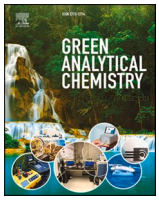
Green Analytical Chemistry 11 (2024) 100158

Contents lists available at ScienceDirect

Green Analytical Chemistry

journal homepage: www.elsevier.com/locate/greeac



White analytical chemistry evaluation of medicines quality screening devices in low- and middle-income countries field settings

Pierre-Yves Sacre [a,1,*], Christelle A. Waffo Tchounga [b,c,1], Charlotte De Bleye [b], Philippe Hubert [b], Roland D. Marini [b], Eric Ziemons [b]

[a] University of Liege (ULiege), CIRM, Research Support Unit in Chemometrics, Department of Pharmacy, 15 Avenue Hippocrate, B36, B-4000 Liège, Belgium
[b] University of Liege (ULiege), CIRM, Vibra-Santé Hub, Laboratory of Pharmaceutical Analytical Chemistry, Department of Pharmacy, 15 Avenue Hippocrate, B36, B-4000 Liège, Belgium
[c] University of Yaoundé I, Faculty of Medicine and Biomedical Sciences, BP 1364 Yaoundé, Cameroon

ARTICLE INFO

Keywords:
Substandard and falsified medicines
Un sustainable development goals
White analytical chemistry
Screening device
Green analytical chemistry

ABSTRACT

Addressing the quality of circulating medicines in low- and middle-Income countries is challenging due to limited access to affordable and rapid laboratory analyses. Moreover, traditional analytical methods generate hazardous waste, posing further environmental and health risks, conflicting with several United Nations Sustainable Development Goals, such as reducing pollution and conserving ecosystems.

In response to these environmental challenges, the concept of green analytical chemistry emerged promoting environmentally friendly practices through principles such as reducing chemical use, minimizing energy consumption, and managing waste. Various green metric tools have been developed to assess the environmental impact of analytical methods, but these often neglect the reliability and applicability of the methods. To address this gap, approaches like White Analytical Chemistry (WAC) have been proposed, integrating criteria for environmental safety, reliability, and cost-effectiveness. WAC's principles are categorized into red (fitness for purpose), green (environmental safety), and blue (cost-efficiency) criteria.

This study aims to use the WAC approach to provide a holistic comparison of different medicines' quality screening devices, helping developing countries with high SF medicine prevalence make informed choices. Three different situations are investigated: *ex-ante* evaluation, *ex-post* evaluation of the qualitative performances and *ex-post* evaluation of the quantitative performances of the most used medicines' quality screening devices in low- and middle-income countries field settings.

1. Introduction

In 2015, the United Nations published the 2030 Agenda for Sustainable Development. This agenda is "a plan of action for people, planet and prosperity" [1]. It contains 17 Sustainable Development Goals (SDGs) whose main objectives are to improve the world welfare while facing urgent global challenges at social, environmental and economical levels. Among these goals is the "universal health coverage including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all" (Target 3.8).

According to the World Health Organization, it is estimated that 10 % of medical products in low- and middle-income countries (LMICs) is either substandard (a medicine that "fails to meet either its quality standards or specifications, or both") or falsified (a medicine that "deliberately/fraudulently misrepresent its identity, composition or source") [2]. Although this average estimate masks the disproportionate impact of SFMs on rural areas and populations with low socio-economic status, it highlights the scale of the scourge represented by SFMs that in turn threatens the achievement of the above-mentioned SDG target.

Abbreviations: API, active pharmaceutical ingredient; CAC, circular analytical chemistry; CS, colour score; GAC, green analytical chemistry; GC, green chemistry; HPLC, high performance liquid chromatography; LMIC, low- and middle-income country; NIR, near infrared; SFM, substandard and falsified medicine; WAC, white analytical chemistry.

