University of New South Wales, Sydney, Australia;
Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands;
Department of Psychiatry, Psychotherapy and Psychosomatic, Medical University Innsbruck, Austria;
Duke-UNC Brain Imaging and Analysis Center, Duke University, Durham, NC, USA;
VA Mid-Atlantic Mental Illness Research, Education and Clinical Center (MIRECC), Durham, NC, USA;
Child Neuropsychology Research Group, Murdoch Children's Research Institute, Melbourne Australia;
Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia;
Department of Psychiatry, University of California, San Diego, USA;
Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam Institute for Addiction Research, Amsterdam, The Netherlands;
School of Psychological Sciences, Monash Institute of Cognitive and Clinical Neurosciences (MICCN), Monash University, Australia;
Department of Psychiatry, Medical University of Warsaw, Warsaw, Poland;
State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China;
School of Psychological Sciences, Turner Institute for Brain and Mental Health, and Monash Biomedical Imaging Facility, Monash University, Melbourne, Victoria, Australia;
Molecular Brain Science Research Section Centre for Addiction and Mental Health, Toronto, Canada;
Research Priority Area Yield, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands;

*Corresponding author
Department of Psychology, University of Amsterdam, Amsterdam, Nieuwe Achtergracht 129B, 1018 WS Amsterdam, The Netherlands.
Email address: y.liu3@uva.nl

**Shared senior authorship

Accepted for publication in Neuroscience & Biobehavioral Reviews, July 2019
Highlights:

• The association between polysubstance use and inhibition is as-yet-unknown
• This association was tested with a mega-analysis using individual participant data
• Only lifetime cannabis use was associated with suboptimal inhibition (stop-task)
• Lifetime cannabis use moderated tobacco’s effect on response inhibition
• In cannabis non-users only, tobacco use was associated with suboptimal inhibition
Abstract

Many studies have reported that heavy substance use is associated with impaired response inhibition. Studies typically focused on associations with a single substance, while polysubstance use is common. Further, most studies compared heavy users with light/non-users, though substance use occurs along a continuum. The current mega-analysis accounted for these issues by aggregating individual data from 43 studies (3610 adult participants) that used the Go/No-Go (GNG) or Stop-signal task (SST) to assess inhibition among mostly “recreational” substance users (i.e., the rate of substance use disorders was low). Main and interaction effects of substance use, demographics, and task-characteristics were entered in a linear mixed model. Contrary to many studies and reviews in the field, we found that only lifetime cannabis use was associated with impaired response inhibition in the SST. An interaction effect was also observed: the relationship between tobacco use and response inhibition (in the SST) differed between cannabis users and non-users, with a negative association between tobacco use and inhibition in the cannabis non-users. In addition, participants’ age, education level, and some task characteristics influenced inhibition outcomes. Overall, we found limited support for impaired inhibition among substance users when controlling for demographics and task-characteristics.

Keywords:
Polysubstance use; Response inhibition; Stop-signal task; Go/No-Go task; Mega-analysis.
Introduction

1.1. Substance Use and Response Inhibition

1.1.1. What is response inhibition and how does it relate to substance use?

Inhibitory control, also known as response inhibition, has been defined as the ability to control one’s attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what is more appropriate or needed (Diamond, 2013). Loss of control over one’s behavior is a defining characteristic of addiction. The DSM-5 lists characteristics such as ‘taking larger amounts or over a longer period than was intended’ and ‘unsuccessful efforts to cut down or control alcohol use’ to define the loss of control over drinking (American Psychiatric Association, 2013). Moreover, inhibitory control has been proposed to play an important role at different stages of the addiction cycle, i.e., 1) initial use of substance; 2) transition from recreational use to heavier use and abuse; 3) continuation of use for those who get addicted; 4) relapse after abstinence (e.g., Garavan, Potter, Brennan, & Foxe, 2015; Koob & Volkow, 2010). Furthermore, the dual process model on addiction proposes that an imbalance between a hyper-sensitized impulsive system, which is responsible for cue-reactivity, and a compromised reflective or control system (including inhibition of impulses) are important in the development of addiction (Bechara, 2005; Gladwin, Figner, Crone, & Wiers, 2011; Volkow, Fowler, Wang, & Swanson, 2004; Volkow, Koob, Mental, Parity, & Act, 2015).

Over the past two decades, multiple studies have focused on the relationship between chronic substance use and response inhibition, but findings have been equivocal. Inhibitory impairment has been associated with chronic use of some substances (e.g., cocaine, ecstasy, methamphetamine, tobacco, and alcohol) but not for others (e.g., opioids, cannabis, see for a meta-analysis, Smith, Mattick, Jamadar, & Iredale, 2014). Results also vary in studies of
single substances. For instance, heavy drinkers have been reported to make more commission errors than light drinkers on the Go/No-Go task (GNG, Kreusch, Quertemont, Vilenne, & Hansenne, 2014), while alcohol-dependent and control participants did not differ significantly on the same measure (Kamarajan et al., 2005). Two main issues might explain these conflicting findings, namely the phenomenon of polysubstance use and the use of extreme group designs (i.e., comparing control participants and problematic or disordered substance users). In addition, sample demographics and task characteristics are often not taken into consideration. In order to address these issues in this mega-analysis, we aimed to investigate the relationship between inhibition and use of multiple substances by analyzing individual-level data, while taking demographics and task characteristics into account. In doing so, we did not exclusively focus on populations diagnosed with substance use disorders (SUD, American Psychiatric Association, 2013).

1.1.2. Experimental paradigms: the Go/No-Go task and the Stop-signal task

Successful suppression of motor responses can involve distinct behavioral processes such as “action restraint” or “action cancellation” (Schachar et al., 2007). Action restraint refers to stopping a prepared but not yet initiated response, which is commonly measured using the GNG and its variants, such as Conners’ continuous performance task (Conners & Sitarenios, 2011; Donders, 1868/1969). These tasks focus on the ability to withhold responding if a no-go stimulus is presented. The main variables of interest are the rate of commission errors (i.e., failures to inhibit a response to no-go targets or false alarms), the rate of omission errors (i.e. failures to respond to go targets, or misses), and the response time (RT) to go stimuli. A relatively high rate of commission errors and a short go RT reflects suboptimal inhibition (Smith et al., 2014).
By contrast, action cancellation refers to stopping a response that is already underway. It is typically measured using the Stop-signal task (SST, Logan, 1994). In this paradigm, each trial starts with the presentation of a go signal that requires an overt response such as a button press. On a subset of trials (typically around 25%), the go signal is followed by a stop signal after a certain interval (stop-signal delay, SSD), upon which participants should inhibit their already initiated go response. Usually, an adaptive tracking algorithm controls the SSD, such that there is a 50% probability of inhibiting the response. A horse-race model, assuming an independent race between the ‘go’ and ‘stop’ processes, affords the estimation of the stop-signal reaction time (SSRT, Logan, 1994). Given that the response could not be withheld on n percent of all stop trials (usually around at 50%), SSRT is calculated by subtracting the mean SSD from the go RT that marks the nth percentile in the go RT distribution.

In contrast to the GNG, the latency of the go response and the latency of the stop process are considered to be independent (Logan & Cowan, 1984). Thus, a longer SSRT reflects an inhibitory deficit, whereas a longer go RT is interpreted as a lack of attention among other influencing factors (preparation, choice, and speed-accuracy trade-off, Lijffijt, Kenemans, Verbaten, & van Engeland, 2005).

In addition to the GNG and the SST, other experimental paradigms, such as the Stroop (Stroop, 1992) and Eriksen Flanker tasks (Eriksen & Eriksen, 1974) have been proposed to measure inhibitory capacities. However, these paradigms measure distractor inhibition rather than motor response inhibition (Nigg, 2000; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004). To keep the present review focused and allow for straightforward comparisons of results, we only included studies using the GNG and SST.
1.2. Research Gaps and Research Needs

1.2.1. Previous meta-analyses and reviews

To date, there are at least nine published meta-analyses or review papers examining the relationship between inhibitory control and long-term substance use or behavioral addiction. In terms of scope, these studies can be classified into three categories. First, literature overviews focusing on a single substance (e.g., alcohol: Aragues, Jurado, Quinto, & Rubio, 2011; Stavro, Pelletier, & Potvin, 2013) or non-substance related disorder (e.g., gambling disorder: Chowdhury, Livesey, Blaszczynski, & Harris, 2017; Moccia et al., 2017). These reviews associated alcohol use with prolonged inhibition impairment, up to one month after abstinence (Stavro et al., 2013) and detoxified alcohol-dependent patients showed poor inhibition compared with healthy controls (Aragues et al., 2011). Polysubstance use was not systematically described or controlled for in either of the review studies on alcohol. Individuals with gambling disorder without comorbid SUD were reported to show large inhibition deficits (Chowdhury et al., 2017), which was attributed to impaired activity in prefrontal areas (Moccia et al., 2017). Second, other reviews focused on drawing general conclusions across multiple substances. For instance, Lipszyc and colleagues found that substance users generally did not differ significantly from controls in SST (Lipszyc & Schachar, 2010) and GNG performance (Wright, Lipszyc, Dupuis, Thayapararajah, & Schachar, 2014). However, such a review does not provide a clear profile for the effects of these substances in isolation or of specific interactions (i.e., greater than additive or compensation effects). A third category of literature reviews included multiple substances and the results were specified by the substance. Examples include a recent systematic review focused on neuroimaging findings (Luijten et al., 2014) and a meta-analysis focused on behavior (Smith et al., 2014). The latter meta-analysis indicated that inhibitory deficits were apparent for heavy use/disorders related to cocaine, ecstasy, methamphetamine, tobacco,
alcohol, and gambling but not for opioids or cannabis, without testing the interaction effect of using multiple substances (Smith et al., 2014). In sum, the current findings and conclusions of reviews and meta-analyses are rather inconsistent. If a conclusion can be drawn, it appears to be the counterintuitive conclusion that reviews and meta-analyses that focused on a specific addictive substance or behavior are more likely to report a significant association with inhibitory control compared to those reporting on multiple substance use. Importantly, none of these reports have considered several key variables that might bias the results, which will be highlighted in the next section.

