Title

A methodological checklist for fMRI drug cue reactivity studies: development and expert consensus

Authors

Ekhtiari, Hamed
Zare-Bidoky, Mehran
Sangchooli, Arshiya
et al.

A methodological checklist for fMRI drug cue reactivity studies: development and expert consensus

A full list of authors and affiliations appears at the end of the article.

Abstract

Cue reactivity is one of the most frequently used paradigms in functional magnetic resonance imaging (fMRI) studies of substance use disorders (SUDs). Although there have been promising results elucidating the neurocognitive mechanisms of SUDs and SUD treatments, the interpretability and reproducibility of these studies is limited by incomplete reporting of participants' characteristics, task design, craving assessment, scanning preparation and analysis decisions in fMRI drug cue reactivity (FDCR) experiments. This hampers clinical translation, not least because systematic review and meta-analysis of published work are difficult. This consensus paper and Delphi study aims to outline the important methodological aspects of FDCR research, present structured recommendations for more comprehensive methods reporting and review the FDCR literature to assess the reporting of items that are deemed important. Forty-five

Related links
Key references using this review
Janes, A. C. et al. *Neuropsychopharmacology* 44, 1631–1638 (2019): https://doi.org/10.1038/s41386-019-0403-y
Zelle, S. L. et al. *Neuroimage* 151, 24–32 (2017): https://doi.org/10.1016/j.neuroimage.2016.03.015
MacNiven, K. H. et al. *JAMA Netw. Open* 1, e186466 (2018): https://doi.org/10.1001/jamanetworkopen.2018.6466
Yuan, K. et al. *Hum. Brain Mapp.* 38, 4644–4656 (2017): https://doi.org/10.1002/hbm.23690
Related content
Lang, P. J. et al. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual* (NIMH, 2005)
Mackey, S. et al. in *Progress in Brain Research* Ch. 10, 203–223 (Elsevier, 2016): https://doi.org/10.1016/bs.pbr.2015.07.026
Nichols, T. E. et al. *Nat. Neurosci.* 20, 299–303 (2017): https://doi.org/10.1038/nn.4500
Poldrack, R. A. et al. *Neuroimage* 40, 409–414 (2008): https://doi.org/10.25490/a97f-egyk
Hamilton, C. M. et al. *Am. J. Epidemiol.* 174, 253–260 (2011): https://doi.org/10.1093/aje/kwr193

FDCR scientists from around the world participated in this study. First, an initial checklist of items deemed important in FDCR studies was developed by several members of the Enhanced NeuroImaging Genetics through Meta-Analyses (ENIGMA) Addiction working group on the basis of a systematic review. Using a modified Delphi consensus method, all experts were asked to comment on, revise or add items to the initial checklist, and then to rate the importance of each item in subsequent rounds. The reporting status of the items in the final checklist was investigated in 108 recently published FDCR studies identified through a systematic review. By the final round, 38 items reached the consensus threshold and were classified under seven major categories: 'Participants' Characteristics', 'General fMRI Information', 'General Task Information', 'Cue Information', 'Craving Assessment Inside Scanner', 'Craving Assessment Outside Scanner' and 'Pre- and Post-Scanning Considerations'. The review of the 108 FDCR papers revealed significant gaps in the reporting of the items considered important by the experts. For instance, whereas items in the 'General fMRI Information' category were reported in 90.5% of the reviewed papers, items in the 'Pre- and Post-Scanning Considerations' category were reported by only 44.7% of reviewed FDCR studies. Considering the notable and sometimes unexpected gaps in the reporting of items deemed to be important by experts in any FDCR study, the protocols could benefit from the adoption of reporting standards. This checklist, a living document to be updated as the field and its methods advance, can help improve experimental design, reporting and the widespread understanding of the FDCR protocols. This checklist can also provide a sample for developing consensus statements for protocols in other areas of task-based fMRI.

Substance-use disorders (SUDs) affect hundreds of millions of individuals and are responsible for a substantial global burden of disease[1]. To improve translational research, as well as treatment and prevention, researchers and clinicians need a better understanding of the underlying neurocognitive mechanisms of SUDs[2]. There is also a need for better brain-based biomarkers to study the course and treatment response in SUDs[3]. A powerful method for investigating brain function among people with SUDs is task-based functional magnetic resonance imaging (fMRI) of drug cue reactivity (FDCR) paradigms[4]. In FDCR studies, subjects are exposed to drug-associated cues in one or more sensory modalities while undergoing fMRI. fMRI cue-reactivity paradigms are popular among researchers, and on the basis of a systematic review, 370 published studies (through April 30, 2021) have used this paradigm (based on a database available at ref. [5]). The results of these studies can help in understanding the neurobiology of SUDs, diagnostic classification of people with SUDs, discovering intervention targets, understanding the temporal evolution of the disease process, and monitoring the effectiveness of treatments and treatment outcomes; for more details, see refs. [6–8]. An overview of typical procedures in an FDCR study is presented in Fig. 1.

Despite the promising results of FDCR studies, the field has been plagued by important limitations. Most studies are cross-sectional[5] rather than longitudinal, which means that it is difficult to get information about cue-induced circuitry changes associated with the many factors that influence drug cue reactivity. In common with other fMRI research, the FDCR literature also suffers from small sample sizes and insufficient power[9,10]. All fMRI experiments can be influenced by random noise that affects study results[11]. It has also been suggested that the low reproducibility of task-based fMRI studies, in general[12],

might be due to a combination of methodological factors, which, if addressed, could improve reproducibility[13,14]. Issues complicating the picture are the sheer methodological complexity of FDCR and researcher discretion in the specification of hypotheses, participant recruitment, FDCR task design, choice of fMRI hardware, analysis pipelines and more. Unless these choices are explicitly and consistently reported across studies, unknown methodological heterogeneities can limit rigor and reproducibility. In turn, this will hinder knowledge production and clinical translation by undermining generalizability and the ability to optimally conduct comparative reports and meta-analyses[7].

There are many sources of potentially significant methodological heterogeneity that probably affect FDCR results, including participants' characteristics, types of cues, durations of cue exposure and analysis methods, such that the field would benefit from the establishment of best/standardized practices for methods reporting to inform the generalizability of specific FDCR study outcomes and guide future research.

There are multiple ways to achieve greater clarity, interpretability and replicability across FDCR studies. They include the following:

1. *Preregistered replicable protocols*. Study protocols define the structure of a study and can include the sequence of different imaging sessions, data acquisition settings and other methodological details[15,16].

2. *Published drug cue databases*. Drug cues in FDCR studies can be validated and standardized in terms of their average effects on arousal and valence, including affect and craving, and activations in relevant brain areas/networks. They can also be matched to control stimuli in multiple respects. One way of achieving this goal would be the sharing and utilization of standardized cue databases[17–21]. For example, the first openly accessible database with 360 cues is a recently validated methamphetamine and opioid cue database[19].

3. *Data-analysis guides and pre-registered and standardized analysis pipelines*. Preprocessing and analysis pipelines have significant effects on fMRI study results[22]. Researchers can use credible recommendations (e.g., by the Committee on Best Practice in Data Analysis and Sharing (COBIDAS)[23]). Pre-registration and open sharing of pipelines would also help in this regard, and moving towards consistent software and toolboxes is recommended[24].

4. *Extant checklists*. Many itemized checklists and recommendations have been developed to address different elements of research design and reporting in fMRI studies in general, with differing degrees of specificity (e.g., see refs. [25–35]). Regarding fMRI analysis specifically, the COBIDAS proposes a checklist with the goal of enhancing the reporting of MRI studies[23]. However, no checklist with clear recommendations for FDCR research design and reporting exists.

Most authoritative research checklists and guidelines represent consortium efforts. This expert consensus development helps to elucidate the research process and its various aspects and clarify opinion on the importance of these aspects. Furthermore, consortium involvement substantiates the claim of the checklist to represent a diversity of opinions in

the field[36]. One of the most common methods of achieving expert consensus is the Delphi technique. In the Delphi process, experts in the field approach consensus on a matter by participating in a series of commenting and/or item rating rounds with feedback[37]. An example of the use of this method in addiction sciences is a 2019 study to determine the significance of Research Domain Criteria (RDoC) in addiction medicine[38].

The purpose of the present study was to develop and validate an itemized checklist of methodological parameters for FDCR researchers to use to clarify methods in future studies. The checklist would include items that are most important in study design and reporting to facilitate the interpretation of study results and data sharing, enable future meta-analyses, increase replicability and validity and improve the transparency of FDCR studies[37]. Using the Delphi consensus technique, we aimed to develop this checklist through an international consensus of FDCR experts. Furthermore, this paper represents the views of experts who participated in the Delphi process, exploring why and how various categories within the checklist affect FDCR research. It should be specifically noted that this checklist does not aim at prescribing the specific methods used in the design of FDCR studies. Instead, it is meant to help researchers explicitly consider and report various study design parameters that may importantly affect the results of their study, and report these methodological decisions when designing and reporting the results of FDCR research.

Methods

Scope of the checklist

The items included in the checklist were predominantly those identified as being methods parameters that are specific to FDCR studies, such as sensory modality of cues. This checklist was developed to act as a standalone tool for describing methodological details considered to influence results of FDCR studies. The authors also detailed additional recommendations for each item that should be considered to increase the quality of reporting. The checklist can be used to increase transparency, support replicability, improve quality of data acquisition, facilitate future data sharing between laboratories and make increasingly sophisticated meta-analyses possible.

Contributors

The contributions to this project were organized on two levels: a steering committee (SC) and a larger expert panel (EP). This method was chosen because it enables a small and collaborative group of leaders to flexibly and rapidly make decisions and resolve conflicts within the SC and lead the project to fruition. This approach also ensured that the voices of a much broader and more diverse group of international experts meaningfully affect the consensus process.

Steering committee—The SC consisted of 14 individuals: Anna Rose Childress, Hamed Ekhtiari, Rita Goldstein, Andreas Heinz, Amy Janes, Jane Joseph, Hedy Kober, F. Joseph McClernon, Martin Paulus, Lara Ray, Rajita Sinha, Elliot Stein, Reagan Wetherill and Anna Zilverstand. This group grew out of the Enhanced NeuroImaging Genetics through Meta-Analyses (ENIGMA) Addiction working group

(https://www.enigmaaddictionconsortium.com) after a series of meetings in which substantial heterogeneity in FDCR studies, poor reporting of methods (insufficient for replication) and disagreements over the importance of various methodological parameters were discussed along with strategies to amend the situation. These discussions led to formation of a group called ENIGMA Addiction Cue Reactivity Initiative (ACRI). Furthermore, the initial members of the SC were asked to identify additional members chosen on the basis of their scientific expertise and contributions to the FDCR literature.

The SC members outlined the scope of the Delphi project[39] and its important questions, developed and approved the initial checklist of important methodological parameters, processed the comments and revisions and led the authorship of this paper, all based on consensus.

Expert panel—The panel of experts for this Delphi study was chosen primarily on the basis of 318 addiction-related FDCR studies published by the end of 2019, from the database of a systematic review[5]. The main inclusion criteria were (i) appearing among the authors of at least four papers in the systematic review database and (ii) holding first, last or corresponding authorship position in at least one of the 318 papers. In addition, the members of the SC were asked to nominate candidates in the field of FDCR for inclusion within the EP. All SC members agreed on the list of experts before the invitation process.

All chosen experts received an email briefly outlining the importance, structure and goals of this Delphi study and were asked to state whether they wished to participate. To invite new participants, each candidate was contacted by email, and if there was no answer, two reminders were sent within roughly 2-week intervals. Those who decided to enroll received a further email with more details about how their feedback would be collected and used in the Delphi study, and then they formally entered the Delphi process. A total of 76 EP candidates were contacted by email, 21 did not respond to the email, 6 had incorrect email addresses, 4 explicitly declined to participate and 45 accepted to join the EP. Providing the study participants with information is not necessary for Delphi studies, which did not rely on explicit information or published data[37,40]. Therefore, in this study, participants were asked to primarily rely on their prior knowledge of FDCR task design and methodology during the Delphi process, although they were provided with the list of the 318 studies included in the aforementioned systematic review, so they could have viewed the relevant articles if needed.

Procedure

A general schematic of the methodology and its various stages is depicted in Fig. 2.

Checklist development phase—To simplify consensus development and facilitate the process of finalizing a comprehensive but concise list of important methodological aspects of FDCR studies, the SC decided to begin the feedback rounds after developing a basic set of categories, items and their associated recommendations. Each item included one concise point of an aspect in the category in which it appeared (the final list of categories and items are available in Tables 1–6 in Results). There could also be some additional recommendations associated with each item. This basic structure evolved on the basis of the initial feedback of the SC and a consideration of the methodological parameters commonly

observed to be important to the studies included in the aforementioned systematic review. Upon completion, the items in the checklist questionnaire were pilot-tested by rating five randomly selected FDCR papers with Yes/No ratings on whether the item was reported in the paper or not. Using data from the pilot-testing analysis, the SC reworded and/or combined items that could not be easily given a Yes/No rating for inclusion in the revision phase.

Checklist revision phase—In the revision phase, 45 EP and 14 SC members were sent the checklist and were asked to add comments and suggest revisions to the existing items and their associated additional recommendations. They were also asked to suggest new items that they feel were overlooked, along with an explanation of why they thought the item should be included. They also were informed that there was no limit to the number of new items they could suggest. 41 members of the EP responded. 10 SC members also added additional comments in this phase. Overall, we reached a response rate of 85% across all participants (EP and SC).

In this revision phase, members of the EP and SC answered a short questionnaire[41] assessing their basic demographic information (age, sex, highest academic degree, country of residence and primary affiliation/place of work), primary field of research (e.g., psychiatry, psychology, pharmacology, neuroscience, cognitive science), primary place of work (e.g., university, hospital, business, independent research institute), length of time spent in addiction medicine and length of time spent specifically researching FDCR. These questions were asked to ensure that we included a diverse field of experts (Supplementary Table 1).

Comments for each item were processed by the SC. During processing, repetitive comments were removed, items with unclear meaning were reworded and those outside the scope of the study were removed[42] so that a list of clear and unique single-point notes extracted from the comments was obtained.

The notes obtained after the processing of comments were of three kinds: first, proposed changes to an existing item or its associated recommendations; second, adding or removing items; and third, general changes or critiques regarding the checklist. The decisions to apply or reject each note were made by the SC.

The modified version was sent once more to the SC and EP, and the members were asked to comment on the new changes. After receiving and applying their comments, the final version was approved by the SC members.

Checklist rating phase—In the second round, participants from the SC and EP were sent the edited checklist along with the newly added items. The participants were asked to rate each item in terms of importance in the methodology of FDCR studies, from 1 to 5 (87.5% completed the entire survey). The exact question was: 'To facilitate visibility, replication and data sharing, how important is it to report this item?'. In addition, for each additional recommendation, we asked: 'Do you support the inclusion of this additional note

as a recommendation to be considered in fMRI drug cue reactivity studies?'. Out of 59 members of the SC and EP, 49 (83%) participated in the rating phase.

To avoid a non-neutral center rating and encourage deliberation, ratings were termed 'not important', 'slightly important', 'moderately important', 'highly important' and 'extremely important'. The participants were allowed not to rate an item if they chose not to do so. The inclusion of each additional recommendation for each item could be rated 'Yes' or 'No'.

Data analysis

All statistical analyses were conducted using RStudio (RStudio version 3.4.1). For the rating phase, the average rating and the number of responses were calculated. On the basis of the distribution of the ratings, it was calculated whether items passed either of two importance thresholds. The more-stringent threshold was a rating of 4 or 5 by 80% of participants (threshold 2, preregistered[43]), and the less-stringent threshold was a rating of 3 by 70% of participants (threshold 1) (dotted lines in Fig. 3). It was decided that items that do not pass the less-stringent threshold would be removed from the checklist, whereas items that pass the less-stringent threshold but not the more-stringent one are included but considered less important than items that pass both thresholds. For additional recommendations, we defined those with a 'Yes' rating by >50% of respondents as key ENIGMA ACRI checklist recommendations.

Reporting state of the checklist items

The state of reporting of the checklist items was assessed among 108 articles (ranging from January 1, 2017 to December 30, 2020) identified through a systematic review[5]. Rating was done by three independent raters (M.Z.-B., A.K.Z., and P.G. A.). An initial pilot rating of 19 articles was conducted and supervised by M.Z.-B., A.S. and H.E. to train the raters. After pilot rating, the remaining 89 articles were assessed by the three raters. Conflicts between raters were resolved by M.Z.-B., A.S. and H.E. in two group meetings, with all raters and supervisors reaching agreement on the final scores. The overall state of the reporting of the checklist items for each of the 108 studies ('reporting score') was calculated as the number of reported items divided by the total number of checklist items, excluding those with a 'not applicable' rating for each study. The inter-rater reliability of the checklist was also assessed on the basis of the three ratings for the 89 articles, using Fleiss' Kappa[43]. To assess whether papers with a better reporting status appear in journals with higher impact factors, whether the reporting status has improved across recent years and whether word-count limitations have an impact on reporting status, the correlations of reporting score with journal word limit, article word count and journal impact factor were also assessed. A number of example papers reporting each item are presented in Supplementary Table 6.

To support the potential utility of the checklist, a list of papers that demonstrate how each checklist item might affect the results of an FDCR study and its importance for interpretability and generalizability is also provided in Supplementary Table 6.

Ethical considerations

To ensure informed autonomy, all contributors were informed about the study's aims and methods in the invitation email. Further notes within the questionnaire and emails during each round provided extra details, although the general study design and purpose remained unchanged. Members of both the SC and EP were invited to view the study's evolving Open Science Foundation (OSF) page[43]. All contributors were informed that they could terminate their participation whenever they wished. To ensure confidentiality, contributors were kept anonymous during both rounds of the Delphi survey, and comments and ratings were anonymized to all except the lead authors. Neither responding to the basic information collected nor commenting on and rating the checklist items was deemed to require the disclosure of personal information.

Results

Characteristics of SC and EP and response rates

Of the original 14 SC members and 45 EP members who accepted the invitation, 51 (86.4%) respondents completed the revision round of the ENIGMA ACRI Delphi questionnaire. In the rating phase, 49 (83%) sent back complete responses. Four members of the EP responded to neither the revision nor the rating phase and therefore, were subsequently removed from the EP.

