PUBLIC HEALTH 171 (2019) 57–65

Available online at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Review Paper

Are oral fluid testing devices effective for the roadside detection of recent cannabis use? A systematic review

S.C.D. Dobri*, A.H. Moslehi, T.C. Davies

Department of Mechanical and Materials Engineering, Queen's University, Kingston, ON, Canada

ARTICLE INFO

Article history:
Received 22 August 2018
Received in revised form
23 February 2019
Accepted 12 March 2019
Available online 15 May 2019

Keywords:
Cannabis
THC
On-site testing
Oral fluid
Roadside

ABSTRACT

Objectives: Although laws related to drug impairment may deter some drivers, enforcement requires effective detection. There are different methods and devices to test for cannabis use, but it is unclear if these devices meet the necessary criteria to be implemented at the roadside. This systematic review synthesized research that investigated on-site oral fluid drug screening devices.

Study design: This is a systematic review.

Methods: Eight databases (PubMed, Web of Science, MEDLINE, Engineering Village, Embase, Compendex, CINAHL, and Scopus) were searched to identify research that had evaluated the effectiveness of oral fluid testing devices. Fifteen articles that used an on-site testing device to detect cannabis use were selected for review.

Results: There is a lack of standardized test protocols with respect to biological matrices used for confirmation analysis (blood and oral fluid), concentration detection cutoff, population sample, and contamination with other drugs (alcohol). There is also a lack of device consistency making it difficult to draw conclusions. Sensitivity, specificity, and accuracy of nine devices showed that none of the current devices meet the minimum requirements suggested by the ROSITA, ROSITA-2, and DRUID projects (80% for all three parameters).

Conclusions: The results of this systematic review indicated that the devices with the ability to detect lower Δ9-tetrahydrocannabinol concentration levels achieved better results with respect to sensitivity, specificity, and accuracy than those with higher detection levels. However, research must be focused on developing a roadside detection oral fluid technique that meets the ROSITA, ROSITA-2, and DRUID projects' guidelines.

* *Corresponding author.*
E-mail address: stephan.dobri@queensu.ca (S.C.D. Dobri).

https://doi.org/10.1016/j.puhe.2019.03.006

Introduction

Cannabis is one of the most commonly used recreational drugs and has been reported as one of the most prevalent involved in impaired driving accidents.[1–3] However, many jurisdictions do not have a means to effectively detect recent cannabis use. As some governments are looking to legalize cannabis in future, for instance, the government of Canada has announced its intention to legalize cannabis by October of 2018[4] making cannabis more accessible to the public, a viable roadside testing method to screen for recent cannabis use is required.

Drug tests can be performed by obtaining body fluid samples from a person suspected of drug use. This screening must be able to identify Δ9-tetrahydrocannabinol (THC)—the major psychoactive constituent of cannabis. For a roadside test, a trained clinician must not be required, the results must be accurate, and testing should take less than 5 min.[5] Oral fluid (OF) has been shown to be a good indicator of recent cannabis use (within the last 4 h), and it is easy to collect by untrained personnel, making it a good testing matrix for roadside detection of recent cannabis use.[1,6,7]

The ROSITA, ROSITA-2, and DRUID are research projects conducted in Europe and the United States in the early 2000s to investigate drug testing devices and their ability to detect driving under the influence of drugs.[5,8,9] Commercially available OF roadside testing devices were evaluated, some with varying concentration cutoffs, to determine the effectiveness of each device. The concentration cutoff of a device is the lowest concentration of THC detectable in a sample and is stated by the manufacturer. The projects recommended that devices should have a minimum sensitivity, specificity, and accuracy of 80% each.[5,8,9] At the time of the projects, none of the evaluated devices met these criteria.[5,8,9] Since the publication of these research studies, newer devices have been developed, and the existing devices have been improved. This systematic review was developed to include the evaluation of newer devices.

Based on the data from applicable studies, this systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (see Appendix 2) in the evaluation of the effectiveness of portable OF testing devices for roadside detection of recent cannabis use.

Methods

Study design

This study is a systematic review.

Eligibility criteria

Papers had to be written in English and published in peer reviewed journals, had to have considered recreational cannabis use only, and had to use an on-site testing device on humans to be included in this review. Articles were excluded if they did not involve OF testing or if they did not use an on-site screening device.