1.2.2. Important factors to consider

1.2.2.1. Polysubstance use

Polysubstance use broadly refers to the consumption of more than one drug over a defined period, either simultaneously or at different times (Connor, Gullo, White, & Kelly, 2014; Subbaraman & Kerr, 2015). This involves different sub-categories, namely using different substances, the dependence of one substance and co-use of other substances or dependence on multiple substances. For instance, tobacco smoking is strongly associated with alcohol and marijuana use (Connor et al., 2014), opioids, and benzodiazepines are often prescribed simultaneously (Jones, Mogali, & Comer, 2012), and stimulants users are more likely to be heavy drinkers (McCabe, Knight, Teter, & Wechsler, 2005). Note that there is some evidence indicating that concurrent use of substances can lead to additionally toxic effects because of a toxic metabolite, as was reported for alcohol and cocaine (Pennings, Leccese, & Wolff, 2002). It is also possible that the use of one substance decreases the negative effect of another substance, as found with alcohol and cannabis (Schweinsburg, Schweinsburg, Nagel, Eyler, & Tapert, 2011). Hence, studying interactions between drugs on neurocognitive functions is important, given the frequent occurrence and possible interaction.
effects. However, studies comparing substance users versus non-users or light users have typically focused on the primary substance of concern, while ignoring secondary substances. Up to now, only a few studies have investigated the relationship between polysubstance use and inhibition (Gamma, Brandeis, Brandeis, & Vollenweider, 2005; Moallem & Ray, 2012; Verdejo-García, Perales, & Pérez-García, 2007). Heavy drinking smokers did not show poorer SST response inhibition than smokers only and heavy drinkers only (Moallem & Ray, 2012). Similarly, ecstasy polysubstance users did not show more strongly disturbed inhibitory brain mechanisms compared with controls (Gamma et al., 2005), and cocaine and heroin polysubstance users showed similar commission error rates as controls in the GNG (Verdejo-García et al., 2007). A limitation of the latter two studies is that the greater-than-additive effect could not be examined without a group of single substance users. The lack of studies calls for a synthesis of research that does take polysubstance use into account.

**1.2.2.2. Substance use as a continuous variable**

All the above-mentioned reviews and meta-analyses included comparisons between a control or light user group and a heavy or problematic user group. Scores retained as a result of such extreme group designs are often coded and analyzed in terms of low versus high, reducing individual differences into a binary code. This practice involves ignoring individual-differences of substance use in favor of creating quasi-arbitrary groups assumed to be homogeneous on the variable of interest (MacCallum, Zhang, Preacher, & Rucker, 2002; Royston, Altman, & Sauerbrei, 2006; Preacher, Rucker, MacCallum, & Nicewander, 2005). In the current study, we aimed to quantify substance use as a continuous variable.

**1.2.2.3. Abstinence**

Studies on long-lasting effects of substance use have generally been conducted by testing recently abstinent users. With respect to response inhibition, some studies have found
that abstinence from cocaine, methamphetamine and heroin normalized inhibitory function (Morie et al., 2014; Schulte et al., 2014), however, one study found sustained suboptimal performance after heroin abstinence (e.g., Fu et al., 2008). In addition, the duration of abstinence appears to moderate the return to normal functioning, which may explain these conflicting findings (Schulte et al., 2014). In order to preclude this as a confounder, we did not include studies on abstinence in (formerly) dependent users. All participants indicated substance use in everyday life, but were requested to refrain from using all substances (in most cases excluding tobacco) 24 hrs to one week before testing.

1.2.2.4. Individual-level and task-level variables

Some individual-level and task-level factors are known to affect inhibitory control and are therefore included in this mega-analysis, including the demographic variables age, sex, and education years. For GNG, six task parameters were controlled for: no-go percentage, number of experimental trials, working memory load (taxed or not), substance-related stimuli (used or not), cued GNG or not, and task complexity. For the SST, five task parameters were controlled for: number of experimental trials, stop-trial percentage, SSD settings, stop-signal modality, and SSRT calculation method. Reasons for controlling these confounders are based on a large primary literature on these tasks and are summarized in Supplementary Materials S1. Except for sex, for which the interaction with substance use was considered, all other factors were only controlled for regarding their main effect.

1.3. Why a Mega-Analysis Rather Than a Meta-Analysis?

A meta-analysis combines the summary statistics (i.e., effect sizes of included studies), while a mega-analysis combines the raw individual data from different studies. The latter method allows studying the combined effect of individual characteristics (cf. Price et al., 2016) and examining the interaction effect of multiple substances used with enhanced
statistical power (Riley, Lambert, & Abo-Zaid, 2010). Therefore, we implemented a mega-analysis with individual-level data.

1.4. The Goal of the Current Study

Our primary goal was to examine the main and interaction effects of various kinds of long-term substance use on response inhibition. As the interaction effects of substance use on inhibition are rarely investigated and reported, we explore these interactions in the current study. We do so while controlling for demographics (e.g., age, sex, education years) and task-related factors (e.g., no-go percentage, number of trials, whether stimuli are substance-related) that likely explain performance variance between studies and individuals. Interactions between substance use and sex were also included. Based on the literature reviewed above, we tested the following hypotheses: 1) According to Smith et al (2014) and other findings (Colzato, van den Wildenberg, & Hommel, 2007; Fillmore & Rush, 2002, Quednow et al., 2007), we assumed that the inhibitory deficit would be more pronounced in users of psychostimulants (e.g., cocaine, ecstasy, methamphetamine, tobacco, and alcohol), especially for cocaine and amphetamines, given the known neuropsychopharmacology of the cortical and subcortical networks underlying impulse control (i.e., the right dorsolateral and inferior frontal cortices, Koob & Volkow, 2010; Smith, Mattick, Jamadar, & Iredale, 2014); 2) Given the literature, and as a validation of our individual-level mega-analysis, we expect some demographics (e.g., age and sex) and task characteristics (e.g., no-go percentage, whether stimuli are substance-related) to be associated with inhibition performance (see for expected directions of effects, Supplementary Materials S1).
2. Method

2.1. Study Identification and Selection

PsycINFO, Medline, EMBASE, Web of Science, CINAHL, and Cochrane Library were searched until 01/03/2016. Search terms and synonyms indicating substance use (alcohol, amphetamine, cocaine, cannabis, heroin, ketamine, methamphetamine, benzodiazepines, gambling, gamer, and internet addiction) were combined with terms indicative of inhibition (go/no-go, inhibitory control, inhibitory process, response inhibition, stop task, etc.). Published meta-analyses and reviews were also checked for additional studies (Horsley, Dingwall, & Sampson, 2011). Although behavioral addictions (e.g., gambling, internet addiction) were initially included, there were too few relevant studies to allow further analyses.

2.1.1. Eligibility criteria

The first author (YL) assessed the eligibility of all records using the following initial inclusion criteria: (a) presented in English; (b) conducted on human participants; (c) reported at least one measure from the following: no-go commission errors or go RT in the GNG; SSRT or go RT in the SST; (d) reported use of at least one kind of substance (e.g., alcohol, tobacco, cannabis, amphetamine, cocaine, ecstasy). Note that we included behavioral data from fMRI/EEG studies if available. In addition, we ran supplementary analyses to investigate whether inhibition performance varied with study type (behavioral/EEG/fMRI). It turned out that study type did not systematically influence behavioral performance (see Supplementary Materials S2). We excluded studies (a) that presented stop signals using a single SSD, as this is known to induce a performance strategy of delayed responding (Logan, 1994); (b) in which the percentage of no-go or stop trials was higher than 50%, as this is known to invalidate the task (Nieuwenhuis, Yeung, & Cohen, 2004; Randall & Smith, 2011);
(c) that focused on the acute effects of substances on inhibition; (d) that recruited participants with a family history of substance dependence; (e) that excluded polysubstance users; (f) with participants that already received treatment for SUD or abstained from substance use; (g) with participants younger than 18. The exclusion of both intoxicated and abstinent consumers may have kept heavily affected/addicted participants from being included in the sample.

After applying the inclusion and exclusion criteria by YL, a second rater (author YG) assessed the eligibility of a random subset (20%) of the records and obtained 100% agreement. Authors of eligible studies were invited via email to contribute raw data. Repeated attempts were made (i.e., four reminders were sent) if no response was received. Corresponding authors of the identified studies were asked to share their raw individual data, following our instructions on data requirements. The ‘essential variables’ included a set of pre-identified variables, including sociodemographic characteristics (e.g., age, sex, and education), typical alcohol and tobacco use (as alcohol and tobacco are two most commonly used substances), and task performance (Table S1a, S1b). ‘Optional variables’ (Supplementary Materials S3) included other demographic information recorded (e.g., race), other substance use (e.g., cocaine, cannabis) and questionnaires administered (e.g., Alcohol Use Disorder Identification Test (AUDIT), Saunders, Aasland, Babor, De la Fuente, & Grant, 1993). The ‘optional variables’ were defined in a more flexible format with open questions. A study was included in our mega-analysis only if information about all ‘essential variables’ could be provided.

2.1.2. Quality assessment and data extraction

As the quality of included studies can influence mega-analysis in unpredictable ways (i.e., shortcomings in original studies will be carried over to the mega-analysis and thus weaken its conclusions, Müller, Brändle, Liechti, & Borgwardt, 2019), a quality assessment
of original studies was conducted. The methodological quality of studies was assessed by two authors (YL and YG) separately. We used the National Heart, Lung, and Blood Institute (NHLBI) Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies, which is widely used and recommended by Cochrane for quality assessment of observational and cross-sectional studies (Table S2, National Heart and Blood Institute, 2014). The total agreement (Good/Fair/Suboptimal) between assessors was high (GNG: 20/24 = 83%, SST: 16/20 = 80%). Inter-rater reliability, measured using Spearman's rank correlation coefficient was high for GNG ($r = 0.84$, $p < 0.001$) and moderate for SST ($r = 0.56$, $p = 0.01$, Kendall, 1938).

All provided data, including predictors (i.e., substance use, demographics, task characteristics) and dependent variables were merged into four datasets separated based on the four dependent variables (i.e., the commission error rate in GNG, go RT in GNG, SSRT in SST, and go RT in SST. As speed-accuracy trade-off is a potential issue in GNG (Zhao, Qian, Fu, & Maes, 2017), a balanced integration score was calculated (Liesefeld & Janczyk, 2019). Main results applying this score as the outcome are presented in Supplementary Materials S4. The first author performed the data merging, which was verified by two authors (RW and WW).