* Corresponding author.
E-mail address: pysacre@uliege.be (P.-Y. Sacre).
[1] These authors contributed equally to the article.

https://doi.org/10.1016/j.greeac.2024.100158
Received 23 August 2024; Received in revised form 24 September 2024; Accepted 26 September 2024
Available online 27 September 2024

Although the presence of SFMs is worldwide, a recent review states that "LMICs bear most of the burden, likely a result of poor surveillance mechanisms, governance, regulations and management of pharmaceutical chains." It also states that "corruption, weak governmental policies and limited technical capacity can also be responsible for enabling the distribution of SFMs in LMICs and decrease the likelihood of detecting SFMs" [3]. These aspects are concerned by the 16th UN SDG and highlight the role of local institutions in the control of illegal trades and traffics. Regarding the quality of the circulating medicines, the national institutions in LMIC are facing several difficulties among which the access to fast and relatively cheap laboratory analysis capacities to assess the quality of the medicines on their territory. Indeed, among the 60 WHO Prequalified Medicines Quality Control Laboratories, only 11 are present on the African continent [4]. Besides the scarcity of the technical capacities, most analytical methods [5] are based on "wet chemistry" principles implying the use of reagents and solvents generating hazardous wastes for the users and/or the environment. The management of these wastes is also challenging in the LMICs context [6] and have an impact on several UN SGDs' targets: T3.9: "reduce of illnesses and death from hazardous chemicals and pollution", T6.1: "provide safe and affordable drinking water" and T15.1: "conserve and restore terrestrial and freshwater ecosystems". These findings point to the need for innovative technologies to support the rapid, environmentally friendly and cost-effective detection of SFMs in LMICs field settings.

These environmental challenges have raised awareness among researchers and chemists. In the beginning of the 1990s, the concept of green chemistry (GC) has emerged as their contribution to make laboratory practices more environmentally friendly. Then, in 1999, GC gave birth to green analytical chemistry (GAC) which is founded on 12 principles and can be achieved in four ways: (1) eliminate or reduce the use of chemical substances (using for example green solvents and solvent-free methods), (2) minimize the use of energy, (3) manage analytical waste properly and, (4) increase the safety of operators [7–9]. Considering these principles, several green metric tools have been proposed among which National Environmental Methods Index (NEMI), Analytical GREEness metric approach (AGREE), Analytical GREEnness metric for sample preparation (AGREEprep), and Complementary Green Analytical Procedure Index (ComplexGAPI) [10–12]. Unfortunately, these tools are not adapted to real life since they focus only on environmental aspects and do not consider the reliability of the analytical method, nor its applicability. As a result, other approaches have been proposed to assess analytical methods. A notable approach is the White Analytical Chemistry (WAC), also founded on 12 principles divided in three categories based on the 3 primary additive colours: red, green and blue. This approach developed by Nowak et al. was first called "RGB-12″ and considers the different facets of WAC, with red assessing the analytical performances, green assessing the safety and eco friendliness, and blue, the cost-efficiency and practical aspects of an analytical method. [13–16]. Another assessment tool was developed by Manousi et al., stemming from the 'blue' principles of WAC: Blue Applicability Grade Index (BAGI) [17].

Typically, these metric tools are only applied to the assessment of different versions of the same method after their development. However, it would be useful to evaluate these criteria beforehand to make an informed choice of the strategy or technology to be used, depending on the final objective envisaged for the method. Recently, Jiménez-Carvelo et al. have proposed, beside the *ex-post* evaluation, an *ex-ante* one, based on a priori knowledge of the method which takes place before its development. According to them, this approach holds the advantage of highlighting potential risks that could be generated before the method is developed, and therefore enables better management of time and resources [18].

A growing number of screening technologies for the detection of SF medicines are expanding and are currently available [19–21]. However, there are few papers offering a holistic comparison of these different tools, with an overall score enabling a rapid and visual comparison. The present study aims to assess the sustainability of medicines screening tools, on the one hand, by implementing the WAC approach for their selection, and on the other hand, by analyzing the evolution of scores of the same screening tool based on its application. Such approach could help decision-makers, especially those from developing countries presenting high prevalence of SFMs, to make an informed choice of screening devices, in line with the UN SDGs Target 17.7: "promote sustainable technologies to developing countries".