The characterization of the SC and EP is provided in Supplementary Table 1, which shows that SC members were older overall than the EP without any significant difference (mean ± s.d.: 51.1 ± 9.1 versus 45.3 ± 9.4); 60% (5 SC and 28 EP) of respondents were male. Most respondents hold a PhD (79% SC and 80% EP) and MD and PhD degrees (21% SC and 10% EP) and reported their primary field of research predominantly in neuroscience (29% SC and 44% EP) and psychiatry (43% SC and 34% EP). The professional affiliations of respondents were primarily universities (57% SC and 80% EP), hospitals (21% SC and 10% EP) and independent research institutes (14% SC and 10% EP). EP and SC members' research involved cue-reactivity studies of many SUD cohorts (e.g., methamphetamine, cocaine, opioid, alcohol, tobacco and gambling).

Delphi process results

A schematic of the entire study process and checklist development stages can be viewed in Fig. 2.

Checklist development phase—After the systematic review of 318 articles, an initial list of suggestions for the overall structure of the checklist and important items was developed. This list consisted of 42 items in 5 categories: 13 General Task Information items, 9 Drug Cue Information items, 9 Control-Cue Information items, 6 Craving Assessment Inside Scanner items and 5 Craving Assessment Outside Scanner items. After the discussions within the SC members, this initial draft was developed into a checklist with 7 categories and 37 items: 8 Participants' Characteristic items, 4 General fMRI Information items, 5 General Task Information items, 6 Cue Information items, 5 Craving Assessment Inside Scanner items, 4 Craving Assessment Outside Scanner items and 5 Pre- and Post-

Scanning Considerations items. In addition, on the basis of the SC inputs, a column with 27 additional recommendations corresponding to the different items was added to this checklist.

Revision phase—On the basis of SC and EP comments on the checklist, one Participants' Characteristic item, one Cue Information item, one Craving Assessment Inside Scanner item, one Craving Assessment Outside Scanner item and two Pre- and Post-Scanning Considerations items were excluded. New items were refined and added to the ENIGMA ACRI checklist following suggestions made by respondents to the 'please suggest extra variable' question. Additional Participants' Characteristic items were 'Psychiatric Profile' and 'Substance Use Profile-Main Drug'. The additional General Task Information items were about 'Temporal Information of the Event/Block Duration' and 'Data and Resource-Sharing'. The additional Pre- and Post-Scanning Considerations item was about 'Other Tasks and Procedures in the Imaging Session'. In addition, one item was split into two items: item 4—Advanced Demographics I and item 5—Advanced Demographics II. Thus, in the rating round, there were 11 Participants' Characteristic items, 4 General fMRI Information items, 7 General Task Information items, 5 Cue Information items, 4 Craving Assessment Inside Scanner items, 3 Craving Assessment Outside Scanner items and 4 Pre- and Post-Scanning Considerations items. The 22 additional recommendations were also expanded to 75, of which 69 were item-specific recommendations and 6 were category-specific recommendations. All the comments received in the revision phase are provided in an anonymized database on the project's OSF page[43].

Rating phase—Rating phase results can be viewed in Fig. 3. Respondents had a high rate of agreement on most checklist items, and all items reached the less-stringent threshold (>70% of participants selected the 'extremely important', 'highly important' or 'moderately important' rating), and no item was excluded due to not reaching the thresholds. Most of the items also met the more-stringent threshold of the consensus (>80% of participants selected the 'extremely important' or 'highly important' rating). The following items (marked with † in Fig. 3) did not reach the most stringent a priori threshold of the consensus: Advanced Demographics I, Advanced Demographics II, Handedness, Substance Use Profile-Main Drug, Substance Use Profile-Other Drug, Data and Resource-Sharing, Sources of Cues-Development, Drug and Neutral/Control Cue Content, Neutral/Control Matching to Drug Cues for Physical Features, Craving Assessment Inside Scanner-Technology, Craving Assessment Outside Scanner-Time Points, Pre-scanning Training and Familiarization, Other Tasks and Procedures in the Imaging Session and Post-scanning Craving Management. The results of the 'Yes/No' rating of the 75 additional recommendations are presented in Fig. 4. The results show that 69 (92%) recommendations reached the 50% threshold, but the following 6 (8%) did not: Interviewer Qualification, Motivation to Quit, Socio-economic Status, Body Mass Index, Menstrual Status and Sleepiness/Alertness. With the exception of revisions for minor grammatical and typographical errors, the checklist was not changed in the rating phase, and no item or category changes were made as a priori planned[43]. The average ratings of the ENIGMA ACRI checklist items and the frequency of 'Yes' ratings for additional recommendations are presented in Tables 1–6.

The short form of the checklist is available in Table 7. The other checklist forms, including both the items and the additional recommendations, are available as PDF or Excel files in Supplementary Tables 2–5.

Reporting state of the checklist items

The consistency of the raters' responses between the three raters resulted in a Fleiss' Kappa of 0.799, indicating that the consistency is between 'substantial agreement' and 'almost perfect agreement'[43]. The Kappa indices for all individual items except 'Other Tasks and Procedures in the Imaging Session' and 'Substance-Use Profile-Other Drugs' items were higher than 0.4, indicating at least a 'moderate agreement' among the raters. The Fleiss' Kappa for each individual item can be found in Extended Data Fig. 1. The reporting status of the ENIGMA-ACRI checklist items ranged from near-universal reporting (99%; Basic Demographic Data) to almost not-reported (8%; post-scanning craving management). Articles also varied widely in terms of their overall reporting score, ranging from reporting only 27% of the checklist items to reporting 92%. On average, 70.4% ± 10.5% (mean ± s.d.) of checklist items were reported by the papers in our database (Fig. 5). Overall, the 'General fMRI Information' section had the highest average reporting across the 108 studies at 90.5% reporting, and the 'pre- and post-scanning considerations' section had the lowest reporting at 44.7%. The highest reporting score was 91.7%, and 10 articles had a score of higher than 80%. The lowest reporting score was 27.3%, and only 6 studies failed to meet a reporting threshold of 50%.

The correlations of study reporting status with journal word limit, article word count and journal impact factor were not significant, and relevant graphs are presented in Extended Data Fig. 2.

Discussion

We developed a checklist resulting from a consensus process that represents the views of participating scientists regarding what they presumed to be important methodological aspects of conducting an FDCR study that would merit universal inclusion as methods details. We also investigated the state of the reporting of these checklist items in the FDCR literature. Key methodological aspects include seven distinct categories of core items and additional recommendations, as enumerated below.

Participants' Characteristics

The Participants' Characteristics section covers data about subjects' demographics, psychiatric profile, handedness, substance-use profile, abstinence status and treatment status. All the items listed in this category were considered important by the experts (Fig. 3 and Table 1), although some such as race or ethnicity and handedness are not frequently reported in the literature (Fig. 5).

Age and sex/gender passed our more-stringent consensus threshold. In terms of age, FDCR studies can typically be divided into two major categories, those involving adolescents/ emerging adults (e.g., refs. [44,45]) and those involving adults (e.g., refs. [46,47]). This distinction is important in part because of the development of the cortical circuitry that provides top-

down control over bottom-up limbic systems that continue to mature throughout adolescence to early adulthood[48]. In addition, it is likely that age is correlated with years of substance use[49], and neurocircuitry adaptations also occur over time, leading to potential confounding. Moreover, although FDCR studies often include participants in specified developmental stages, not much is known about the association of age (in years) with FDCR in each developmental category, perhaps partly due to restriction of participant age range. In addition, older adults have been routinely excluded from MRI studies that do not focus on aging and the shared neurodegenerative impacts of addiction and biological aging[50], and there is relatively little known about FDCR among the elderly. In terms of sex/gender, multiple studies have demonstrated sex-/gender-related differences in FDCR, particularly in participants who smoke cigarettes[51,52], individuals with cocaine dependence[53,54] and those with gambling[55] and gaming disorders[56–58], which may depend, in part, on menstrual cycle phase in women[59].

Additional demographics that passed the less-stringent consensus threshold included education/intelligence, handed-ness and race/ethnicity. These were rated as relatively less important than age and sex/gender partly because of a lack of published evidence for their association with FDCR.

It is perhaps not surprising that education/intelligence has not been found to be reliably associated with FDCR, given the often-low cognitive demands of a typical FDCR task (i.e., passively perceiving sensory stimuli). However, education/intelligence might be an important factor in FDCR in populations with intellectual disabilities[60]. Seventy-two percent of the assessed studies reported a measure of intelligence or education. Although handedness can be a critical consideration in fMRI studies of cognition (e.g., language and memory[61]), it does not appear to play a major role in the lateralization of FDCR, and only 41% of the 108 FDCR studies reported a measure of handedness.

In the case of race and ethnicity, it is possible that the literature as a whole has not provided sufficient opportunity to detect associations between FDCR and participant ethnicity or race (which could be driven entirely by unmodeled environmental/contextual variables), because studies have historically contained too few non-white/Hispanic participants to provide adequate statistical power to detect such associations. Only 40% of the reviewed FDCR studies reported participants' race or ethnicity. Some racial and ethnic differences in brain activation during fear processing[62] and social evaluation[63] have been noted in the literature, but the importance of these differences in FDCR remains largely unknown.

In terms of clinical characteristics, the pattern/severity of substance use, addiction treatment status, last use and abstinence status, psychiatric profile and study inclusion/exclusion criteria passed our more-stringent consensus threshold. All of these items were reported in 75% of the assessed FDCR studies, with the exception of abstinence status, which was reported in only 59% of the studies. The importance of all of these items has been discussed previously. For example, in people who use cocaine, greater FDCR has been positively associated with addiction severity[8,46,64] and could be predictive of relapse[8,65,66]. Perhaps unsurprisingly, self-reported craving has also been associated with FDCR across various drugs[8,16].

Although both treatment seekers and non-treatment seekers demonstrate similar activation to drug cues in the ventral striatum[67], treatment seekers have lower activation to drug cues in various non-limbic (e.g., frontal, cingulate and temporal) brain regions than non-treatment seekers[49]. This difference may be attributable to the expected availability of drug reward after cue exposure[68,69], an additional variable of potential interest to consider for future consensus checklists.

Abstinence has also been associated with increased drug cue reactivity (e.g., in dorsolateral PFC and occipital cortex) in cigarette smokers[70] and (e.g., in the midbrain) in individuals with cocaine use disorder[71] but needs further study. Although individuals with acute psychiatric illness co-occurring with SUDs are typically excluded in FDCR studies, studies could collect information on lifetime histories of psychiatric illness and present subclinical symptoms of psychiatric disorders like depression and anxiety and investigate the interaction of past psychopathology or present subclinical symptoms on FDCR[72–74]. Researchers should consider explicitly stating whether individuals were assessed for the existence of subclinical symptoms of psychiatric disease, even if the assessment was performed as part of the inclusion or exclusion criteria. If individuals with subclinical symptoms are included, the impact of psychiatric symptoms on FDCR parameters and the sensitivity of the analyses to their presence may be estimated.

Finally, all study inclusion/exclusion criteria, including those already discussed, must be carefully considered. As just one example, psychiatric medications have been shown to alter FDCR[75]; information concerning psychiatric medications should be provided to readers in a standardized manner (e.g., in chlorpromazine equivalents for neuroleptic medication), and attempts should be made to prevent or at least examine the potential impact of all medication classes on FDCR via appropriate randomization and/or analytic strategies.

Additional clinical characteristics that passed our less-stringent consensus threshold included substance administration method and the co-occurring use of other drugs.

FDCR studies often isolate participants by route of drug administration either purposefully or through convenience sampling (e.g., demographic homogeneity due to geographic location of participant recruitment). Nonetheless, care (e.g., in cue representation and covariate analysis) should be taken when combining groups of individuals who use the same drug (e.g., opioids) but self-administer it via different routes (e.g., intravenous versus oral[76]) within the same sample or study. In our sample of FDCR studies, 75% reported the route of drug administration, although this is partly because some substances commonly investigated in FDCR studies (such as alcohol) have only a single plausible administration route, and in these cases the studies were not required to explicitly report the administration route for a 'Yes' rating.

Although researchers typically aim to isolate a single or 'primary' drug in FDCR studies, the use of other drugs should also be considered, because sensory cues of the 'primary' drug may nonetheless trigger neurobehavioral responses to multiple drugs, particularly when such drugs are commonly used simultaneously (e.g., cannabis and alcohol[77]). Only 17% of studies failed to report the use of other drugs.

Another potentially important participants' characteristic is genetics. This factor was not considered important for inclusion in this checklist by our participating experts, perhaps because the influence of genes on various aspects of FDCR remains understudied. Nonetheless, polymorphisms in dopaminergic, GABAergic, glutamatergic, cholinergic, opioidergic and other genes may affect FDCR results (e.g., refs. [78–91]). As FDCR methods are harmonized and more data sharing can occur, we suggest that FDCR studies consider banking subject DNA for future genotyping so that DNA will be available to support analyses such as those involving polygenic risk scoring. Prospective use of genetic data could involve explicit informed consent or a waiver of informed consent from independent review boards to use deidentified data.

General fMRI Information

This section covers general details for the reporting of methods for fMRI acquisition details (hardware and software), data analytic procedures and scanning results in FDCR studies (Fig. 3 and Table 2). These items were considered extremely important to report by >80% of raters, and the category overall had the highest mean rating of all seven reporting categories. Similarly, for additional recommendation items (Fig. 4 and Table 2), the General fMRI Information category had the highest proportion of elements (89%) recommended by 75% of raters. This strong consensus is not surprising because these FDCR elements robustly influence data quality and variability. Nearly all of the 108 assessed studies reported all except the more specific 'fMRI data reporting' item, the requirements for which were met in 65% of the studies (Fig. 5). Below, we discuss selected items in each subcategory (acquisition, preprocessing, processing and reporting) to illustrate key points.

It was recommended with near unanimity that FDCR data acquisition details be reported using detailed checklists (e.g., COBIDAS Report[23] and/or ref. [92]). Detailed reporting can increase experimental design consistency, assist investigators new to the field in implementing robust methods, and increase FDCR replicability and enable data sharing and meta-analyses. For example, it is very important to report hardware details that could affect fMRI signals in different ways across the brain, such as the number of head-coil channels (e.g., 32 versus 8).

Indeed, a 'coil-bias' effect has been documented by several studies: one study determined that a 32-channel coil was more sensitive than an 8-channel coil for detecting cortical surface signals during a finger-tapping paradigm but less sensitive for detecting subcortical activations[93]. A more recent and comprehensive study investigating coil bias determined that head-coil channel number affects volumetric and diffusion measures as well as resting-state BOLD signal measures, with channel number strongly affecting BOLD signals in posterior visual and default mode network areas[94].

In addition, although most current FDCR studies are conducted on 3-Tesla (T) systems, other factors will need to be considered in future as more studies are conducted at higher magnetic field strengths. For example, a preliminary (bioRxiv) communication compared fMRI results on a monetary incentive task in eight subjects scanned both at 7 and 3 T[95]. The study reported that 7-T scans yielded higher effects than 3-T scans in small subcortical

nuclei relevant to FDCR studies, including the substantia nigra, ventral tegmentum and locus coeruleus.

Detailed reporting of preprocessing parameters using the structured checklists noted above was unanimously endorsed. Preprocessing parameters such as the spatial smoothing Full-Width Half Maximum value should be reported because they affect statistical inferences. In this regard, a meta-analysis of fMRI tasks involving rewarding stimuli revealed that the spatial smoothing value affects apparent nucleus accumbens volumes and anatomical positions[96].

There was near unanimity in the endorsement of reporting of artifact detection methods and motion thresholds for data exclusion.

There was substantial but lower agreement (79%) regarding reporting of group motion parameters during FDCR drugversus neutral-cue blocks, which, if differing by group, could confound data analyses. This version of the checklist did not explicitly include denoising protocols, which when applied can affect task-related fMRI data by reducing noise and signal[97]. Future checklist versions might consider including denoising procedures, which hopefully will evolve to more selectively attenuate noise.

For data processing pipeline procedures, there was near unanimity (98–100%) for most elements, including recommendations to report on single-subject and group-level processing steps, nature of GLM analyses (random, mixed and fixed), whether covariates or demeaning are used, software tools used, multiple comparisons corrections applied and regions of interest specifications, if applicable (e.g., manually drawn, atlas-based or dataset-determined).

Reporting of the pre-registration of data-processing methods and reporting of effect sizes were considered important but with lower priorities. This lower priority does not mean that the checklist contributors did not believe that reporting the effect size matters. However, it should be noted that the focus of the survey was on the consideration and reporting of methodological factors, not details of the results. This might explain why effect sizes have been de-prioritized by survey respondents. The sample sizes commonly used in task-based fMRI research tend to generate small-to-medium effect sizes (Cohen's $d < 0.8$ [98]). However, it seems likely that effect size reporting will be considered a higher priority in the future.

There was greater variability across fMRI data-reporting elements, with >80% of raters endorsing detailed reporting of second-level maps or activation foci within groups, whole-brain contrasts, beta-weights during craving and neutral conditions and inclusion of whole-brain maps even in studies not using standard analytic methods, to facilitate data comparisons across studies.

Other reporting elements were considered somewhat lower priorities, including providing non-thresholded statistical maps and stating whether data have been or will be deposited in publicly available repositories, which can be challenging given inconsistencies in repository reporting requirements. Most (78%) raters recommended that reporting go beyond the use of checklists by providing as much experimental detail as possible. Undoubtedly, over time, as

more data are aggregated in meta-analyses and as additional factors are determined to affect FDCR data effect sizes, such factors will be added to the reporting checklist.

General Task Information

While FDCR tasks are often straightforward cue-presentation paradigms, an adequate description of the task design, task components, requested subject engagement and precise temporal information is essential to assess the appropriateness of analytical procedures and interpret the results. As such, it is not surprising that experts considered this category to be almost as important as the 'Participants' Characteristics' and 'General fMRI Information' sections (Figs. 3 and 4 and Table 3), and three of the seven items were reported by almost all of the assessed FDCR studies (Fig. 5). Because of its fundamental implications for modeling and design efficiency, it is necessary to report the exact temporal structure of the task, specifically the order, the onset, the spacing and the duration of stimuli, and it is not sufficient to merely report whether stimuli were presented in blocks or an event-related or mixed design was used. The temporal pattern of stimulation also significantly influences the amplitude of the evoked hemodynamic response.