2.1.3. Publication bias check

To examine whether significant findings in the original papers are indicative of evidential value, a $p$-curve was calculated and plotted (Simonsohn et al., 2015). In a $p$-curve, the x-axis represents $p$-values below 0.05, and the y-axis represents the percentage of studies yielding such a $p$-value. A right-skewed $p$-curve indicates evidential value, whereas a left-skewed $p$-curve, many $p$-values just below 0.05, may be indicative of flexibility in data analysis (Simonsohn et al., 2015). If the data did not indicate evidential value, a 33% power
test is performed to examine whether the absence of evidential value is due to insufficient power. A $p$-curve disclosure table was added in Supplementary Materials (Table S3) according to Simmons and Nelson (2015). $P$-curves and corresponding analyses were conducted using the $p$-curve app 4.06 (http://www.p-curve.com/app4, 2018).

### 2.2. Individual Participant Data Meta-Analysis

The analysis was conducted in the following steps: 1) apply additional exclusion criteria to the merged datasets; 2) standardize all continuous independent variables; 3) determine substance-related one-way variables; 4) dummy code all discrete variables; 5) determine and generate substance-related interaction variables; 6) multiple imputations of the missing values using all main and interaction variables; 7) build the linear mixed regression model with fixed effects of all predictors and a random intercept; 8) variable selection by stepwise backward elimination. These eight steps are outlined in more detail below.

#### 2.2.1. Construction of the database

**2.2.1.1. Individual and group exclusion criteria**

The data from the included studies were stacked into a single data file for each dependent variable, with unique identifiers for each study and for each participant. We further applied some minimal exclusion criteria to the individuals. That is, we excluded a participant if (1) he/she was younger than 18 years old; (2) he/she had missing data on all indices of substance use; (3) the dependent variable of current analysis (e.g., commission error rate) was missing; (4) SSRT was negative.

A group of substance users from a certain study was excluded if the substance was not included as a predictor in the model. This happened when there was limited data provided for that substance (see criteria in 2.2.1.3.1.). For example, if it was concluded that opiate use was assessed insufficiently across all studies, we did not add opiate as a predictor. Consequently,
opiate users were excluded from the analysis. The excluded cases and groups from each study are listed in Table 1 and Table 2.

2.2.1.2. **Standardization of independent variables**

2.2.1.2.1. **Continuous variables**

Demographics like age and education level were transformed respectively into continuous variables years and years of education according to the education system in the country where the study was conducted. Task characteristics such as no-go percentage and number of trials in both tasks were also treated as continuous variables.

Alcohol consumption was converted into the continuous variable grams of ethanol per month. Data on alcohol consumption were provided in two different ways. Most researchers provided data based on timeline follow-back (TLFB). These data were either already in grams per month or could be transformed by making use of standard drinks adjusted for country (Cooper, 1999). Some studies only had data from more general questionnaires. For instance, three studies (de Ruiter, Oosterlaan, Veltman, van den Brink, & Goudriaan, 2012; Luijten, O’Connor, Rossiter, Franken, & Hester, 2013; Rossiter, Thompson, & Hester, 2012) provided the raw data of the AUDIT (Saunders et al., 1993). In that case, we multiplied midpoints of item 1 (frequency), midpoints of item 2 (drinking days per month) and standard drinks in the country where the study took place. Similarly, four studies (Littel et al., 2012; Luijten et al., 2011; Luijten, Meerkerk, Franken, van de Wetering, & Schoenmakers, 2015; Luijten et al., 2013) provided Quantity Frequency Variability (QFV) score (Lemmens, Tan, & Knibbe, 1992). Again, items of quantity, frequency, and standard drinks were multiplied together. Smoking was coded as cigarettes per day. Two studies (Moallem & Ray, 2012; Rossiter et al., 2012) only had data from the Fagerström Test for Nicotine Dependence (FTND, Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991). In these cases, the midpoint
of the answer to item “How many cigarettes a day do you smoke” was used for daily cigarette use. One study used a self-developed 7-point Likert scale for the past 6 months tobacco consumption, for which we estimated daily cigarette use with the midpoint scores (Ames et al., 2014). Alcohol and tobacco use were standardized across the full dataset. All the other substance use variables had to be treated as dichotomous variables, as insufficient information was provided for treating it as a continuous variable in the model (see details below).

2.2.1.2.2. Dichotomous variables

For interpretability, dichotomous variables were effect-coded with value +1 or -1. Except for alcohol and tobacco use, other substances were coded as ‘lifetime use (yes = 1/no = -1)’.

Four dummy task-characteristics were defined to classify the GNG studies: ‘working memory load (low/high)’, ‘substance-related (yes/no)’ ‘cued GNG (yes/no)’, and ‘task complexity (low/high)’. High working memory load, substance-related, cued GNG versions and complicated tasks were assigned the value of 1 (otherwise -1). Tasks with high working memory load were also assigned a value of 1 for task complexity as the association between stimuli and response was more complicated in these tasks.

Similarly, for the SST, three dummy task characteristics were extracted, including ‘stop-signal modality (visual/auditory)’, ‘SSD (fixed/staircase-tracking)” and ‘SSRT calculation (integration/others)’. These variables were assigned a value of 1 if auditory stop signals were used; staircase-tracking procedure for SSD; and integration method for SSRT calculation (otherwise -1).
2.2.1.3. Identification and generation of substance-related variables

Except for alcohol use and tobacco use, other kinds of substances had missing data as not all studies provided information. Data provided varied in the level of detail, the way questions were asked, and the substances of main interest. For instance, depending on the primary substance of interest, some studies provided detailed information for cannabis use but no information on cocaine use (Bidwell et al., 2013), with an opposite pattern for others (Colzato, van den Wildenberg, & Hommel, 2007). In the following section, we explain the criteria for including substance-related variables in the model.

2.2.1.3.1. One-way variables

Due to missing data, a criterion was needed to include a variable in the model. We decided on a minimum of 100 participants per cell for a substance (which comes down to a power of 0.94 for the effect size of 0.5). As a result, final models for the GNG (both commission error rate and go RT) included cannabis, cocaine, amphetamine, ecstasy, and hallucinogens, in addition to alcohol and tobacco. For the SST (both SSRT and go RT), the final models included cannabis, cocaine, and ecstasy in addition to alcohol and tobacco.

2.2.1.3.2. Two-way variables

There were two types of two-way variables; the interaction of sex × substance and substance1 × substance2. Variables of sex × substance were created by multiplying sex with substance directly. For the second type, in order to evaluate whether there was sufficient data to assess these interactions, we again applied a criterion for inclusion. For example, dummy coding cannabis and cocaine use yielded a two by two table cannabis (yes/no) × cocaine (yes/no). The corresponding interaction was only entered into the model if all four cells had more than 20 entries. For alcohol and tobacco use, we dichotomized the data by a median split for table construction only. We performed an
additional analysis to test whether the number of substances used was a predictor of inhibition performance, and this was not the case (see Supplementary Materials S5). The list of included two-way variables can also be found in Supplementary Materials (Table S4a-S4d). Demographics (in addition to sex) and task parameters could further moderate the relationship between substance use and inhibition. This, however, was not the focus of the current paper. In order to explore this potential issue, we analyzed interactions between alcohol on the one hand and demographics and task parameters on the other (see Supplementary Materials S6).

2.2.1.3.3. Three-way variables

Three-way variables were generated based on the substance1 × substance2 variables combined with sex. The corresponding variables were entered into the model only when all the eight cells in the three-way table sex (male/female) × substance1 (yes/no) × substance2 (yes/no) consisted of at least 10 entries. The list of three-way variables can be found in Supplementary Materials (Table S4a-4d).

2.2.2. Missing data for independent variables and their interactions

In the analysis of GNG commission error rate, the percentage of missing values ranged from 0 to 68.2% (highest: alcohol × hallucinogens × sex) and in the GNG go RT analysis, it ranged from 0 to 69.6% (highest: alcohol × hallucinogens × sex). For the SST, the percentage of missing values ranged from 0 to 84% for the SSRT (highest: tobacco × ecstasy × sex) and from 0 to 83.2% for the go RT (highest: tobacco × ecstasy × sex), a full list of missing data per variable can be found in Table S4a-s4d).

In order to deal with these missing data, we used multiple imputations (Rubin, 2004). The default imputation option in SPSS was chosen. It first scans the data and determines the
suitable method for imputation (Monotone or Fully Conditional Specification, FCS; Dong & Peng, 2013). All variables in the mixed regression model, including the main and interactive predictors and the dependent variable, were used for imputation. Apart from that, the discrete variable of ‘tobacco lifetime use’ was also used, as some studies assessed tobacco use dichotomously (smokers/non-smokers). It has been suggested that the number of imputations should be similar to the percentage of cases that are incomplete (White, Royston, & Wood, 2011) and the precision improves by increasing the number of imputations (Bodner, 2008). Therefore, 100 complete data sets were generated, which were combined into a pooled result using the method proposed by Rubin (Rubin, 2004) and Schafer (Schafer, 1997).

2.3. Statistical Analysis

Backward elimination was used for variable selection. Initially, each imputed dataset was analyzed with a linear mixed model including all the above-mentioned main, second order, and third order effects as fixed effects and a random intercept (for which a model summary can be found in Tables S4a-S4d). We did not include random slopes and thus assumed that predictors had similar effects in each study. The fixed effects that were least significant (i.e., the one with the largest p-value) were removed and the model was refitted. Each subsequent step removed the least significant variable in the model until all remaining variables or its higher order variables had p-values smaller than 0.05 (Draper & Smith, 2014). For instance, if the variable alcohol \( \times \) tobacco was significant, then variables of alcohol and tobacco would also be included in the model, irrespective of their independent significance.