Information sources and search

A search was conducted between February and April 2017 among eight databases (PubMed, Web of Science, MEDLINE, Engineering Village, Embase, Compendex, CINAHL, and Scopus) to identify research that had evaluated the effectiveness of OF testing devices. The search was performed with the following search string: (oral fluid test OR saliva test OR mouth swab OR drug test OR detection OR residue) AND (marijuana OR cannabis) AND (roadside OR road-side OR road side OR on-site OR on site)

Study selection

Titles were searched first to eliminate any papers that did not meet the inclusion criteria, followed by abstract review. The titles were searched by three investigators, and the abstracts were reviewed by at least two authors.

For each paper that was included, at least two investigators assessed the paper using the McMaster Quantitative Review Form (http://srs-mcmaster.ca/research/evidence-based-practice-research-group/). Using this form, any potential biases (i.e. subject, measurement, and intervention biases), study design, study participants, ethics procedure, outcome measures, intervention, and study results were identified. A total of 136 articles were identified. After removing the duplicates, 79 articles were obtained and reviewed. After the complete selection process, 15 articles were considered in this review. The selection process is illustrated in Fig. 1.

Data collection process

Each study compared performance of an on-site OF test with a laboratory test. If the results from both the device and laboratory were positive, the result would be interpreted as a true positive. If both results were negative, the results would be recorded as a true negative. If the result from on-site was positive, but the laboratory result was negative, the result would be a false positive (FP). Finally, if the result from on-site was negative, but the result from the laboratory was positive, the device's result would be considered as a false negative (FN). The true and false negative and positive results were then used to calculate the sensitivity, specificity, and accuracy of the devices and were reported in each article. These were calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \times 100$$

$$Specificity = \frac{TN}{TN + FP} \times 100$$

$$Accuracy = \frac{TP + TN}{Total\ number\ of\ tests} \times 100$$

The sensitivity indicates the rate at which people were correctly identified as having recently used cannabis. The

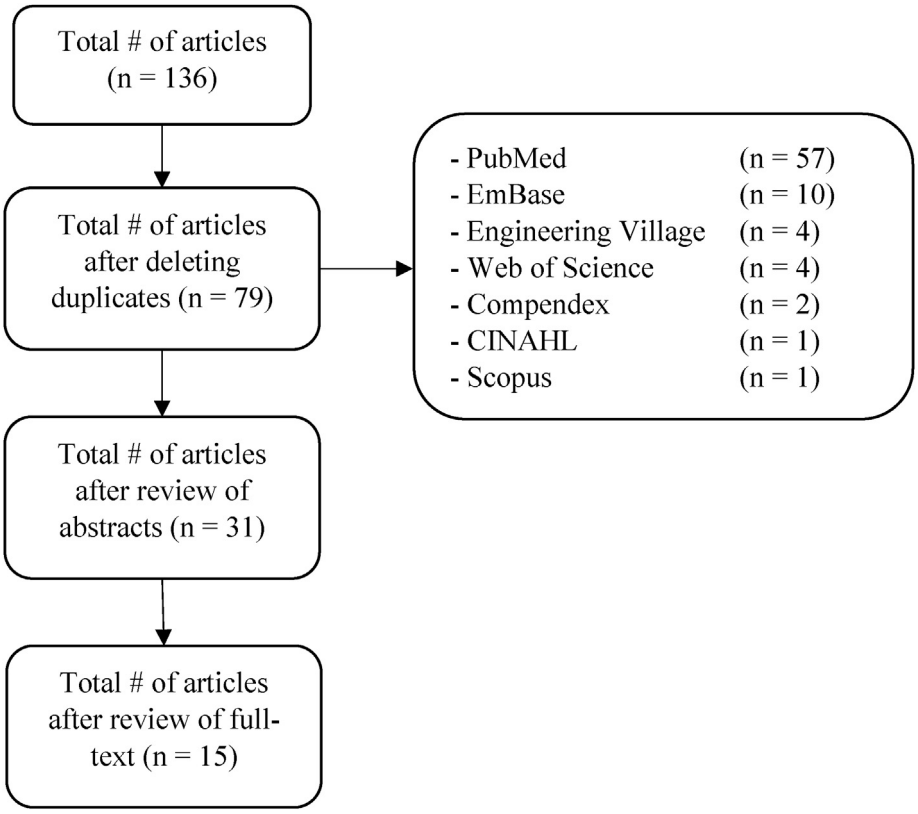


Fig. 1 – Flowchart showing the process of article selection for this systematic review.

specificity indicates the rate at which people were correctly identified as not having recently used cannabis, and the accuracy is a measure of both. Devices with a low sensitivity would not detect all the impaired drivers who were tested, while devices with a low specificity would identify high numbers of sober drivers as being impaired.[9] Each article calculated and reported the sensitivity, specificity, and accuracy of the devices tested.

