SUURJ: Seattle University Undergraduate Research Journal

Volume 8 Article 12

2024

Can the Results Be Trusted? Assessing the Reliability of In-Home Drug Tests for the Detection of THC in Urine

Éabha R. Finn
Seattle University, Seattle

Jennifer Huffman

Madison Calvert
Seattle university, Seattle

Kristen Skogerboe
Seattle University

Can the Results Be Trusted? Assessing the Reliability of In-Home Drug Tests for the Detection of THC in Urine

Éabha Finn, Chemistry
Jennifer Huffman, Chemistry and Criminal Justice
Madison Calvert, Forensic Science

Faculty Mentor:
Kristen Skogerboe, PhD, Chemistry

Faculty Content Editor:
Renata Everett, PhD, Chemistry

Student Editors:
Krishna Rakotonirina and William Smolskis

Abstract

Drug testing, a crucial tool in clinical, occupational, parental, and forensic settings, aims to detect and deter illicit substance use. The accuracy of a urine drug test is determined by both *sensitivity*, the ability to correctly identify true positives, and *specificity*, the ability to correctly identify true negatives. This study assesses the accuracy of five different lateral flow immunoassay (LFI) testing devices for marijuana usage on 22 anonymous urine samples (12 positives and 10 negatives) as performed by students in Seattle University's forensic science laboratory. Students on site interpreted the results of their tests, and photographs of the tests were interpreted by four blind participants. Overall, 10 of the 22 samples had a false result with at least one of the LFI tests. Of the five test devices that were investigated, only one demonstrated 100% accuracy with sensitivity and specificity. The other four devices had sensitivities ranging from 45 to 100% and specificities ranging from 80 to 100%. These findings underscore the considerable variability in sensitivity and specificity, factors that are contingent on the test kit itself and the person interpreting the test results. Four different types of errors were identified to account for these inaccurate results: 1) false invalid interpretation, 2) switching the sign of a positive and negative result, 3) difficulty in visual confirmation of the presence of a test line, and 4) sample matrix variability, which occurs when constituents in the sample, other than the substance being measured, impact the result. The primary consumers of drug tests—healthcare workers, employers, parents, law enforcement agencies, and forensic experts—have tremendous influence on our society; the results of this study generate a valuable, challenging discussion about the need for standardized interpretation and enhanced accuracy within the drug testing industry.

Introduction

There has been an increased reliance on over-the-counter test kits to detect a variety of different health concerns, including pregnancy, viral infection, and drug use within domestic spaces, workplaces, and healthcare facilities. Drug testing is a common practice in the United States. One study estimated that 46% of all workers have undergone employment-related drug testing, and approximately 30 million people have been subjected to workplace urine drug testing (Carpenter 2007; Oh et al. 2023). The Bureau of Labor Statistics estimated that 70% of employers with more than 1000 employees conduct workplace drug testing (Hartwell et al. 1996).

Urine drug screenings are conducted outside of clinical facilities for a variety of reasons. For example, individuals can test themselves at home to ensure they are likely to pass upcoming drug screenings. Parents can use LFI tests to monitor their children's potential drug use, and drug treatment centers will often use them to assess clients' sobriety. Parole boards rely on screenings to detect the presence of drugs and alcohol in individuals legally required to maintain sobriety (Del Carmen and Sorenson 1988). In these scenarios, the typical test kit is a test kit based on a lateral flow immunoassay (LFI) platform. These test kits assess a sample by binding the substance of interest (such as a protein or hormone) to an antibody. In the case of drug testing, it is the actual drug (or its metabolite) binding to the antibody (Koczula and Gallotta 2016).

Figure 1 shows a typical LFI test for drug detection in urine. When this testing surface is exposed to the urine sample the liquid flows over the antibody region via capillary action. If the concentration of the drug is below the target threshold concentration, dyes bind to the antibody which produces a visible test line, indicative of a negative sample. If there is a sufficient drug present above the threshold concentration, the drug instead binds to the antibody and blocks dye from this binding site. In this case, no line is produced, yielding a positive sample.

