PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A national cross-sectional survey of U.S. adults to assess the reliability of current and lifetime cannabis smoking
AUTHORS	Lum, Emily; Tang, Janet; Ryder, Annie; Vali, Marzieh; Cohen, Beth; Keyhani, Salomeh

VERSION 1 - REVIEW

REVIEWER NAME	*Connor, Jason*
REVIEWER AFFILIATION	The University of Queensland
REVIEWER CONFLICT OF INTEREST	None
DATE REVIEW RETURNED	31-Oct-2023

| GENERAL COMMENTS | This is an important research aim. Accurate measurement of cannabis exposure is critical to understanding the progression of cannabis use disorder, other substance use (via the much-debated gateway hypothesis) and physical and mental health problems. To date, measurement of cannabis use is poor, particularly lifetime use. More robust measurement in nicotine occurs with valid and reliable "packet years" which allow assessment of both cumulative and type of use (heavy, light, infrequent, binge). This is critical to understanding risk of disease and disease progression. Similarly, robust short term self-report measures are available for alcohol (e.g., Timeline Follow-back) but less so for longer term self-report. Only a small number of tools have been developed that attempt to measure lifetime alcohol exposure accurately and all were developed decades ago. The lifetime drinking history (LDH) is the most applied but compared to timeline follow back, it is used less frequently and has fewer validation studies. In non-clinical populations, underreporting for cannabis use can occur due to response biases driven by social desirability and stigma of heavy cannabis use. In treatment seeking patients with more stereotypical patterns of use, these biases are less pronounced and cannabis use recall via standardised short-term methods such as Timeline Follow-back are more robust. Long term cannabis recall remains underdeveloped. To address potential bias, measures often apply 'anchoring' to recall around significant life events. There remains a gap in the the valid measurement of lifetime cannabis exposure that this study proposes to examine.

Comments on study:
1. Psychometric validation of a scale (or a set of items) typically involves both validity (is it measuring what it intends to measure?) and reliability (does it measure consistently?) testing. There are many, but the key validity tests are construct validity (via EFA/CFA, how well a set of indicators represent or reflect the construct?) and concurrent validity (do the scores correlate well to other similar criterion measures?). This study does not test validity so the risk is |

that although test-retest reliability might be sound (consistent measurement over time), it may not be accurately measuring the construct of interest. In the worst case, it is reliably measuring an inaccurate representation of self-reported cannabis use. As a minimum, internal reliability (Cronbach alpha) should also be conducted. This is a simple analysis with the 6 items used.

2. Risk of sample bias, which is recognized by authors. The abstract does not refer to a sample size, but methods suggest originally 60,000 civilian, non-institutionalized U.S. adults were surveyed in 2017. 9,003 follow-up survey (2020 survey) > 957 survey respondents > 557 reliable survey respondents > 435 current cannabis smokers > "420 excluded were excluded due to missing data". The actual sample size used for the final analysis is unclear. A consort style flow chart would greatly assist in understanding the denominator for the current study, and the reasons for the considerable attrition potentially contributing to selection bias and therefore generalizability.

3. The authors note that changes to North American cannabis policy have diversified cannabis products, making measuring use more challenging. Conversely, in states where cannabis is legal or where it is approved for medical purposes, more transparent sales regulations and labeling may make historical measurement more reliable? Further, with the changes to recreational and medicinal legislation, potentially there is reduced stigma in these jurisdictions. An interesting question, perhaps not for this study, although is it probably powered to do so.

Summary:
Important research question but design and analysis limited in scope (to test-test reliability) and carries potential significant selection bias in sample eventually used for analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comments to the Author:

This is an important research aim. Accurate measurement of cannabis exposure is critical to understanding the progression of cannabis use disorder, other substance use (via the much-debated gateway hypothesis) and physical and mental health problems. To date, measurement of cannabis use is poor, particularly lifetime use. More robust measurement in nicotine occurs with valid and reliable "packet years" which allow assessment of both cumulative and type of use (heavy, light, infrequent, binge). This is critical to understanding risk of disease and disease progression. Similarly, robust short term self-report measures are available for alcohol (e.g., Timeline Follow-back) but less so for longer term self-report. Only a small number of tools have been developed that attempt to measure lifetime alcohol exposure accurately and all were developed decades ago. The lifetime drinking history (LDH) is the most applied but compared to timeline follow back, it is used less frequently and has fewer validation studies.
In non-clinical populations, underreporting for cannabis use can occur due to response biases driven by social desirability and stigma of heavy cannabis use. In treatment seeking patients with more stereotypical patterns of use, these biases are less pronounced and cannabis use recall
via standardised short-term methods such as Timeline Follow-back are more robust. Long term

cannabis recall remains underdeveloped. To address potential bias, measures often apply 'anchoring' to recall around significant life events. There remains a gap in the valid measurement of lifetime cannabis exposure that this study proposes to examine.

