Received: 5 April 2022 | Accepted: 26 October 2022

DOI: 10.1111/add.16085

RESEARCH REPORT

ADDICTION SSA

Diagnostic accuracy for self-reported methamphetamine use versus oral fluid test as the reference standard in a methamphetamine-dependent intervention trial population

Gregory Carter[1] | Matthew J. Spittal[2] | Linda Glowacki[3] | Dimitri Gerostamoulos[3] | Paul Dietze[4,5] | Barbara Sinclair[6] | Shalini Arunogiri[7] | Michael Berk[8,9] | Dan I. Lubman[10] | Victoria Manning[7,10] | Peter Higgs[4,11] | Brendan Quinn[4,12] | Amanda Baker[1] | Olivia M. Dean[8,9] | Alyna Turner[8] | Rebecca McKetin[13]

[1]College of Health, Medicine and Wellbeing, School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Melbourne School of Population and Global Health, University of Melbourne, Australia

[3]Victorian Institute of Forensic Medicine, Southbank, Victoria, Australia

[4]Behaviours and Health Risks Program, Burnet Institute, Melbourne, Australia

[5]National Drug Research Institute and enable Institute, Curtin University, Melbourne, Australia

[6]Graduate School of Medicine, University of Wollongong, Australia

[7]Monash Addiction Research Centre, Eastern Health Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Australia

[8]Deakin University, IMPACT – the Institute for Mental and Physical Health and Clinical Translation, School of Medicine, Barwon Health, Geelong, Australia

[9]Florey Institute for Neuroscience and Mental Health, University of Melbourne, Melbourne, Australia

[10]Turning Point, Eastern Health, Richmond, Australia

[11]Department of Public Health, La Trobe University, Bundoora, Australia

[12]Australian Institute of Family Studies, Melbourne, Australia

[13]National Drug and Alcohol Research Centre, UNSW, Sydney, Australia

Abstract

Aims: Treatment of methamphetamine dependence requires monitoring of recent use or abstinence. Self-report is commonly used for routine monitoring, but the accuracy of self-report is not established. For the treating clinician, the key accuracy statistic is the negative predictive value (NPV). The study aim was to estimate the NPV of self-reported non-use of methamphetamine compared with an oral fluid reference standard.

Design, Setting and Participants: This study was a secondary (subgroup) analysis from a randomized controlled pharmacotherapy trial. Three Australian outpatient addiction services took part. Particpants were 139 people dependent on methamphetamine.

Measurements: Weekly oral fluid samples over 12 weeks to determine methamphetamine (and amphetamine) concentrations were used as the reference standard. Self-report of any methamphetamine use in the previous 7 days by the time-line follow-back method was the index test. Standard diagnostic accuracy statistics were calculated for all available paired episodes (n = 1134). Three NPV values were calculated: unadjusted NPV and NPV adjusted for clustering of observations through logistic regression and generalized estimating equation (GEE). We also calculated the NPVs for a range of prevalence rates of methamphetamine use, for the calculated levels of sensitivity and specificity.

Findings: Sensitivity was 96.4% [95% confidence interval (CI) = 95–97.5], specificity was 63.7% (95% CI = 57.3–69.8) and positive predictive value (PPV) was 90.8% (95% CI = 88.8–92.6). The unadjusted NPV was 82.7% (95% CI = 76.5–87.9), adjusted NPV by logistic regression 82.7% (95% CI = 73.9–91.5) and GEE 76.8% (95% CI = 66.8–86.8). At a methamphetamine use prevalence of 5%, the estimated NPV would be 99.7% (95% CI = 99.6–99.9) and at 95% prevalence, 48.2% (95% CI = 39.6–57.0).

Conclusions: Self-report of no recent methamphetamine use appears to be sufficiently accurate to be clinically useful at the expected prevalence rates of methamphetamine use in clinical treatment settings. If generalizable to clinical settings, where these tests

Correspondence
Gregory Carter, College of Health, Medicine and Wellbeing, School of Medicine and Public Health, University of Newcastle, Callaghan, NSW, Australia.
Email: gregory.carter@newcastle.edu.au

Funding information
Australian National Health and Medical Research Council: Project Grant (1128147), Senior Principal Research Fellowships (1059660 and 1156072); Senior Research Fellowships (1136908 and 1135901); Leadership Fellowship (1196892); and R.D. Wright Biomedical Career Development Fellowship (APP1145634). Australian Research Council Future Fellowship (FT180100075). None of the funding bodies had a role in the design of the study, or the collection, analysis, and interpretation of data, or in writing the manuscript.

are routinely conducted, this may permit a reduction in the frequency and cost of oral fluid assays.

KEYWORDS
Addiction, clinical trial, diagnostic accuracy, methamphetamine, oral fluid, psychiatry, saliva, self-report, substance use disorders

INTRODUCTION

Globally, increased availability and use of methamphetamine has been associated with greater rates of regular use and associated harms [1], with a similar pattern reported in Australia [2]. Oral fluid testing is used in occupational and policing settings due to adverse impacts of methamphetamine use on driving [3] and injury risk [1]. The use of oral fluid to detect illicit drug use has expanded over the past decade, particularly for roadside drug testing [4–8]. Oral fluid testing has also increasingly been used to assess substance use and related harms in select populations (e.g. occupational settings) [9], drug treatment attendees [10, 11] and entertainment event patrons [12–14].

In Australia (2019–20), amphetamines were the second most common principal drug of concern for the fourth consecutive year, accounting for more than one-third of all treatment episodes [15]. Clinicians treating clients using methamphetamine often rely upon client self-report of recent use, because more objective tests of recent use (e.g. formal confirmative quantitative analysis in blood, urine, saliva or oral fluid) or longer-term use (hair samples) may be unavailable, inconvenient, costly or have a long lag time in reporting results. Urine and oral fluid testing are the most practical due to non-invasive sample collection, although urine collection is more inconvenient, and each method has limitations in technical execution and interpretation of results [16]. Liquid chromatography coupled with single-stage or tandem mass spectrometry is the gold standard method for quantification in clinical and forensic toxicology [17].