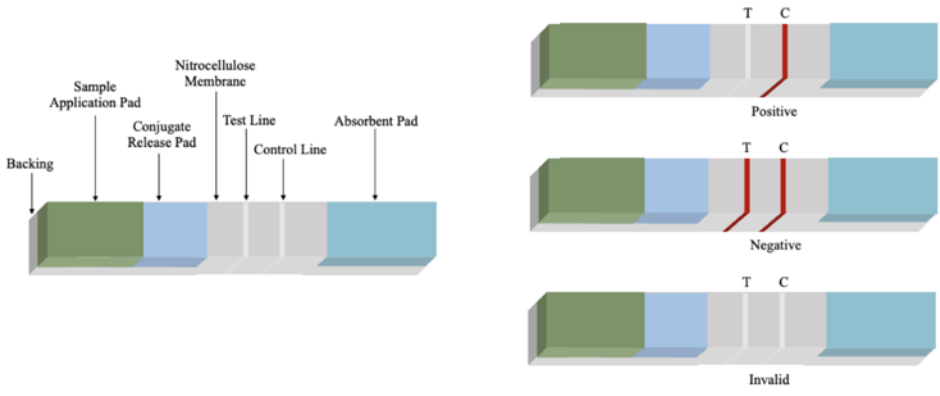


Figure 1 Mechanism of an LFI device for drug testing with key of possible results.

It is important to note that drug testing LFI devices have an opposite format for positive and negative results than LFI kits for COVID-19, which many people are familiar with (Arshadi et al. 2022). In COVID tests, a positive result is denoted by the *presence* of a test line, while a positive LFI drug test is denoted by the *absence* of a test line. This difference increases the likelihood of false results due to unintentional misinterpretation by users familiar with COVID testing procedures.

The Food and Drug Administration (FDA), under the Clinical Laboratory Improvement Amendments (CLIA), is responsible for overseeing products that are used to measure chemicals, specifically in bodily fluids such as urine (CLIA 2020). CLIA is a set of federal regulations that were passed in 1988 to ensure oversight of laboratory practices and reduce diagnostic errors within laboratory settings. The FDA has since declared that rapid test kits can be used at home, by the untrained user, for 142 different substances (CLIA 2019). This home use approval, known as "CLIA-waived," is an FDA designation used for simple testing methods with a purportedly low rate of error. LFI kits for home and workplace drug testing are CLIA-waived and may be performed by untrained users as long as they read and follow the instructions.

There is concern about error when using CLIA-waived drug tests, particularly in legal or workplace contexts. While the FDA deems error rates for CLIA-waived tests to be small, LFI screening tests are known to have false positives and negatives (Algren and Christian 2015; Kapur 2012; Stellpflug et al. 2020). In settings with legal ramifications, positive screening tests for drugs are always followed up with confirmatory tests, but this still does not guarantee an accurate result, especially if there are errors with manufacturing or test interpretation.

LFI kits do not actually test marijuana's psychoactive ingredient, delta-9-tetrahydrocannabinol; instead, LFI kits measure a secondary metabolite, 11-nor-9-carboxy-THC (THC-COOH). A positive LFI test is expected if the sample has a THC-COOH concentration above 50ng/mL, the threshold to produce a positive result. The chemical structures of THC and THC-COOH are shown in Figure 2.

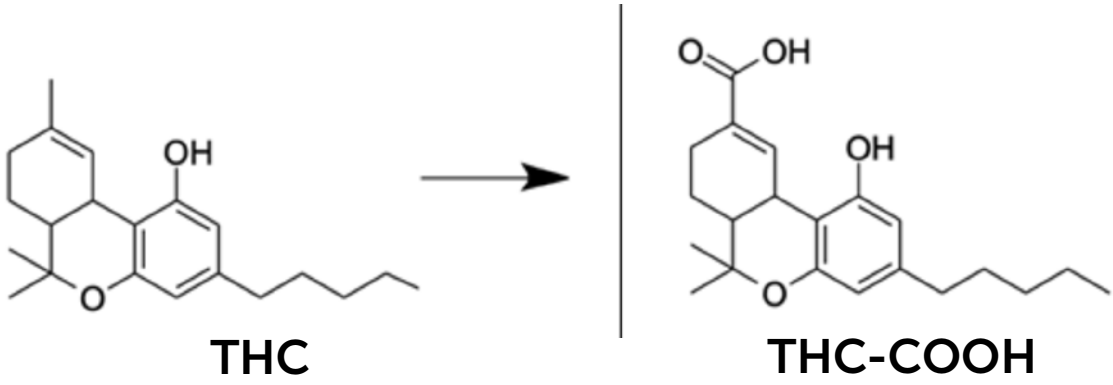


Figure 2 The metabolization of THC to THC-COOH, the metabolite detected using LIF at > 50 ng/mL for a positive result for all kits used in this study.