In addition to simple cue-presentation experiments, sophisticated tasks with complex trial structures are increasingly used to investigate the interactions between various affective and cognitive trial components, such as attentional bias[99] or response inhibition during the presentation of drug cues[49]. In these cases, a detailed description of the timing of stimulus presentations and participant responses within trials and blocks and the related modeling approach can be especially necessary to understand and assess the experimental procedure. To optimally sample hemodynamic responses in event-related designs and also decrease the predictability of stimulus presentation, the interstimulus interval (ISI) is often jittered, resulting in random ISIs across the task duration. The formulations used to obtain jittered intervals and the distribution of the resulting ISIs are important to assess design efficiency and should be described in detail[100,101].

Beyond this micro-timing information, information like the overall duration of the scanner session, the duration of the experimental paradigm, the start in relation to the onset of the scanning session and the position within the order of possible additional paradigms are also of interest because multiparadigm fMRI experiments are known to be prone to carry-over and order effects[16].

Reporting should further mention whether and how the order and timing of stimulus presentation were optimized. If appropriate, all of this information could be provided in compact and understandable ways by means of graphic displays (e.g., see refs. [44,102–104]). Most of the assessed FDCR studies report at least some information regarding these items, with the least frequently reported item being the 'Temporal Information of the Task' item at 80% reporting. In the interest of a complete description of the experimental setup, we also suggest that the technical details of stimulation procedures and parameters and the equipment used be reported, especially if a less-common sensory modality was targeted. For example, studies using gustatory cues (e.g., alcoholic beverages) could report substance concentration and temperature, whether cues were preceded with another stimulus, potential latencies in substance delivery and the equipment and material that were used.

Whether participants are instructed to interact passively or actively with the cue, to allow or to regulate craving, is an important component of instruction, influencing the experimental setting. To enable the reader to judge the clarity of the instruction, the verbatim instructions given to the participant should be included. Especially in passive tasks, additional processes such as mind wandering and attentional drift could occur[105], potentially harming the specificity of statistical analyses. Therefore, the chosen activity level and possible attempts to quantify participants' compliance, attention and vigilance should be described in detail. For instance, some studies include trials to assess participant attention or use eye-tracking technologies (e.g., see refs. [106–108]. Over 39% of the rated studies failed to report this crucial item.

Although 58% of the panel experts were of the opinion that the task code and stimuli-sharing item (Table 3) should be included in the checklist, its importance was rated lower (3.31) compared to the other items. This is particularly surprising given the intense contemporary discussion about reproducibility in fMRI research[98]. In our opinion, authors should still report whether they have used an open scientific platform to provide task-related data (stimuli and software) to the imaging community. Therefore, the manuscript should include, where appropriate, information on access points and conditions of access (e.g., see refs. [109,110]), in accordance with the FAIR principles for data exchange (https://www.force11.org/fairprinciples). This item was the least frequently included in the rated FDCR studies, with only 6% of the 108 papers sharing their task-related data and resources.

Cue Information

The drug and control cues used in FDCR research fall under a number of different sensory modalities, can be developed and parametrized depending on modality and preferably validated and matched in terms of their important characteristics. This checklist category includes information regarding important features of the utilized cues and their origin, validity and content, and several items and recommendations received near-unanimous support (Figs. 3 and 4 and Table 4). Item rating means ranged from 4.07 (for the description of the validation extent of the cues) to 4.77 (for the description of the sensory modality of cues).

Multiple drug- or control-cue–related aspects of FDCR studies may affect study outcomes[16]. The most important factor may be the description of the sensory modality of drug and control cues, which was also reported in 97% of the rated FDCR studies (Fig. 5). Although cues in different sensory modalities often induce distinct brain activation profiles[111], some studies do not clearly describe the sensory modality of their utilized cues. Depending on the sensory modality, there are various parameters that may need to be further considered and specified for drug cues and control stimuli. For instance, for pictorial cues, it is recommended that authors provide details regarding picture luminance, complexity (including human presence), hue and saturation. For auditory cues, it is important to consider factors such as intensity and frequency (loudness and pitch)[19,21,112]. Only half of the 108 assessed FDCR studies reported their choices regarding cue matching (i.e., trying to control for both physical features like size and color and content features in the substance and control cues).

Furthermore, these parameters may be used to 'match' drug cues and control stimuli (or those belonging to other cue categories in a study). Matching is done to minimize the effects of these other factors on the differential activation patterns elicited by different cue types. In addition, cues can be matched on the basis of their standardized arousal, valence or craving induction scores[19,112,113].

Another important but often overlooked factor limiting replicability and interpretation of FDCR studies is confusion over the sources of utilized cues, how they were obtained or developed and whether they have been validated (i.e., shown to elicit a certain range of arousal, affective or craving-related responses in individuals). Experts considered providing cue-validation details to be very important, but the reporting of cue-development details was not rated as highly. Nevertheless, there was near-unanimous support that researchers should consider reporting the exact source of their cues and how their cues were developed from this source, where applicable, which suggests that the participating experts broadly considered this a significant aspect of an FDCR study. Even in cases where authors are using cues developed or validated in another published study, it is still desirable to provide minimal development and validation details in addition to references. A notable gap between the aggregated expert opinion and reporting status in the reviewed literature was also observed, with 72% of FDCR studies containing information on cue development but only 28% reporting any cue-validation processes.

Although not always optimal, using cues from already validated and widely used cue databases may save researchers considerable resources and improve consistency across studies. There have been recent attempts to develop large pictorial cue databases to address these issues[19,112]. These databases include cues that have been developed in a methodologically consistent manner and whose craving and arousal elicitation effects have been formally studied. The best FDCR cue databases include neutral stimuli as well as drug cues that are matched according to various characteristics[21,114]. Newer databases with a greater focus on drug cue–reactivity studies have become available in recent years[17], and large developing cue banks may even contain multiple drug cues and control stimuli types[19].

The exact content of cues can also influence multiple dimensions of cue reactivity. Drug cues may depict the drugs themselves, drug paraphernalia, individuals preparing or using drugs or spaces where drug use is likely. Differences in the content of cues (drug versus drug-use tools versus drug-use actions) may recruit different brain areas, and this may have implications for how these cues link to drug-seeking behavior[115]. It may be important to consider this aspect of cue selection when designing studies, because certain cue contents may be more appropriate for testing some, but not all, hypotheses.

In addition, among recommendations in this category, there was widespread agreement on the importance of describing substance-delivery methods in studies in which a substance is administered as a cue, prior cue exposure, and cue tailoring. Studies in which a substance is directly administered (usually in small amounts) remain relatively rare in the field of FDCR as a whole. However, given the popularity of these paradigms in some fields (such as in tobacco use disorder and alcohol use disorder) and the large variety of substance-delivery mechanisms used, it is recommended that researchers describe their delivery

mechanisms in detail and cite the relevant literature when possible[116–118]. Prior exposure of participants to cues is also important. Some brain regions may rapidly habituate to specific drug cues, decreasing their reactivity to them, even in the absence of a reduction in self-reported craving[119]. Lastly, personalized tailoring of cues presents unique challenges and opportunities in FDCR studies. Although it potentially leads to maximal cue reactivity in all participants, it also leads to heterogeneous cues that present problems for generalizability and interpretation. It is recommended that authors specify whether tailoring was conducted (if there is room for misunderstanding) and present precise details for how tailoring was conducted for each participant. Although all individual cues in a study may be tailored[120], tailoring can be particularly applied on the basis of the participant drug of choice in samples of individuals who use multiple drugs[109]. Tailoring of drug-related messages meant to encourage drug-use cessation is another possibility[121]. Tailoring for gender/race/ethnicity is another area that is not well explored yet.

Task-Related Assessments

This section includes items regarding the inside- and outside-scanner assessment of the subject's craving, including when and how the craving was assessed. Integration of self-report, behavioral or physiological measures as part of FDCR is commonplace[122–124]. Yet, perhaps because fMRI is the primary focus of these papers, the methodological details of other task-related assessments (e.g., self-reported drug use and craving/urge) that would be standard to report in behavioral research papers are sometimes excluded. Details of items, ratings and recommendations are presented in Figs. 3 and 4 and Table 5. A recent review of opioid-craving measurement identified many different questionnaires for assessing opioid craving; however, many had not been tested for reliability and validity[125]. Harmonization and validation of the questionnaires used for subjective reporting of drug craving should be considered a priority in the field. As an example, a systematic review is ongoing to develop an extensive map of every instrument used to assess craving in clinical trials[126].

The timing of additional task-related assessments received high ratings of importance overall, with universal agreement that reporting the time period considered for in-scanner tasks (i.e., urges while viewing the image versus afterward) is important. Assessment time points were reported by ~90% of the rated FDCR studies for craving assessments both inside and outside the scanner. This information is critical for proper interpretation of the nature and magnitude of the response. There is evidence that the effects of imagery-based cue procedures on urge may persist for extended periods of time (e.g., 15–30 min)[127,128], but the duration of effects from the brief image presentations commonly used in FDCR are largely unknown. Indeed, given that many FDCR paradigms rely on random/pseudo-random presentation of interleaved images from varying categories, an implicit assumption of most research is that the duration of these effects is brief. Continued research on this topic examining the validity of this assumption is critical and could conceivably lead to the development of formal guidelines for such assessments depending upon the nature of the study, the cue modality used and the specific question being asked.

As with timing, there was near-universal agreement that detailed reporting of the contents of both in-scanner and out-of-scanner assessments is important. This is perhaps particularly

critical for in-scanner assessments, for which research has historically relied more heavily on single-item measures and may not have been subjected to the same rigorous examination of psychometric properties common for traditional self-report measures[129–131]. Although the general construct is frequently reported (e.g., urge or liking), reporting the exact phrasing is less common despite long-standing recognition that subtle differences in wording can affect participant interpretation and study outcomes[132,133]. This issue will be particularly important as research continues to explore covariation of constructs with brain activation. Indeed, research has already shown that patterns of activation may be at least partly dependent upon urge strength[134]. It should be recognized, however, that subjective 'craving/ urge' is highly variable and situation specific (e.g., scanner versus bar). As such, brain activation to cues during fMRI might be less variable and, in fact, was one of the reasons for the initial development of FDCR paradigms.

There was also agreement about the importance of reporting hardware (e.g., button box and response pad) used for collection of these assessments. This may be particularly critical for research in which response time is examined as a primary or secondary outcome. An extensive body of literature documents the existence of substantial variability in the accuracy of data-collection devices outside the scanner[135–137]. To our knowledge, no similar evaluation of variability in the accuracy of common MRI-compatible devices has been conducted. However, the importance of reporting utilized hardware in fMRI research[138] and using similar and calibrated hardware in multi-site fMRI studies[139] has been noted in the literature.

Comparatively fewer experts (61%) recommended the inclusion of other physiological measures relative to other topics under consideration. One likely reason is that to date, these measures have rarely been included in FDCR studies. Nonetheless, examination of heart rate, skin conductance and other peripheral physiological measures are standard in the broader drug cue reactivity[140]. It is certainly plausible that changes in peripheral physiology could influence findings, particularly for certain types of imaging (e.g., arterial spin labeling). Moreover, inclusion of peripheral signals as covariates is becoming standard in resting scans in light of evidence showing it can alter connectivity maps[141], and there is little reason to believe that these concerns should not extend to task-based scans. Although it may be premature to make formal recommendations for inclusion of peripheral measures at this time, continued exploration of this topic is critical and may reveal a need for inclusion in later instances.

Pre- and Post-Scanning Considerations

This section covers the items that have to be considered before and after the scanning session, which includes training and familiarization, pre-scanning substance consumption, other tasks and procedures besides cue reactivity and post-scanning craving consumption. Of the pre-/post-scanning considerations, pre-scanning drug and smoking consumption was the only metric rated as moderately to extremely important by all reviewers (Figs. 3 and 4 and Table 6). This is probably because of the impact that both abstinence and recent substance use can have on cue-induced craving and brain function. The length of abstinence also matters, because studies generally support the idea that short-term abstinence enhances cue

reactivity relative to satiety[142–146], which mirrors preclinical findings[147]. In contrast, longer-term abstinence is associated with reduced cue reactivity[146]. Furthermore, deprivation and cue presentations may have independent, interactive effects on subjective reports of craving[148], supporting the need to clearly indicate the conditions under which cue reactivity is evaluated. There is also a need to report the recency of other substance use and medications because they may influence subjective cue responses and the physiology underlying the fMRI signal, but this was reported by only 54% of the 108 rated FDCR studies.

Other recommendations include indicating whether participants have had prior cue exposure in the context of the study. This is important because habituation to emotionally evocative stimuli has been identified in specific brain regions[149], yet not in all participant groups, particularly those who may be more reactive to the cue content[150]. While within-session habituation is a potential confounder[119,149], cues continue to elicit subjective craving and comparable brain activity patterns over repeated sessions separated by longer durations (2–3 weeks)[151–153]. However, this finding has not been supported in all studies[119], thus supporting the need to clearly report details surrounding previous cue presentations. Reporting drug expectancy is also recommended, because recent work suggests that participant expectations influence cue reactivity and related circuitry[154–156].

Several elements of pre/post scanning considerations did not reach a stringent consensus. Pre-scanning training and familiarization were ranked as highly important by ~60% of respondents, because some reviewers felt this was such a fundamental aspect of good scientific procedures that it was assumed that study participants were familiarized in some way with the task, and only 25% of the assessed FDCR studies reported this item. In addition, most cue reactivity tasks involve passive exposure to cues, which, unlike complex behavioral tasks, do not require extensive pre-scan training. However, such familiarization may also affect potential habituation and expectancy, which would support the need to report on the basis of the discussion points above. The need to report other tasks and procedures in the imaging session was similarly ranked and did not reach a stringent consensus. It is plausible that the lack of reporting of other tasks may imply a singular focus on cue reactivity, with no potential influence for the other tasks. That said, reporting tasks that have the potential to influence cue reactivity is considered best practice. Post-scanning craving management was rated the lowest element, with <35% of the respondents ranking it as extremely/highly important, perhaps because it is viewed as more of an ethical consideration that would be considered by local institutional review boards rather than a factor that would affect cue reactivity directly. Given the potential ethical importance of craving management, it may be concerning that it was included in only 8% of the FDCR study sample. However, the ethical implications of this element depend on the nature of the specific study, because the consequences of inducing craving are more profound when assessing a cohort in treatment for opiate use disorder than when assessing a community sample of nicotine-dependent individuals not seeking treatment.

Conclusion and future directions

As demonstrated by the consensus of the experts participating in this study and the review of the literature, FDCR studies have a vast methodological parameter space in which many impactful choices regarding study design and reporting can be made. The lack of methodological transparency complicates replication and generalizability and hampers data synthesis and clinical translation, necessitating further harmonization in reporting methodological details. Focusing primarily on representing expert opinion on best reporting practices in the field, this initial checklist is envisioned as a starting point to gain further empirical insight into the effect of methodological details in FDCR research. Importantly, this checklist was derived from FDCR researcher estimations of what methods parameters are likely to substantially affect FDCR study results. However, uniform and thorough reporting of these parameters in future studies is necessary to enable sensitivity analyses (e.g., meta-analyses) to confirm or refute the ostensible importance of these factors, yielding critical mechanistic insights into cue reactivity in the process. We hope that the development of this checklist will set an initial standard for research practices and encourage scientific authorities in other areas of task-based fMRI to promote harmonization and transparency in reporting methodological details across different areas of functional human brain mapping[38]. As a secondary effect, journal reviewers and editors may consider aspects of this checklist during the peer review of relevant FDCR articles.

This paper presents the results of an international effort to develop an initial checklist of important items and recommendations that FDCR researchers can use to plan future studies or assess past work. The itemized and hierarchical structure of the checklist is meant to help researchers read and consider various parts as needed, and the ratable format makes it possible to use the checklist to score an FDCR study. In addition, a list of papers that appropriately report checklist items is provided in the supplementary materials and can be consulted when using the checklist. Our ultimate hope is that this checklist will be used widely within the field to foster transparency in FDCR research and facilitate data syntheses. Crucially, the checklist is not meant to limit variance and flexibility in study design, but rather to invite attention to various methodological aspects of an FDCR study, in particular under-reported elements such as abstinence status/recent drug use, participant task familiarization and compliance/attention, cue validation and matching and how they bear on the obtained results, wherever they might be applicable in the context of a particular project.

This is merely the first iteration of the checklist. Considering the rapid rate of progress in the field and based on feedback from the FDCR academic community, the checklist will be revised in later editions and is now an open-source project at https://osf.io/gwrh6/ for public commenting and discussion. To ensure the feasibility of the checklist application, we suggest considering and reporting the 'items' as a 'must' in FDCR studies and the use of 'additional recommendations' as suggestions to improve the methodological design and reporting of FDCR studies. The extent to which the checklist is adopted by journal editors/reviewers and FDCR researchers around the world will determine its influence in the long term.