3. Results

3.1. Study Selection

3.1.1. Summary of authors’ responsiveness
Applying the inclusion and exclusion criteria resulted in a sample of 153 potentially eligible studies (Fig. 1). Out of these targeted papers, 4 researchers responded that they no longer had access to the datasets, 21 declined to participate, 52 did not respond to our invitation and 11 did not have all the basic information we asked for. In total, we obtained raw data from 65 studies. Out of these, 22 had to be excluded because the authors could not provide all the ‘essential variables’, such as data on monthly alcohol use in grams was unavailable (9 studies), missing data of tobacco use (5 studies), participants were abstaining from substance use (3 studies), participants were younger than 18 years old (2 studies), uncommon tasks were used (2 studies) and unsuitable outcome measures (1 study, provided stop latency instead of SSRT). The full list can be found in Supplementary Materials S7. The final dataset for the GNG comprised of 23 independent datasets from 24 papers (in some cases, more than one paper was published with the same dataset). For the SST, 19 datasets from 20 papers were included. In addition, one study administered both GNG and SST; therefore 43 unique studies were included in total.

The final list of eligible studies was slightly different from the list of studies included in Smith and colleagues meta-analysis on summary statistics (Smith et al., 2014). For the GNG, there were 11 studies in common. For the SST, there were 6 studies in common. These discrepancies were related to different research questions. Since we aimed to assess the unique and combined effects of different substances, while Smith and colleagues focused on the unique effect of a single substance, some studies that were excluded by Smith and colleagues were included here and vice versa. In addition, individual data mega-analysis typically has a lower response rate compared to traditional meta-analysis, as it requires more work from the researchers (Riley et al., 2010; Riley, Simmonds, & Look, 2007).
3.1.2. Study description

Table 1 and Table 2 present descriptive characteristics of the included GNG and SST studies before imputation, respectively.

3.1.3. Findings in original studies

For GNG, out of the 24 studies included, 9 (37.5%) reported that (heavy/problematic) substance users/excessive gamers made more commission errors than controls/light users (3 for alcohol, 2 for tobacco, 1 for ecstasy, 1 for inhalant and 2 for excessive gamers), 1 (4.2%) reported opposite findings (i.e., opiate users made fewer commission errors compared to controls), 11 (45.8%) reported no significant differences (5 for alcohol, 2 for tobacco, 1 for ecstasy, 1 for inhalant and 2 for polysubstance use), and 3 (12.5%) didn’t have such an analysis (See Table 1 footnote). For the SST, out of the 20 studies, 5 (25%) reported substance users/gamblers had longer SSRT than controls (2 alcohol, 2 cocaine and 1 pathological gambling), 1 (5%) reported the opposite direction (alcohol), 8 (40%) reported no difference (3 alcohol, 2 tobacco, 1 cannabis, 1 cocaine, and 1 pathological gambling) and 6 (30%) did not provide such an analysis (see Table 2 footnote).

3.2. Quality Assessment

We rated the methodological quality of the studies according to the NHLBI assessment tool (see Tables 3a and 3b). For the GNG, most (58.3%) of the studies were of intermediate quality, 37.5% of high quality and 4.2% of suboptimal quality. For the SST, 40% of studies were of high quality and another 60% of intermediate quality. The main limitations were small sample size, especially for the studies focused on neuroimaging findings, and insufficient control of confounders such as the history of other kinds of drug use. For a few studies, the population was not fully described, lacking information of where and when the participants were recruited. To explore whether different study types differ in methodological
quality, we did a chi-square test based on Table 3. The results indicate that the percentages of studies of good, fair and suboptimal quality did not differ between behavioral (10/23, 13/23, 0/23), EEG (4/8, 3/8, 1/8) and fMRI (3/12, 9/12, 0/12) studies ($\chi^2$ (4, N = 44) = 6.51, $p = 0.15$).

3.3. Publication Bias Check

To examine evidential value in the original studies, a $p$-curve was created (Supplementary Materials Fig. S1). Out of the 31 effect sizes (unavailable for some studies), 11 were statistically significant ($p < 0.05$), with 8 $p < 0.025$. The $p$-curve analysis on the association between substance use and response inhibition indicated no evidential value (full $p$-curve $z = -0.98, p = 0.16$; half $p$-curve $z = 0.58, p = 0.72$). However, this was likely due to a lack of power (33% power test, full $p$-curve $z = -0.95, p = 0.17$).

3.4. Main Outcomes

3.4.1. GNG: no-go commission errors

None of the substance-related variables or their interactions had a significant effect on the commission error rate. Among all other variables, two demographic variables and three task characteristics significantly predicted commission error rates. Age significantly predicted commission error rate ($\beta = -0.01, p < 0.01, 95\% \text{ CI } [-0.02, 0.00]$), indicating that older participants showed decreased commission error rates. Education years also significantly predicted commission error rate ($\beta = -0.01, p = 0.03, 95\% \text{ CI } [-0.02, 0.00]$), indicating that the higher the educational level, the lower the commission error rates. The nominal variable working memory load had a significant effect on commission error rate ($\beta = 0.10, p < 0.01, 95\% \text{ CI } [0.07, 0.14]$), indicating that when working memory load was high, participants made more commission errors. The no-go percentage had a significant effect on commission error rate ($\beta = -0.04, p < 0.01, 95\% \text{ CI } [-0.07, -0.02]$), such that the higher the no-go percentage,
the lower the rate of commission errors. The number of trials also had a significant effect on commission error rate ($\beta = 0.04, p < 0.01, 95\% \text{ CI } [0.02, 0.07]$), indicating higher commission error rates when there were more trials.

### 3.4.2. SST: SSRT

Lifetime cannabis use significantly predicted SSRT, with users showing longer SSRT than non-users ($\beta = 5.59, p = 0.03, 95\% \text{ CI } [0.41, 10.77]$). Tobacco use was positively, although not significantly, associated with SSRT ($\beta = 3.21, p = 0.06, 95\% \text{ CI } [-0.13, 6.55]$), indicating that the more tobacco was consumed, the longer SSRT. The tobacco $\times$ cannabis interaction also had a significant effect on SSRT ($\beta = -4.19, p = 0.03, 95\% \text{ CI } [-8.03, -0.37]$, Fig. 2). Post-hoc analyses were performed by splitting the imputed data sets and fitting the same restricted model without the interaction term. These analyses revealed that for the cannabis non-users, higher tobacco use was associated with longer SSRT ($\beta = 6.44, t = 2.70, p < 0.01$). For cannabis users, no effect of tobacco use on SSRT was observed ($\beta = -0.15, t = -0.05, p = 0.96$). When split based on cigarette smoking (median-split of $z$-score), the following effects were obtained: for low tobacco users, cannabis lifetime users did not differ significantly from cannabis non-users in SSRT ($\beta = 7.62, t = 1.90, p = 0.06$). A similar finding was observed among high tobacco users ($\beta = 4.80, t = 1.74, p = 0.08$).

Education years also significantly predicted SSRT ($\beta = -9.33, p < 0.01, 95\% \text{ CI } [-12.88, -5.80]$), indicating that the higher the education level, the shorter the SSRT. Age significantly predicted SSRT ($\beta = 13.46, p < 0.01, 95\% \text{ CI } [9.29, 17.63]$), with an increase in SSRT along with an increase in age. The number of trials also significantly predicted SSRT ($\beta = -17.44, p < 0.01, 95\% \text{ CI } [-30.60, -4.28]$), indicating a decrease in SSRT when there were more trials. In addition, stop-signal modality had an effect on SSRT ($\beta = -28.58, p = 0.01, 95\% \text{ CI } [-50.61, -6.56]$), indicating that auditory stop signals induced shorter SSRT.
compared to visual stop signals. SSD also had a significant effect on SSRT ($\beta = -33.29$, $p = 0.04$, 95% CI [-64.61, -1.96]), indicating that the staircase-tracking procedure resulted in shorter SSRT compared to the fixed SSD procedure.

For both SSRT and commission error rate, models including the interaction between alcohol use on the one hand and demographics and task parameters on the other resulted in largely comparable findings as presented here\(^1\). Only in the GNG, an interaction between alcohol use and age appeared ($\beta = 0.01$, $p = 0.02$, 95% CI [0.001, 0.02]). For light drinkers, older people made less commission errors ($\beta = -0.02$, $t = -2.56$, $p = 0.01$), which was in line with the main effect of age. Whereas for heavy drinkers, this relationship was absent ($\beta = -0.01$, $t = -1.50$, $p = 0.14$). All other interactions with alcohol were found to be non-significant (Supplementary Materials S6).

Outcomes for go RT in GNG and SST can be found in Supplementary Materials S8. Briefly, older people had longer go RT in both GNG and SST. Higher educated people had shorter go RT in SST. Although the interaction between cocaine and tobacco had an effect on go RT in SST, post-hoc analysis revealed no significant simple effect.

4. Discussion

Previous individual studies, reviews, and meta-analyses investigating inhibitory control deficits in relation to long-term substance use and SUD have provided mixed results (Luijten et al., 2014; Smith et al., 2014; Wright et al., 2014). These inconsistent findings might at least partly be due to insufficient control of frequently occurring polysubstance use. In addition,

---

\(^1\) In the model including interactions with demographics and task-parameters, tobacco and cannabis use were both positively associated with SSRT. However, their interaction was not significant, but the three-way interaction with sex was. Post-hoc tests indicated that, only for male non-cannabis users, tobacco use was positively associated with SSRT (see in Supplementary Materials S6).
studies differed in sample demographics and task-related variables and used extreme group designs. The current mega-analysis aggregated data of 3610 individuals, from 43 studies, in which polysubstance use, demographics, and task parameters were included in the prediction of inhibition performance by means of an imputed multilevel analysis. Most of the included studies were of medium to high quality, which validates the overall conclusions drawn. Surprisingly, our overall pattern of results indicated that most types of substance use did not show an association with response inhibition. While for most substances no effects were found, lifetime cannabis use was found to be associated with impaired inhibition, as indexed by an increased SSRT in the SST. Tobacco use was also associated with impaired inhibition as indexed by the same variable. In addition, an interaction between lifetime cannabis and tobacco use was found on SSRT, which indicated a strong positive relationship between daily tobacco use and SSRT in participants who did not use cannabis (indicating poorer inhibition), and the absences of such a relationship in users smoking cannabis. In addition, demographic factors such as age and years of education and task characteristics such as no-go percentage, affected inhibition performance in the expected direction, strengthening the credibility of the other results.