2. Material and methods

2.1. Medicines quality screening device's purpose

A medicines quality screening device, as suggested by its name is intended to perform an early detection of quality defects in medicines. Since the "quality of a medicine" is a vague concept and implies several aspects, one may rely on the different quality items proposed by the pharmacopoeias. As an example, the "oral drug products—product quality tests" monograph of the USP [22] states that these tests are divided in two categories:

- product quality test: identification, strength (assay), impurities, dose content uniformity, pH, minimum fill, alcohol content, volatile content, and microbial content
- product performance tests: designed to assess in vitro drug release from dosage forms

Obviously, not all these items may and should be tested during a "screening" investigation. One may agree that the most important ones to ensure before testing any of the other are the identity and strength of the active ingredient(s) (APIs). Another choice to make is the kind of dosage form to be tested. Indeed, drugs may be administered in several routes (oral, topical, nasal, mucosal, inhaled, injected, etc.). Each of these have specific product quality attributes and their testing may require specific instrumentations that will not be covered by this paper. As oral drug (tablets and capsules) is the most sold and consumed dosage form of medicines, we will mainly focus on the ability of screening devices to test for qualitative (identity) and quantitative (strength) aspects of oral drug products.

The screening devices are intended to be used as close to the field as possible to enable the fastest possible response to the detection of poor-quality products, and to decrease the workload of confirmatory laboratories. In addition, as the word "screening" suggests, the poor-quality character of the products must be confirmed in laboratory with "confirmatory" tests. It is therefore important not to destroy the sample. Finaly, in LMICs, the cost-efficiency is an important matter.

Considering all these aspects, one may derive ideal properties. A screening device should be able to provide both qualitative and quantitative information on the sample in a non-destructive manner. It should also be able to be manipulated by a minimally trained user directly in the field considering the available resources at a minimal cost. Finally, the ideal device should return immediately an answer and should be as versatile as possible (analysis of multiple products, dosage forms, brands, active ingredients etc.).

Unfortunately, these ideal properties are not all met by the screening devices currently available. It is therefore useful to compare the available devices considering the needs and constraints of the final user. Some guidelines exist to conduct evaluation of screening devices such as the ⟨1850⟩ USP monograph [23]. Although comprehensive, this evaluation may be long and tedious to implement and lacks a fast, visual and easy comparison of the results. This may prevent analysts to use it. We believe that the approach followed in this work may help compelling and gathering the results in an easy way to compare the different devices.

2.2. Included devices

Six screening devices have been included in the present study (see Table 1).

These devices have been selected because of their widespread use as devices or technologies for medicines quality screening in LMIC. Though it can be used for screening purpose, HPLC is not a screening device *per se* but rather a confirmatory technology and serves here as a benchmark.

As one may notice, the screening devices compared in this study include very different technologies and scientific principles. However, as the purpose is the same: detecting poor quality medicines in LMIC field settings, it appeared legitimate to compare them. Indeed, if one must choose a screening device, the described approach may be used to perform an informed choice while having a holistic view of the device's performances and characteristics.

2.3. White analytical chemistry (WAC) approach

The WAC is a relatively new concept whose main idea is the additive combination of the three primary colours: Red, Green and Blue. Each primary colour represents an aspect of the analytical method:

- Red refers to the analytical performances of the method and is generally assessed by validation criteria (accuracy, precision, etc.);
- Green stands for the Green Analytical Chemistry (GAC) principles [9] and refers to the environmental impact of the method;
- Blue represents the practical and economic efficiency of the method.

An ideal method that fulfils completely each item is therefore considered as White ($W = R + G + B$). This "whiteness" score can be viewed as a global score of the method under investigation considering its intended purpose. As is the case with most studies applying GAC and WAC concepts, the analysis and scoring is performed once the method is developed since each item requires precise values obtained during validation (Redness) and a fixed experimental protocol to compute the greenness of the method. However, the WAC concept may also be useful when selecting methods/technologies before their implementation or development.