Two studies evaluating the reliability of non-CLIA-waived, FDA-regulated immunoassay testing indicated that positive screening tests for THC performed by trained toxicology technologists were accurate 99 to 100% of the time (Stuck 1996) with a false positivity rate of 4% (UTDM 1983). Since these studies were conducted in accredited labs using FDA-approved clinical tests, these statistics represent the best-case scenario for LFI testing. But how often do truly positive drug samples yield false negatives with CLIA-waived texts?

The results from this study raise concerns about the accuracy of CLIA-waived LFI drug kits in comparison to tests performed by experienced technicians in controlled, monitored environments. This study seeks to evaluate the reliability of CLIA-waived LFI kits that determine cannabinoid content in urine. The reliability of five LFI kits (purchased from online vendors) was evaluated by members of a college laboratory course in Forensic Science. The test's *sensitivity*, the ability to correctly identify urine samples that contain THC, and *specificity*, the ability to correctly identify urine samples that do not contain THC, were calculated using the equations given in Figure 3.

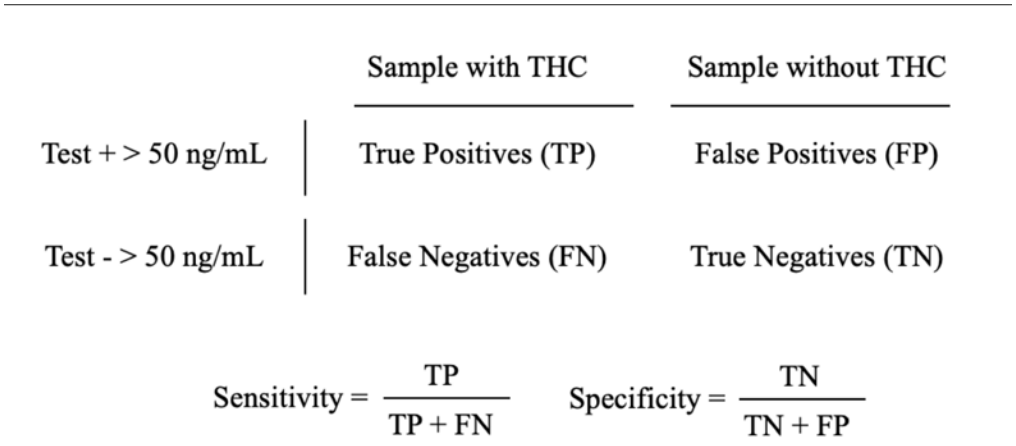


Figure 3 Calculations and information for specificity and sensitivity analysis.

Additional assessment of the tests was undertaken to determine if results were impacted by the reliability of visual interpretation rather than inherent problems with the sample or test kit. This was accomplished by having four independent chemists blindly interpret photographs of each test result.

Methods and Materials

Sample Acquisition and Personnel Performing the Testing

Students enrolled in Seattle University's Forensic Science laboratory, CRJS 4895, performed the tests on the samples. The students were junior and senior standing and had

taken coursework in chemistry, biology, physics, and forensic science. On the day of lab, fresh urine samples were collected from 11 anonymous donors in sterile urine collection cups.