4.1. Response Inhibition and Substance Use

The main significant finding of our mega-analysis was that lifetime cannabis use was associated with prolonged response inhibition in the SST. One possible explanation is that this could (partly) involve subacute effects of cannabis use (i.e. lasting 7 hours to 4 weeks after last cannabis use, Gruber & Yurgelun-Todd, 2005; Pope & Yurgelun-Todd, 1996; Schulte et al., 2014). Acute cannabis use (i.e., 0-6 hours after last cannabis use) has been consistently reported to impair response inhibition in the SST (Metrik et al., 2012; Ramaekers et al., 2006). In contrast, findings of its long-term effect (i.e., 3 weeks or longer after last cannabis use) were mixed (Crean, Crane, & Mason, 2011), with some confirming an
impairing effect (Moreno et al., 2012), while others did not (Tapert et al., 2007). To have a closer look at the effect of cannabis, we compared cannabis daily users with less frequent users. A linear mixed regression model was built with the fixed effect of ‘cannabis daily users (yes/no)’ and a random intercept. It indicated that cannabis daily users did not differ from less frequent users on their stopping latency (i.e., SSRT., $\beta = -6.42$, $p = 0.90$, 95% CI [-114.27, 127.10]), which does not support the hypothesis of subacute cannabis effects. Despite conflicting behavioral findings of the relationship between cannabis use and response inhibition, abnormalities in neural activation have often and more consistently been reported in relation to acute as well as chronic cannabis use compared with non-users (systematic review: Wrege et al., 2014). Age of onset may have a moderating effect on the neural effects of cannabis (Hester, Nestor, & Garavan, 2009), but we did not have sufficient data to test this hypothesis.

In line with previous findings, tobacco use tended to impair inhibition. Participants with a higher level of tobacco dependence demonstrated a lower level of response inhibition capacities (Billieux et al., 2010), and smokers performed worse than non-smokers in a smoking-related GNG (Luijten et al., 2011). However, it should be noted that the main effect of tobacco use was qualified by a significant interaction with cannabis use, indicating a negative effect of tobacco use only in non-cannabis users. Another study reported that co-administration of cannabis and tobacco attenuated the impairment in delayed recall memory caused by cannabis alone (Hindocha, Freeman, Xia, Shaban, & Curran, 2017), and other reports have indicated weaker impairment on some measures after polysubstance use (e.g., alcohol and cannabis, Schweinsburg et al., 2011). One possible interpretation of these findings is that cannabis has a protective effect when used together with other substances such as alcohol and tobacco (cf., Viveros, Marco, & File, 2006). Due to the high co-occurrence of cannabis and tobacco use (Badiani et al., 2015; Leatherdale, Ahmed, &
Kaiserman, 2006), and the fact that concurrent tobacco use contributes to cannabis
dependence symptoms (Ream, Benoit, Johnson, & Dunlap, 2008), further studies of the
combined and single effects on response inhibition are warranted to elucidate these findings.

What could explain the low evidence for a relationship between (most) long-term
substance use and inhibition? On closer inspection, only 30% of studies included reported
evidence for negative associations between substance use (or gambling) and response
inhibition (Tables 1 and 2). In contrast, other studies reported evidence for positive
associations between substance use and inhibition performance in GNG and SST (significant:
Glass et al., 2009; nonsignificant: Galván, Poldrack, Baker, McGlennen, & London, 2011;
In light of this, it is less surprising that the integrated results indicated overall largely null
findings (most of the confidence intervals ranged around zero). Similarly, only one out of the
five studies included in a recent review (Carbia, López-Caneda, Corral, & Cadaveira, 2018)
reported impaired response inhibition—as measured by SST and GNG tasks—in binge
drinkers compared with controls (Czapla et al., 2015).

One explanation is that chronic recreational substance use without a diagnosis of SUD
is not associated with response inhibition impairment. In other words, a threshold effect
rather than a linear effect might exist between substance use and response inhibition
performance. Alternatively, there might be a linear relationship, albeit shallow and we only
see the effects when comparing very extreme groups (e.g., healthy controls vs. SUD in
clinical samples). As a result of our exclusion criteria, Fig. S2a and S3a indicate that only a
minority of the participants reached the level of SUD (either reported in individual paper or
categorized based on questionnaire score), and most others were still within the normal range
of use. It is conceivable that inhibition is only impaired in SUD (Bjork, Hommer, Grant, &
Danube, 2004; Fernández-Serrano, Pérez-García, & Verdejo-García, 2011; Noël et al., 2007;
Alternatively, inhibition problems may play a role in the transition from heavy use to SUD. In the SST sample, there were more people diagnosed with tobacco dependence (about 10%, Fig. S3a), which might explain why a positive (although not significant) association of SSRT and tobacco use was found.

A second possibility is that substance use is actually associated with impaired inhibition, but we were unable to detect this. Possible reasons include: sample characteristics (as was discussed in the last paragraph), the type of tasks included, outcome measures (i.e., effects may only be visible in biological markers but not in behavior), and statistical power. Regarding tasks included, there is the possibility that (heavy) use of psychoactive substances does not lead to a general inhibition problem, but only to a specific problem in the domain of substance use (hence an interaction between an appetitive process and suboptimal control, Jones, Duckworth, Kersbergen, Clarke, & Field, 2018). A related explanation can be that self-control failures like maladaptive substance use may reflect a reduced mobilization of inhibitory control in substance-related contexts rather than generally impaired inhibitory control competencies (Krönke et al., 2018; Krönke, Wolff, Benz, & Goschke, 2015; Wolff et al., 2016). However, in a secondary analysis, we did not find that substance-related GNG moderated the relationship between alcohol and commission error rate (see details in 4.2.).

Furthermore, the SST and GNG measure stimulus-driven (exogenous) inhibition, which may not closely match real-world ‘loss of control’ behavior related to substance use (e.g., an initial intention to have one drink escalating into a binge-drinking session, failed suppression of craving, etc). These examples reflect a different type of inhibition, namely endogenous or intentional rather than exogenous inhibition. Intentional inhibition paradigms such as the Marble task (Schel et al., 2014) could be considered in future research. Regarding outcome measures, it is possible that biological but not behavioral markers might be more sensitive to inhibition impairments among substance users (Garrison & Potenza, 2014). Relatedly, some
of the included MRI studies reported specific group-related abnormalities in brain activation but not in behavioral outcomes (e.g., Claus, Ewing, Filbey, & Hutchison, 2013; de Ruiter et al., 2012; Galván, Poldrack, Baker, McGlennen, & London, 2011; Karoly, Weiland, Sabbineni, & Hutchison, 2014; Luijten et al., 2013; Roberts & Garavan, 2010). In addition, a recent study indicated that resting state fMRI connectivity might serve as a promising biomarker of alcohol use disorder severity (Fede, Grodin, Dean, Diazgranados, & Momenan, 2019; see further, Steele, Ding, & Ross, 2019 for additional recent approaches to identifying biomarkers for addiction). Alternatively, Kwako, Bickel, and Goldman (2018) suggested a dimensional approach to biomarkers in terms of executive functions (inhibitory control, working memory, etc.), which includes measuring neuropsychological tests and epigenetic changes in relevant genes (e.g., COMT). With respect to statistical power, polysubstance use was coarsely defined, such that substances other than alcohol and tobacco had to be coded in a binary lifetime use variable. It is still possible that (heavy) use of a specific combinations of substances at the same time (e.g., cocaine and alcohol, Schulte et al., 2014) does have a negative impact, which did not emerge from our analysis here using binary variables. In addition, the total author response rate was low, which we discuss as a limitation. Currently, it remains an open question whether substance use is not associated with a motor inhibition impairment or if we were incapable of detecting such an impairment.

4.2. Demographics and Task Parameters

Our results indicate that age is a significant predictor of performance. In the GNG-task, the age-related increase in accuracy is most likely due to the strategic slowing of responses (confirmed by longer go RTs). In the SST, SSRT increased with age. Education was positively correlated with inhibition capability in both tasks. There was not a significant effect of sex on inhibition, nor any interactions between sex and substance use. In the GNG, higher working memory load, lower no-go percentages, and a higher number of experimental
trials resulted in more commission errors. These effects are in line with the primary literature on these tasks and are further discussed in Supplementary Materials S1. Somewhat surprisingly, we did not obtain an effect of substance-related GNG on performance measures compared to classical task versions. This is in line with a recent meta-analysis, where the main effect of appetitive cues was not observed after correction for publication bias, and where drinking status (light vs. heavy drinkers) also did not moderate this effect (Jones, Duckworth, Kersbergen, Clarke, & Field, 2018). In a small exploratory analysis, we examined the alcohol × substance-related task interaction effect, which was not a significant predictor of commission error rates in GNG (Supplementary Materials S6). Still, since our conclusion is based on only 5 out of 23 included studies, future research should address this question. In the SST, visual (vs. auditory) stop signals, fewer number of trials and fixed SSDs (vs. staircase-tracking procedure) induced prolonged SSRT (elaboration in Supplementary Materials S1).

4.3. **Implications**

Our results showed no relationship between the use of most substances and impaired response inhibition, except for a relationship between cannabis use and impaired inhibition, and in non-cannabis users an association between cigarette use and impaired inhibition. What are the theoretical implications? First, these findings could be of relevance for the current debate on the question whether addiction should be considered a chronic brain disease or not (Heather et al., 2017; Leshner, 1997; Field, 2015; Volkow, Koob, Mental, Parity, & Act, 2015). The current findings do not support the idea that long-term recreational substance use leads to irreparable problems in inhibition, although it cannot be excluded that inhibition problems are present in (a subgroup of) people diagnosed with SUD. Second, in many dual process models of addiction, suboptimal inhibition of stimulus-driven appetitive processes (cue-reactivity) plays an important role in the escalation of use (e.g., Baler & Volkow, 2006;
Wiers et al., 2007). An alternative perspective does not emphasize the competition between stimulus-driven and goal-directed processes, but rather between different goal-directed processes (Moors, Boddez, & de Houwer, 2017). Individuals learn to mobilize and allocate resources strategically according to goal saliency and importance (Köpetz, Lejuez, Wiers, & Kruglanski, 2013). In this way, the inhibition capability of substance users is expected to fluctuate moment-to-moment (i.e., state-like) based on the external and internal context. Note again that the current findings do not exclude the possibility that in severe addiction(s), chronic inhibition problems of stimulus-driven processes do play a role. It merely underscores the goal-directed nature of (heavy) substance use. Third, impaired response inhibition as an immediate consequence of substance consumption may be more important than general inhibitory impairments in the long term. Compared with long-term (non-dependent) substance use, acute use is more consistently related to impaired inhibitory control that enhances further consumption (Gan et al., 2014).