Extended Data
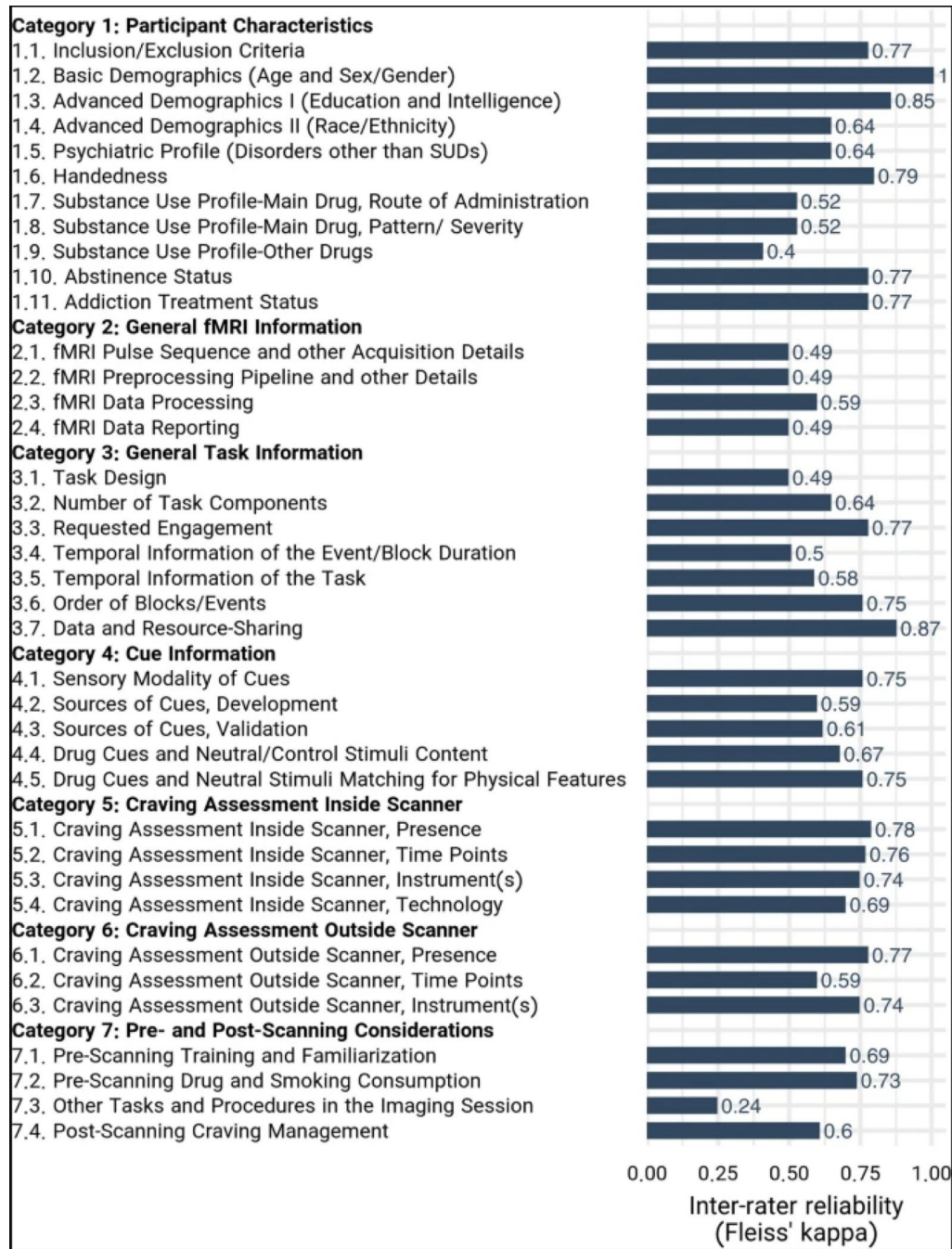
Extended Data Fig. 1 |. Inter-rater reliability for individual checklist items.
Inter-rater reliability assessed by Fleiss' Kappa for each ENIGMA ACRI checklist item, calculated on the basis of the assessment of reporting status of the checklist items among 108 papers by three independent raters.
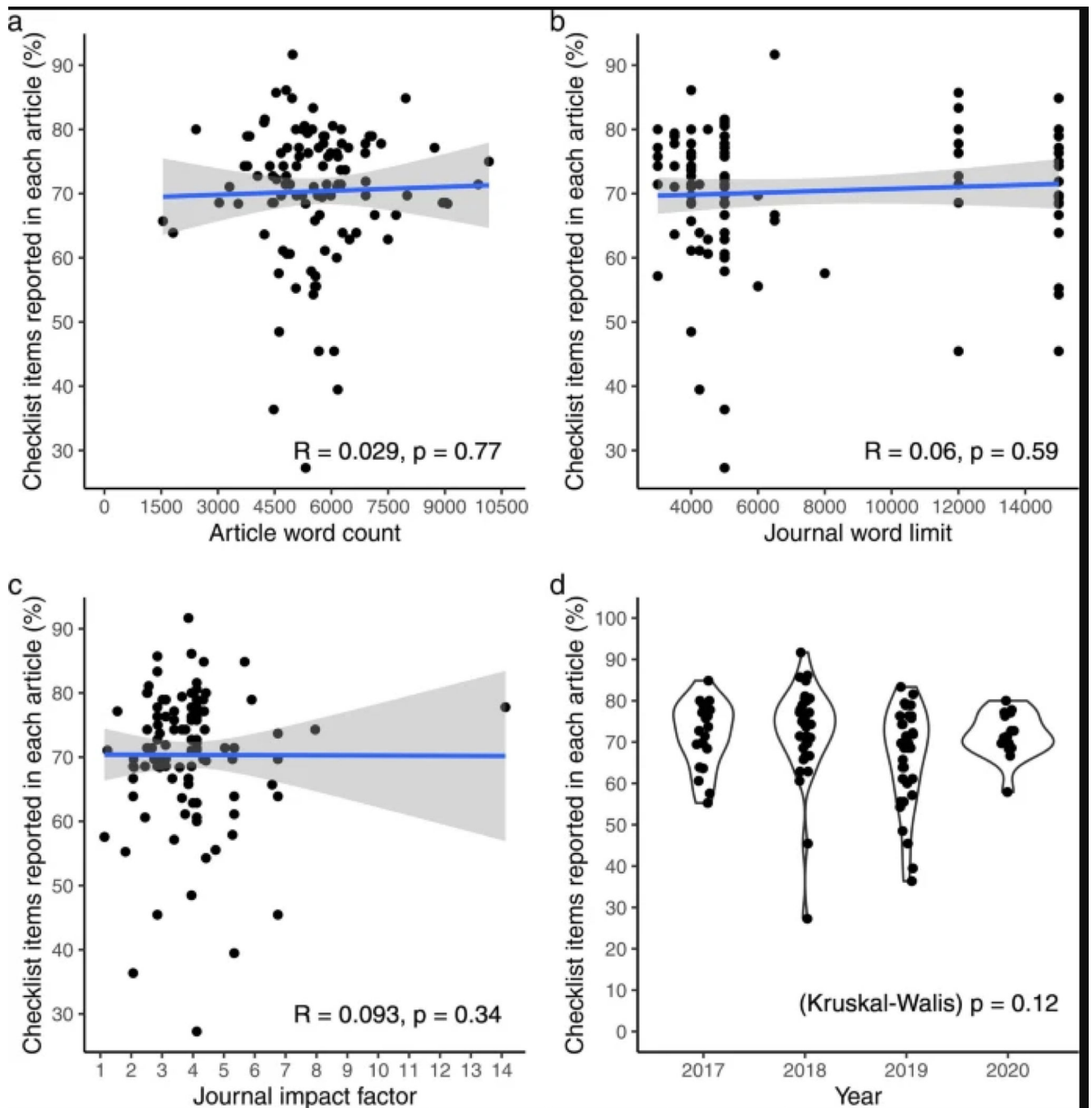
Extended Data Fig. 2 |. Relationships between reporting score and publication context.
a, Relation between the reporting score of each article and its word count. (Note that article word count is not exactly accurate, because it is measured by counting the words from the beginning of the introduction to the end of the discussion; thus, it might include the running title of each page, footnotes and the captions of figures and tables.) **b**, Relation between the reporting score of each article and its journal word limit. (Note that the word limitation for journals with no word limitation is counted as 15,000.) **c**, Relation between the reporting score of each article with journal impact factor. **d**, Article reporting scores across

the years. The relations in panels **a**, **b** and **c** were assessed using linear regressions, whereas a Kruskal-Wallis test was performed for panel **d**.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Hamed Ekhtiari[1,2], Mehran Zare-Bidoky[3,4,50], Arshiya Sangchooli[3,50], Amy C. Janes[5], Marc J. Kaufman[5], Jason A. Oliver[6,7,8], James J. Prisciandaro[9], Torsten Wüstenberg[10], Raymond F. Anton[9], Patrick Bach[11], Alex Baldacchino[12], Anne Beck[10,13], James M. Bjork[14], Judson Brewer[15], Anna Rose Childress[16], Eric D. Claus[17], Kelly E. Courtney[18], Mohsen Ebrahimi[3], Francesca M. Filbey[19], Dara G. Ghahremani[20], Peyman Ghobadi Azbari[3,21], Rita Z. Goldstein[22], Anna E. Goudriaan[23], Erica N. Grodin[20], J. Paul Hamilton[24,25], Colleen A. Hanlon[26], Peyman Hassani-Abharian[27], Andreas Heinz[10], Jane E. Joseph[28], Falk Kiefer[11], Arash Khojasteh Zonoozi[3,29], Hedy Kober[30], Rayus Kuplicki[1], Qiang Li[31], Edythe D. London[20], Joseph McClernon[6], Hamid R. Noori[32,33], Max M. Owens[34], Martin P. Paulus[1], Irene Perini[24,25], Marc Potenza[30,35,36,37], Stéphane Potvin[38], Lara Ray[20], Joseph P. Schacht[39], Dongju Seo[30], Rajita Sinha[30], Michael N. Smolka[40], Rainer Spanagel[41], Vaughn R. Steele[30], Elliot A. Stein[42], Sabine Steins-Loeber[43], Susan F. Tapert[18], Antonio Verdejo-Garcia[44], Sabine Vollstädt-Klein[11], Reagan R. Wetherill[16], Stephen J. Wilson[45], Katie Witkiewitz[46], Kai Yuan[47], Xiaochu Zhang[48,49], Anna Zilverstand[2]

Affiliations

[1]Laureate Institute for Brain Research, Tulsa, OK, USA.

[2]Department of Psychiatry & Behavioral Sciences, University of Minnesota, Minneapolis, MN, USA.

[3]Iranian National Center for Addiction Studies (INCAS), Tehran University of Medical Sciences, Tehran, Iran.

[4]Shahid-Sadoughi University of Medical Sciences, Yazd, Iran.

[5]Department of Psychiatry, McLean Hospital, Harvard Medical School, Belmont, MA, USA.

[6]Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA.

[7]TSET Health Promotion Research Center, Stephenson Cancer Center, Oklahoma City, OK, USA.

[8]Department of Psychiatry & Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, USA.

[9]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, USA.

[10]Department of Psychiatry and Neurosciences, Charité Campus Mitte, Charité–Universitätsmedizin Berlin, Berlin, Germany.

[11]Department of Addictive Behaviour and Addiction Medicine, Central Institute of Mental Health (CIMH), Heidelberg University, Mannheim, Germany.

[12]Division of Population Studies and Behavioural Sciences, St Andrews University Medical School, University of St Andrews, Scotland, UK.

[13]Faculty of Health, Health and Medical University, Campus Potsdam, Potsdam, Germany.

[14]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA.

[15]Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, USA.

[16]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[17]Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, USA.

[18]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA.

[19]Center for BrainHealth, School of Behavioral and Brain Sciences, University of Texas at Dallas, Dallas, TX, USA.

[20]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, USA.

[21]Department of Biomedical Engineering, Shahed University, Tehran, Iran.

[22]Departments of Psychiatry & Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[23]Department of Psychiatry, Amsterdam University Medical Center, University of Amsterdam and Amsterdam Neuroscience, Amsterdam, The Netherlands.

[24]Center for Social and Affective Neuroscience, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden.

[25]Center for Medical Image Science and Visualization, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden.

[26]Department of Cancer Biology, Wake Forest School of Medicine, Winston-Salem, NC, USA.

[27]Brain and Cognition Clinic, Institute for Cognitive Science Studies, Tehran, Iran.

[28]Department of Neuroscience, Medical University of South Carolina, Charleston, SC, USA.

[29]Mashhad University of Medical Sciences, Mashhad, Iran.

[30]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA.

[31]Department of Radiology, Tangdu Hospital, Fourth Military Medical University, Xi'an, China.

[32]International Center for Primate Brain Research, Center for Excellence in Brain Science and Intelligence Technology (CEBSIT)/Institute of Neuroscience (ION), Chinese Academy of Sciences, Shanghai, China.

[33]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA.

[34]Department of Psychiatry, University of Vermont, Burlington, VT, USA.

[35]Connecticut Mental Health Center, New Haven, CT, USA.

[36]Connecticut Council on Problem Gambling, Wethersfield, CT, USA.

[37]Department of Neuroscience, Child Study Center and Wu Tsai Institute, Yale School of Medicine, New Haven, CT, USA.

[38]Centre de recherche de l'Institut Universitaire en Santé Mentale de Montréal, University of Montreal, Montreal, Canada.

[39]University of Colorado Anschutz Medical Campus, Aurora, CO, USA.

[40]Department of Psychiatry, Technische Universität Dresden, Dresden, Germany.

[41]Institute of Psychopharmacology, Central Institute of Mental Health, Mannheim, Germany.

[42]Intramural Research Program, National Institute on Drug Abuse, Baltimore, MD, USA.

[43]Department of Clinical Psychology and Psychotherapy, Otto-Friedrich-University of Bamberg, Bamberg, Germany.

[44]Turner Institute for Brain and Mental Health, Monash University, Clayton, Australia.

[45]Department of Psychology, The Pennsylvania State University, University Park, PA, USA.

[46]Department of Psychology, University of New Mexico, Albuquerque, NM, USA.

[47]School of Life Science and Technology, Xidian University, Xi'an, China.

[48]Department of Psychology, School of Humanities and Social Science, University of Science and Technology of China, Anhui, China.

[49]Department of Radiology, First Affiliated Hospital of USTC, Hefei National Laboratory for Physical Science at the Microscale and School of Life Science, Division of Life Science and Medicine, University of Science and Technology of China, Anhui, China.

[50]These authors contributed equally: Mehran Zare-Bidoky, Arshiya Sangchooli.

Acknowledgements

References

1. Degenhardt L. et al. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Psychiatry 5, 987–1012 (2018). [PubMed: 30392731]

2. Ekhtiari H. et al. Functional neuroimaging for addiction medicine: from mechanisms to practical considerations. Prog. Brain Res 224, 129–153 (2016). [PubMed: 26822357]

3. Moeller SJ & Paulus MP Toward biomarkers of the addicted human brain: using neuroimaging to predict relapse and sustained abstinence in substance use disorder. Prog. Neuropsychopharmacol. Biol. Psychiatry 80, 143–154 (2018). [PubMed: 28322982]

4. Ekhtiari H. et al. Neuroscience of drug craving for addiction medicine: from circuits to therapies. Prog. Brain Res 223, 115–141 (2016). [PubMed: 26806774]

5. Ekhtiari H. & ACRI Secretariat. A systematic review on fMRI drug cue reactivity studies (OSF, 2020). https://osf.io/eb972/(2020).

6. Bough KJ et al. Biomarkers for the development of new medications for cocaine dependence. Neuropsychopharmacology 39, 202–219 (2014). [PubMed: 23979119]

7. Carmichael O. et al. The role of fMRI in drug development. Drug Discov. Today 23, 333–348 (2018). [PubMed: 29154758]

8. Zilverstand A. et al. Neuroimaging impaired response inhibition and salience attribution in human drug addiction: a systematic review. Neuron 98, 886–903 (2018). [PubMed: 29879391]

9. Cremers HR, Wager TD & Yarkoni T. The relation between statistical power and inference in fMRI. PloS One 12, e0184923 (2017).

10. Turner BO et al. Small sample sizes reduce the replicability of task-based fMRI studies. Commun. Biol 1, 62 (2018). [PubMed: 30271944]

11. Liu TT Noise contributions to the fMRI signal: An overview. Neuroimage 143, 141–151 (2016). [PubMed: 27612646]

12. Elliott ML et al. What Is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychol. Sci 31, 792–806 (2020). [PubMed: 32489141]

13. Korucuoglu O. et al. Test-retest reliability of fMRI-measured brain activity during decision making under risk. NeuroImage 214, 116759 (2020).

14. Kragel P. et al. Functional MRI can be highly reliable, but it depends on what you measure: a commentary on Elliott et al. (2020). Psychol. Sci. 32, 622–626 (2021). [PubMed: 33685310]

15. Casey B. et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54 (2018). [PubMed: 29567376]

16. Jasinska AJ et al. Factors modulating neural reactivity to drug cues in addiction: a survey of human neuroimaging studies. Neurosci. Biobehav. Rev 38, 1–16 (2014). [PubMed: 24211373]

17. Billieux J. et al. The Geneva Appetitive Alcohol Pictures (GAAP): development and preliminary validation. Eur. Addict. Res. 17, 225–230 (2011). [PubMed: 21654175]

18. Holla B. et al. Visual image-induced craving for ethanol (VICE): development, validation, and a pilot fMRI study. Indian J. Psychol. Med. 36, 164–169 (2014). [PubMed: 24860218]

19. Ekhtiari H. et al. Methamphetamine and Opioid Cue Database (MOCD): development and validation. Drug Alcohol Depend. 209, 107941 (2020).

20. Khazaal Y, Zullino D. & Billieux J. The Geneva Smoking Pictures: development and preliminary validation. Eur. Addict. Res 18, 103–109 (2012). [PubMed: 22285929]

21. Lang PJ, Bradley MM & Cuthbert BN International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. (NIMH, Center for the Study of Emotion & Attention, 2005).

22. Lindquist M. Neuroimaging results altered by varying analysis pipelines. Nature 582, 36–37 (2020). [PubMed: 32433631]

23. Nichols TE et al. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci 20, 299–303 (2017). [PubMed: 28230846]

24. Esteban O. et al. Analysis of task-based functional MRI data preprocessed with fMRIPrep. Nat. Protoc 15, 2186–2202 (2020). [PubMed: 32514178]

25. Bossuyt PM et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351, h5527 (2015).

26. Collins GS et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. Circulation 131, 211–219 (2015). [PubMed: 25561516]

27. Gagnier JJ et al. The CARE guidelines: consensus-based clinical case reporting guideline development. J. Med. Case Rep 7, 223 (2013). [PubMed: 24228906]

28. Moher D. et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6, e1000097 (2009).

29. Bossuyt PM et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin. Chem 61, 1446–1452 (2015). [PubMed: 26510957]

30. Collins GS, Reitsma JB, Altman DG & Moons KG Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Br. J. Surg 102, 148–158 (2015). [PubMed: 25627261]

31. Moher D. et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int. J. Surg 8, 336–341 (2010). [PubMed: 20171303]

32. O'Brien BC et al. Standards for reporting qualitative research: a synthesis of recommendations. Acad. Med. 89, 1245–1251 (2014). [PubMed: 24979285]

33. Schulz KF, Altman DG & Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Trials 11, 32 (2010). [PubMed: 20334632]

34. Tong A, Sainsbury P. & Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int. J. Qual. Health Care 19, 349–357 (2007). [PubMed: 17872937]

35. Von Elm E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull. World Health Organ 85, 867–872 (2007). [PubMed: 18038077]

36. Hsu C-C & Sandford BA The Delphi technique: making sense of consensus. Pract. Assess. Res. Eval 12, 10 (2007).

37. Jorm AF Using the Delphi expert consensus method in mental health research. Aust. N. Z. J. Psychiatry 49, 887–897 (2015). [PubMed: 26296368]

38. Yucel M. et al. A transdiagnostic dimensional approach towards a neuropsychological assessment for addiction: an international Delphi consensus study. Addiction 114, 1095–1109 (2019). [PubMed: 30133930]

39. Diamond IR et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J. Clin. Epidemiol 67, 401–409 (2014). [PubMed: 24581294]

40. Forsman AK et al. Research priorities for public mental health in Europe: recommendations of the ROAMER project. Eur. J. Public Health 25, 249–254 (2015). [PubMed: 25678606]

41. Chipchase L. et al. A checklist for assessing the methodological quality of studies using transcranial magnetic stimulation to study the motor system: an international consensus study. Clin. Neurophysiol 123, 1698–1704 (2012). [PubMed: 22647458]

42. Hasson F, Keeney S. & McKenna H. Research guidelines for the Delphi survey technique. J. Adv. Nurs 32, 1008–1015 (2000). [PubMed: 11095242]

43. Ekhtiari H. Methodological Checklist for fMRI Drug Cue Reactivity Studies: Development and Consensus. Accessed 27 Jan 2020. https://osf.io/gwrh6/ (2020).