Sample Preparation and Testing

Within an hour of collection, the course faculty (professor and teaching assistant) screened the samples for THC using a multilevel THC lateral flow immunoassay (U-Catch 5-multilevel test kit; Amazon). The samples were split into two 100 mL volumes and given separate identifiers. One of each split sample was spiked with 100 mL of an aqueous stock solution of (-) 11-Nor-9-carboxy-THC (Sigma Aldrich; T-018; suitable for immunoassay) to yield a final concentration of approximately 100 ng/mL, as measured by the multilevel THC test. This process created 22 urine samples. Although an initial test for THC had already been performed, the samples were not labeled to indicate those results, so testers performed subsequent tests blind. All 22 samples were placed in the lab in numerical order, and teams of two students were assigned a test kit from Table 1. All kits were still valid for testing, as they were used before the manufacturer's expiration date. Each pair of testers used their kit to assess all 22 samples.

Aside from the multilevel drug kit, the evaluated devices were not named fully for proprietary reasons. Each sample was given a 2 to 4 letter code name denoting the first letters of major syllables or words in the product name. The product code names, distributors, expiration dates, and costs per unit are listed in Table 1. Each team used the assigned test kit to determine if THC-COOH was present above the 50 ng/mL threshold specified for each kit. After 5 minutes each team then read the LFI as either positive, negative, or invalid. Results were then photographed using a personal cell phone camera under the lighting provided in the laboratory with no specific photographic conditions.

Results for the tests were recorded according to standard laboratory record-keeping practices and then submitted on a testing worksheet that included the photographs taken of each sample. The on-site results represent the day of testing interpretation of the data by the student team for each sample, as reported on the students' worksheet and compiled by the teaching assistant. This data was used to determine the on-site report sensitivity and specificity of each test kit.

Table 1 Lateral Flow Immunoassay devices used in this study.

Abbreviated Vendor Name	Distributor	Expiration date	Cost/test ($)
AL[a]	ToxTests; Dayton, MT	07/31/23	0.80
IVB[b]	Amazon; Seattle, WA	04/21/24	0.37
EZH[a]	Amazon; Seattle WA	05/19/24	0.58
EZLV[a]	Amazon; Seattle WA	08/06/24	0.78
WF[b]	Amazon; Seattle WA	07/11/23	0.40
Drug Exam (Multi) [a]	Amazon; Seattle WA	08/15/24	2.99

[a] indicates a product with a solid cassette and [b] is a product that is paper without strong backing

Interpretation of Photographic Data

Photographs of each brand of test kit for 22 samples were submitted by the student teams. These were compiled and separately evaluated as positive or negative by each of the authors, who were assigned the ID of Photo Interpreter 1 (PI-1), Photo Interpreter 2 (PI-2), Photo Interpreter 3 (PI-3), and Photo Interpreter 4 (PI-4).

Data Analysis

Data was collected in spreadsheets, and sensitivity and specificity were calculated using Microsoft Excel.

Results and Discussion

For the 22 samples tested using five different test kits, 109 different results were recorded and photographed. One sample (#21; EZLV) was inadvertently not tested by one of the testing teams so no photo was taken. There were also no discrepant results for sample #21 with any of the other test kits. Examples of true positive and true negative results are shown for each of the five test kits in Table 2 (see Appendix). False positive and false negative examples are provided in Table 3 (see Appendix). Two samples [#2 (WF) and #14 (WF)] yielded on-site invalid test results due to the lack of appearance of a control line. Photographs of these invalid results were submitted to the photo interpreters and there was 100% agreement with the on-site report. Sample #20 (EZH) was deemed valid for an on-site negative report by the photo interpreters. Two photo interpreters disagreed, calling the control line invisible, thus an invalid test. The remaining photo interpreters read sample #20 as positive.

Details of these invalid results and other false results described below are provided in Table 4 (see Appendix).

Overall, there were 106 on-site test results with a visible positive result. These were the results used to assess if the test yielded a true value based on spiking (e.g., true positive, true negative) and calculating the sensitivity and specificity of the on-site report for each kit.

There were nine samples with valid test results for all kits where false results were present. In aggregate, ten of the 22 samples (45.5%) had a false result. For four of these ten samples, the on-site report and the photo interpreters were in 100% agreement. The remaining six samples with false results had variable agreement between the interpreters. Of these 10 false results, there was one false positive and nine false negatives.