The use of oral fluid to detect methamphetamine is well established in various clinical settings and may be more sensitive than urine detection [18]. However, usefulness is limited in non-clinical populations with low prevalence of drug use [5]. Single-stage or tandem mass spectrometry (LC–MS, LC–MS/MS) of oral fluid samples has several advantages (sample volumes, time and cost) over gas chromatography procedures and are becoming increasingly important in routine analysis, despite several technical limitations [17]. Some LC–MS procedures for detecting methamphetamine in oral fluid are particularly time-efficient and adequately sensitive [19]. The study assay has been successfully used for analysis of oral fluid for law enforcement procedures, forensic results and epidemiological prevalence in the relevant populations. The fast and reliable detection of a broad range of drugs and subsequent automated data processing provides the opportunity for high throughput and fast turnaround times for forensic toxicology [20].

Even when confidentiality is assured, the use of illicit substances may be denied or under-reported. The non-disclosure of illicit drug use can result from embarrassment, fear of punishment, social disapproval [21] or cultural issues [22]. 'Over-reporting' is uncommon for self-reported drug use, and in the Australian criminal justice system, the under-reporting of illicit drug use was more common by people who use cocaine or methamphetamine than those using heroin when verified against a urinalysis test [23]. Thus, the key question for clinicians is, if the client reports no recent methamphetamine use, how likely is this report to be correct?

Although the full range of diagnostic accuracy statistics is important to understand for a given diagnostic test [24], for clinical utility and to answer the clinician's key question the most relevant diagnostic accuracy statistics for self-report of recent methamphetamine use would be the negative predictive value (NPV) and the likelihood ratio negative (LR−). The NPV tells a clinician what proportion of client reports of no recent use are correct [25], and the LR− shows the change from pre- to post-test probability for the clinician to judge if this additional change in probability is clinically useful [26]. There is no absolute NPV level that is considered to be clinically useful and clinician judgement is required to take into account the clinical circumstances, including the consequences of accepting a negative test that may be a false negative. Perhaps the most important consideration is the prevalence of the condition in the population being examined because of the impact on NPV [24]. Conversely, a LR− value of < 0.1 is usually considered to be a 'good' diagnostic test, and the LR− is not affected by prevalence of the condition of interest [24]. Another important difference concerns the interpretation of results from

repeated testing required in monitoring for drug use, which means that a clinician has a more complex task to interpret the accuracy of an individual test in the context of repeated testing in an individual, with a resulting lack of independence for each observation. Diagnostic tests for most diseases are intended to identify a disease that is either present or absent, although some tests may be used for relapsing conditions—for example, the dexamethasone suppression test for depression [27]—so often require only a single test or a small number of sequential tests, while methamphetamine or other drug use can be intermittent and require multiple regular testing.

However, very few studies report any accuracy statistics of client self-reported methamphetamine use when compared with any oral fluid reference standard in clinical populations, and there is considerable heterogeneity in client characteristics, study design, analysis and reporting of diagnostic accuracy statistics, which make any comparisons difficult. Moderate accuracy was reported for self-reported methamphetamine use (previous 30 days) versus oral fluid detection (kappa = 0.66, sensitivity 60%) [28], poor PPV for amphetamine or methamphetamine use at 47.1% for 3 days and 60.4% for 90 days against oral fluid [29] and poor agreement (kappa = 0.25) for current intoxication with methamphetamine or amphetamine against a Drug Swipe 6S oral fluid screen [14]. None of these studies were in treatment-seeking populations or used repeated measures.

In this study, we examined the range of diagnostic accuracy statistics, with a special focus on NPV and LR−, for self-reported methamphetamine use in the previous 7 days (index test) versus assays of oral fluid methamphetamine (and amphetamine) as the reference standard in participants who were dependent upon methamphetamine and were tested weekly during a 12-week RCT.

AIMS

Participant self-report of recent methamphetamine as the index test use was compared with oral fluid testing as the reference standard to estimate:

1. the full range of diagnostic accuracy statistics;
2. the NPV unadjusted and adjusted for multiple testing (i.e. within-person clustering); and
3. the range of possible NPVs for the range of possible prevalence values of methamphetamine use.

METHODS

Setting and study population

Data for the current study were obtained from participants who were enrolled into the N-acetylcysteine (NAC) for methamphetamine dependence (N-ICE) trial, and full details of the study design are available elsewhere [30]. In short, the trial was conducted in three Australian outpatient settings in Melbourne (Turning Point), Geelong (Barwon Health) and Wollongong (Illawarra Drug and Alcohol Service). Inclusion criteria were as follows: aged 18–60 years, met DSM-IV criteria for methamphetamine dependence in the past year, seeking to reduce methamphetamine use but not currently enrolled in substance use treatment, not in need of acute care for psychiatric or other medical conditions, no primary psychotic disorder, not pregnant or lactating and with no contraindications for the trial medication (N-acetylcysteine).

The N-ICE trial was designed primarily as an intervention trial. The diagnostic accuracy analyses used in this paper were not included in the pre-registered study protocol [31], so the results arising from these analyses should be considered exploratory.

We used the Standards for Reporting of Diagnostic Accuracy (STARD-2015) to report the results [32] and the STARD-2015 diagram to report flow of participants and tests through the study (see Fig. 1). In keeping with the STARD-2015 terminology, we have used the term 'diagnostic accuracy' throughout the paper, although it should be recognized that the condition of interest, methamphetamine use (even in a methamphetamine-dependent population), is not a 'diagnosis' of a specific illness or disorder.

Procedures

Participants underwent weekly assessments for 12 weeks, conducted at a local health service or at a public venue when convenient to the participant (e.g. cafe).