Summary measures

In cases where the same on-site drug screening device was used across multiple studies, the weighted mean and variance of sensitivity, specificity, and accuracy were calculated. These values were then used to calculate the effect size (Cohen's d) between different devices. The weighted means, standard deviations, and Cohen's d's were calculated as follows:

$$\textit{Weighted mean} : \bar{x} = \frac{\sum N_i x_i}{\sum N_i}$$

$$\textit{Weighted Variance} : s_x^2 = \frac{\sum \left[N_i (x_i - \bar{x})^2\right]}{\sum N_i}$$

$$95\%\ \textit{Confidence Interval} : CI = \bar{x} \pm 1.96\sqrt{\frac{s_x^2}{k}}$$

$$\textit{Standard Deviation} : s = \sqrt{\frac{(N_t - 1)s_t^2 + (N_c - 1)s_c^2}{N_t + N_c}}$$

$$\textit{Cohen's d} : d = \frac{\bar{x}_t - \bar{x}_c}{s}$$

where N is the number of participants in a group; k is the number of studies a device was used in; and the subscripts t and c represent 'test' and 'control' (i.e. the two devices that are being compared), respectively. The sign of the effect size is dependent on which device is chosen as 'test' or 'control', and the order has been noted in the Results and discussion section.

Performance of each device was reported only as sensitivity, specificity, and accuracy in each article, so there was not enough statistical information to complete a typical meta-analysis.

Results and discussion

Performance of on-site screening devices

Table A in Appendix 1 summarizes the data reported from all included studies. Table 1 shows that the currently available

Table 1 – Summary of sensitivity, specificity, and accuracy of different on-site OF drug screening devices.

Device	Device cutoff (ng/mL)	Sensitivity (%)	Specificity (%)	Accuracy (%)
OraLine[7]	4	69	92	74
Cozart DDS 806[10]	31	22	100	71
BIOSENSE Dynamic[10]	Unknown	50	Not reported	51
OraLab 6[10]	50	16	99	61
OrAlert[10]	100	11	100	78
Oratect III[10]	40	32	100	41
RapidSTAT[a,11,12]	15	58 (7)	85 (17)	76 (6)
DrugWipe[a,6,11–14]	15–30	51 (11)	88 (4)	82 (5)
DrugTest 5000[a,3,10–12,15]	5–25	75 (8)	73 (14)	78 (6)

[a] A weighted average was taken for the sensitivity, specificity, and accuracy in this table. Weighted standard deviations are in brackets.

on-site devices do not have sufficient sensitivity and accuracy based on the ROSITA, ROSITA-2, and DRUID projects' recommendations. The sensitivities of all devices are 75% or less, whereas the ROSITA and DRUID projects recommended a minimum value of 80%.[5,8,9] Furthermore, as shown in Table 1, the results for the DrugWipe and DrugTest 5000 are highly variable across studies. This variability may be due to different sample populations, device and laboratory cutoffs, and reference specimens (i.e. blood or OF).

The weighted means and 95% confidence intervals for all devices' sensitivities, specificities, and accuracies are shown in Fig. 2. The weighted means and variances were used to calculate the effect size, Cohen's d. The effect size allows more accurate comparisons to be drawn between results from various studies by accounting for the variance and sample size of each device being compared. The effect sizes are shown in Fig. 3.

The effect sizes support the conclusions that would be drawn directly from the weighted means. Of the three devices compared across studies, the Dräger DrugTest 5000 was the most sensitive device, and the DrugWipe was the least sensitive, whereas it had the highest specificity and accuracy. The DrugTest had the lowest specificity, and the RapidSTAT had the lowest accuracy. Again, none of the tested devices met the ROSITA, ROSITA-2, or DRUID projects' recommendations; however, the DrugWipe outperformed the other two devices on two of the three outcome measures.