4.4. Limitations and Suggestions for Future Study

There are several limitations of the current study worth considering. First, the response rate was rather low. Although more than 100 studies met our inclusion and exclusion criteria, authors of only 65 studies provided raw data. The reasons for this include inaccessibility of the data, data could not be shared due to regulations, and a lack of success in contacting the authors. The low response rate is an obstacle encountered commonly in mega-analyses (Riley et al., 2010, 2007). We calculated and compared the effect sizes of studies that were included, studies that provided data but that were not included, and studies did not provide data. It was found that these three kinds of studies did not differ significantly on effect size (Fig. S4, see statistics in Supplementary Materials S9). In light of this, an open science framework is recommended in order to increase the transparency and availability of data for future research. Despite these obstacles, we received raw data from 3610 participants,
which should provide sufficient power to test effects on inhibition of substance use. Second, and relatedly, we noticed that the original studies did not score the use of every substance, for example, data on opiates were scarce. Although we tried to remedy this by means of multiple imputations, the analyses on the effects of these substances might have been underpowered. Third, except for alcohol and tobacco use, other substances could only be coded as a binary ‘lifetime use’ variable. It would be optimal if a standard way of assessing all substances could be used in the future when assessing the relationship between substance use and inhibition (or other neuropsychological functions). Guidelines for experimental protocols and assessment of substance use would facilitate future multicenter comparisons, which could be stimulated by funding agencies requiring a standard assessment of all commonly used substances in a uniform format. Fourth, studies did not focus on poly-substance use. Studies recruited individuals taking one substance and recorded one/several other substances. Therefore, the samples are highly selective and not representative of poly-substance users. In addition, future studies are suggested to include a standard index of trait impulsivity (e.g., Eysenck’s personality inventory, Eysenck & Eysenck, 1965; BIS-11, Patton, Stanford, & Barratt, 1995) as it is possible that within-sample variability on this dimension is obscuring common effects of drug exposure, or has stand-alone effects, especially for stimulant users (Ersche et al., 2012). Last, the effects of age and education years should be considered in the analysis and explanation of results. Task characteristics like stop trial percentage that consistently influence task performance should also be considered when comparing across studies.

5. Conclusions

The current mega-analysis aggregated raw data from 3610 participants in 43 studies on long-term (mostly) light to moderate substance use and response inhibition. The main finding is that limited evidence was found for impaired response inhibition in substance users, with two exceptions: lifetime cannabis use, and cigarette smoking in people who do not use
cannabis. The validity of these findings is underscored by expected findings for demographics (e.g., age, education level) and task characteristics (e.g., stop percentage). Broad assessment, standardized recording and reporting of substance use are highly needed in future studies.
Acknowledgments

We thank Janneke Staaks at the library of the University of Amsterdam for providing support for the literature search. We thank Dr. Jorien Treur for her feedback on the manuscript. We thank Lauren Kuhns for her proofreading.

We thank Dr. Rebecca L. Ashare, Dr. Elliot T. Berkman, Dr. Craig R. Colder, Dr. Pike Erika, Dr. Mark T. Fillmore, Dr. Rual Gonzalez, Dr. Bernice Porjesz, Dr. Olga Rass, and Dr. Craig R. Rush who contributed raw data without co-authorship.

Yang Liu thanks the China scholarship council (CSC) (No. 201506990019) for fellowship support.

HMH is supported by a VICI grant awarded by the Netherlands Organization of Scientific Research (NWO) [grant number 453-12-005]

MY was supported by a National Health and Medical Research Council of Australia Fellowship (#APP1117188) and the David Winston Turner Endowment Fund.

Funding:

China scholarship council (CSC) (No. 201506990019)

VICI grant awarded by the Netherlands Organization of Scientific Research (NWO) [grant number 453-12-005]

Declarations of interest: None

Supplementary Materials: see attachment
Reference

References marked with an asterisk indicate studies included in the mega-analysis


https://doi.org/10.1093/jama/9780195176339.022.529


https://doi.org/10.1016/j.bbr.2014.08.039


http://doi.org/https://doi.org/10.1016/j.drugalcdep.2015.02.015


http://doi.org/10.1016/j.molmed.2006.10.005


https://doi.org/10.1016/j.pbb.2017.09.009


*Bø, R., Aker, M., Billieux, J., & Landrø, N. I. (2016). Binge drinkers are fast, able to stop—but they fail to adjust. Journal of the International Neuropsychological Society, 22(1), 38–46. https://doi.org/10.1017/s1355617715001204


[https://doi.org/10.1159/000367939](https://doi.org/10.1159/000367939)


state connectivity best predicts alcohol use severity in moderate to heavy alcohol users.

*NeuroImage: Clinical, 22*(October 2018), 101782.

https://doi.org/10.1016/j.nicl.2019.101782


https://doi.org/10.1016/j.neubiorev.2010.04.008


*Hendershot, C. S., Wardell, J. D., Strang, N. M., Markovich, M. S. D., Claus, E. D., &


*Kamarajan, C., Porjesz, B., Jones, K. A., Choi, K., Chorlian, D. B., Padmanabhapillai,
https://doi.org/10.1016/j.biopsycho.2004.08.004


https://doi.org/10.2307/2332226


https://doi.org/10.1016/j.ijpsycho.2014.08.001


https://doi.org/10.1111/j.1369-1600.2011.00418.x

https://doi.org/10.1016/b978-0-12-813838-0.00006-6


http://doi.org/10.1007/s00213-007-0823-y


Volkow, N. D., Fowler, J. S., Wang, G.-J., & Swanson, J. M. (2004). Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. *Molecular Psychiatry, 9*(6), 557. [https://doi.org/10.1038/sj.mp.4001507](https://doi.org/10.1038/sj.mp.4001507)


Figure captions
**Figure 1:** PRISMA for the mega-analysis detailing our search and selection decisions.

**Figure 2:** The interaction between cannabis and tobacco use on SSRT. Only for cannabis non-users, the more tobacco a person smoked on a daily basis, the longer his/her stopping latency. For cannabis users, a mild positive association was found between tobacco use and SSRT.
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size (reserved)</th>
<th>Age</th>
<th>Males</th>
<th>Education years</th>
<th>Main substance in the original paper</th>
<th>criteria for the heavy/problematic substance use group</th>
<th>Other substance use info provided</th>
<th>Trial number</th>
<th>No-go percentage</th>
<th>Substance related</th>
<th>Working memory load</th>
<th>Task complexity</th>
<th>Cue GNG</th>
<th>No-go commission error</th>
<th>Go RT</th>
<th>M (SD)</th>
<th>Main behavioral findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ames et al, (2014)</td>
<td>41</td>
<td>20.46 (1.27)</td>
<td>41</td>
<td>Missing</td>
<td>Alcohol</td>
<td>21 heavy drinker with AUDIT score&gt;5, binge drinker (once/week and 15 drinks/female 8/week)</td>
<td>83 participants were diagnosed with alcohol dependence according to DSM-5</td>
<td>200</td>
<td>20</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>10 (6.22)</td>
<td>459 (48)</td>
<td>There was no difference between light and heavy drinker on commission error rate, and mean go RT. There was no correlation between alcohol use disorder severity and inhibition performance</td>
<td></td>
</tr>
<tr>
<td>Claus et al, (2013)</td>
<td>144</td>
<td>32.64 (9.65)</td>
<td>69</td>
<td>14.2 (2.25)</td>
<td>Alcohol</td>
<td>All participants at least binged drink once in the past month</td>
<td>All participants were diagnosed with alcohol dependence according to DSM-5</td>
<td>624</td>
<td>6.41</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>59 (16.37)</td>
<td>335 (59)</td>
<td>Response inhibition was worsened following the rising limb of blood alcohol concentration (BAC), which pattern increased during BAC plateau. Only baseline data (without alcohol intake) were used in the current study.</td>
<td></td>
</tr>
<tr>
<td>Hendershot et al, (2015)*</td>
<td>83</td>
<td>19.86 (0.81)</td>
<td>48</td>
<td>12.99 (1.34)</td>
<td>Alcohol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kamarajan et al, (2005)</td>
<td>59</td>
<td>29.4 (7.14)</td>
<td>53</td>
<td>13.46 (2.89)</td>
<td>Alcohol</td>
<td>30 participants were alcoholics, patients according to SDM-5</td>
<td></td>
<td></td>
<td>10 (11.02)</td>
<td>297 (20)</td>
<td>There was no difference between alcoholics and controls in commission error rate and go RT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>N</td>
<td>Age (SD)</td>
<td>Gender</td>
<td>Substance</td>
<td>Description</td>
<td>Score</td>
<td>No.</td>
<td>No.</td>
<td>No.</td>
<td>No.</td>
<td>No.</td>
<td>z</td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>------</td>
<td>----------</td>
<td>--------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kreusch et al, (2014)</td>
<td>30</td>
<td>21.47 (3.01)</td>
<td>47</td>
<td>Alcohol</td>
<td>15 heavy drinkers with AUDIT &gt;11</td>
<td>100</td>
<td>25</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>4.45</td>
<td>0.0336</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Littel et al, (2012)</td>
<td>56</td>
<td>21.91 (4.17)</td>
<td>61</td>
<td>Game</td>
<td>25 excessive gamers had a Videogame Addiction Test (VAT) score&gt;2.5</td>
<td>636</td>
<td>11.6</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>4.19</td>
<td>0.0371</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>López-Canedo et al, (2014)</td>
<td>57</td>
<td>18.74 (0.55)</td>
<td>46</td>
<td>Alcohol</td>
<td>Binge drinkers binge drink at least once a week OR binge drink once a month with at least three drinks per hour for at least two years.</td>
<td>150</td>
<td>50</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>4.06</td>
<td>0.0416</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luijten et al, (2011)</td>
<td>78</td>
<td>21.46 (2.05)</td>
<td>72</td>
<td>Tobacco</td>
<td>Smokers smoked at least 10 cigarettes per day for at least two years.</td>
<td>196</td>
<td>25</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>3.01</td>
<td>0.0026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luijten, O’Connor et al, (2013)</td>
<td>32</td>
<td>25.25 (5.21)</td>
<td>63</td>
<td>Tobacco</td>
<td>Smokers smoked at least 15 cigarettes per day for at least two years.</td>
<td>160</td>
<td>12.5</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>2.46</td>
<td>0.0139</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the letter GNG task, heavy drinkers made more commission errors than light drinkers, while no difference on go RT.