Examination of the false results indicates that there are various contributing factors and there may be four different causes of error. The first possible error, False Invalid Interpretation, FII, is a result of incorrectly reading the test as invalid when it is not, due to a control line's presence. The second type of error, Inverse Interpretation Error, IIE, involves misreading of the drug LFI testing device or wrongfully interpreting the presence of a test line as a positive or the lack of a test line as a negative. This error is possible, especially in an environment where many people are familiar with LFI test kits for pregnancy and COVID, that rely on the opposite format from THC testing for interpretation. The third possible type of error, Visualization Variation Error (VVE), occurs when there is variation among the observers interpreting the data in identifying the presence or absence of a test line. The fourth potential error, Unknown Sample/Test error (USTE), denotes an unknown cause of the false result, which could be attributed to a matrix effect or a problem with an individual test kit.

Each of these errors may have occurred in this study as detailed in Table 5 (see Appendix). There was one instance where a positive result was interpreted as an invalid, FII: EZH (#20). There were two samples where the on-site report was in error, seemingly due to the use of the inverse of the correct interpretation, IIE: #3 EZH and #10 WF. There were four unknown sample or test kit errors where all on-site and photo interpreters yielded false results, USTE: EZLV #18 and #22; WF #5; and IVB #3. The basis of these errors is not known but could be attributed to a sporadic unreliable test kit or a matrix interferent in the sample that caused determinant error. The most common error identified in this study was variation in the interpretation of the test when inverse interpretation error was not thought to occur. This error was identified in multiple samples with three of the test kits where there was disagreement amongst two or more of the interpretations, VVE: EZLV #17; WF #8, #12, #13, #18, and #22; and IVB #18 and #22.

The sensitivity and specificity of each test kit were calculated for the on-site report and all photo interpreters and were compiled in Table 6 (see Appendix). One of 11 samples was found positive above 100 ng/mL of 11-nor-9-carboxy-THC. The sensitivity results for

the test kits ranged from 45.5% to 100%, and the AL kit was the only product where both the sensitivity and specificity were 100% for all interpreters. A 50% sensitivity equates to a coin flip, whereas 100% sensitivity is excellent. The WF test kit produced the most disputed and ambiguous results, with an on-site report correctly identifying less than 50% of true positives. False positives were not common in this study, as evidenced by a specificity range of 80% to 100%.

The lower sensitivity and specificity and the various types of errors observed raise concern about the overall accuracy of the testing. Because most of the errors were false negatives, the biggest challenge in relying on this test would be missing true positive samples. Parents, drug testing programs, and probationary programs using LFI assays to identify drug usage would miss up to 55% of positives depending on the user and which test kit was used. However, few false positives and high specificities were observed for all the test kits. Thus, a positive result could be trusted at least 80% of the time, signifying that a positive result with any of these test kits most likely indicated recent cannabinoid use.

The results affirm the importance of reading the package insert and instructions. Even though the testing was done by individuals with significant college laboratory experience, there were still interpretation errors. The most common error type that led to lower sensitivity was visualization variability error. Being able to ascertain if a test line is observable seemed to be influenced by the packaging of the test kit and correlated with the cost of the test. The AL test kit was the most expensive device. Two products, WF and IVB test kits, which were not encased in plastic and appeared to be paper strips, were the lowest performers when it came to accuracy. The variable results of the photo interpreters indicate that it is often not clear if there is a test line or not. All the package inserts indicated that even a faint line is a negative result, meaning even a shadow would be negative. Impregnated chemicals along the test line seemed to be particularly susceptible to shadows when the kit was made of thin paper without substantial backing. The variability among photo interpreters also suggests that the implementation of a standardized method for interpretation could improve accuracy since the interpreter's determination holds a lot of power in the outcome for the tested individual. Previous studies have evaluated a computer-automated method for reading LFI drug testing strips, involving an automated analysis of LFIs using machine learning algorithms, which has been found to improve the reliability of results (Kim et al. 2017). Algorithmic interpretation of each apparatus would lessen the uncertainty of results and standardize the analysis. If this approach is not available or too expensive, a second or third interpreter should look at the result and discrepant interpretations discarded or retested.