Participant self-report of methamphetamine use (index test)

Days of methamphetamine use in the past 7 days was assessed using the time-line follow-back (TLFB) method [33]. Self-report of any methamphetamine use in the previous 7 days indicated a positive response, and self-report of no methamphetamine use in the period indicated a negative response.

Oral fluid assay (reference standard)

At each assessment, an oral fluid sample was taken using a Quantisal Oral Fluid Collection Device™, which was analysed for methamphetamine and amphetamine using tandem LC–MS/MS by the Victorian Institute of Forensic Medicine. The fast and reliable detection of a broad range of drugs and subsequent automated data processing provides the opportunity for high throughput and fast turn-around times for forensic toxicology [20].

Methamphetamine and amphetamine were separated over 5 minutes (with an additional 0.5-minute re-equilibration). Analytes were detected using a Sciex® API 4500 Q-Trap LC–MS/MS system with positive ESI in MRM mode monitoring three transitions per analyte. The method was fully validated in accordance with

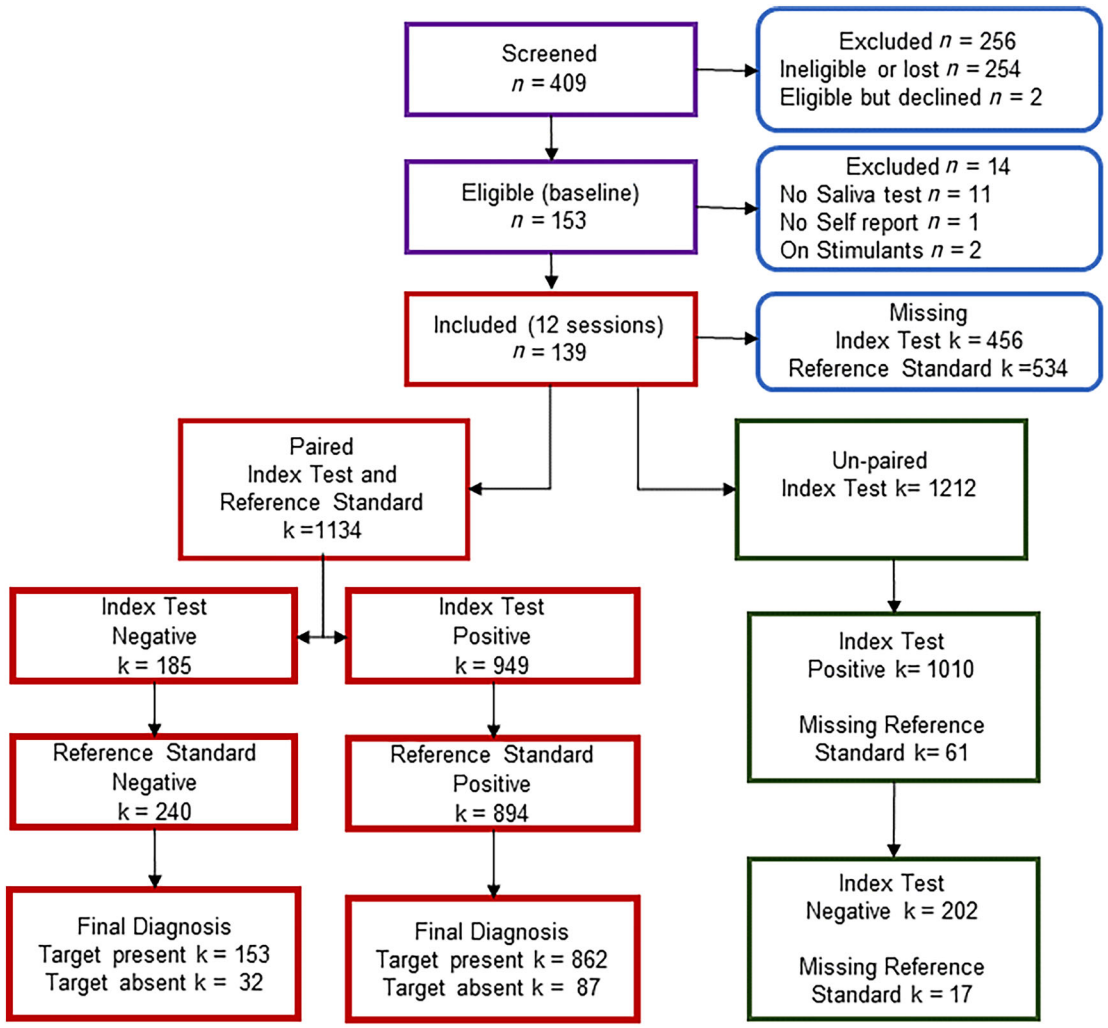

FIGURE 1 Standards for Reporting of Diagnostic Accuracy (STARD)-2015: participant and tests flow throughout the study; *k* = number of tests; *n* = number of participants

international guidelines and also monitored carbon-13 isotopes of 3,4-methylenedioxy-methamphetamine (MDMA) and methamphetamine to reduce detector saturation effects, allowing for confirmation of large concentrations of these compounds without the need for dilution or re-analysis [20]. There were no pre-screening tests, and the concentration was determined in all samples. The lower limit of quantitation was 1 ng/ml of methamphetamine and 1 ng/ml of amphetamine. Australian standards for LC–MS/MS were used; confirmations of methamphetamine > 25 ng/ml were considered a positive result.

Blindness

The raters of the self-reported index test (TLFB) collected the oral fluid samples but were blind to results of the oral fluid reference standard (LC–MS/MS), and the toxicologists responsible for the reference standard were blind to clinical information and the index test results.

Ethical approval

The study was approved by Human Research Ethics Committees: Eastern Health (E21-2017), Barwon Health (17/202), University of Wollongong and the Illawarra Shoalhaven Local Health District Health (2017/549), UNSW Human Research Ethics Committee (17/202), Deakin University (2019-211) and Curtin University (HRE2018-0205), Australia.