Excessive gamers made more commission errors than controls.

There was no difference between binge drinkers and controls in go RT and commission error rate.

Smokers made more commission errors than controls, while there was no correlation between daily cigarette consumption and commission error rate. And there was no group difference of go RT.

Smokers did not differ from controls in commission error rate and go RT.
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Mean Age (SD)</th>
<th>Prevalence</th>
<th>Substance Use</th>
<th>Commission Errors</th>
<th>Reaction Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luijten, Veltman et al, (2013)</td>
<td>48</td>
<td>22.17 (2.42)</td>
<td>67</td>
<td>Tobacco</td>
<td>Smokers smoked at least 15 cigarettes per day for at least three years.</td>
<td>Cannabis, cocaine, amphetamine, ecstasy, hallucinogens</td>
</tr>
<tr>
<td>Luijten et al, (2015)</td>
<td>16</td>
<td>21.38 (3.03)</td>
<td>100</td>
<td>Gamer</td>
<td>Problem gamers scored more than 2.5 on VAT.</td>
<td>Cannabis, cocaine, amphetamine, ecstasy, hallucinogens</td>
</tr>
<tr>
<td>Mahmood et al. (2013)</td>
<td>36</td>
<td>18.64 (0.34)</td>
<td>72</td>
<td>No specific</td>
<td>High frequency substance users had any drug use over 180 occasions.</td>
<td>Cannabis, cocaine, amphetamine, ecstasy, hallucinogens</td>
</tr>
<tr>
<td>Pett et al, (2012)</td>
<td>35</td>
<td>21.29 (1.98)</td>
<td>51</td>
<td>Alcohol</td>
<td>Heavy social drinkers had an average 20 drinks per week, and with AUDIT&gt;11.</td>
<td></td>
</tr>
<tr>
<td>Paz et al, (2018)</td>
<td>203</td>
<td>21.06 (1.87)</td>
<td>48</td>
<td>No specific</td>
<td>Binge drink was assessed with the last three items of alcohol use questionnaire (AUD).</td>
<td>Cannabis, cocaine, ecstasy</td>
</tr>
</tbody>
</table>

Smokers made more commission errors and also had longer go RT compared with non-smokers.

Problem gamers made more commission errors than controls, while there was no group difference in go RT.

There was no difference in commission error rate between high and low-frequency substance users.

Heavy drinkers made more commission error than light drinkers when the background picture is alcohol-related.

The correlation between the commission error rate and binge score was not reported.
There was no control group and all participants reported cocaine use for the past month. Cannabis, amphetamine, hallucinogens

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Age Mean (SD)</th>
<th>Duration</th>
<th>Control</th>
<th>Cocaine</th>
<th>Cocaine Use &amp; Severity</th>
<th>Amphetamine</th>
<th>Hallucinogens</th>
<th>Ecstasy Use &amp; Severity</th>
<th>Smokers</th>
<th>Ecstasy Use &amp; Severity</th>
<th>Tobacco Use &amp; Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pike et al. (2015)</td>
<td>91</td>
<td>39.09 (8.28)</td>
<td>64 days</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>10 (12.13)</td>
<td>Yes</td>
<td>25 (12.25)</td>
</tr>
<tr>
<td>Quassnow et al. (2007)</td>
<td>51</td>
<td>24.29 (4.75)</td>
<td>100 days</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>25 (12.35)</td>
<td>No</td>
<td>25 (12.25)</td>
</tr>
<tr>
<td>Rass et al. (2014)</td>
<td>82</td>
<td>25.29 (3.36)</td>
<td>48 days</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>25 (12.25)</td>
<td>Yes</td>
<td>25 (12.25)</td>
</tr>
<tr>
<td>Roberts et al. (2010)</td>
<td>39</td>
<td>22.38 (2.93)</td>
<td>51 days</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>45 (17.51)</td>
<td>Yes</td>
<td>316(43)</td>
</tr>
<tr>
<td>Author(s) et al.</td>
<td>Year</td>
<td>Sample Size</td>
<td>Mean Age</td>
<td>Standard Deviation</td>
<td>Reason for Inclusion</td>
<td>Ecstasy &amp; Poly</td>
<td>Cannabis, Cocaine, Amphetamine, Ecstasy, Hallucinogens</td>
<td>D-Prime Scores</td>
<td>Opiate Dependents</td>
<td>Opiate Dependents’ Average SDS Score</td>
<td>Controls</td>
<td>Controls vs. Opiate Dependents</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------</td>
<td>----------------------</td>
<td>----------------</td>
<td>------------------------------------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>----------------------------------</td>
<td>---------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Roberts et al.</td>
<td>2013</td>
<td>59</td>
<td>23.26</td>
<td>2.99</td>
<td>Missing</td>
<td>Ecstasy group needs to take ecstasy for at least five occasions.</td>
<td>Cannabis, cocaine</td>
<td>240</td>
<td>25</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rossiter et al.</td>
<td>2012</td>
<td>124</td>
<td>26.43</td>
<td>6.79</td>
<td>Alcohol</td>
<td>The harmful alcohol use group had an AUDIT score no less than 16.</td>
<td>Alcohol,</td>
<td>160</td>
<td>12.5</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Takagi et al.</td>
<td>2011, 2014</td>
<td>30</td>
<td>20.49</td>
<td>1.48</td>
<td>Inhalant &amp; cannabis</td>
<td>Inhalant users had inhalants daily or almost daily use for more than 12 months.</td>
<td>Inhalant, cocaine, amphetamine, ecstasy</td>
<td>300</td>
<td>10</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Verdejo-Garcia et al.</td>
<td>2012</td>
<td>19</td>
<td>28.68</td>
<td>7.92</td>
<td>Opiate</td>
<td>Opiate dependents had an average score on SDS (Severity of Dependence Scale) of 8.3.</td>
<td>Opiate,</td>
<td>300</td>
<td>23.33</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Wetherell et al.</td>
<td>2013</td>
<td>18</td>
<td>19.40</td>
<td>0.99</td>
<td>Alcohol</td>
<td>Heavy drinkers at least had 4 drinks per occasion, less than once per month but more than once per year.</td>
<td>Alcohol,</td>
<td>180</td>
<td>32</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[^1]: For alcohol use, the harmful alcohol use group had an AUDIT score no less than 16. For cannabis, cocaine, amphetamine, ecstasy, and hallucinogens, the inhalant group had lower d-prime scores compared with controls.
[^2]: Controls made more commission errors compared with opiate dependents.
Note: go RT = correct go trials reaction time; M = mean; SD = standard deviation.

*Unpublished dataset at time of searching literature

Why comparison between substance users and controls could not be obtained from the original paper

1 interested in the difference between the increasing and decreasing limb of BAC but we only used baseline data when participants were sober

2 the correlation between commission error rate and binge score was not reported

3 focused on the experimental effect (different kinds of cued GNG) instead of the individual difference
Table 2
Description of included SST studies (dependent variable is SSRT)

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size (reserved)</th>
<th>Age</th>
<th>Male</th>
<th>Education years</th>
<th>Main substance in the original paper</th>
<th>Criteria for the heavy/problematic substance use group</th>
<th>Other substance use info provided</th>
<th>Task parameters</th>
<th>Dependent variables</th>
<th>Number of cases excluded (including the whole group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bidwell et al. (2013)</td>
<td>150</td>
<td>21.56 (3.16)</td>
<td>64</td>
<td>Missing</td>
<td>Cannabis</td>
<td>All participants used marijuana at least once a week in the past month and at least 10 times in the past 6 months.</td>
<td>Missing</td>
<td>Auditory</td>
<td>SSRT</td>
<td>112 (66)</td>
</tr>
<tr>
<td>Bø et al. (2016)</td>
<td>119</td>
<td>21.71 (2.12)</td>
<td>5</td>
<td>14.95 (1.56)</td>
<td>Alcohol</td>
<td>All participants use alcohol on a regular basis, binge score was calculated based on the last three items of the Alcohol Use Questionnaire.</td>
<td>Missing</td>
<td>Auditory</td>
<td>SSRT</td>
<td>104 (54)</td>
</tr>
<tr>
<td>Bø et al. (2017)*</td>
<td>186</td>
<td>36.22 (12.8)</td>
<td>32</td>
<td>16.45 (2.7)</td>
<td>Depression</td>
<td>No special requirement for substance use Cannabis, cocaine</td>
<td>Missing</td>
<td>Auditory</td>
<td>SSRT</td>
<td>187 (50)</td>
</tr>
<tr>
<td>Colzato et al. (2007)</td>
<td>24</td>
<td>29.33</td>
<td>83</td>
<td>Missing</td>
<td>Cocaine</td>
<td>Recreational cocaine users should consume cocaine 1 to 4 grams per month by snorting route for a minimum of two years</td>
<td>Missing</td>
<td>Visual</td>
<td>SSRT</td>
<td>24 (27)</td>
</tr>
<tr>
<td>Courtney et al. (2012, 2013)</td>
<td>304</td>
<td>37.15 (10.81)</td>
<td>7</td>
<td>13.29 (3.28)</td>
<td>Alcohol</td>
<td>All participants were problem drinkers, with a maximum of 46 standard drinks per month</td>
<td>64</td>
<td>Auditory</td>
<td>SSRT</td>
<td>241 (90)</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Mean Age (SD)</td>
<td>Gender</td>
<td>Diagnosis/Condition</td>
<td>Measure</td>
<td>SSRT Mean (SD)</td>
<td>Go RT Mean (SD)</td>
<td>Other Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>--------</td>
<td>---------------------</td>
<td>---------</td>
<td>----------------</td>
<td>-----------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Ruiter et al. (2012)</td>
<td>35</td>
<td>34.2 (9.25)</td>
<td>1</td>
<td>Problem gamblers and heavy smokers</td>
<td>Gambling &amp; Tobacco</td>
<td>Problem gamblers were diagnosed by SDM-5. Heavy smokers smoked at least 15 cigarettes per day.</td>
<td>360 (32)</td>
<td>Visual</td>
<td>Staircase</td>
<td>Other</td>
</tr>
<tr>
<td>Filbey et al. (2013)</td>
<td>74</td>
<td>24.14 (7.2)</td>
<td>74</td>
<td>Problem gamblers, heavy smokers, and controls did not differ in SSRT and go RT</td>
<td>Cannabis</td>
<td>All participants were cannabis users with at least 4 uses per week for at least 6 months prior. Among them, 44 were diagnosed with cannabis dependents according to SDM-5.</td>
<td>384 (25)</td>
<td>Auditory</td>
<td>Staircase</td>
<td>Integration</td>
</tr>
<tr>
<td>Fillmore et al. (2002)</td>
<td>44</td>
<td>40.27 (6.66)</td>
<td>61</td>
<td>Cocaine users</td>
<td>Cocaine</td>
<td>Participants in the cocaine use group need to score ≥4 on the Drug and Abuse Screening Test (DAST), habitual cocaine use for a minimum of 6 months and past week cocaine use.</td>
<td>176 (27)</td>
<td>Auditory</td>
<td>Fixed</td>
<td>Integration</td>
</tr>
<tr>
<td>Galván et al. (2011)</td>
<td>59</td>
<td>19.49 (1.3)</td>
<td>61</td>
<td>Daily smokers</td>
<td>Tobacco</td>
<td>Daily smokers should smoke daily for at least 6 months.</td>
<td>256 (25)</td>
<td>Auditory</td>
<td>Staircase</td>
<td>Integration</td>
</tr>
</tbody>
</table>