Conclusions

The use of LFI in a variety of different contexts provides helpful information to many users. Since there is often a medical, legal, or quasi-legal reason behind such a screening, it is important that testing rises to a standard of accuracy to avoid decisions made on false results. The results of this study indicate that on-site reports of false negative THC results are not uncommon, as these were found in four of the five different test kits evaluated. False positives were much less common, found for two of the test devices. The fact that specificity is better than sensitivity for each tested device is fortuitous because, in most circumstances, false positive results would have broader-reaching legal implications. False results raise concern for the reliance on these tests and suggest that more costly tests, built upon more sturdy materials, are more accurate; thus, the old caveat, "let the buyer beware" applies as consumers utilize CLIA-waived devices to monitor the presence of THC-COOH.

References

Algren AD, Christian MR. 2015. Buyer beware: pitfalls in toxicology laboratory testing. Missouri Med 12(3):206–210. https://pubmed.ncbi.nlm.nih.gov/26168592/.

Arshadi M, Fardsanei F, Deihim B, Farshadzadeh Z, Nikkhahi F, Khalili F, Sotgiu G, Shahidi Bonjar AH, Centis R, Migliori GB, et al. 2022. Diagnostic accuracy of rapid antigen tests for COVID-19 detection: a systematic review with meta-analysis. Front Med 9:870738. doi:10.3389/fmed.2022.870738.

Carpenter CS. 2007. Workplace drug testing and worker drug use. Health Serv Res. 42(2):795–810. doi:10.1111/j.1475-6773.2006.00632.x.

[CLIA] Clinical laboratory improvement amendments, currently waived analytes. 2019. US Food and Drug Administration. https://www.accessdata.fda.gov/scripts/cdrh/ cfdocs/cfClia/analyteswaived.cfm.

[CLIA] Clinical laboratory improvement amendments (CLIA). 2020 Mar 5. Centers for Disease Control and Prevention. https://www.cdc.gov/clia/index.html.

Del Carmen RV, Sorensen JR. 1989. Legal issues in drug testing probation and parole clients and employees. Office of Justice Programs; [accessed 2024 Feb 21]. https://www. ojp.gov/ncjrs/virtual-library/abstracts/legal-issues-drug-testing-probation-and-parole-clients-and.

Hartwell, TD, Steele, PD, French, MT, Rodman, N. 1996 Nov. Prevalence of drug testing in the workplace. Monthly Labor Review: U.S. Bureau of Labor Statistics; [accessed 2024 Feb 21]. https://www.bls.gov/opub/mlr/1996/article/prevalence-of-drug-testing-in-the-workplace.htm.

Kapur BM. 2012. False positive drugs of abuse immunoassays. Clin Biochem. 45(9):603–604. doi:10.1016/j.clinbiochem.2012.05.004.

Kim SY, Kim H, Park Y, Lim J, Kim J, Koo SH, Kwon GC. 2017. Evaluation of an automated reader and color interpretation-based immunoassays for multiplexed drug-of-abuse testing in urine. J Anal Toxicol. 41(5): 412–420. doi:10.1093/jat/ bkx014.

Koczula KM, Gallotta A. 2016. Lateral flow assays. Essays Biochem. 60(1):111–120. doi:10.1042/ebc2015001.

Oh S, Hodges J, Salas-Wright C, Smith B, Goings TC. 2023. Ethnoracial differences in workplace drug testing and policies on positive drug tests in the United States. Drug Alcohol Depen. 247:109898. doi:10.1016/j.drugalcdep.2023.109898.

Stellpflug SJ, Cole JB, Greller HA. 2020. Urine drug screens in the emergency department: the best test may be no test at all. J Emerg Nurs. [accessed 2023 Feb 22]; 46(6):923–931. doi:10.1016/j.jen.2020.06.003.

Stuck MM. 1996 Jul 1. Comparison of positive screening and confirmatory results from federally mandated drug testing of urine. Old Dominion University Community & Environmental Health Theses & Dissertations; [accessed 2024 Feb 21]. doi:/10.25777/ w6h1-0j84.

[UTDM] Urine testing for detection of marijuana: an advisory. 1983 Sep 16. Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/preview/mmwrhtml/ 00000138.htm.

Appendix

Table 2 Representative True Positive and Negative results.