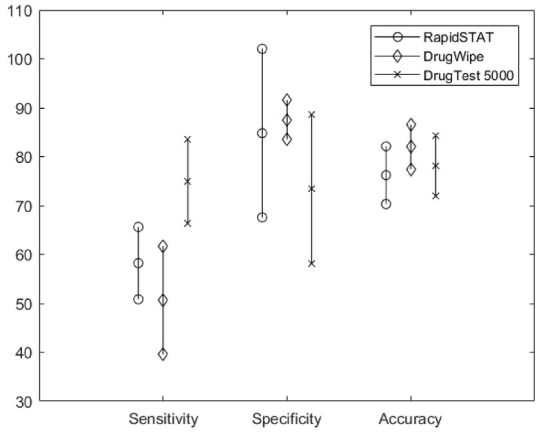


Fig. 2 – Weighted mean and 95% confidence intervals for device sensitivity, specificity, and accuracy across studies.

It should be noted that Fig. 2 does not include the confidence intervals usually found on forest plots because of insufficient information to perform these calculations within the studies.

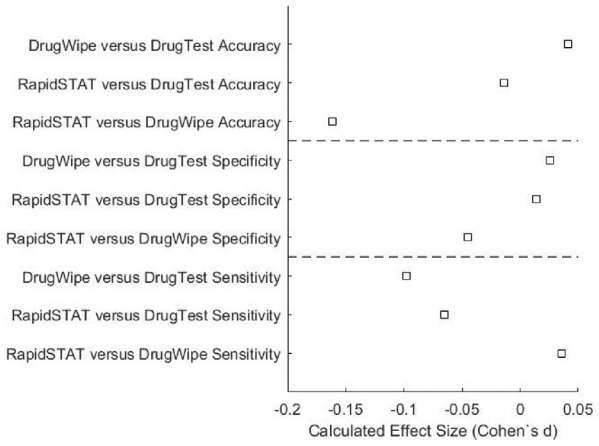
In most studies, the laboratory test cutoff used to confirm the on-site testing results was different than that of the portable testing device. This difference can impact the performance of the device so devices with lower cutoffs generally outperformed devices with higher cutoffs.

Performance of on-site devices overtime

Two of the reviewed studies investigated the performance of on-site devices overtime.[1,10] In the study by Bosker et al.,[1] 20 heavy cannabis smokers (aged 24.3 [1.4] years) smoked cigarettes with THC concentrations normalized to their body weight (400 μg/kg). OF samples were collected and analyzed every 15 min in the first hour, and subsequently, every 30 min, to 4 h after smoking using the Dräger DrugTest 5000 and Securetec Drugwipe 5. With the Securetec Drugwipe 5, OF samples were collected from both the cheek and tongue, and the results from each type of sample were reported separately. Blood samples were also collected as reference specimens for confirmatory laboratory analysis.

Fig. 4, which was derived from the data presented by Bosker et al.,[1] depicts the sensitivity results of both on-site devices up to 4 h after smoking. As shown in this figure, the sensitivity of the Dräger DrugTest 5000 tends to stay constant for the first two and a half hours after smoking, whereas the sensitivity of the DrugWipe 5 is significantly reduced 45 min after smoking and is always lower than the Dräger DrugTest 5000. The difference in sensitivity between the two devices could be due to the difference in their THC cutoff values (5 versus 30 ng/mL). Because the participants were never drug-free during the study, the number of FPs was not calculated.

In Wille et al.,[10] ten chronic cannabis smokers between the ages of 18 and 40 years smoked two cannabis cigarettes 75 min apart containing THC concentrations normalized to their body weight (i.e. 300 and 150 μg/kg). OF samples were collected prior to smoking the first cigarette, 5 min after smoking each cigarette, and 80 min after smoking the second cigarette. The OF samples were analyzed on-site using a newer version of Securetec DrugWipe 5S with a THC cutoff of 15 ng/mL (compared to 30 ng/mL in the previous model). Both OF and blood samples were used as reference specimens in the laboratory confirmatory analysis with THC cutoffs of 1 and 10 ng/



Devices	Sensitivity Effect Size	Specificity Effect Size	Accuracy Effect Size
RapidSTAT and DrugWipe	0.036	-0.045	-0.162
RapidSTAT and DrugTest	-0.065	0.014	-0.014
DrugWipe and DrugTest	-0.098	0.026	0.041

The order in which the devices are written is the order they were put into the equation for Cohen's d.