Statistical analyses

We first calculated the accuracy statistics using all paired episodes of self-report and oral fluid results. We developed a contingency table of the reference standard (LC–MS/MS-based oral fluid) against the index test (TLFB self-report) and reported the raw scores for each cell. We calculated sensitivity, specificity, positive predictive values (PPV), NPV, likelihood ratios positive (LR+), LR−, diagnostic odds ratios and diagnostic effectiveness.

We then calculated the adjusted NPVs in two additional ways, using logistic regression with cluster-adjusted standard errors, and with the GEE approach as an alternative method to account for the repeated measurements within an individual. Our GEE model used the binomial family with a logit link and an exchangeable correlation structure.

We undertook a sensitivity analysis to account for missing reference standard or self-reported data. We performed this by imputing 50 data sets using chained equations and re-estimating NPV values on the imputed data. Imputation samples were generated using the following predictors: positive oral fluid result for methamphetamine, self-reported methamphetamine use in the past week, treatment arm, number of days of methamphetamine use at intake, consuming methamphetamine via injection (yes, no), age at intake, methamphetamine use in years, use of cocaine in the past 4 weeks (at intake) and use of antidepressants in the past 4 weeks (at intake). This procedure resulted in 1668 paired results from 139 individuals being available for analysis.

Because NPVs are influenced by the underlying prevalence of the condition (methamphetamine use), which will vary in different populations, we used Bayes' rule to calculate NPVs, using the sensitivity and specificity of the test, for different levels of prevalence of the condition. This allows clinicians with a knowledge (or estimate) of the prevalence of the condition in their clinical population to identify a relevant NPV.

We used Stata version 16.1 (Copyright 1985–2021; StataCorp LLC, College Station, TX, USA) for analyses.

RESULTS

Participants in the study

From the 153 participants recruited to the N-ICE intervention trial, 14 were excluded from this study; 11 had no saliva samples, one had no data on self-reported methamphetamine use and two had been prescribed medications containing amphetamine-type stimulants during the trial. The study sample consisted of n = 139 participants for 12 sessions during the intervention phase. From a maximum of k = 1668 potentially paired index tests and reference standards, there were k = 1134 paired self-report and oral fluid samples [median = nine per person, interquartile range (IQR) = 5–10 pairs] for analysis. The details are shown in Fig. 1.

Participant characteristics

Detailed characteristics have been previously reported for all trial participants [30]. For this subsample, 59% of participants were male, with a mean age of 38 [standard deviation (SD) = 8] years and a median of 23 (IQR = 15–28) days of methamphetamine use in the 4 weeks prior to trial baseline. Polysubstance use in the 4 weeks prior to baseline consisted mainly of cannabis (41%; median = 16 days), alcohol (53%; median = 4 days) and tobacco (83%; median = 28 days), with lower levels of other substance use (cocaine 4%, heroin 4%, ecstasy 3% and hallucinogens 2%).

Diagnostic accuracy of self-report of recent methamphetamine use versus oral fluid assay

The majority of oral fluid samples had methamphetamine present, with a prevalence of 79% (95% CI = 76.0–81.2%) considered positive in all samples. The unadjusted diagnostic accuracy results were sensitivity = 96.4% (95% CI = 95.0–97.5%), specificity = 63.7% (95% CI = 57.3–69.8%), PPV = 90.8% (95% CI = 88.8–92.6%), NPV = 82.7% (95% CI = 76.5–87.9%), LR+ 2.66 = (95% CI = 2.25–3.15) and LR− = 0.06 (95% CI = 0.04–0.8) (see Table 1).

Unadjusted and adjusted NPV

The NPV in the unadjusted (naive method ignoring clustering) model was 82.7% (95% CI = 76.5–87.9%), while in the adjusted models it was 82.7% (95% CI = 73.9–91.5%) and 76.8% (95% CI = 66.8–86.8%) for the logistic regression and GEE models, respectively. The sensitivity analysis using imputed data showed similar results (see Table 2).

TABLE 1 Self-reported methamphetamine use versus methamphetamine oral fluid samples (reference standard): 2 × 2 contingency table

	Self-reported methamphetamine use				
Methamphetamine oral fluid result	**Yes**	**No**	**Total**	**Diagnostic accuracy statistics**	
Positive	862	32	894	Se 96.4%	LR+ 2.66
Negative	87	153	240	Sp 63.8%	LR− 0.06
Total	949	185	1134		
Diagnostic accuracy statistics	PPV 90.8%	NPV 82.7%	Diagnostic OR 47.37 Diagnostic effectiveness 116.3		

Diagnostic OR = diagnostic odds ratio; LR+ = likelihood ratio-positive; LR− = likelihood ratio-negative; NPV = negative predictive value; PPV = positive predictive value; Se = sensitivity; Sp, = specificity.

TABLE 2 Negative predictive values for self-reported methamphetamine use using three methods

	NPV (%)	(95% CI)
Primary analysis (original values)		
Naive method (ignoring within-person clustering)	82.7	76.5–87.9
Logistic regression (adjusting the SE for clustering)	82.7	73.9–91.5
GEE (exchangeable correlation matrix)	76.8	73.9–91.5
Sensitivity analysis (imputed values)		
Naive method (ignoring within-person clustering)	82.4	76.7–88.0
Logistic regression (adjusting the SE for clustering)	82.4	75.0–89.7
GEE (exchangeable correlation matrix)	79.3	70.9–87.7

GEE = general estimating equation; NPV = negative predictive value; SE = standard error; CI = confidence interval.

TABLE 3 NPV calculated using Bayes' rule at different prevalence levels (Se = 96.4%%, Sp = 63.7%)

Assumed prevalence of methamphetamine use (%)	NPV (%)	95% CI
5	99.7	99.6–99.8
10	99.4	99.1–99.6
20	98.6	98.0–99.0
30	97.6	96.7–98.3
40	96.4	94.9–97.4
50	94.7	92.6–96.2
60	92.2	89.2–94.4
70	88.3	84.2–91.5
80	81.6	75.7–86.3
90	66.3	58.0–73.7
95	48.2	39.6–57.0

NPV = negative predictive value; Se = sensitivity; Sp = specificity.