Table 1
Main characteristics of the devices included in the study.

Device name	Manufacturer	Working principle	
NIR-S-G1	Innospectra Corp	• Near infrared dispersive spectrophotometer • Diffuse reflection mode • Spectral range: 900- 1700 nm	[23]
NIR-S-T2	Innospectra Corp	• Near infrared dispersive spectrophotometer • Transmission mode • Spectral range: 900- 1700 nm	[24]
Truscan RM	ThermoFisher Scientific	• Backscattering Raman spectrophotometer • Excitation laser wavelength: 785 nm	[25]
Paper analytical devices (PADs)	University of Notre Dame	• Paper-based colour test	[26]
Minilab	Global Pharma Health Fund E.V.	• Visual inspection • Weight verification • Disintegration test • Thin layer chromatography	[27]
High Pressure Liquid Chromatography	Multiple manufacturers	• Separation of sample constituents based on their relative affinity for the stationary and the mobile phase • Several detection options (e.g. UV–Vis spectroscopy, mass spectrometry etc.)	/

Recently, Jiménez-Carvelo proposed a two-step evaluation [18]. First, an *ex-ante* evaluation is performed based on data available from the literature or from preliminary tests. Then, a second phase results in the *ex-post* evaluation of the method with the exact values coming from the method's validation/development.

Both Nowak et al. and Jiménez-Carvelo et al. insist on the fact that the scoring of a method depends on its intended purpose. This implies that a same technology may have different scores depending on its specific application. This also requires that the evaluated items are designed for the intended purpose. Indeed, possibly less important for the green and red items, the intended use of the method will directly drive the choice of the blueness items to be evaluated.

This paper proposes an Excel sheet as template for the *ex-ante* and *ex-post* evaluation of a medicine's quality screening device. It follows the general and easy-to-use approach of Jiménez-Carvelo et al.: for the assessment of greenness and blueness, several items are proposed and, within each of these, several sub-items have to be evaluated [18]. A score is attributed to each sub-items summing up to 100 % for each item. Contrarily to the original paper that was based on a simple "yes" or "no" response for each sub-items, we adopted a score between 0 and 1 enabling a more flexible answer. This will be exemplified in the results section.

The following sections will detail how each Red, Green, Blue and White scores are computed.

2.3.1. Redness

As stated above, the Redness criterion represents the analytical performances of the method. Several quantitative criteria may be used for the redness rating depending on the qualitative (classification) or quantitative purpose of the method.

For the *ex-ante* evaluation of Redness, Jiménez-Carvelo et al. propose to use the same "default" values for each device. Although this might be a convenient approach when no data is available, we had the opportunity to have a collection of papers that evaluated the different devices on a similar basis at our disposal (see [20,24–27]).

In this paper, for the *ex-ante* evaluation, we based our rating on the sensitivity and specificity of each device as reported by Zambrzycki et al. [25]. Sensitivity is defined as the percentage of true positives over the sum of true positives and false negatives, and specificity as the percentage of true negatives over the sum of true negatives and false positives. A true positive was defined as the sample being good quality with the device correctly giving a pass result. A true negative was defined as the sample being poor-quality with the device correctly giving a fail result. The poor-quality medicines of the study were both samples containing no or the wrong API and samples containing only 50 % or 80 % of the declared amount of API covering both substandard and falsified cases. These values were available for NIR-S-G1, Truscan RM, PADs and the Minilab.

The sensitivity and specificity values of HPLC were set to 100 % by default since it is a "gold standard" method. The values for the NIR-S-T2 were obtained from the study of Ciza et al. [28] for the analysis of quinine samples in NIR transmission spectroscopy. This study applied a qualitative analysis of the sample's solution to separate the quinine samples from related cinchona bark alkaloids (quinidine and cinchonine) and from placebos.