Fig. 3 – Forest plot of Cohen's d effect sizes between devices. The exact effect size for sensitivity, specificity, and accuracy is also shown.

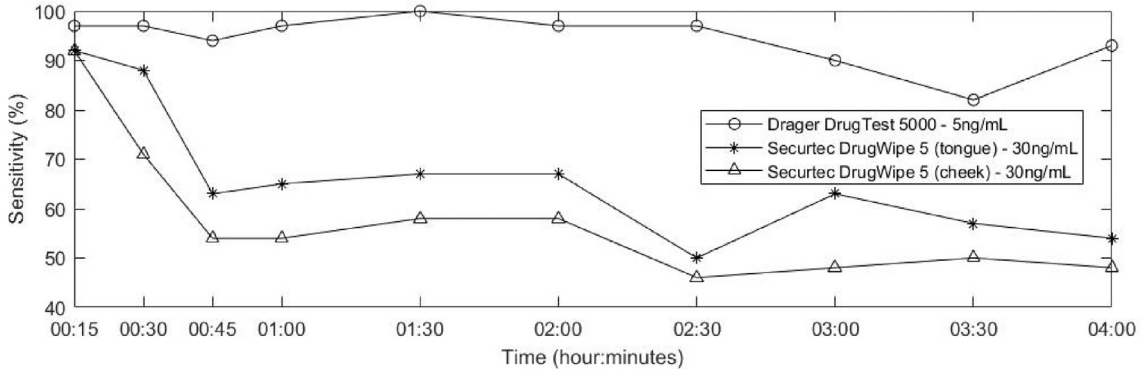


Fig. 4 – Device sensitivity for Dräger DrugTest 5000 and Securtec DrugWipe 5 from 15 min up to 4 h after smoking cigarettes containing THC.[1]

mL for OF and 1 and 5 ng/mL for blood samples. The device's sensitivity and accuracy declined with time after smoking the cigarette and were higher when OF was used as a reference specimen compared to blood. These results indicate that on-site OF testing is a viable option to determine recent cannabis use but becomes less effective as time from consumption increases.

Device limitations

The devices tested by each study had several limitations. These limitations include the volume of OF required for detection, device usability, and cold-weather constraints.

One of the main limitations was the collection of a sufficient OF volume for the testing devices.[11] An insufficient sample of OF could lead to incorrect test results. Some devices (the Cozart DDS 806, DrugTest 5000, and Oratect III) have built-in indicators to show when a full sample has been collected, decreasing the chance of collecting incorrect sample volumes.[11] Also, smoking cannabis can cause dry mouth, making it difficult to collect a full OF sample from a suspected driver. Two studies found that drinking a beverage before collecting OF either did not impact the THC concentration or would slightly increase the THC concentration (10% increase in 7 of 41 participants).[13,14] These results indicate that it could be beneficial to give a suspected driver a small drink of water (less than 300 mL) prior to taking an OF sample. The small drink would ensure that a sufficient amount of OF would be obtained.

Device usability was another major limitation. For instance, the on-site testing times ranged from 2 to 30 min.[11] As previously mentioned, a roadside test should

take less than 5 min to perform. Devices that took less than 5 min included the BIOSENSE Dynamic, Cozart DDS 806, and the DrugTest 5000.[11] Some devices, such as the Oratect III, had extremely high variability in testing time (5–30 min), making them difficult to implement on the roadside.[11]

Another issue with device usability was the rate of failed tests. Failed tests arose when the results could not be interpreted or if control lines did not appear on a test strip.[11]

There were also difficulties reading different devices, including the DrugWipe 5A, as the test line was often very weak and delayed.[15] Difficulties reading the device could lead to a high rate of impaired drivers not being properly identified.

Finally, the cold weather affected the ability of a device to detect and display a result. Devices such as the DrugTest 5000, RapidSTAT, DrugWipe, and DrugWipe 5^{+} were reported to have issues in cold weather.[12,15] This would limit the places and times the devices could be used at the roadside.

Review limitations

A major limitation of the studies is the lack of standardized testing. Different studies used different biological matrices (blood plasma or OF) for confirmation analysis with different cutoff concentrations ranging from 1 to 30 ng/mL. Different matrices and cutoff levels could drastically impact device performance, which would impact the summarized results. Additionally, none of the reviewed articles discussed correlations between THC concentration in OF and the level of impairment. A known concentration linked to impairment, as exists for alcohol, would set a precedent for standardized testing of these devices.