NPV estimates for range of prevalence of methamphetamine use

For a range of prevalence estimates (95% CI = 5–95%) for the condition of interest (recent methamphetamine use) the NPVs ranged from 99.7 to 48.2%, respectively. NPVs were highest when prevalence was lowest (see Table 3).

DISCUSSION

Main findings

The NPV estimates from our study (unadjusted and adjusted) were approximately 80% (i.e. self-report of no recent use correct approximately four times in five), which we would consider to be a clinically useful level of accuracy for self-report of no recent methamphetamine use (using TLFB). However, there are no absolute standards for the NPV to be considered clinically useful and clinician judgement is required, which would be context- and consequence-dependent. We also showed that at prevalence levels of recent methamphetamine use of 80% or lower that the NPV improved from 82 to 99.7%, a level of accuracy that would clearly be considered clinically useful. We would anticipate the prevalence of recent methamphetamine use encountered in clinical treatment settings to be less than 80%.

Likelihood ratios are used to express a change in odds (from pre- to post-test) and with further calculation used to understand a change in probability [34]. A LR− of 0.1 would indicate a 10-fold decrease in the odds of having methamphetamine use in the past 7 days in the presence of a negative self-report, while a LR− of 0.05 would be a 20-fold decrease. A LR− of 0.1 would reduce the probability of recent methamphetamine use by approximately 45% [34]. In our study, we estimated the LR− to be 0.06, which would be considered to be clinically useful.

Diagnostic accuracy statistics and prevalence of the condition

NPV describes the probability of not having a condition with a negative index test result, and predictive values are reasonably dependent upon the prevalence of the condition in the study population. NPV decreases with an increase in the prevalence of the condition in a population, and NPV is somewhat more weakly influenced by the condition prevalence than is PPV [24]. Our study population (participants in a pharmacotherapy trial not currently involved in clinical treatment) had a high prevalence (79%) of the condition (recent methamphetamine use), while many treatment settings might have lower prevalence (due to greater range of clinical severity, treatment response, inclusion of 'stable' responders or lower frequency of daily use) and higher NPVs would result.

Likelihood ratios

Likelihood ratios should be an optimal choice for reporting diagnostic accuracy, and it is important to report paired measures (predictive values and likelihood ratios) for clinically meaningful thresholds [35]. Specificity and sensitivity are used to calculate the likelihood ratios, so neither LR+ nor LR− directly depend upon condition prevalence, although the spectrum of clients and condition severity effect sensitivity and specificity and may thereby be indirectly related to prevalence of the condition [36]. LR− values are directly linked to post-test probabilities; the lower the LR−, the stronger the evidence for absence of the condition [35]. In our study, a LR− of 0.06 implies a substantial improvement in post-test probability, which implies that the self-report of no recent methamphetamine use (index test) would

be useful to the clinician relying upon that client self-report to monitor treatment response, whatever the prevalence of the condition in that clinical population.

'Over-reporting' methamphetamine use

We found 949 occasions of positive self-reported methamphetamine use in the previous 7 days, with 87 (9.2%) of these occasions recording a negative test of oral fluid using LC–MS/MS. It is likely that most of these negative tests of oral fluid came about because of the effective detection window of methamphetamine in oral fluid, which varies based on the assay used and populations studied: 5–48 hours after ingestion (non-clinical and clinical populations) [37] or 24–72 hours for the assay used in our study [18, 38], which is shorter (at both ends of the detection window) than the self-report period of 7 days used in our study. We believe it is unlikely that 'over-reporting' of methamphetamine use would have been common in the context of this intervention trial.

Power calculations and sample sizes for NPV estimates

This study was a secondary analysis from an intervention trial and the sample size was determined by the requirements of the original study. Moreover, our estimates of NPV involved repeated measures in the same participants, which is appropriate to the clinical study setting and the research question of interest. Formulae and tables for the calculations for sample sizes required for sensitivity and specificity estimates have been developed [39]; however, these do not account for prevalence of the condition of interest and so alternate formulae are required for the calculation of PPV or NPV, which are the statistics of relevance in the context of assay accuracy and rare disorders [40]. In a *post-hoc* calculation, we estimated the power for our NPV analyses following recommended methods [41] (details are available from the authors). In our study, there were 153 of 185 samples where negative self-reported methamphetamine use corresponded with a negative reference-standard oral fluid test, yielding 95% power to detect a NPV of 82.7% with a 95% confidence interval of $\pm$ 10%. This calculation does not account for these samples being nested within individuals.

Strengths of the study

Little is known regarding the diagnostic accuracy of client self-report of recent methamphetamine use in clinical treatment populations, even when using a structured assessment of self-reported use with good psychometric properties for the index test (e.g. TLFB) [33]. We followed the STARD 2015 standards [32] for reporting this study. Index test and reference standards were conducted independently, with research and clinical staff blinded to reference standard and toxicology staff blinded to index test. We reported all unadjusted classical accuracy statistics and then focused upon two measures of diagnostic accuracy, NPV and LR−, which are of key relevance to the clinical issue of using patient self-report in order to determine 'true' non-use and hence inform treatment response. We also calculated NPVs after adjustment for clustering (repeated measures on individuals), which produced similar NPV estimates to the unadjusted values, suggesting that there was little effect of repeated measures. We reported the range of NPVs for various prevalence levels of methamphetamine use to provide a clinically meaningful context and to demonstrate the higher NPVs at lower prevalence rates, which would be typical of many treatment settings.