For the *ex-post* evaluation of qualitative methods, since the objective is a binary classification as "good-quality" or "poor-quality", rather than using the sensitivity, specificity, accuracy and precision, we propose to use the Matthews Correlation Coefficient as a single item for Redness [29]. If only the target class is available, which is often the case when a new medicines identification method is developed, only the sensitivity might be tested [30].

Regarding the *ex-post* evaluation of quantitative methods, quality metrics such as R^2 of linearity, standard error of validation, bias, accuracy, precision, etc. are generally used. In our case, the methods

evaluated in the source papers were validated following the "combined accuracy and precision" approach of the USP ⟨1210⟩ [31] and ICH Q2R2 [32]. Rozet et al. proposed an integrated score of the validation criteria to show the overall quality of the method over the concentration range studied called the "accuracy index" [33]. We used this item to evaluate the Redness of quantitative methods.

2.3.2. Greenness

The "green" character of an analytical method is a vague concept that may be interpreted in different ways. A perspective paper by Nowak [15] states than "*Greenness is a measure of the destructive impact that humans have on the environment and themselves.*" From this definition, several interpretations may be drawn, and a lot of different implementations are possible. However, as stated by the author, no method can be qualified as "green" since it implies not to cause any destructive impact on the environment and humans, which is practically unfeasible.

Regarding this definition of greenness, a new, holistic concept may also be of interest: the Circular Analytical Chemistry (CAC) [34]. However, while honourable, the evaluation of the real and complete impact of a method, its components or a procedure is difficult to implement in practice. Indeed, considering portable spectroscopy, the method may generally be considered as "green" since it enables non-destructive analysis of samples with no or limited use of reagents etc. But how circular is it? How were its elements produced, how many rare earth elements does it contain, how recyclable is it, etc. These questions are complex and difficult to answer. In our opinion, this is probably not the responsibility of the analytical chemist to spend time and efforts to evaluate these elements. Hopefully, these elements will form part of the equipment specification in the future.

In the frame of this paper, the main idea was to have an evaluation and comparison of different screening devices to enable an informed choice based on the end-user's particular points of attention. Therefore, it is our responsibility to define clearly what we mean by greenness: the different items evaluated are the use of chemicals, the use of resources, the safety of the operator and the generation of analytical wastes. These different items are the same as the ones proposed in [18] and cover several aspects of the greenness definition.

2.3.3. Blueness

As described in Section 2.1, the practical and economical aspects of a screening device are critical. Therefore, the Blueness item is the most developed criterion with a total of 8 items.

2.3.3.1. Sample throughput. This item represents the time needed to analyse a sample. It is divided into 4 sub-items: <1 h, <30 min, <15 min and <5 min per sample. Each sub-item has the same score of 22.5 with a baseline of 10. For example, if a technique takes <5 min, it has a score of 100 (10+22.5+22.5+22.5+22.5) but if it takes between 30 min and 1 h, its score is of 32.5 (10+22.5). The reported analysis times per sample for the different devices come from the following publication [24] except for HPLC and NIR-S-T2 whose analysis times are based on the author's experience.

2.3.3.2. Cost effectiveness. This item reports how much costs the analysis of a sample. It is based on the average annual cost considering an analysis of 1000 samples over 5 years that is considered as the average depreciation period for each device. Its detailed computation is described in the supplementary Table S1.

2.3.3.3. Sample destruction. As stated above, the destruction of a sample prevents its subsequent analysis by a confirmatory technique. However, this is crucial in the frame of legal investigations to justify the recall or batch destruction. The non-destruction of a sample is understood as the fact that the sample remains in the same physical and chemical state after measurement. Some recent devices also enable what is called "non-invasive" analysis. In other words, the integrity of the sample remains intact since the analysis is performed without opening of the packaging and the medicine may even be consumed after analysis. This is clearly an advantage when considering that high demand for a drug increases the risk of falsification (e.g. chloroquine during the COVID-19 pandemic) and puts pressure on available stocks.