44. Karoly HC et al. Investigating a novel fMRI cannabis cue reactivity task in youth. Addict. Behav 89, 20–28 (2019). [PubMed: 30243035]

45. Rubinstein ML et al. Smoking-related cue-induced brain activation in adolescent light smokers. J. Adolesc. Health 48, 7–12 (2011). [PubMed: 21185518]

46. Claus ED et al. Association between nicotine dependence severity, BOLD response to smoking cues, and functional connectivity. Neuropsychopharmacology 38, 2363–2372 (2013). [PubMed: 23708507]

47. Filbey FM et al. Marijuana craving in the brain. Proc. Natl. Acad. Sci. U. S. A. 106, 13016–13021 (2009).

48. Casey BJ, Jones RM & Hare TA The adolescent brain. Ann. N. Y. Acad. Sci 1124, 111–126 (2008). [PubMed: 18400927]

49. Prisciandaro JJ et al. The relationship between years of cocaine use and brain activation to cocaine and response inhibition cues. Addiction 109, 2062–2070 (2014). [PubMed: 24938849]

50. Cheng GLF et al. Heroin abuse accelerates biological aging: a novel insight from telomerase and brain imaging interaction. Transl. Psychiatry 3, e260 (2013). [PubMed: 23695235]

51. McClernon FJ, Kozink RV & Rose JE Individual differences in nicotine dependence, withdrawal symptoms, and sex predict transient fMRI-BOLD responses to smoking cues. Neuropsychopharmacology 33, 2148–2157 (2008). [PubMed: 17987060]

52. Wetherill RR et al. The impact of sex on brain responses to smoking cues: a perfusion fMRI study. Biol. Sex Differ 4, 9 (2013). [PubMed: 23628003]

53. Joseph JE et al. Neural correlates of oxytocin and cue reactivity in cocaine-dependent men and women with and without childhood trauma. Psychopharmacology (Berl.) 10.1007/s00213-019-05360-7 (2019).

54. Potenza MN et al. Neural correlates of stress-induced and cue-induced drug craving: influences of sex and cocaine dependence. Am. J. Psychiatry 169, 406–414 (2012). [PubMed: 22294257]

55. Kober H. et al. Brain activity during cocaine craving and gambling urges: an fMRI study. Neuropsychopharmacology 41, 628–637 (2016). [PubMed: 26119472]

56. Dong G. et al. Gender-related differences in cue-elicited cravings in Internet gaming disorder: the effects of deprivation. J. Behav. Addict 7, 953–964 (2018). [PubMed: 30556781]

57. Dong G. et al. Gender-related differences in neural responses to gaming cues before and after gaming: implications for gender-specific vulnerabilities to Internet gaming disorder. Soc. Cogn. Affect. Neurosci. 13, 1203–1214 (2018). [PubMed: 30272247]

58. Dong G. et al. Gender-related functional connectivity and craving during gaming and immediate abstinence during a mandatory break: implications for development and progression of internet gaming disorder. Prog. Neuropsychopharmacol. Biol. Psychiatry 88, 1–10 (2019). [PubMed: 29684536]

59. Franklin TR et al. Menstrual cycle phase modulates responses to smoking cues in the putamen: preliminary evidence for a novel target. Drug Alcohol Depend 198, 100–104 (2019). [PubMed: 30901743]

60. van Duijvenbode N. et al. Problematic alcohol use and mild intellectual disability: standardization of pictorial stimuli for an alcohol cue reactivity task. Res. Dev. Disabil 33, 1095–1102 (2012). [PubMed: 22502834]

61. Cuzzocreo JL et al. Effect of handedness on fMRI activation in the medial temporal lobe during an auditory verbal memory task. Hum. Brain Mapp 30, 1271–1278 (2009). [PubMed: 18570207]

62. Moriguchi Y. et al. Specific brain activation in Japanese and Caucasian people to fearful faces. Neuroreport 16, 133–136 (2005). [PubMed: 15671862]

63. Greer TM, Vendemia JM & Stancil M. Neural correlates of race-related social evaluations for African Americans and white Americans. Neuropsychology 26, 704–712 (2012). [PubMed: 23106117]

64. Volkow ND et al. Cocaine cues and dopamine in dorsal striatum: mechanism of craving in cocaine addiction. J. Neurosci 26, 6583–6588 (2006). [PubMed: 16775146]

65. Kosten TR et al. Cue-induced brain activity changes and relapse in cocaine-dependent patients. Neuropsychopharmacology 31, 644–650 (2006). [PubMed: 16123763]

66. Prisciandaro JJ et al. Prospective associations between brain activation to cocaine and no-go cues and cocaine relapse. Drug Alcohol Depend 131, 44–49 (2013). [PubMed: 23683790]

67. Chase HW et al. The neural basis of drug stimulus processing and craving: an activation likelihood estimation meta-analysis. Biol. Psychiatry 70, 785–793 (2011). [PubMed: 21757184]

68. Wertz JM & Sayette MA Effects of smoking opportunity on attentional bias in smokers. Psychol. Addict. Behav 15, 268–271 (2001). [PubMed: 11563808]

69. Wilson SJ et al. Carry-over effects of smoking cue exposure on working memory performance. Nicotine Tob. Res. 9, 613–619 (2007). [PubMed: 17454718]

70. Engelmann JM et al. Neural substrates of smoking cue reactivity: a meta-analysis of fMRI studies. Neuroimage 60, 252–262 (2012). [PubMed: 22206965]

71. Moeller SJ et al. Neural correlates of drug-biased choice in currently using and abstinent individuals with cocaine use disorder. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 3, 485–494 (2018). [PubMed: 29735157]

72. Coffey SF et al. Craving and physiological reactivity to trauma and alcohol cues in posttraumatic stress disorder and alcohol dependence. Exp. Clin. Psychopharmacol 18, 340–349 (2010). [PubMed: 20695690]

73. Potvin S. et al. Increased ventro-medial prefrontal activations in schizophrenia smokers during cigarette cravings. Schizophr. Res 173, 30–36 (2016). [PubMed: 27005897]

74. Wiers CE et al. Effects of depressive symptoms and peripheral DAT methylation on neural reactivity to alcohol cues in alcoholism. Transl. Psychiatry 5, e648 (2015). [PubMed: 26418276]

75. Goudriaan AE et al. Neurophysiological effects of modafinil on cue-exposure in cocaine dependence: a randomized placebo-controlled cross-over study using pharmacological fMRI. Addict. Behav. 38, 1509–1517 (2013). [PubMed: 22591950]

76. McHugh RK et al. Cue-induced craving to paraphernalia and drug images in opioid dependence. Am. J. Addict. 25, 105–109 (2016). [PubMed: 26848719]

77. Clayton RB, Bailey RL & Liu J. Conditioned "cross fading": the incentive motivational effects of mediated-polysubstance pairings on alcohol, marijuana, and junk food craving. J. Health Commun 24, 319–327 (2019). [PubMed: 30963814]

78. Bach P. et al. The effects of single nucleotide polymorphisms in glutamatergic neurotransmission genes on neural response to alcohol cues and craving. Addict. Biol 20, 1022–1032 (2015). [PubMed: 26289945]

79. Blaine S. et al. TACR1 genotypes predict fMRI response to alcohol cues and level of alcohol dependence. Alcohol Clin. Exp. Res 37, E125–E130 (2013). [PubMed: 23078527]

80. Chen J. et al. CREB-BDNF pathway influences alcohol cue-elicited activation in drinkers. Hum. Brain Mapp 36, 3007–3019 (2015). [PubMed: 25939814]

81. Filbey FM et al. Differential neural response to alcohol priming and alcohol taste cues is associated with DRD4 VNTR and OPRM1 genotypes. Alcohol Clin. Exp. Res 32, 1113–1123 (2008). [PubMed: 18540916]

82. Janes AC et al. Association between CHRNA5 genetic variation at rs16969968 and brain reactivity to smoking images in nicotine dependent women. Drug Alcohol Depend 120, 7–13 (2012). [PubMed: 21764527]

83. Jorde A. et al. Genetic variation in the atrial natriuretic peptide transcription factor GATA4 modulates amygdala responsiveness in alcohol dependence. Biol. Psychiatry 75, 790–797 (2014). [PubMed: 24314346]

84. Kareken DA et al. A polymorphism in GABRA2 is associated with the medial frontal response to alcohol cues in an fMRI study. Alcohol Clin. Exp. Res 34, 2169–2178 (2010). [PubMed: 20698837]

85. Kuhn AB et al. FTO gene variant modulates the neural correlates of visual food perception. Neuroimage 128, 21–31 (2016). [PubMed: 26767945]

86. McClernon FJ et al. DRD4 VNTR polymorphism is associated with transient fMRI-BOLD responses to smoking cues. Psychopharmacology (Berl.) 194, 433–441 (2007). [PubMed: 17611740]

87. Moeller SJ et al. Gene x abstinence effects on drug cue reactivity in addiction: multimodal evidence. J. Neurosci 33, 10027–10036 (2013).

88. Schacht JP et al. Predictors of naltrexone response in a randomized trial: reward-related brain activation, OPRM1 genotype, and smoking status. Neuropsychopharmacology 42, 2640–2653 (2017). [PubMed: 28409564]

89. Schacht JP et al. Dopaminergic genetic variation influences aripiprazole effects on alcohol self-administration and the neural response to alcohol cues in a randomized trial. Neuropsychopharmacology 43, 1247–1256 (2018). [PubMed: 29362512]

90. Xu K. et al. A variant on the kappa opioid receptor gene (OPRK1) is associated with stress response and related drug craving, limbic brain activation and cocaine relapse risk. Transl. Psychiatry 3, e292 (2013). [PubMed: 23962922]

91. Yang BZ et al. A preliminary study of DBH (encoding dopamine beta-hydroxylase) genetic variation and neural correlates of emotional and motivational processing in individuals with and without pathological gambling. J. Behav. Addict 5, 282–292 (2016). [PubMed: 27194378]

92. Poldrack RA et al. Guidelines for reporting an fMRI study. Neuroimage 40, 409–414 (2008). [PubMed: 18191585]

93. Albrecht J. et al. Potential impact of a 32-channel receiving head coil technology on the results of a functional MRI paradigm. Clin. Neuroradiol 20, 223–229 (2010). [PubMed: 20857080]

94. Panman JL et al. Bias introduced by multiple head coils in MRI research: an 8 channel and 32 channel coil comparison. Front. Neurosci 13, 729 (2019). [PubMed: 31379483]

95. Colizoli O. et al. Comparing fMRI responses measured at 3 versus 7 Tesla across human cortex, striatum, and brainstem. Preprint at 10.1101/2020.05.12.090175v1 (2020).

96. Sacchet MD & Knutson B. Spatial smoothing systematically biases the localization of reward-related brain activity. Neuroimage 66, 270–277 (2013). [PubMed: 23110886]

97. Mayer AR et al. A comparison of denoising pipelines in high temporal resolution task-based functional magnetic resonance imaging data. Hum. Brain Mapp 40, 3843–3859 (2019). [PubMed: 31119818]

98. Poldrack RA et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci 18, 115–126 (2017). [PubMed: 28053326]

99. Lorenz RC et al. Cue reactivity and its inhibition in pathological computer game players. Addict. Biol 18, 134–146 (2013). [PubMed: 22970898]

100. Dale AM Optimal experimental design for event-related fMRI. Hum. Brain Mapp 8, 109–114 (1999). [PubMed: 10524601]

101. Josephs O. & Henson RN Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos. Trans. R Soc. Lond. B Biol. Sci 354, 1215–1228 (1999). [PubMed: 10466147]

102. Holla B. et al. Brain functional magnetic resonance imaging cue-reactivity can predict baclofen response in alcohol use disorders. Clin. Psychopharmacol. Neurosci 16, 290–301 (2018). [PubMed: 30121979]

103. Vollstädt-Klein S. et al. Initial, habitual and compulsive alcohol use is characterized by a shift of cue processing from ventral to dorsal striatum. Addiction 105, 1741–1749 (2010). [PubMed: 20670348]

104. Li X. et al. The neural mechanisms of immediate and follow-up of the treatment effect of hypnosis on smoking craving. Brain Imaging Behav 14, 1487–1497 (2020). [PubMed: 31028603]

105. Gusnard DA, Raichle ME & Raichle ME Searching for a baseline: functional imaging and the resting human brain. Nat. Rev. Neurosci 2, 685–694 (2001). [PubMed: 11584306]

106. Janes AC et al. Quitting starts in the brain: a randomized controlled trial of app-based mindfulness shows decreases in neural responses to smoking cues that predict reductions in smoking. Neuropsychopharmacology 44, 1631–1638 (2019). [PubMed: 31039580]

107. Dean AC et al. No effect of attentional bias modification training in methamphetamine users receiving residential treatment. Psychopharmacology 236, 709–721 (2019). [PubMed: 30415277]

108. Kang O-S et al. Individual differences in smoking-related cue reactivity in smokers: an eye-tracking and fMRI study. Prog. Neuropsychopharmacol. Biol. Psychiatry 38, 285–293 (2012). [PubMed: 22542509]

109. Hanlon CA et al. Cortical substrates of cue-reactivity in multiple substance dependent populations: transdiagnostic relevance of the medial prefrontal cortex. Transl. Psychiatry 8, 186 (2018). [PubMed: 30194288]

110. Mondino M. et al. Effects of repeated transcranial direct current stimulation on smoking, craving and brain reactivity to smoking cues. Sci. Rep 8, 8724 (2018). [PubMed: 29880873]

111. Yalachkov Y. et al. Sensory modality of smoking cues modulates neural cue reactivity. Psychopharmacology (Berl.) 225, 461–471 (2013). [PubMed: 22890475]

112. Manoliu A. et al. SmoCuDa: a validated smoking cue database to reliably induce craving in tobacco use disorder. Eur. Addict. Res 27, 107–114 (2021). [PubMed: 32854096]

113. Macatee RJ et al. Development and validation of a cannabis cue stimulus set. Addict. Behav 112, 106643 (2021).

114. Stritzke WG et al. Assessment of substance cue reactivity: advances in reliability, specificity, and validity. Psychol. Addict. Behav 18, 148–159 (2004). [PubMed: 15238057]

115. Zeng H. et al. The action representation elicited by different types of drug-related cues in heroin-abstinent individuals. Front. Behav. Neurosci 12, 123 (2018). [PubMed: 30013467]

116. Lindsey KP et al. Nicotine content and abstinence state have different effects on subjective ratings of positive versus negative reinforcement from smoking. Pharmacol. Biochem. Behav 103, 710–716 (2013). [PubMed: 23219727]

117. Schacht JP, Anton RF & Myrick H. Functional neuroimaging studies of alcohol cue reactivity: a quantitative meta-analysis and systematic review. Addict. Biol 18, 121–133 (2013). [PubMed: 22574861]

118. Wall MB et al. Investigating the neural correlates of smoking: feasibility and results of combining electronic cigarettes with fMRI. Sci. Rep 7, 1–8 (2017). [PubMed: 28127051]

119. Ekhtiari H. et al. It is never as good the second time around: brain areas involved in salience processing habituate during repeated drug cue exposure in treatment engaged abstinent methamphetamine and opioid users. Neuroimage 238, 118180 (2021).

120. Seo D. et al. Disrupted ventromedial prefrontal function, alcohol craving, and subsequent relapse risk. JAMA Psychiatry 70, 727–739 (2013). [PubMed: 23636842]

121. Chua HF et al. Self-related neural response to tailored smoking-cessation messages predicts quitting. Nat. Neurosci 14, 426–427 (2011). [PubMed: 21358641]

122. McClernon FJ et al. Hippocampal and insular response to smoking-related environments: neuroimaging evidence for drugcontext effects in nicotine dependence. Neuropsychopharmacology 41, 877–885 (2016). [PubMed: 26179147]

123. McClernon FJ et al. Abstinence-induced changes in self-report craving correlate with event-related FMRI responses to smoking cues. Neuropsychopharmacology 30, 1940–1947 (2005). [PubMed: 15920499]

124. Li Q. et al. Craving correlates with mesolimbic responses to heroin-related cues in short-term abstinence from heroin: an event-related fMRI study. Brain Res 1469, 63–72 (2012). [PubMed: 22759909]

125. Kleykamp BA et al. Craving and opioid use disorder: a scoping review. Drug Alcohol Depend 205, 107639 (2019).

126. Ekhtiari H. Craving as an Outcome Measure in Clinical Trials: A Systematic Review on Craving Assessment Instruments in Clinical Trials for Substance Use Disorders. Accessed 25 Aug 2021. https://osf.io/vk9ug/ (2021).