Limitations of the study

First, data were derived from a randomized trial study design of a pharmacological intervention with repeated measures of the reference and index standards. The most commonly used study design for diagnostic accuracy would be a cross-sectional design in a relevant population, showing the full range of the severity of the condition at the prevalence rate characteristic of that population, although other study designs are possible [42]. In our population and clinical setting, a cross-sectional design at a single time-point would be less relevant than repeated testing on each weekly clinic visit. The criteria for entry into an intervention trial will produce a sampling bias so that the study population would not be likely to represent the full spectrum of severity (frequency and severity of recent use) of the condition of interest, so our estimates of sensitivity, specificity and LRs might be affected. The prevalence of the condition of interest in the population (at 79%) was probably higher than in most treatment populations, and as higher prevalence populations will yield relatively lower NPV estimates, our reported estimates for NPV might be conservative.

The second set of limitations concern the appropriateness of the reference standard and time-period, specifically the ability to detect methamphetamine used in the previous 7 days. We used an assay based on oral fluid samples, with the detection window of 24–72 hours [18, 38]. The third set of limitations concern potential response bias by participants, who understood that their oral fluid samples would be tested on a weekly basis. For the whole sample this regular monitoring could influence more accurate self-reports. Those who believed themselves to be on active treatment or placebo (despite blindness of allocation) may have a differential response to the TLFB. Under-reporting could occur where participants unknowingly ingested methamphetamine (e.g. as a contaminant to a different drug) during a reported methamphetamine-abstinent period.

The fourth set of limitations are relevant to the oral fluid testing. Estimates of sensitivity, specificity and, hence, likelihood ratios may be affected by imperfect reference standards [36]. All methods of estimating the presence of illicit drug use are subject to procedural issues [43] and methamphetamine is no exception. Direct observation of oral fluid collection (as performed in the current study) reduces the risk of adulteration or substitution [38], reducing oral fluid pH (e.g. by

chewing gum to increase saliva production) can reduce methamphetamine concentrations [44], rinsing may reduce detection, compliance with abstaining from food or beverage consumption prior to testing can be impractical and difficult to control in a non-experimental setting, and people who use methamphetamine may have reduced production of saliva [45]. Although methamphetamine is reasonably stable in oral fluid, transport and storage arrangements also have the potential to affect the concentration of methamphetamine detected [46]. There is potential for high inter- and intra-individual variability in the relationship between blood and oral fluid levels, so oral fluid levels are a proxy for plasma levels [44, 47]. Confirmatory analysis using LC/MS can be expensive and has restricted availability in some settings. There is no current evidentiary screening device, which can rapidly confirm the presence of methamphetamine in oral fluid. The issue of mass testing (at scale) with immediate results is of particular relevance for the impaired driving situation. However, oral fluid screening for methamphetamine has been shown to be reliable, with confirmation rates of various screening devices greater than 95%, and serves as a potential tool to detect methamphetamine use in this context [48].

The fifth set of limitations concerns external validity. We used a population that volunteered and met specific criteria for an intervention trial. There are also particular circumstances relevant to a clinical trial that would not occur in regular clinical practice: the detailed collection of self-report data using the daily TLFB method, double-blinding of oral fluid assay results and shielding from negative consequences of use. Generalization of these results to other clinical populations—for example, those in stable treatment settings or people who use methamphetamine in the community—should be conducted with caution.

CONCLUSIONS

Participant self-report of no recent methamphetamine use in this RCT was accurate enough to be clinically useful, whether considering the variously calculated NPVs (prevalence-dependent) or by considering the LR− values (independent of prevalence). This level of accuracy of self-reported use may permit some cautious reduction in the frequency and cost of routine (i.e. weekly or each clinic visit) oral fluid (or other) assays in clinical settings routinely using these assays. Any resulting relative increase in the use of patient self-report in place of biomarker testing in methamphetamine use disorder might reduce stigma and increase trust between providers and patients. However, the accuracy of self-report may vary depending on context, and the current findings, which were based on a clinical trial setting, may not generalize to other clinical settings and are unlikely to generalize to settings where revealing illicit drug use may result in negative consequences (e.g. criminal justice settings). Where treatment units do not routinely use these assays, generalization of these results should be performed cautiously, and the diagnostic accuracy in other populations should be independently evaluated. Future research activities could also include a self-report period for any methamphetamine use in the previous 48 hours to generate more exact estimates of diagnostic accuracy when compared with oral fluid samples analysed by the LC–MS/MS method.

ACKNOWLEDGEMENTS

Thanks go to members of our Data Safety Monitoring Board (Jason White, Matthew Spittal, Deborah Kerr, Grant Sara, Juanita Koeijers), Long Nguyen and Wenbin Liang for assistance with programming and trial randomization, Steven Shoptaw for advice on trial methods, other trial staff (including Nicole Edwards, Nina te Pas, Margaret Kent, Ramez Bathish, Davinia Rizzo, Ellie Brown, Bruno Agustini, Behrooz Maylie, Olalekan Ogunleye, Scott Hall and Vicky Phan), agencies and individuals who have assisted with recruitment efforts and the trial participants. Open access publishing facilitated by The University of Newcastle, as part of the Wiley - The University of Newcastle agreement via the Council of Australian University Librarians.

DECLARATION OF INTERESTS

G.C. has received educational and travel support from Servier, Astra Zeneca, Otsuka Australia, Merck Sharp & Dohme and Janssen-Cilag in the past 5 years. O.D. has received grant support from the Brain and Behavior Foundation, Simons Autism Foundation, Stanley Medical Research Institute, Deakin University, Lilly, NHMRC and ASBDD/Servier. She has also received support in kind from BioMedica Nutraceuticals, NutritionCare and Bioceuticals. D.L. has provided consultancy advice to Lundbeck and Indivior and has received travel support and speaker honoraria from Camurus, Indivior, Janssen, Lundbeck, Servier and Shire. He has received research grants from Camurus and Seqirus. P.D. has received investigator-initiated funding from Gilead Sciences, an untied educational grant from Indivior, and is an unpaid member of an Advisory Board for Mundipharma for work unrelated to this study. M.B. has received Grant/Research Support from the NIH, Cooperative Research Centre, Simons Autism Foundation, Cancer Council of Victoria, Stanley Medical Research Foundation, Medical Benefits Fund, National Health and Medical Research Council, Medical Research Futures Fund, Beyond Blue, Rotary Health, A2 milk company, Meat and Livestock Board, Woolworths, Avant and the Harry Windsor Foundation, has been a speaker for Abbot, Astra Zeneca, Janssen and Janssen, Lundbeck and Merck and served as a consultant to Allergan, Astra Zeneca, Bioadvantex, Bionomics, Collaborative Medicinal Development, Eisai, Janssen and Janssen, Lundbeck Merck, Pfizer and Servier. P.H. has received investigator-initiated research funded by Gilead Sciences and Abbvie. S.A. has received speaker honoraria from Camurus, Gilead, Janssen and Servier.