Different testing protocols were used across studies. Many collected data at the roadside, while others collected samples in controlled settings or in public places where drugs were used (such as a cannabis coffee shop in the Netherlands and night clubs in Rome).[11,15] Different populations could yield different results, which could skew the sensitivity, specificity, and accuracy results of such a test.

Finally, most devices were evaluated in single studies. Only the RapidSTAT, DrugWipe, and DrugTest 5000 were assessed in enough studies to have the effect sizes calculated in this review. Other devices appeared in one study, and results could not be compared. Typically, an effect size is calculated in each study, and so, variance in effect size can be calculated to generate a full forest plot. Had different studies used and compared multiple testing devices and calculated relevant effect sizes, the confidence intervals could have been calculated and included in this review. Results were also only reported as sensitivity, specificity, and accuracy without additional statistical data, which limited analysis in this review.

Recommendations

Statistical information on device performance, such as repeatability of results, should be determined. Knowing statistical information on device performance would allow studies to calculate effect sizes between tested devices, as well as allow future reviews to complete more in-depth comparisons between devices. Repeatability of devices could be determined by taking multiple samples from the same participant and comparing the results of the device across these samples.

Emerging technologies

Two new prototypes were identified, which are still in the development stage. Plouffe and Murthy[16] used fluorescent nanoparticles to determine the concentration of THC in an OF sample. In preliminary testing, the device was able to correctly determine the THC concentration in OF samples within 10% of the actual concentration.[16] This device is still in early stages of development and will require more testing before it could be implemented.

The second device created by Wanklyn et al.[2] is a screen-printed sensor which uses electrochemical detection for THC. The device was created to be portable, fast, and easy to use. Initial tests took roughly 30 s; however, the sensitivity, specificity, and accuracy were 28%, 99%, and 52%, respectively. The researchers noted a very high number of FNs but a small number of FPs.

Conclusions

OF has been deemed a suitable biological matrix for the detection of recent cannabis use by the studies included in this review. OF can be obtained in an easy and non-invasive manner. However, the methods of measuring OF at the roadside need to be improved. New research should focus on these gaps to allow for enforcement of impaired driving laws.

Author statements

Acknowledgments

The authors would like to thank A. Limoges-Friend for his help with the conception and design of this article and the acquisition of data.

Ethical approval

Ethical approval was not needed for this systematic review as it analyzed existing published data and did not perform new experiments or obtain information from new or previous participants of the studies synthesized.

Funding

This work was supported through funding from the Natural Sciences and Engineering Research Council grant number 513272-17.

Competing interests

None declared.