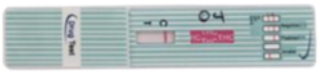
Test ID	Sample ID	On-Site Report	Photographic Result
AL	4	positive	
WF	4	positive	
EZLV	4	positive	
IVB	4	positive	
EZH	4	positive	
AL	7	negative	
WF	7	negative	
EZLV	7	negative	
IVB	7	negative	
EZH	7	negative	

Table 3 Representative False Positive and False Negative THC results.

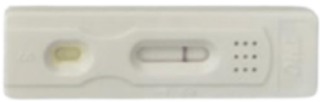
Test ID (cost)	Sample ID (True Result)	On-Site Report	Photographic Result
AL	18 (positive)	positive	
WF	18 (positive)	negative	
EZLV	18 (positive)	negative	
IVB	18 (positive)	negative	
EZH	18 (positive)	positive	
AL	3 (negative)	negative	
WF	3 (negative)	negative	
EZLV	3 (negative)	negative	
IVB	3 (negative)	positive	
EZH	3 (negative)	positive	

Table 4 Twelve of 22 samples had invalid, false positive or false negative THC results. AL had no false results so is not included in the table. PI= Photo Interpreter; OR= On-Site Report; Samples with observed disagreement from true are denoted in red. Results where the control line did not appear are denoted as Ø.

Sample ID	True Value	IVB					WF					EZH					EZLV				
		PI 1	PI 2	PI 3	PI 4	OR	PI 1	PI 2	PI 3	PI 4	OR	PI 1	PI 2	PI 3	PI 4	OR	PI 1	PI 2	PI 3	PI 4	OR
3	–	+	+	+	+	+	–	–	–	–	–	+	+	+	+	–	–	–	–	–	–
5	+	+	+	+	+	+	–	–	–	–	–	+	+	+	+	+	+	+	+	+	+
8	+	+	+	+	+	+	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+
10	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+
12	+	+	+	+	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+	+
13	+	+	+	+	+	+	+	+	–	+	–	+	+	+	+	+	+	+	+	+	+
17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
18	+	+	+	+	–	–	–	–	+	+	–	+	+	+	+	–	–	–	–	–	–
22	+	+	+	+	–	+	–	–	+	+	–	+	+	+	+	–	–	–	–	–	–
2	+	+	+	+	+	+	Ø	Ø	Ø	Ø	Ø	+	+	+	+	+	+	+	+	+	+
14	–	–	–	–	–	–	Ø	Ø	Ø	Ø	Ø	–	–	–	–	–	–	–	–	–	–
20	+	+	+	+	+	+	+	+	+	+	+	+	Ø	Ø	+	+	+	+	+	+	+

Table 5 Possible errors in the testing of THC with LFI kits and samples where these errors were observed.

#	Error Name (Abbreviation)	Result revealing this type of error Test Kit: Sample #
1	False Invalid Interpretation (FII)	EH: 20
2	Inverse Interpretation Error (IIE)	EL: 3 WF: 10
3	Visualization Variation Error (VVE)	EL: 17 WF: 8,12,13,18,22 IVB: 18, 22
4	Unknown Sample/Test Error (USTE)	EL: 18, 22 WF: 5 IVB: 3

Table 6 Sensitivity and specificity of 5 LFI devices from laboratory observation and photographic interpretation.

Test Kit Name (Cost per test)	On-Site Report	PI #1	PI # 2	PI # 3	PI #4
AL- ($1.00) Sensitivity Specificity	100% 100%	100% 100%	100% 100%	100% 100%	100% 100%
EZLV ($0.78) Sensitivity Specificity	83.3% 100%	83.3% 100%	83.3% 100%	83.3% 100%	75.0% 100%
IVB ($0.38) Sensitivity Specificity	92.3% 100%	92.3% 100%	92.3% 100%	92.3% 100%	83.3% 90.0%
EZH ($0.58) Sensitivity Specificity	100% 80.0%	100% 80.0%	100% 88.9%	100% 88.9%	80.0% 90.0%
WF- ($0.40) Sensitivity Specificity	45.5% 100%	63.6% 100%	63.6% 100%	72.7% 100%	90.9% 100%