AUTHOR CONTRIBUTIONS

Gregory Carter: Conceptualization; formal analysis; funding acquisition; investigation; methodology; writing-original draft; writing-review and editing. **Matthew J Spittal:** Conceptualization; formal analysis; investigation; methodology; writing-original draft; writing-review and editing. **Linda Glowacki:** Investigation; methodology; resources; writing-review and editing. **Dimitri Gerostamoulos:** Investigation;

methodology; resources; writing-review and editing. **Paul Dietze:** Conceptualization; funding acquisition; methodology; resources; writing-review and editing. **Barbara Sinclair:** Conceptualization; data curation; funding acquisition; resources; supervision; writing-review and editing. **Shalini Arunogiri:** Data curation; project administration; supervision; writing-review and editing. **Michael Berk:** Funding acquisition; project administration; resources; supervision; writing-review and editing. **Dan I Lubman:** Funding acquisition; project administration; resources; writing-review and editing. **Victoria Manning:** Data curation; project administration; supervision; writing-review and editing. **Peter Higgs:** Conceptualization; data curation; funding acquisition; resources; writing-review and editing. **Brendan Quinn:** Conceptualization; data curation; funding acquisition; project administration; resources; writing-review and editing. **Amanda Baker:** Conceptualization; funding acquisition; supervision; writing-review and editing. **Olivia Dean:** Conceptualization; data curation; funding acquisition; project administration; resources; writing-review and editing. **Alyna Turner:** Conceptualization; data curation; funding acquisition; project administration; resources; writing-review and editing. **Rebecca McKetin:** Conceptualization; data curation; funding acquisition; methodology; project administration; resources; supervision; writing-original draft; writing-review and editing.

ORCID

Gregory Carter https://orcid.org/0000-0001-8180-1798
Dan I. Lubman https://orcid.org/0000-0002-6747-1937
Rebecca McKetin https://orcid.org/0000-0003-2833-4830

REFERENCES

1. Farrell M, Martin NK, Stockings E, Bórquez A, Cepeda JA, Degenhardt L, et al. Responding to global stimulant use: challenges and opportunities. Lancet. 2019;394:1652–67.
2. Degenhardt L, Sara G, McKetin R, Roxburgh A, Dobbins T, Farrell M, et al. Crystalline methamphetamine use and methamphetamine-related harms in Australia. Drug Alcohol Rev. 2017;36:160–70.
3. Bosanquet D, MacDougall HG, Rogers SJ, Starmer GA, McKetin R, Blaszczynski A, et al. Driving on ice: impaired driving skills in current methamphetamine users. Psychopharmacology. 2013;225:161–72.
4. Davey J, Armstrong K, Martin P. Results of the Queensland 2007–2012 roadside drug testing program: the prevalence of three illicit drugs. Accid Anal Prev. 2014;65:11–7.
5. Gjerde H, Langel K, Favretto D, Verstraete AG. Detection of illicit drugs in oral fluid from drivers as biomarker for drugs in blood. Forensic Sci Int. 2015;256:42–5.
6. Matthews AJ, Bruno R, Dietze P, Butler K, Burns L. Driving under the influence among frequent ecstasy consumers in Australia: trends over time and the role of risk perceptions. Drug Alcohol Depend. 2014;144:218–24.
7. Matzopoulos R, Lasarow A, Bowman B. A field test of substance use screening devices as part of routine drunk-driving spot detection operating procedures in South Africa. Accid Anal Prev. 2013;59:118–24.
8. Simonsen KW, Steentoft A, Hels T, Bernhoft IM, Rasmussen BS, Linnet K. Presence of psychoactive substances in oral fluid from randomly selected drivers in Denmark. Forensic Sci Int. 2012;221:33–8.
9. Griffiths A, Leonars R, Hadley L, Stephenson M, Teale R. Smoke on the water—oral fluid analysis at sea. Forensic Sci Int. 2017;278:361–6.
10. Miller KL, Puet BL, Roberts A, Hild C, Carter J, Black DL. Urine drug testing results and paired oral fluid comparison from patients enrolled in long-term medication-assisted treatment in Tennessee. J Subst Abuse Treat. 2017;76:36–42.
11. Kunkel F, Fey E, Borg D, Stripp R, Getto C. Assessment of the use of oral fluid as a matrix for drug monitoring in patients undergoing treatment for opioid addiction. J Opioid Manag. 2015;11:435–42.
12. Johnson MB, Voas RA, Miller BA, Holder HD. Predicting drug use at electronic music dance events: self-reports and biological measurement. Eval Rev. 2009;33:211–25.
13. Miller BA, Furr-Holden D, Johnson MB, Holder H, Voas R, Keagy C. Biological markers of drug use in the club setting. J Stud Alcohol Drugs. 2009;70:261–8.
14. Miller P, Curtis A, Jenkinson R, Droste N, Bowe SJ, Pennay A. Drug use in Australian nightlife settings: estimation of prevalence and validity of self-report. Addiction. 2015;110:1803–10.
15. Australian Institute of Health and Welfare (AIHW). Alcohol and other drug treatment services in Australia annual report Canberra, Australia: AIHW; 2021.
16. Allen KR. Screening for drugs of abuse: which matrix, oral fluid or urine? Ann Clin Biochem. 2011;48:531–41.
17. Maurer HH. Advances in analytical toxicology: the current role of liquid chromatography-mass spectrometry in drug quantification in blood and oral fluid. Anal Bioanal Chem. 2005;381:110–8.
18. Vindenes V, Yttredal B, Øiestad EL, Waal H, Bernard JP, Mørland JG, et al. Oral fluid is a viable alternative for monitoring drug abuse: detection of drugs in oral fluid by liquid chromatography-tandem mass spectrometry and comparison to the results from urine samples from patients treated with methadone or buprenorphine. J Anal Toxicol. 2011;35:32–9.
19. Wood M, de Boeck G, Samyn N, Morris M, Cooper DP, Maes RA, et al. Development of a rapid and sensitive method for the quantitation of amphetamines in human plasma and oral fluid by LC-MS-MS. J Anal Toxicol. 2003;27:78–87.
20. Di Rago M, Chu M, Rodda LN, Jenkins E, Kotsos A, Gerostamoulos D. Ultra-rapid targeted analysis of 40 drugs of abuse in oral fluid by LC-MS/MS using carbon-13 isotopes of methamphetamine and MDMA to reduce detector saturation. Anal Bioanal Chem. 2016;408:3737–49.
21. Swadi H. Validating and improving the validity of self-reports in adolescent substance misuse surveys. J Drug Issues. 1990;20:473–86.
22. Pearce LA, Homayra F, Dale LM, Moallef S, Barker B, Norton A, et al. Non-disclosure of drug use in outpatient health care settings: findings from a prospective cohort study in Vancouver, Canada. Int J Drug Policy. 2020;84:102873.
23. McGregor K, Makkai T. Self-reported drug use: how prevalent is under-reporting? Canberra, Australia: Australian Institute of Criminology; 2003.
24. Simundic AM. Measures of diagnostic accuracy: basic definitions. eJIFCC. 2009;19:203–11.
25. Trevethan R. Sensitivity, specificity, and predictive values: foundations, Pliabilities, and pitfalls in research and practice. Front Public Health. 2017;5:307.
26. Attia J. Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. Aust Prescr. 2003;26:111–3.
27. Carroll BJ. Dexamethasone suppression test for depression. Adv Biochem Psychopharmacol. 1984;39:179–88.
28. Fendrich M, Mackesy-Amiti ME, Johnson TP. Validity of self-reported substance use in men who have sex with men: comparisons with a general population sample. Ann Epidemiol. 2008;18:752–9.
29. Rendon A, Livingston M, Suzuki S, Hill W, Walters S. What's the agreement between self-reported and biochemical verification of drug use? A look at permanent supportive housing residents. Addict Behav. 2017;70:90–6.