REFERENCES

1. Bosker WM, Theunissen EL, Conen S, Kuypers KPC, Jeffery WK, Walls HC, et al. A placebo-controlled study to assess Standardized Field Sobriety Tests performance during alcohol and cannabis intoxication in heavy cannabis users and accuracy of point of collection testing devices for detecting THC in oral fluid. cited 2017 Feb 8 *Psychopharmacology (Berl) [Internet]* 2012 Oct;**223**(4):439–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22581391.
2. Wanklyn C, Burton D, Enston E, Bartlett C-A, Taylor S, Raniczkowska A, et al. Disposable screen printed sensor for the electrochemical detection of delta-9-tetrahydrocannabinol in undiluted saliva. *Chem Cent J* 2016 Jan;**10**(1).
3. Desrosiers NA, Lee D, Schwope DM, Milman G, Barnes AJ, Gorelick DA, et al. On-site test for cannabinoids in oral fluid. cited 2017 Feb 8 *Clin Chem [Internet]* 2012 Oct;**58**(10):1418–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22912396.
4. Health Canada. *Canada takes action to legalize and strictly regulate cannabis - Canada.ca [Internet]*. 2017 [cited 2017 May 23]. Available from: https://www.canada.ca/en/health-canada/news/2017/04/canada_takes_actiontolegalizeandstrictlyregulatecannabis.html.
5. Verstraete A. ROSITA Project: Roadside Testing Assessment. [cited 2017 Nov 15]; Available from: http://www.transport-research.info/sites/default/files/project/documents/rositarep.pdf.
6. Pehrsson A, Gunnar T, Engblom C, Seppä H, Jama A, Lillsunde P. Roadside oral fluid testing: comparison of the results of drugwipe 5 and drugwipe benzodiazepines on-site tests with laboratory confirmation results of oral fluid and whole blood. cited 2017 Feb 8 *Forensic Sci Int [Internet]* 2008 Mar 5;**175**(2–3):140–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17640837.
7. Cirimele V, Villain M, Mura P, Bernard M, Kintz P. Oral fluid testing for cannabis: on-site OraLine IV s.a.t. device versus GC/MS. cited 2017 Feb 8 *Forensic Sci Int [Internet]* 2006 Sep 12;**161**(2–3):180–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16854544.
8. Verstraete AG, Raes E. *ROSITA-2 project*. Final report. Ghent; 2006.
9. Schulze H, Schumacher M, Urmeew R, Auerbach K, Alvarez J, Bernhoft IM, et al. Findings from the DRUID project Driving Under the Influence of Drugs, Alcohol and Medicines in Europe — findings from the DRUID project. [cited 2017 Nov 15]; Available from: http://www.emcdda.europa.eu/attachements.cfm/att_192773_EN_TDXA12006ENN.pdf.
10. Wille SMR, Di Fazio V, Toennes SW, van Wel JHP, Ramaekers JG, Samyn N. Evaluation of Delta(9)-tetrahydrocannabinol detection using DrugWipe5S((R)) screening and oral fluid quantification after Quantisal (TM) collection for roadside drug detection via a controlled study with chronic cannabis users. *Drug Test Anal* 2015 Mar;**7**(3):178–86.
11. Blencowe T, Pehrsson A, Lillsunde P, Vimpari K, Houwing S, Smink B, et al. An analytical evaluation of eight on-site oral fluid drug screening devices using laboratory confirmation results from oral fluid. cited 2017 Feb 8 *Forensic Sci Int [Internet]* 2011 May 20;**208**(1–3):173–9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S037907381000527X.
12. Wille SMR, Samyn N, Ramírez-Fernández M del M, De Boeck G. Evaluation of on-site oral fluid screening using Drugwipe-5(+), RapidSTAT and Drug Test 5000 for the detection of drugs of abuse in drivers. cited 2017 Feb 8 *Forensic Sci Int [Internet]* 2010 May 20;**198**(1–3):2–6. Available from: http://linkinghub.elsevier.com/retrieve/pii/S037907380900423X.
13. Hartman RL, Anizan S, Jang M, Brown TL, Yun K, Gorelick DA, et al. Cannabinoid disposition in oral fluid after controlled vaporizer administration with and without alcohol. *Forensic Toxicol* 2015 Jul;**33**(2):260–78.
14. Toennes SW, Schneider K, Wunder C, Kauert GF, Moeller MR, Theunissen EL, et al. Influence of ethanol on the pharmacokinetic properties of Δ9-tetrahydrocannabinol in oral fluid. cited 2017 Feb 8 *J Anal Toxicol [Internet]* 2013 Apr 1;**37**(3):152–8. Available from: https://academic.oup.com/jat/article-lookup/doi/10.1093/jat/bkt002.
15. Gentili S, Solimini R, Tittarelli R, Mannocchi G, Busardo FP. A study on the reliability of an on-site oral fluid drug test in a recreational context. *J Anal Methods Chem* 2016;**2016**:1234581. 10 pages. https://doi.org/10.1155/2016/1234581.
16. Plouffe BD, Murthy SK. Fluorescence-based lateral flow assays for rapid oral fluid roadside detection of cannabis use. cited 2017 Feb 27 *Electrophoresis [Internet]* 2017 Feb;**38**(3–4):501–6. Available from: http://doi.wiley.com/10.1002/elps.201600075.
17. Pehrsson A, Blencowe T, Vimpari K, Langel K, Engblom C, Lillsunde P. An evaluation of on-site oral fluid drug screening devices DrugWipe 5+ and Rapid STAT using oral fluid for confirmation analysis. cited 2017 Feb 8 *J Anal Toxicol [Internet]* 2011 May;**35**(4):211–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21513614.
18. Laloup M, Fernandez M, del MR, Wood M, De Boeck G, Maes V, Samyn N. Correlation of Delta 9-tetrahydrocannabinol concentrations determined by LC-MS-MS in oral fluid and plasma from impaired drivers and evaluation of the on-site Drager DrugTest (R). *Forensic Sci Int* 2006 Sep;**161**(2–3):175–9.