127. Heishman SJ et al. Prolonged duration of craving, mood, and autonomic responses elicited by cues and imagery in smokers: effects of tobacco deprivation and sex. Exp. Clin. Psychopharmacol. 18, 245 (2010). [PubMed: 20545389]

128. Heishman SJ, Saha S. & Singleton EG Imagery-induced tobacco craving: duration and lack of assessment reactivity bias. Psychol. Addict. Behav 18, 284–288 (2004). [PubMed: 15482084]

129. Franken IH, Hendriks VM & Van den Brink W. Initial validation of two opiate craving questionnaires: The Obsessive Compulsive Drug Use Scale and the Desires for Drug Questionnaire. Addict. Behav 27, 675–685 (2002). [PubMed: 12201376]

130. Heishman SJ, Singleton EG & Liguori A. Marijuana Craving Questionnaire: development and initial validation of a self-report instrument. Addiction 96, 1023–1034 (2001). [PubMed: 11440613]

131. Tiffany ST & Drobes DJ The development and initial validation of a questionnaire on smoking urges. Br. J. Addict. 86, 1467–1476 (1991). [PubMed: 1777741]

132. Kozlowski LT et al. "Cravings" are amibiguous: ask about urges or desires. Addict. Behav 14, 443–445 (1989). [PubMed: 2782125]

133. Kozlowski LT & Wilkinson DA Use and misuse of the concept of craving by alcohol, tobacco, and drug researchers. Br. J. Addict 82, 31–36 (1987). [PubMed: 3470042]

134. Wilson SJ & Sayette MA Neuroimaging craving: urge intensity matters. Addiction 110, 195–203 (2015). [PubMed: 25073979]

135. Plant RR, Hammond N. & Whitehouse T. How choice of mouse may affect response timing in psychological studies. Behav. Res. Methods Instrum. Comput 35, 276–284 (2003). [PubMed: 12834085]

136. Plant RR & Turner G. Millisecond precision psychological research in a world of commodity computers: new hardware, new problems? Behav. Res. Methods 41, 598–614 (2009). [PubMed: 19587169]

137. Segalowitz SJ & Graves RE Suitability of the IBM XT, AT, and PS/2 keyboard, mouse, and game port as response devices in reaction time paradigms. Behav. Res. Methods Instrum. Comput 22, 283–289 (1990).

138. Schwarz AJ et al. A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #1: processes and requirements. Drug Discov. Today 16, 583–593 (2011). [PubMed: 21635967]

139. Sutton BP et al. Investigation and validation of intersite fMRI studies using the same imaging hardware. J. Magn. Reson. Imaging 28, 21–28 (2008). [PubMed: 18581342]

140. Carter BL & Tiffany ST Meta-analysis of cue-reactivity in addiction research. Addiction 94, 327–340 (1999). [PubMed: 10605857]

141. Chang C, Cunningham JP & Glover GH Influence of heart rate on the BOLD signal: the cardiac response function. Neuroimage 44, 857–869 (2009). [PubMed: 18951982]

142. Gloria R. et al. An fMRI investigation of the impact of withdrawal on regional brain activity during nicotine anticipation. Psychophysiology 46, 681–693 (2009). [PubMed: 19490513]

143. Janes AC et al. Brain fMRI reactivity to smoking-related images before and during extended smoking abstinence. Exp. Clin. Psychopharmacol 17, 365–373 (2009). [PubMed: 19968401]

144. Lou M. et al. Cue-elicited craving in heroin addicts at different abstinent time: an fMRI pilot study. Subst. Use Misuse 47, 631–639 (2012). [PubMed: 22329835]

145. McClernon FJ et al. 24-h smoking abstinence potentiates fMRI-BOLD activation to smoking cues in cerebral cortex and dorsal striatum. Psychopharmacology (Berl.) 204, 25–35 (2009). [PubMed: 19107465]

146. Parvaz MA, Moeller SJ & Goldstein RZ Incubation of cue-induced craving in adults addicted to cocaine measured by electroencephalography. JAMA Psychiatry 73, 1127–1134 (2016). [PubMed: 27603142]

147. Lu L. et al. Incubation of cocaine craving after withdrawal: a review of preclinical data. Neuropharmacology 47, 214–226 (2004). [PubMed: 15464139]

148. Bailey SR, Goedeker KC & Tiffany ST The impact of cigarette deprivation and cigarette availability on cue-reactivity in smokers. Addiction 105, 364–372 (2010). [PubMed: 19922514]

149. Fischer H. et al. Brain habituation during repeated exposure to fearful and neutral faces: a functional MRI study. Brain Res. Bull 59, 387–392 (2003). [PubMed: 12507690]

150. Siegle GJ et al. Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. Biol. Psychiatry 51, 693–707 (2002). [PubMed: 11983183]

151. Franklin T. et al. Effects of varenicline on smoking cue-triggered neural and craving responses. Arch. Gen. Psychiatry 68, 516–526 (2011). [PubMed: 21199958]

152. LaRowe SD et al. Reactivity to nicotine cues over repeated cue reactivity sessions. Addict. Behav 32, 2888–2899 (2007). [PubMed: 17537583]

153. Schacht JP et al. Stability of fMRI striatal response to alcohol cues: a hierarchical linear modeling approach. Neuroimage 56, 61–68 (2011). [PubMed: 21316465]

154. McBride D. et al. Effects of expectancy and abstinence on the neural response to smoking cues in cigarette smokers: an fMRI study. Neuropsychopharmacology 31, 2728–2738 (2006). [PubMed: 16598192]

155. Perry RN et al. The impacts of actual and perceived nicotine administration on insula functional connectivity with the anterior cingulate cortex and nucleus accumbens. J. Psychopharmacol 33, 1600–1609 (2019). [PubMed: 31542980]

156. Gu X. et al. Belief about nicotine modulates subjective craving and insula activity in deprived smokers. Front. Psychiatry 7, 126 (2016). [PubMed: 27468271]
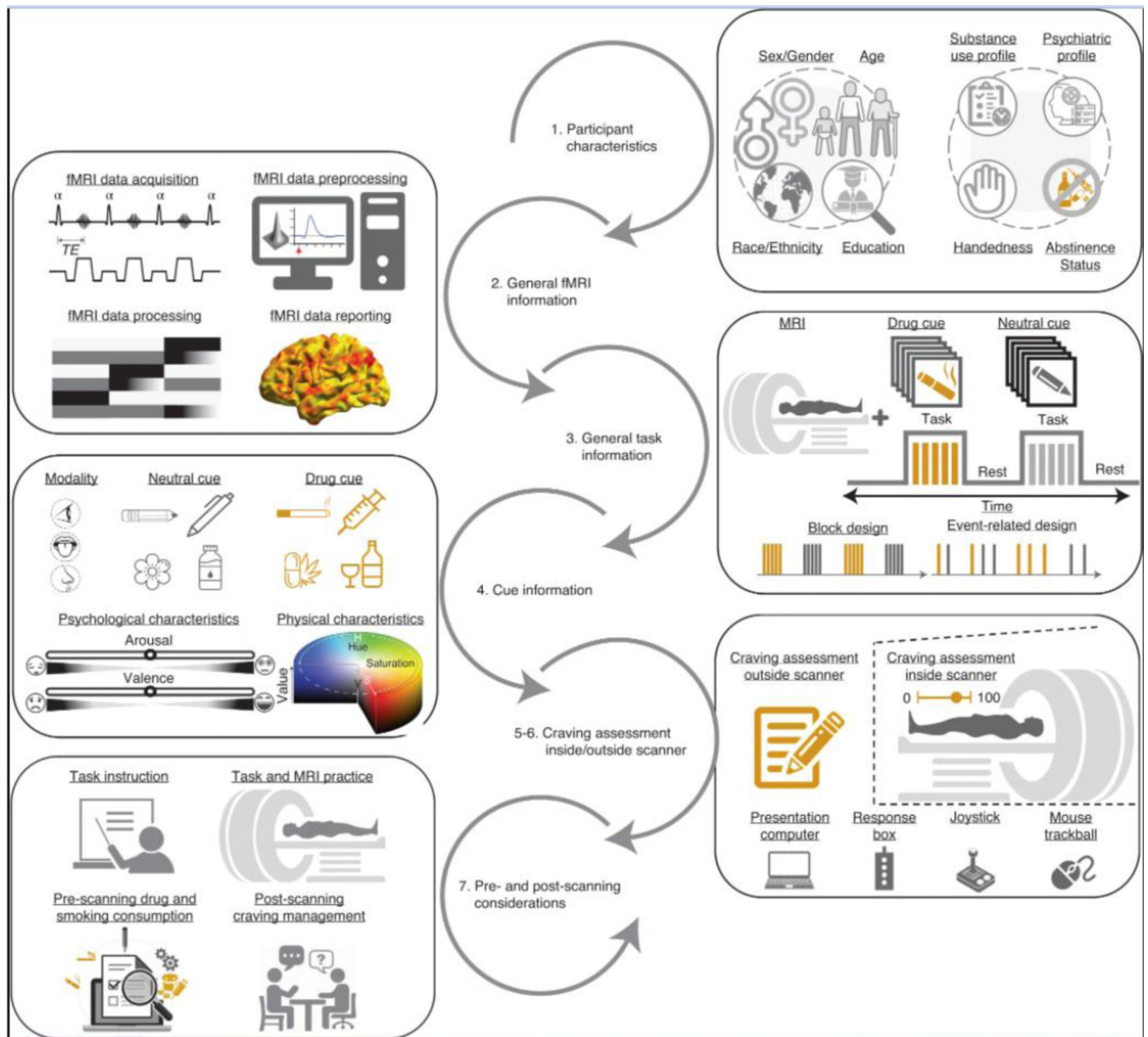
Fig. 1 |. Schematic representation of key reportable aspects of an fMRI drug cue reactivity study.
1. Participants are recruited on the basis of explicit criteria, and baseline data are collected
on participant demographics, handedness, psychiatric history and substance use history.
2. Participants undergo fMRI scanning with carefully selected hardware and software
parameters, and data are analyzed through specified preprocessing and analysis pipelines
for statistical inference. 3. Participants engage with drug and neutral cues during fMRI
scanning, with cues of specified durations presented in events and/or blocks with a chosen
temporal architecture. 4. These cues stimulate one or more sensory modalities and are
typically matched in terms of psychological characteristics, such as induced arousal or
valence, and/or physical characteristics, such as saturation and hue for pictorial cues. 5.
and 6. Participants provide craving self-reports outside and/or inside the scanner, using

various short and long-form instruments and hardware such as response boxes or joysticks.
7. In addition to pre-scanning sources of between-study variance such as task instructions and scanner familiarization, there are important post-scanning safety procedures such as craving-management interventions and additional assessments before participants leave the imaging center.
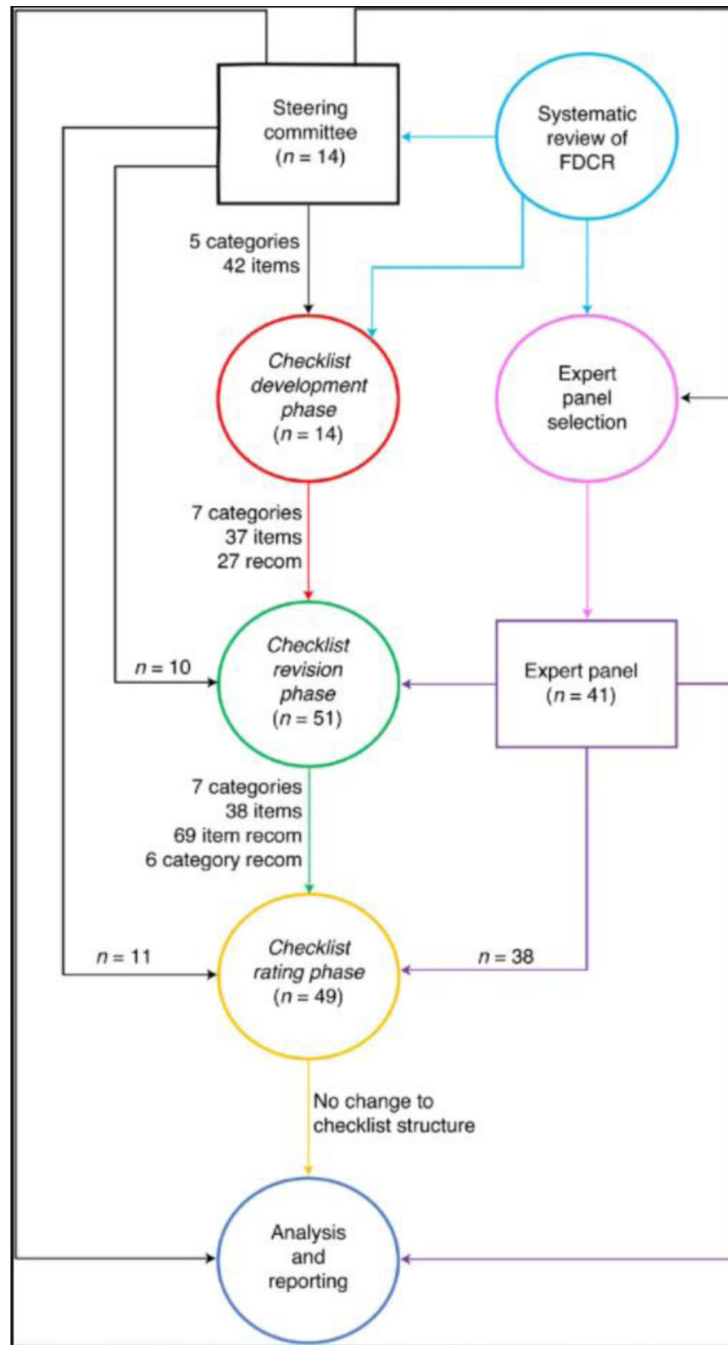
Fig. 2 |. A schematic of the entire Delphi study methodology.

The process has been roughly divided into distinct stages: the selection of the SC (in black) using the results of an earlier mentioned systematic review to choose the initial checklist items and expert committee candidates (in pink), checklist development phase (in red), expert panel selection (in purple), checklist commenting and revision phase (in green), checklist rating phase (in yellow) and data analysis and Delphi process finalization (in blue). The number of contributors to each section is displayed by '$n =$'. To the left of the