30. McKetin R, Dean OM, Turner A, Kelly PJ, Quinn B, Lubman DI, et al. N-acetylcysteine (NAC) for methamphetamine dependence: a randomised controlled trial. eClinicalMedicine. 2021;38:101005.
31. McKetin R, Dean OM, Turner A, Kelly PJ, Quinn B, Lubman DI, et al. A study protocol for the N-ICE trial: a randomised double-blind placebo-controlled study of the safety and efficacy of *N*-acetylcysteine (NAC) as a pharmacotherapy for methamphetamine ('ICE') dependence. Trials. 2019;20:325.
32. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
33. Fals-Stewart W, O'Farrell TJ, Freitas TT, McFarlin SK, Rutigliano P. The timeline followback reports of psychoactive substance use by drug-abusing patients: psychometric properties. J Consult Clin Psychol. 2000;68:134–44.
34. McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002;17:646–9.
35. Eusebi P. Diagnostic accuracy measures. Cerebrovasc Dis. 2013;36:267–72.
36. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. Can Med Assoc J. 2013;185:E537–44.
37. Verstraete AG. Detection times of drugs of abuse in blood, urine, and oral fluid. Ther Drug Monit. 2004;26:200–5.
38. Huestis MA, Cone EJ. Methamphetamine disposition in oral fluid, plasma, and urine. Ann NY Acad Sci. 2007;1098:104–21.
39. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193–204.
40. Steinberg DM, Fine J, Chappell R. Sample size for positive and negative predictive value in diagnostic research using case-control designs. Biostatistics. 2009;10:94–105.
41. Zhou X-H, Obuchowski NA, McClish DK. Statistical Methods in Diagnostic Medicine New York, NY: John Wiley & Sons, Inc.; 2011.
42. Chasse M, Fergusson DA. Diagnostic accuracy studies. Semin Nucl Med. 2019;49:87–93.
43. Kintz P, Samyn N. Use of alternative specimens: drugs of abuse in saliva and doping agents in hair. Ther Drug Monit. 2002;24:239–46.
44. Drummer OH. Drug testing in oral fluid. Clin Biochem Rev. 2006;27:147–59.
45. Götrick B, Giglio D, Tobin G. Effects of amphetamine on salivary secretion. Eur J Oral Sci. 2009;117:218–23.
46. Turfus SC, Chu M, Gerostamoulos D, Drummer OH. An assessment of the stability of MDMA, methamphetamine and THC in oral fluid. Aust J Forensic Sci. 2014;46:397–410.
47. Gjerde H, Verstraete AG. Estimating equivalent cutoff thresholds for drugs in blood and oral fluid using prevalence regression: a study of tetrahydrocannabinol and amphetamine. Forensic Sci Int. 2011;212:e26–30.
48. Beirness DJ, Smith DAR. An assessment of oral fluid drug screening devices. Can Soc Forensic Sci J. 2017;50:55–63.

How to cite this article: Carter G, Spittal MJ, Glowacki L, Gerostamoulos D, Dietze P, Sinclair B, et al. Diagnostic accuracy for self-reported methamphetamine use versus oral fluid test as the reference standard in a methamphetamine-dependent intervention trial population. Addiction. 2023;118(3):470–9. https://doi.org/10.1111/add.16085