Visual Staircase
Auditory Staircase
Integration
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Age (Mean ± SD)</th>
<th>Gender Ratio</th>
<th>Diagnoses</th>
<th>Outcome Measures</th>
<th>Other</th>
<th>SSRT/Go RT</th>
<th>Alcoholism Severity</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass et al. (2009)</td>
<td>495</td>
<td>44.1 (4.97)</td>
<td>47</td>
<td>Al &amp; To</td>
<td>Auditory Staircase</td>
<td>Other</td>
<td>13.9 (2.27)</td>
<td>44.1 (4.97)</td>
<td>Both SSRT and go RT had a significant negative correlation with alcoholism severity.</td>
</tr>
<tr>
<td>Karoly et al. (2014)</td>
<td>53</td>
<td>28.3 (6.91)</td>
<td>47</td>
<td>Al</td>
<td>Auditory Staircase</td>
<td>Integration</td>
<td>15.55 (1.85)</td>
<td>28.3 (6.91)</td>
<td>The relationship between SSRT/go RT and alcohol use was not reported in the paper.</td>
</tr>
<tr>
<td>Kräplin et al. (2015)</td>
<td>75</td>
<td>26 (7.92)</td>
<td>39</td>
<td>Gam &amp; To</td>
<td>Visual Staircase</td>
<td>Integration</td>
<td>11.74 (0.76)</td>
<td>26 (7.92)</td>
<td>PG lead to prolonged SSRT compared with controls. There is no difference between PG and ND, ND and PG comorbid ND with regard to SSRT.</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Age Mean (SD)</td>
<td>Gender</td>
<td>Smoking Status</td>
<td>CONCENTRATION</td>
<td>-task</td>
<td>TASK</td>
<td>Note</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>--------</td>
<td>----------------</td>
<td>---------------</td>
<td>------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Moallem et al. (2012)</td>
<td>287</td>
<td>30.97 (10.61)</td>
<td>73</td>
<td>Smokers</td>
<td>Alcohol &amp; Tobacco</td>
<td>14.68 (2.59)</td>
<td>64</td>
<td>25</td>
<td>Auditory Staircase Other 223(88) 509(90) Smokers should smoke cigarettes no less than 10 per day and had less than 3 months' smoking abstinence in the past year. Heavy drinkers were defined by National Institute on Alcohol Abuse and Alcoholism (NIAAA), i.e. drinks per week &gt;14 (women &gt; 7) or drinks per occasion ≥ 5 (≥ 4 for women) at least once per month over the past year.</td>
</tr>
<tr>
<td>Papachristou et al. (2012a)</td>
<td>42</td>
<td>25.5 (9.66)</td>
<td>24</td>
<td>Missing</td>
<td>Alcohol</td>
<td></td>
<td>256</td>
<td>25</td>
<td>Auditory Staircase Other 222(50) 344(63) All participants were light to moderate social drinkers with an average AUDIT score of 7.7. Heavy and light social drinkers were classified by the cut-off score of 11 of AUDIT.</td>
</tr>
<tr>
<td>Papachristou et al. (2012b)</td>
<td>75</td>
<td>23.29 (5.2)</td>
<td>33</td>
<td>Missing</td>
<td>Alcohol</td>
<td></td>
<td>256</td>
<td>25</td>
<td>Auditory Staircase Other 203(32) Missing Heavy drinkers had similar SSRT. Heavy and light social drinkers were classified by the cut-off score of 11 of AUDIT.</td>
</tr>
<tr>
<td>Paz et al. (2018)</td>
<td>182</td>
<td>21.15 (1.83)</td>
<td>49</td>
<td>Not specific</td>
<td>Cannabis, cocaine, ecstasy</td>
<td></td>
<td>256</td>
<td>25</td>
<td>Auditory Staircase Integration 227(47) 694(175) Binge drink was assessed with the last three items of alcohol use questionnaire (AUQ). The relationship between SSRT and binge score was not reported.</td>
</tr>
<tr>
<td>Tsaur et al. (2015)</td>
<td>21</td>
<td>34.73 (12.47)</td>
<td>76</td>
<td>Tobacco</td>
<td>Tobacco</td>
<td>15.9(1.18)</td>
<td>192</td>
<td>25</td>
<td>Auditory Staircase Other 252(52) 560(112) Only baseline data was used. The relationship between SSRT and binge score was not reported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>
Cocaine dependence was diagnosed with DSM-IV. All cocaine users should have cocaine as the primary used illegal drug, cocaine use of >0.5 g per month, and abstinence duration of ≤6 months.

Pathological gambling (PG) was diagnosed with ICD-5 and a score ≥5 on the SOGS (South Oaks Gambling Screen).

<table>
<thead>
<tr>
<th>Study</th>
<th>Group 1</th>
<th>Group 2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vonmoos et al. (2013)</td>
<td>163</td>
<td>30.03 (8.18)</td>
<td>0.03</td>
</tr>
<tr>
<td>Cocaine</td>
<td>71</td>
<td>10.45 (1.74)</td>
<td></td>
</tr>
</tbody>
</table>

| Zack et al. (2015)             | 12      | 33.75 (11.23) | 1       |
| Gambling                       | 1      | 15.92 (0.52)  |

Note: DV: dependent variable; SSD = stop-signal delay; SSRT = stop-signal reaction time; go RT = correct go trials reaction time; M = mean; SD = standard deviation.

Unpublished dataset at time of searching literature

Why comparison between substance users and controls could not be obtained from the original paper

1. did regression analysis
2. only reported MRI results
3. focused on experimental effect rather than individual difference with a within-subject design
4. the correlation between commission error rate and binge score was not reported
5. longitudinal study along substance abstinence, only baseline data were used

Two cocaine use group (recreational users and dependent users) and the control group had similar SSRT and go RT.
### Table 3a

Quality assessment scores of included GNG studies according to the NHLBI Quality Assessment Tool

<table>
<thead>
<tr>
<th>Study</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
<th>Q12</th>
<th>Q13</th>
<th>Q14</th>
<th>Quality Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ames et al. (2014)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Claus et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Hendershot et al. (2015)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Kamarajan et al. (2005)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
<td></td>
</tr>
<tr>
<td>Kreusch et al. (2014)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Littel et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>López-Caneda et al. (2014)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Luitjen et al. (2011)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
<td></td>
</tr>
<tr>
<td>Luitjen, O’Connor et al. (2013)</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>CD</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Luitjen, Veltman et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>CD</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Luitjen et al. (2015)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Mahmood et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Petit et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Paz et al. (2018)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>no</td>
<td>fair</td>
</tr>
<tr>
<td>Pike et al. (2015)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Quednow et al. (2007)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Rass et al. (2014)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Roberts et al. (2010)</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Roberts et al. (2013)</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>suboptimal</td>
</tr>
<tr>
<td>Rossiter et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
</tr>
<tr>
<td>Takagi et al. (2011)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
</tr>
<tr>
<td>Takagi et al. (2014)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>no</td>
</tr>
<tr>
<td>Verdejo-Garcia et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Wetherill et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
</tbody>
</table>

Note: CD: cannot determine; NA: not applicable; NR: not reported; Meanings of criteria Q1-Q14 can be found in Table S2.
### Table 3b

**Quality assessment scores of included SST studies according to the NHLBI Quality Assessment Tool**

<table>
<thead>
<tr>
<th>Study</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
<th>Q12</th>
<th>Q13</th>
<th>Q14</th>
<th>Quality Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bidwell et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Bø et al. (2016)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Bø et al. (2017)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Colzato et al. (2007)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Courtney et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Courtney et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>de Ruiter et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Filbey et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Fillmore et al. (2002)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Galván et al. (2011)</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Glass et al. (2009)</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
<td></td>
</tr>
<tr>
<td>Karoly et al. (2014)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>NA</td>
<td>no</td>
<td>fair</td>
</tr>
<tr>
<td>Kräplin et al. (2015)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Moallem et al. (2012)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Papachristou et al. (2012a)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
</tr>
<tr>
<td>Papachristou et al. (2012b)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Par et al. (2018)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>no</td>
<td>fair</td>
</tr>
<tr>
<td>Tsaur et al. (2015)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>CD</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>yes</td>
<td>fair</td>
</tr>
<tr>
<td>Vonmoos et al. (2013)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>CD</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>good</td>
</tr>
<tr>
<td>Zack et al. (2015)</td>
<td>yes</td>
<td>yes</td>
<td>NR</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>NR</td>
<td>NA</td>
<td>yes</td>
<td>fair</td>
</tr>
</tbody>
</table>

Note: CD: cannot determine; NA: not applicable; NR: not reported; Meanings of criteria Q1-Q14 can be found in Table S2.