main graph, an overview of the structure of the checklist at each stage is presented. recom, recommendations.
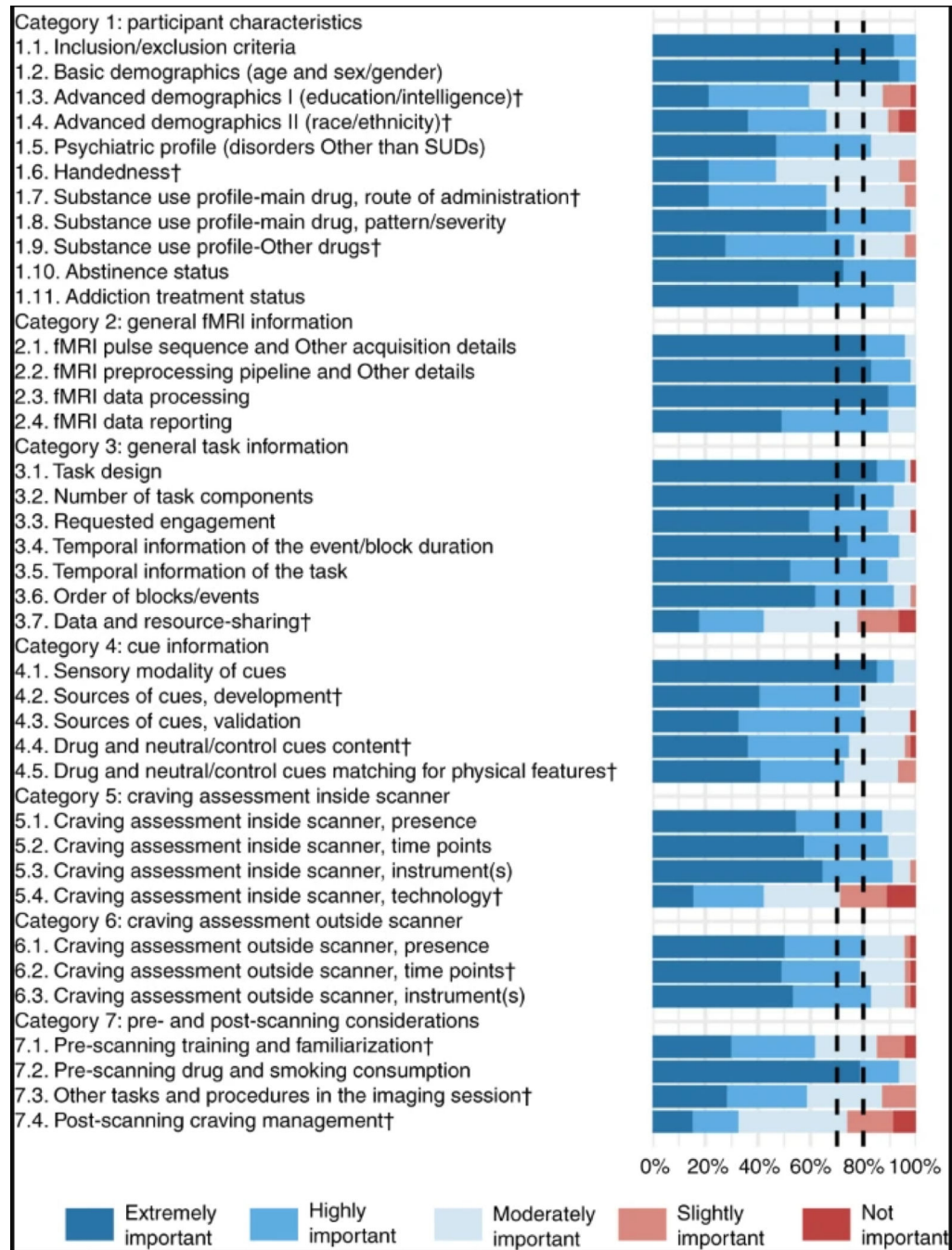
Fig. 3 |. Ratings for 38 items in seven categories.
This figure depicts the rating of 49 raters (11 from the steering committee and 38 from the expert panel) for the checklist items. Each item was rated from 1 to 5 (not important to extremely important). All the items met threshold 1 and were rated as moderately, highly or extremely important by >70% of the raters. In addition, 24 items reached the more-stringent threshold 2 of being rated as either highly or extremely important by 80% of raters (the ones that did not reach this threshold are marked with '†'). Items are represented by their summary in the figure. Full text of the items is provided in Tables 1–6.
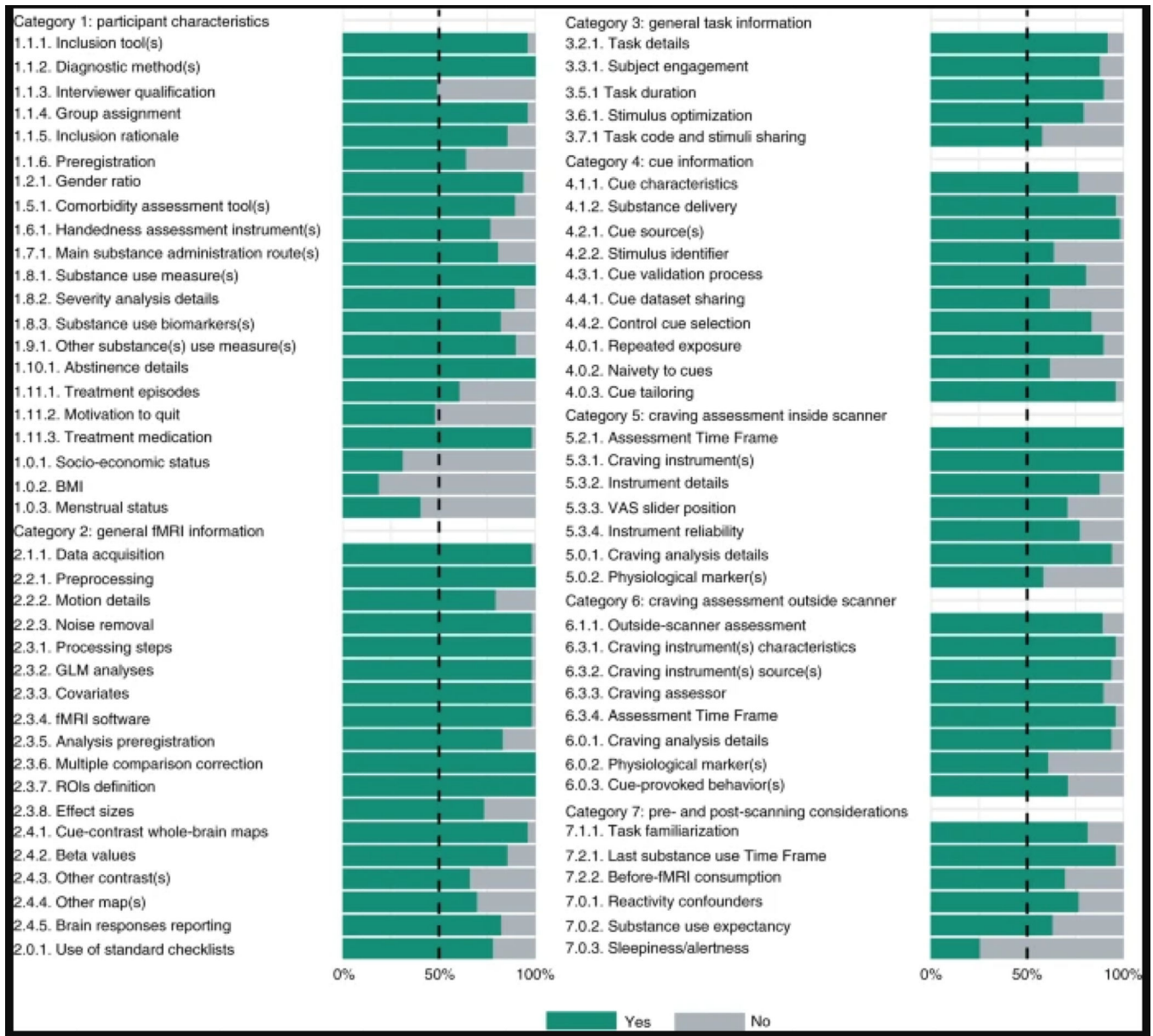
Fig. 4 |. Ratings for 75 additional recommendations in seven categories.
This figure depicts the rating of 49 raters (11 from the steering committee and 38 from the expert panel) for the checklist additional recommendations. Each additional recommendation was rated either 'Yes' or 'No' on the question of whether it should be included as a recommendation. Recommendations are represented by their summary in the figure. Full text of the recommendations is provided in Tables 1–6.
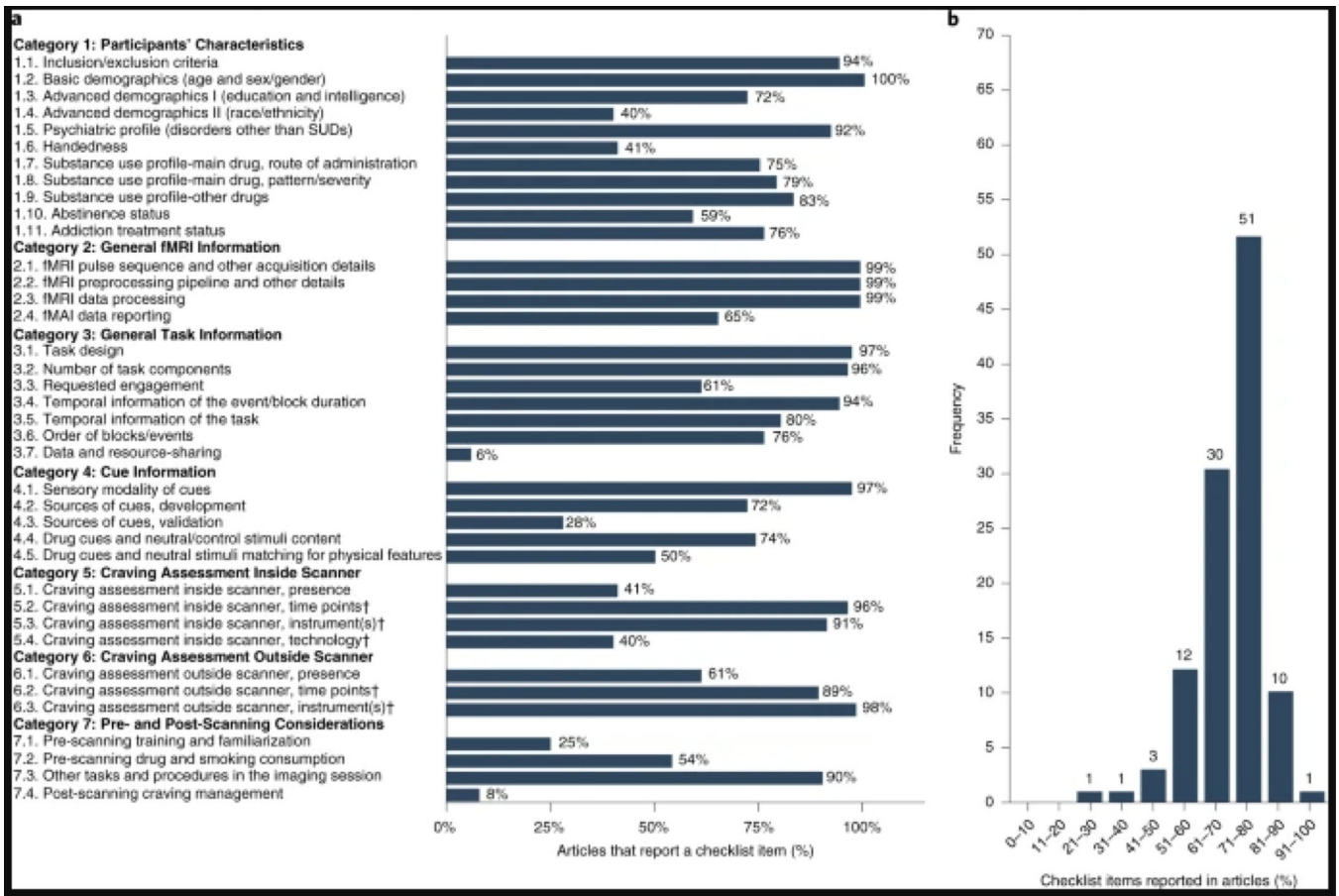
Fig. 5 |. State of reproducibility/transparency in fMRI drug cue reactivity research in the context of the ENIGMA-ACRI checklist.

Assessments by three independent raters on the basis of 108 FDCR articles. **a**, Percentage of articles that reported each checklist item. Note that the percentages are calculated out of applicable items for each article. For example, craving-rating technology was not applicable for an article without craving rating. **b**, Percentage of overall reporting status of articles.

Table 1 |

Items to report and recommendations in the Participants' Characteristics category (category 1) of the checklist

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
Inclusion/Exclusion Criteria	1.1. Inclusion and exclusion criteria for all participant groups	4.91 (0.28)	1.1.1. Include specific diagnostic criteria/ measurement tools for conditions that were included and those that were excluded	46 (96%)
			1.1.2. Clearly specify methods used to assess any diagnostic/dimensional criteria (e.g., SCID, MINI and their versions)	48 (100%)
			1.1.3. Report the qualification of the person who has applied these criteria (e.g., clinical psychologist, institute secretary or psychiatrist)	23 (49%)
			1.1.4. Report how participants were assigned to different groups in studies in which participants are assigned to more than one group	46 (96%)
			1.1.5. Explain the rationale for the criteria selected for recruitment (e.g., if only males are included)	41 (85%)
			1.1.6. Report whether methods for any additional subgroups and adjusted analyses were preregistered before or not (e.g., protocol paper, registration websites)	30 (64%)
Basic Demographics	1.2. Age and sex/gender for all participant groups	3.83 (0.82)	1.2.1. Report the number of males/ females in the sample included in the reported analyses. There are studies that have reported the ratio in the recruited sample without reporting the ratio in the sample included in the analyses	44 (94%)
Advanced Demographics I	1.3. Education or a measurement of intelligence for all participant groups	4.64 (0.53)	-	-
Advanced Demographics II	1.4. Race or ethnicity for all participant groups	4 (0.81)	-	-
Psychiatric Profile	1.5. Any categorical or dimensional measurement of psychopathologies other than substance use disorder	4.72 (0.45)	1.5.1. Report psychiatric comorbidities using diagnostic criteria (e.g., DSM) or questionnaires to assess the level of psychiatric comorbidities (e.g., a quantitative assessment of depression or anxiety using various questionnaires)	42 (89%)
Handedness	1.6. Handedness for all participant groups	4.47 (0.65)	1.6.1. Use validated handedness inventories like the Edinburgh Handedness Inventory. The effect of handedness in the laterality of fMRI drug- cue reactivity and its significance is still unclear. However, this effect can be explored with reproducible reporting of the handedness in the shared databases	36 (77%)
Substance Use Profile-Main Drug	1.7. Route(s) of administration for the main substance (if it is obvious, it does not need to be reported; i.e., there is only one route of administration for cigarette smokers or alcohol drinkers)	4.91 (0.28)	1.7.1. Report the breakdown of the main drug by type and route	37 (80%)
	1.8. Current and lifetime use pattern/ severity for the main drug of use for all participant groups	4.94 (0.25)	1.8.1. Report the exact measures and instruments used to assess current (e.g., last few days, last month, last 3 months) and lifetime substance use (e.g., questions, questionnaires or laboratory tests)	46 (100%)
			1.8.2 Report whether/how derived variables from these severity measures have been used in fMRI drug-cue reactivity analysis (e.g., whether they are used as variables of interest or as a regressed-out variable)	41 (89%)
			1.8.3 Include biological markers of drug use/severity (if available)	36 (82%)

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
Substance Use Profile-Other Drugs	1.9. Measures of current or lifetime use pattern/severity for drugs, other than the main drug of use, for all participant groups	4.3 (0.75)	1.9.1. Report the current and lifetime patterns and severity of use of other substances and potential use disorders	43 (90%)
Abstinence Status	1.10. Days/hours/minutes since last use (duration of abstinence) and how abstinence was verified for all participant groups	3.62 (0.9)	1.10.1. Report a clear definition of abstinence, its assessment methods (e.g., timeline follow-back, urine toxicology, monitoring (e.g., breathalyzer or CO measures), clinical interviews) and the reference time point (i.e., recruitment or scanning)	47 (100%)
Addiction Treatment Status	1.11. Treatment status for all participant groups (e.g., non-treatment-seeking active users, treatment-seeking active users, undergoing active treatment, treated and abstinent, relapsed after treatment)	3.83 (0.82)	1.11.1 Specify the number and the nature of treatment episodes if participants have undergone multiple unsuccessful treatment episodes	29 (60%)
			1.11.2 Report the level of motivation to discontinue substance use for active drug users	22 (48%)
			1.11.3 Report whether they are on medication to treat their SUD	47 (98%)
General Recommendations	-	-	1.0.1. Probe and report a measure of income or sociodemographic status; however, the effect of this demographic dimension in fMRI drug-cue reactivity is not explored yet	15 (31%)
			1.0.2. Report BMI for all participant groups	9 (19%)
			1.0.3. Report the menstrual status (e.g., days since the first day of last menstrual period or menstrual phase/ status) in female participants	19 (40%)

Ratings for items (1–5) are reported as mean (s.d.) in the 'Item importance' column, and ratings for recommendations (Yes) are reported as frequency (percentage%) of 'Yes' reports in the 'Recommendation inclusion' column. BMI, body mass index; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; MINI, Mini International Neuropsychiatric Interview; SCID, Structured Clinical Interview for DSM-5.

Table 2 |

Items to report and recommendations in the General fMRI Information category (category 2) of the checklist

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
fMRI Pulse Sequence and Other Acquisition Details	2.1. fMRI data-acquisition details	4.77 (0.52)	2.1.1. Report fMRI data acquisition details based on the available checklists (e.g., COBIDAS). fMRI data acquisition details might have explicit effects on drug-cue reactivity results, for example, number of head coil channels, because higher channels (32 compared to 8) might be associated with better SNR in cortex, with the cost of losing signal in the deep parts of the brain	47 (98%)
fMRI Preprocessing Pipeline and Other Details	2.2. fMRI preprocessing details	4.81 (0.45)	2.2.1. Report fMRI preprocessing details based on the available checklists (e.g., COBIDAS). There are items in the preprocessing steps that might have an effect on fMRI drug-cue reactivity results. For example, higher FWHM might be related to the loss of signal in small nuclei 2.2.2 Report motion differences between participant groups (i.e., individuals with an SUD vs. controls), because higher motion during the drug-related blocks compared to neutral blocks might act as a confounder 2.2.3 Report quality-control measures, artefact detection methods and the threshold to exclude participants with heavy movement	48 (100%) 38 (79%) 47 (98%)
fMRI Data Processing	2.3. fMRI analyses and statistical modeling details	4.89 (0.31)	2.3.1. Report fMRI single-subject-level and group-level processing steps on the basis of the standard checklists (e.g., COBIDAS) 2.3.2. Report whether GLM analyses are random, mixed, or fixed effects for inclusion in future metaanalyses 2.3.3 Report all covariates used for each model and whether or not demeaning was done for covariates of interest 2.3.4 Report any publicly available tool/software use (e.g., SPM, AFNI and FSL) 2.3.5 Report any attempt for preregistration of data processing methods 2.3.6 Report methods that are used to control for multiple comparisons error and spatial autocorrelations 2.3.7 Report the definition of the ROIs for studies using an ROI approach 2.3.8 Provide effect sizes for all reported statistics	47 (98%) 47 (98%) 47 (98%) 47 (98%) 39 (83%) 47 (100%) 47 (100%) 33 (73%)
fMRI Data Reporting	2.4. Basic whole-brain response to drug cues	4.38 (0.68)	2.4.1 Report the second-level maps or activation foci therein of each study group singly, as well as a group-difference map (e.g., between the clinical group and the control group) (if applicable) in the results or the supplements as a figure or table (foci coordinates and stats) with details on the thresholding measures and quantities. Even if the paper has other analyses (e.g., task-based connectivity), the whole-brain maps of the craving>neutral contrast should be reported for comparison with other studies and future metaanalyses 2.4.2 Report beta values for both conditions (craving and neutral), because an 'activation' in the mPFC during craving could be explained by a de-activation in the control condition 2.4.3 Report the contrast map for other included conditions (e.g., multiple drug stimuli, affective images and other active control) if other conditions are included 2.4.4. Provide the effect size map, the non-thresholded statistical map and the data in an accessible repository (e.g., OSF, NIMH/NIAAA data archive, GitHub or Neurovault) 2.4.5. It is understandable that researchers who are not using conventional whole-brain GLM-based methods (e.g., ICA, Graph Theory, PPI connectivity, ROI only analysis) or developing other innovative and non-conventional methods might face difficulties in reporting 'whole-brain response to drug cues'. It is still recommended for these studies to consider strategies for reporting whole-brain responses to drug cues to make data/results aggregation and comparison possible	45 (96%) 41 (85%) 31 (66%) 32 (70%) 37 (82%)
General Recommendations	-	-	2.0.1. Refer to standard checklists (e.g., COBIDAS) for items in this category. Items in the ENIGMA ACRI checklist are designed to be dichotomous (Yes or No); however, there is a continuum for the details to be reported. Provide as much detail as available	28 (78%)

Ratings for items (1–5) are reported as mean (s.d.) in the 'Item importance' column, and ratings for recommendations (Yes) are reported as frequency (percentage%) of 'Yes' reports in the 'Recommendation inclusion' column. AFNI, Analysis of Functional Neuroimages; FSL, FMRIB Software Library; GLM, generalized linear model; ICA, independent component analysis; mPFC,

medial prefrontal cortex; NIAAA, National Institute on Alcohol Abuse and Alcoholism; NIMH, National Institute of Mental Health; OSF, Open Science Foundation; PPI, psychophysiological interaction; ROI, region of interest; SNR, signal-to-noise ratio; SPM, statistical parametric mapping.