Appendix 1. Data from included papers

Table A – Sensitivity, specificity, and accuracy of on-site OF drug screening devices reported across different studies.

Study	Device	Device cutoff (ng/mL)	Laboratory cutoff (ng/mL)	Sensitivity (%)	Specificity (%)	Accuracy (%)	# of participants
[7]	OraLine	4	1	69	92	74	27
[11]	Cozart DDS 806	31	1	22	100	71	138
[11]	BIOSENSE Dynamic	Unknown	1	50	Not reported	51	39
[11]	OraLab 6	50	1	16	99	61	248
[11]	OrAlert	100	1	11	100	78	95
[11]	Oratect III	40	1	32	100	41	28
[11]	RapidSTAT	15	1	56	90	78	333
[12]	RapidSTAT	15	2 blood	71	55	66	58
[17]	DrugWipe 5/5^{+}	30	1	43	87	82	1807
[6]	DrugWipe 5	30	2	52	91	85	266
[6]	DrugWipe 5	30	1 blood	68	88	85	266
[15]	DrugWipe 5A	30	0.6 ng/pad	29	88	53	83
[12]	DrugWipe 5	30	2 blood	71	50	63	46
[14]	DrugWipe 5^{+} (cheek)	30	Unknown	88	94	88	112
[14]	DrugWipe 5^{+} (tongue)	30	Unknown	89	94	89	144
[11]	DrugTest 5000	5	1	59	96	82	218
[18]	DrugTest 5000	20	0.5	49	100	55	127
[18]	DrugTest 5000	20	0.5 blood	51	93	56	127
[18]	DrugTest 5000	20	2.0 blood	58	88	66	127
[12]	DrugTest 5000	25	2 blood	72	50	68	76
[12]	DrugTest 5000	5	2 blood	93	71	90	48
[14]	DrugTest 5000	5	Unknown	94	15	91	282
[14]	DrugTest 5000	5	1 blood	88	71	80	108
[3]	DrugTest 5000	5	2	91	75	88	66
[3]	DrugTest 5000	5	1	88	78	86	66
[3]	DrugTest 5000	5	0.5	86	75	85	66
[3]	DrugTest 5000	10	10	93	76	86	66
[3]	DrugTest 5000	10	2	82	100	85	66
[3]	DrugTest 5000	10	1	77	100	80	66
[3]	DrugTest 5000	10	0.5	76	100	79	66

If blood was used as reference specimen in laboratory, it is noted under 'Laboratory cutoff' column.

Appendix 2. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Title page
Abstract			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	1
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1, 2
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	2
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g. web address), and, if available, provide registration information including registration number.	N/A

(continued)

Section/topic	#	Checklist item	Reported on page #
Eligibility criteria	6	Specify study characteristics (e.g. PICOS, length of follow-up) and report characteristics (e.g. years considered, language, publication status) used as criteria for eligibility, giving rationale.	Appendix 1: Detailed Search Strategy
Information sources	7	Describe all information sources (e.g. databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	2, 3 and Appendix 1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e. screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Appendix 1
Data collection process	10	Describe method of data extraction from reports (e.g. piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	3 and Appendix 1
Data items	11	List and define all variables for which data were sought (e.g. PICOS, funding sources) and any assumptions and simplifications made.	3, 4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Appendix 1
Summary measures	13	State the principal summary measures (e.g. risk ratio, difference in means).	4, 5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. I^2) for each meta-analysis.	4
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g. publication bias, selective reporting within studies).	Appendix 1
Additional analyses	16	Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.	N/A
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	3
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g. study size, PICOS, follow-up period) and provide the citations.	6–8
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12).	Appendix 1
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9, 10
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9, 10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Appendix 1
Additional analysis	23	Give results of additional analyses, if done (e.g. sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
Discussion			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g. healthcare providers, users, and policy makers).	5–14
Limitations	25	Discuss limitations at the study and outcome level (e.g. risk of bias), and at the review level (e.g. incomplete retrieval of identified research, reporting bias).	12–14
Conclusions	26	Provide a general interpretation of the results in the context of other evidence and implications for future research.	14
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (e.g. supply of data); role of funders for the systematic review.	Summary information on article submission

From Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. https://doi.org/10.1371/journal.pmed1000097.
For more information, visit: www.prisma-statement.org.