Comments on study:

1. Psychometric validation of a scale (or a set of items) typically involves both validity (is it measuring what it intends to measure?) and reliability (does it measure consistently?) testing. There are many, but the key validity tests are construct validity (via EFA/CFA, how well a set of indicators represent or reflect the construct?) and concurrent validity (do the scores correlate well to other similar criterion measures?). This study does not test validity so the risk is that although test-retest reliability might be sound (consistent measurement over time), it may not be accurately measuring the construct of interest. In the worst case, it is reliably measuring an inaccurate representation of self-reported cannabis use. As a minimum, internal reliability (Cronbach alpha) should also be conducted. This is a simple analysis with the 6 items used.

Where appropriate throughout the manuscript, we have added details about testing the internal reliability. To summarize, we found the question "On how many of the past 30 days did you smoke marijuana in a joint, pipe, or bong?" to have internal reliability ($\alpha=0.94$). The other current cannabis use question, "On those days, how many joints, pipes, or bongs did you smoke per day?" did not reach the threshold of acceptable internal reliability ($\alpha=0.26$). The questions, "Over the entire period you were smoking marijuana, about how many years did you smoke marijuana on a daily or near daily basis?" and "During the years that you smoked on a daily or near daily basis, in which form did you most often smoke marijuana?" demonstrated internal reliability ($\alpha=0.91$ and $\alpha=0.87$, respectively). "During the __ years that you smoked [joints/pipes/bongs] on a daily or near daily basis, how many [joints/pipes/bons] did you smoke per day?" did not reach the threshold acceptable for internal reliability ($\alpha=0.67$). The categorical question, "Which category best describes the total number of times you've smoked marijuana over your lifetime?" also demonstrated internal reliability ($\alpha=0.88$)."

2. Risk of sample bias, which is recognized by authors. The abstract does not refer to a sample size, but methods suggest originally 60,000 civilian, non-institutionalized U.S. adults were surveyed in 2017. 9,003 follow-up survey (2020 survey) > 957 survey respondents > 557 reliable survey respondents > 435 current cannabis smokers > "420 excluded were excluded due to missing data". The actual sample size used for the final analysis is unclear. A consort style flow chart would greatly assist in understanding the denominator for the current study, and the reasons for the considerable attrition potentially contributing to selection bias and therefore generalizability.

We have added a flow chart as Figure A1 (Additional file 3) to demonstrate how many participants were included in the analyses for each question.

We have also explained in the Statistical Analysis portion of the Methods how the sample size for the analysis for the cannabis use questions was arrived upon. The Statistical Analysis section now reads as follows, "Figure A1 (Additional file 3) shows how many respondents were included in the analysis of each question for current and lifetime cannabis use. Respondents were dropped from the analysis if they either did not answer the question in both the 2020 survey and the reliability survey or if they answered the question only in one of the surveys. As such, the sample size used to conduct the analyses varied for each question. The sample size used for the analysis of each question is shown in both Figure A1 and Table 2."

We have added that the generalizability may be limited due to the small sample sizes used in the analyses for the current and lifetime cannabis smoking questions. The section now includes the following, "Generalizability of the study may also be limited due to the small final sample sizes used in the analyses for the current and lifetime cannabis smoking questions. Future work should aim to

include larger sample sizes to decrease the effect of selection bias, and thus, increase generalizability.

3. The authors note that changes to North American cannabis policy have diversified cannabis products, making measuring use more challenging. Conversely, in states where cannabis is legal or where it is approved for medical purposes, more transparent sales regulations and labeling may make historical measurement more reliable? Further, with the changes to recreational and medicinal legislation, potentially there is reduced stigma in these jurisdictions. An interesting question, perhaps not for this study, although is it probably powered to do so.