Table 3 |

Items to report and recommendations in the General Task Information category (category 3) of the checklist

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
Task Design	3.1. Task structure (event, block or mixed (events in blocks))	4.77 (0.7)	-	-
Number of Task Components	3.2. Number of runs (if more than one), blocks (for block-designed studies),and events (including drug cues, control cues, fixations, etc.)	4.68 (0.63)	3.2.1. Explicitly define terms such as 'block', 'event', 'session', and 'run' with reference to standard checklists (e.g., COBIDAS), given the ambiguity surrounding these terms	44 (92%)
Requested Engagement	3.3. Instructions to the study participants on how to engage with the cues	4.45 (0.83)	3.3.1. Report the details of the given instructions on how to engage (interact) with cues and provide the exact text of the instruction. The interactions may be passive viewing (if there was explicitly no instruction or if they were asked to do nothing), free craving, attentive viewing, rating or classifying each cue, spatial cueing, inhibiting craving, etc	42 (88%)
Temporal Information of the Event/Block Duration	3.4. Duration of each cue (for both event and blocked-design tasks) *and* the total block duration (for blocked-design tasks)	4.67 (0.6)	-	-
Temporal Information of the Task	3.5. Total task duration	4.41 (0.69)	3.5.1. Report the duration of all sections of the task between the cues/events/blocks and within them	43 (90%)
Order of Blocks/Events	3.6. Order of block types (e.g., drug and control) (for blocked-designs) or event types (e.g., drug and control) (for event-related designs) (The order can be fully randomized (randomized and different between subjects), pseudorandomized (identical between subjects, but randomized once for the order of events/blocks) or not randomized (fixed order like neutral-drug-neutral-drug for all subjects)	4.51 (0.72)	3.6.1. Report if the stimulus presentation was optimized using any software (e.g., genetic algorithm or optseq)	38 (79%)
Data and Resource-Sharing	3.7. Sharing the behavioral task code or source images	3.31 (1.14)	3.7.1. Provide the task code and the code used for generating these sequences (i.e., GitHub or OSF platforms)	26 (58%)

Ratings for items (1–5) are reported as mean (s.d.) in the 'Item importance' column, and ratings for recommendations (Yes) are reported as frequency (percentage%) of 'Yes' reports in the 'Recommendation inclusion' column.

Table 4 |

Items to report and recommendations in the Cue Information category (category 4) of the checklist

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
Sensory Modality of Cues	4.1. Modality(ies) of utilized drug and neutral/control cues (The modalities can be word, picture, smell, taste, tactile, audio script, written script, imagination, silent video, audiovisual video, paraphernalia, substance itself or mixed.)	4.77 (0.6)	4.1.1. Provide an overview of the range of values for important characteristics of chosen cues. In the case of visual cues, this could be in the form of describing the complexity, luminance, and hue of cues. For auditory cues, this could consist of describing the volume and frequency, and for scripts, it could be font and typeface 4.1.2. Report the amount of the substance and its method of delivery (i.e., oral or i.v.), if the substance itself is administered as a cue (e.g., very small amounts of alcohol or cigarette smoke)	36 (77%) 46 (96%)
Sources of Cues, Development	4.2. Source of drug and neutral/control cues	4.09 (0.84)	4.2.1. Report the exact source of acquiring the cues if the cues are newly developed, or cite the relevant references if they are from other already developed sources. If the stimulus set is newly developed, criteria used for stimulus selection should be specified (e.g., exclusion of people in images and paraphernalia only). If a subset of developed sources was used, indicate what criteria were used for selecting this subset (could be a random selection) 4.2.2. List the stimulus identifiers in the appendix or supplementary material of the paper, if the cue sources include stimulus identifiers	47 (98%) 30 (64%)
Sources of Cues, Validation	4.3. Extent of prior validation of drug and neutral/control cues used in the study (Drug and neutral/ control cues in a study might be not validated, validated by assessing the craving induction of each cue individually using simple-item craving instruments like single-item VAS, or using standardized instruments of craving assessment and emotion or stress reactivity)	4.07 (0.95)	4.3.1. Provide the details of the validation process. Even if the validation has been done in another study, the validation study should be cited, and then the validation process of the cues should be briefly introduced as well	37 (80%)
Drug and Neutral/ Control Cue Content	4.4. Content of drug cues and its relationship to the targeted drug (These include stimulus related to the drug, stimulus related to instruments of drug use, stimulus related to various stages of drug use (e.g., 'beginning' or 'end' stimuli (lit cigarette vs. ashtray)), stimulus related to drug intake, stimulus related to typical drug-consumption environments, stimulus related to preparation of drug, stimulus related to purchasing the drug)	4.07 (0.93)	4.4.1. Explicitly report if they are willing to share their drug and neutral/control cue database/task in the published paper. Providing a reliable link (like GitHub or other open science repositories) to a shared database inside the paper is the ideal scenario; however, facing copyright concerns for the drug cues collected from the web or other copyright-protected resources might limit this potential. All too often, links are provided in papers that are broken a few years after publication 4.4.2. Explain the nature of neutral/ control cues and why they were chosen, as they might belong to several types in terms of their content	29 (62%) 40 (83%)
Drug to Neutral/ Control Cues Matching for Physical Features	4.5. Factors for which drug and neutral/ control cues have been matched (e.g., color, brightness, hue, content, complexity, scrambled drug cue)	4.09 (0.84)	-	-
General Recommendations			4.0.1. Report the characteristics of the cue sets used when a task is repeated if a study involves a longitudinal design 4.0.2. Control and report being naïve to drug-cue exposure or previous experiences	42 (89%) 29 (62%) 45 (96%)

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
			of cue exposure before the target study. Recent evidence shows that participants will respond differently to drug cues in the second exposure. However, asking people to report cue exposure outside of the target study might be complex 4.0.3. Report whether and how drug and neutral/control cues were tailored for each participant. Drug -and neutral/control-cue tailoring could involve asking participants to choose cues from a cue database or developing participant-specific cues based on consultation with individual participants. Details of the individualization protocol should be provided	

Ratings for items (1–5) are reported as mean (s.d.) in the 'Item importance' column, and ratings for recommendations (Yes) are reported as frequency (percentage%) of 'Yes' reports in the 'Recommendation inclusion' column. VAS, visual analog scale.

Table 5 |

Items to report and recommendations in the Craving Assessment categories (categories 5 and 6) of the checklist

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
Craving Assessment Inside Scanner, Presence	5.1. Craving assessment inside the scanner	4.41 (0.72)	-	-
Craving Assessment Inside Scanner, Time Points	5.2. Description of the time points at which craving-related a assessment is performed inside the scanner (e.g., before and/or after each cue/event/block/trial/scan/ run/session) (Yes/No/Not Applicable (in cases when there is no assessment inside the scanner))	4.47 (0.69)	5.2.1. Report the time frame of craving ssessment (i.e., now (after cue presentation) or during cue presentation)	48 (100%)
Craving Assessment Inside Scanner, Instrument(s)	5.3. Description of the instrument (s) used to assess craving and craving-related constructs inside the scanner (Yes/No/Not Applicable (in cases when there is no assessment inside the scanner))	4.53 (0.73)	5.3.1. Report the exact characteristics of the instrument(s) used to assess craving and craving constructs (e.g., urge, desire, interest, like vs. want) inside the scanner, including number of items, range of possible responses, whether it was VAS or Likert, internal consistency and whether any transformations were applied to the instrument and its scores before the data collection and analysis 5.3.2. Cite any relevant sources of instruments, and whenever possible, provide the exact wording of the craving question(s) 5.3.3. Provide information on the start position of the slider, when using VAS or other continuous scales with a slider (e.g., in the middle or lateral ends of the scales) 5.3.4. Report information on the reliability of the instrument if the instrument(s) was administered repeatedly before/during/after scanning	48 (100%) 42 (88%) 34 (71%) 37 (77%)
Craving Assessment Inside Scanner, Technology	5.4. Description of the hardware used to obtain participant responses, with specifications of models and brands of devices, if necessary (e.g., response box and fiber-optic pad) (Yes/No/Not Applicable (in cases when there is no assessment inside the scanner))	4.53 (0.73)	-	-
General Recommendations			5.0.1. Report analyses related to the craving measurements, i.e., whether they differed between the main group and control(s) or from pre- to post-scan 5.0.2. Probe and report physiological correlates of craving (i.e., skin conductance, heart rate, temperature, respiration and blood volume pulse amplitude) before/during/after cue presentation	45 (94%) 28 (58%)
Craving Assessment Outside Scanner, Presence	6.1. Any craving-related assessment outside the scanner	4.24 (0.95)	6.1.1. Probe and report craving assessment outside or inside the scanner in FDCR tasks. The assessment of cue-induced craving is of great relevance to the validity of the FDCR task. Thus, the authors should at least clarify whether they have considered including a craving assessment inside/ outside the scanner, even if they have finally decided not to report the results	40 (89%)
Craving Assessment Outside Scanner, Time Points	6.2. Description of the time points at which craving-related assessment is performed outside the scanner (e.g., immediately before entering the scanner or immediately after exiting the scanner) (Yes/No/Not	4.21 (0.95)	-	-

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
	Applicable (in cases when there is no assessment outside the scanner))			
Craving Assessment Outside Scanner, Instrument(s)	6.3. Description of the instrument (s) used to assess craving and craving-related constructs outside the scanner (Yes/No/Not Applicable (in cases when there is no assessment outside the scanner))	4.3 (0.93)	6.3.1. Report the exact characteristics of the instrument(s) used to assess craving and craving constructs (e.g., urge, desire, interest, like vs. want) outside the scanner, including number of items, range of responses, internal consistency whether it was VAS or Likert and whether any transformations were applied to the instrument and its scores before the data collection and analysis	45 (96%) 44 (94%) 42 (89%) 45 (96%)
			6.3.2 Cite any relevant sources of instruments, and whenever possible, provide the exact wording of the craving question(s)	
			6.3.3 Report that the instrument is self-assessed or experimenter assessed	
			6.3.4 Report the time frame of craving assessment (i.e., now or during the scan)	
General Recommendations			6.0.1. Report analyses related to the craving measurements, i.e., whether they differed between the main group and control(s) or from pre- to post-scan	44 (94%) 28 (61%) 32 (71%)
			6.0.2. Probe and report physiological correlates of craving (i.e., skin conductance, heart rate, temperature, respiration and blood volume pulse amplitude) before/during/after cue presentation	
			6.0.3. Probe cue-provoked behaviors (e.g., drug-seeking or using behaviors) after scanning, whenever possible	

Ratings for items (1–5) are reported as mean (s.d.) in the 'Item importance' column, and ratings for recommendations (Yes) are reported as frequency (percentage%) of 'Yes' reports in the 'Recommendation inclusion' column.

Table 6 |

Items to report and recommendations in the Pre- and Post-Scanning Considerations category (category 7) of the checklist

Subcategories	Main items to report	Item importance (1–5)	Specific recommendation to report	Recommendation inclusion
Pre-scanning training and familiarization	7.1. Procedure to train/familiarize participants with the task/scanner before the scanning	3.72 (1.14)	7.1.1. Report both task training and scanner familiarization procedures before scanning. Familiarization can be done using various methods (e.g., describing the situation for the participants or placing the subjects in mock scanners). Training can be done by letting the participants do the task outside the scanner	39 (81%)
Pre-scanning drug and smoking consumption	7.2. Whether participants were allowed to smoke or use other drugs before scanning	4.72 (0.58)	7.2.1 Report the time interval between the last use of nicotine and other drugs and scanning 7.2.2 Consider and report the consumption of caffeine, prescribed medications or food eaten on the basis of the context of the study (e.g., controlling the time and the quantity of consumption). For instance, for the assessment of alcohol craving, it is essential to control for liquid intake before scanning because drinking high amounts of water can blunt alcohol craving	45 (96%) 32 (70%)
Other tasks and procedures in the imaging session	7.3. Presence and order of other tasks and procedures (e.g., resting fMRI or DTI before drug-cue reactivity or familiarization) in the imaging session	3.13 (1.15)	-	-
Post-scanning craving management	7.4. Steps taken to reduce participant craving after performing the task	3.72 (1.14)	-	-
General recommendations			7.0.1. Report the elements that might change the fMRI drug-cue reactivity as potentially partially partially state-dependent data, across the study days (i.e., time of scanning during the day considering the diurnal variation in responding to cues) or between studies (i.e., sequence of imaging tasks/protocols) to make sure that the result is representing data unconfounded by procedural differences 7.0.2. Explicitly report the participants' drug use expectancy, even though the potential for having access or expectation of drug use after the cue-exposure process is usually implicit in the study setting and inclusion/exclusion criteria. It has been shown that the participants' expectancy for drug use might influence cue reactivity. Participants who expect that they might have access to drugs after cue exposure will respond differently to cues compared to those who are sure that there is no access to drugs after cue exposure (e.g., being back in an in-patient or residential setting). In addition, explicitly discuss how they considered the influence of expectancy and whether they attempted to modulate or control for it in the study 7.0.3. Report a measure of sleepiness or alertness before fMRI drug-cue reactivity	36 (77%) 29 (63%) 12 (26%)

Ratings for items (1–5) are reported as mean (s.d.) in the 'Item importance' column, and ratings for recommendations (Yes) are reported as frequency (percentage%) of 'Yes' reports in the 'Recommendation inclusion' column. DTI, diffusion tensor imaging.

Table 7 |

ENIGMA ACRI Checklist, short form

Categories	Subcategories	Main items
Participant Characteristics	Inclusion/Exclusion Criteria	Inclusion and exclusion criteria for all participant groups
	Basic Demographics	Age and sex/gender for all participant groups
	Advanced Demographics I	Education or a measurement of intelligence for all participant groups
	Advanced Demographics II	Race or ethnicity for all participant groups
	Psychiatric Profile	Any categorical or dimensional measurement of psychopathologies other than substance use disorder
	Handedness	Handedness for all participant groups
	Substance Use Profile-Main Drug	Route(s) of administration for the main substance (if it is obvious (i.e., there is only one route of administration for cigarette smokers or alcohol drinkers), then it does not need to be reported.)
	Substance Use Profile-Main Drug	Current and lifetime use pattern/severity for the main drug of use for all participant groups
	Substance Use Profile-Other Drugs	Measures of current or lifetime use pattern/severity for drugs, other than the main drug of use, for all participant groups
	Abstinence Status	Days/hours/minutes since last use (duration of abstinence) and how abstinence was verified for all participant groups
	Addiction Treatment Status	Treatment status for all participant groups (e.g., non-treatment-seeking active users, treatment-seeking active users, undergoing active treatment, treated and abstinent, relapsed after treatment)
General fMRI Information	fMRI Pulse Sequence and Other Acquisition Details	fMRI data acquisition details
	fMRI Preprocessing Pipeline and Other Details	fMRI preprocessing details
	fMRI Data Processing	fMRI analyses and statistical modeling details
	fMRI Data Reporting	Basic whole-brain response to drug cues
General Task Information	Task Design	Task structure (event, block or mixed (events in blocks))
	Number of Task Components	Number of runs (if more than one), blocks (for block-designed studies) and events (e.g., drug cues, control cues, fixations)
	Requested Engagement	Instructions to the study participants on how to engage with the cues
	Temporal Information of the Event/ Block Duration	Duration of each cue (for both event and blocked-design tasks) and the total block duration (for blocked-design tasks)
	Temporal Information of the Task	Total task duration
	Order of Blocks/Events	Order of block types (e.g., drug and control) (for blocked-designs) or event types (e.g., drug and control) (for event-related designs) (The order can be fully randomized (randomized and different between subjects), pseudorandomized (identical between subjects but randomized once for the order of events/blocks) or not randomized (fixed order like neutral-drug-neutral-drug for all subjects))
	Data and Resource-Sharing	Sharing the behavioral task code or source images
Cue Information	Sensory Modality of Cues	Modality(ies) of the utilized drug and neutral/control cues (The modalities can be word, picture, smell, taste, tactile, audio script, written script, imagination, silent video, audiovisual video, paraphernalia, substance itself or mixed.)
	Sources of Cues, Development	Source of drug and neutral/control cues

Categories	Subcategories	Main items
	Sources of Cues, Validation	Extent of prior validation of drug and neutral/control cues used in the study (Drug and neutral/control cues in a study might be not validated or validated by assessing the craving induction of each cue individually using simple-item craving instruments like a single-item VAS or using standardized instruments of craving assessment and emotion or stress reactivity)
	Drug and Neutral/Control Cue Content	Content of drug cues and its relationship to the targeted drug (These include stimulus related to the drug, stimulus related to instruments of drug use, stimulus related to various stages of drug use (e.g., 'beginning' or 'end' stimuli (lit cigarette vs. ashtray)), stimulus related to drug intake, stimulus related to typical drug consumption environments, stimulus related to preparation of drug, stimulus related to purchasing the drug)
	Neutral/Control Matching to Drug Cues for Physical Features	Factors for which drug and neutral/control cues have been matched (e.g., color, brightness, hue, content, complexity, scrambled drug cue)
Craving Assessment Inside Scanner	Craving Assessment Inside Scanner, Presence	Craving assessment inside the scanner
	Craving Assessment Inside Scanner, Time Points	Description of the time points at which craving-related assessment is performed inside the scanner (e.g., before and/or after each cue/event/ block/trial/scan/run/session) (Yes/No/Not Applicable (in cases when there is no assessment inside the scanner))
	Craving Assessment Inside Scanner, Instrument(s)	Description of the instrument(s) used to assess craving and craving- related constructs inside the scanner (Yes/No/Not Applicable (in cases when there is no assessment inside the scanner))
	Craving Assessment Inside Scanner, Technology	Description of the hardware used to obtain participant responses, with specifications of models and brands of devices if necessary (e.g., response box and fiber-optic pad) (Yes/No/Not Applicable (in cases when there is no assessment inside the scanner))
Craving Assessment Outside Scanner	Craving Assessment Outside Scanner, Presence	Any craving-related assessment outside the scanner
	Craving Assessment Outside Scanner, Time Points	Description of the time points at which craving-related assessment is performed outside the scanner (e.g., immediately before entering the scanner or immediately after exiting the scanner) (Yes/No/Not Applicable (in cases when there is no assessment outside the scanner))
	Craving Assessment Outside Scanner, Instrument(s)	Description of the instrument(s) used to assess craving and craving- related constructs outside the scanner (Yes/No/Not Applicable (in cases when there is no assessment outside the scanner))
Pre- and Post-Scanning Considerations	Pre-Scanning Training and Familiarization	Procedure to train/familiarize participants with the task/scanner before the scanning
	Pre-Scanning Drug and Smoking Consumption	Whether participants were allowed to smoke or use other drugs before scanning
	Other Tasks and Procedures in the Imaging Session	Presence and order of other tasks and procedures (e.g., resting fMRI or DTI before drug-cue reactivity or familiarization) in the imaging session
	Post-Scanning Craving Management	Steps taken to reduce participant craving after performing the task

This form contains only the main items and excludes the additional recommendations.