We have included the following statement in the Conclusions section, "Additionally, given that legalization of cannabis for recreational use has become more widespread, regulation in terms of labeling may allow for more accurate measurements of use and future work should include questions assessing that gather information on quantities such as tetrahydrocannabinol and cannabidiol content and serving size."

VERSION 2 – REVIEW

REVIEWER NAME	*Connor, Jason*
REVIEWER AFFILIATION	The University of Queensland
REVIEWER CONFLICT OF INTEREST	n/a
DATE REVIEW RETURNED	24-Apr-2024

GENERAL COMMENTS	bmjopen-2023-078245.R1
	The authors have responded adequately to some of these comments, but not all.
	1. In particular they were silent on the original assessment of: "Psychometric validation of a scale (or a set of items) typically involves both validity (is it measuring what it intends to measure?) and reliability (does it measure consistently?) testing. There are many, but the key validity tests are construct validity (via EFA/CFA, how well a set of indicators represent or reflect the construct?) and concurrent validity (do the scores correlate well to other similar criterion measures?). This study does not test validity so the risk is that although test-retest reliability might be sound (consistent measurement over time), it may not be accurately measuring the construct of interest. In the worst case, it is reliably measuring an inaccurate representation of self-reported cannabis use."
	It is hard to know how to proceed without the issue being addressed.
	2. Internal reliability returned a Cronbach Alpha of 0.26 and test-test of 0.16. This non-significant result was not reported in sub-heading "Internal reliability" (line 269). But included in Table 2. No information was provided on how this was managed. Removed? Limitation?
	3. Abstract should include reliability coefficients.

Reviewer Report:

Reviewer: 1
Dr. Jason Connor, The University of Queensland

Comments to the Author:
bmjopen-2023-078245.R1

The authors have responded adequately to some of these comments, but not all.

1. In particular they were silent on the original assessment of:
"Psychometric validation of a scale (or a set of items) typically involves both validity (is it measuring what it intends to measure?) and reliability (does it measure consistently?) testing. There are many, but the key validity tests are construct validity (via EFA/CFA, how well a set of indicators represent or reflect the construct?) and concurrent validity (do the scores correlate well to other similar criterion measures?). This study does not test validity so the risk is that although test-retest reliability might be sound (consistent measurement over time), it may not be accurately measuring the construct of interest. In the worst case, it is reliably measuring an inaccurate representation of self-reported cannabis use."

It is hard to know how to proceed without the issue being addressed.

We have added the following sentence to the Limitations section: "We did not include any items to validate the measures of cannabis use in the survey. However, the question on frequency of cannabis use in the past 30 days has been validated in other settings by our team. Days of cannabis use in the past month has been associated with stroke and myocardial infarction with more frequent use associated with worse outcomes." We also mention in the Conclusions section that future work should examine the validity of the measures included in our study.

Included reference: Jeffers AM, Glantz S, Byers AL, Keyhani S. Association of Cannabis Use With Cardiovascular Outcomes Among US Adults. J Am Heart Assoc. 2024 Mar 5;13(5):e030178. doi: 10.1161/JAHA.123.030178. Epub 2024 Feb 28. PMID: 38415581; PMCID: PMC10944074.

2. Internal reliability returned a Cronbach Alpha of 0.26 and test-test of 0.16. This non-significant result was not reported in sub-heading "Internal reliability" (line 269). But included in Table 2. No information was provided on how this was managed. Removed? Limitation?

These values have been included in the appropriate section in the Results. We've also addressed them in the Limitations section by adding the following: "One of the measures we included ("On those days, how many joints, pipes, or bongs did you smoke per day?") demonstrated unacceptable test-retest and internal reliability. This may be due to day-to-day differences in the amount of cannabis smoked or differences resulting from different 30-day windows for the the 2020 survey and reliability survey.

3. Abstract should include reliability coefficients.

We have added the reliability coefficients to the abstract, as well as the classification of these results. We have also added this to the Results section.

Reviewer: 1
Competing interests of Reviewer: n/a

VERSION 3 – REVIEW

REVIEWER NAME	*Connor, Jason*
REVIEWER AFFILIATION	The University of Queensland
REVIEWER CONFLICT OF INTEREST	no
DATE REVIEW RETURNED	23-Jul-2024

GENERAL COMMENTS	In the absence of data to provide validation assessments, the Limitations section now adequately outlines this measurement weakness. Other more minor changes have been corrected in this revision.