2.3.3.4. Versatility. This item first explores the possibility of carrying out qualitative and quantitative analysis. Some devices enable "semi-quantitative" analyses and will therefore be attributed a 0.5 for the "quantitative analysis" sub-item. The practical difficulty to analyse certain dosage forms is expressed as a "dosage forms score". It is inspired from Caillet et al. [26] who compares, for several devices, the degree of difficulty to analyse different medicines formulations relative to the analysis of a tablet. The detail is provided in table S2. The rationale for this choice is that most devices are designed for the analysis of solid samples among which tablets are the most common ones. Compared to these, capsules, liquids, powders, creams or gels may be easier or more difficult to analyse depending on the technology considered. This is an easy and smart way to evaluate the impact and restrictions on dosage forms for each device. Finally, the possibility to analyse several APIs with the same device is assessed.

Obviously, this item is removed from the *ex-post* analyses since the latter refer to a specific application of the device.

2.3.3.5. Formation. As stated in paragraph 2.1, screening devices should ideally be used by a "minimally trained" user. However, this minimal training depends on the technology and the initial formation of the user. Therefore, we considered the time needed to have a basic formation to the device enabling its routine use for a non-technical user. Following the classification used in [35], a user is considered as having "no-technical experience" if he has no prior laboratory experience and no background in one of the physical sciences (e.g. chemistry, physics, …). This item comprises 3 levels: less than half a day, less than one day and more than one day getting 50, 30 and 20 points each respectively.

2.3.3.6. Portability. As most screening devices are intended to be used in the field or very close to it, it is supposed to be transported. However, the ease of transport may be a key aspect facilitating or preventing its use in some situations. To assess the portability of the device, we propose the following classification inspired from Leary et al. [36] classifying devices as:

- Transportable: typically packaged in suitcases and can weigh >20 kg. It must typically be transported by car and is performed in "fixed mobile laboratories".
- Portable: typically, ≤ 50 dm^3 and weigh between 3 and 20 kg.
- Handheld: typically, ≤3 dm^3 and weigh between 0.5–3 kg.
- Miniature: all devices ≤0.75 dm^3 and weighing <0.5 kg.

The HPLC is a special case since it is not portable and considered as benchtop device and we have not investigated portable versions of HPLC. Its score is fixed at the baseline score of 10. The Size, Weight, and Power (SWaP) analysis of each device is provided in Table S3.

2.3.3.7. Usability. The usability or "user-friendliness" of a device is an important feature that will impact its acceptability and finally its use in routine by the final user. Three sub-items are evaluated: automatic interpretation of results, the necessity to analyse a reference sample or substance and the possibility to "pre-calibrate" the device.

- **Automation of results analysis**: To enable an average user to rapidly take actions, the screening device should report an easy-to-understand result such as "Pass" or "Fail" possibly with a confidence score. For example, colorimetric results may be difficult to

interpret or even non readable for colour-blind users. Therefore, the availability of e.g. smartphone applications to transform the colour response in a reportable result may be of great help. However, this possibility is not always immediately available or even not possible in a short term. This availability or possibility to have it in a short term is evaluated in this sub-item which is then combined with two other sub-items to evaluate the usability.

The necessity to have a reference sample/substance at disposal to analyse the test samples is less practical. Indeed, these references must be transported alongside the device and respecting certain storage precautions (light protection, temperature control etc.). The non-respect of these precautions might hinder or lower the quality of the obtained results. Another point considers the possibility to "pre-calibrate" the devices. This point is especially important for spectroscopic techniques and is discussed in the next paragraph.

2.3.3.8. Calibration/maintenance. Some devices (notably spectrophotometers), must be "pre-calibrated" before going in the field. This calibration clearly constitutes an obstacle to their widespread use. Indeed, this calibration phase generally requires the acquisition of spectra of reference samples representing the natural variability of the product and the different conditions in which the analyses will be performed. In addition, the spectral data must be analysed by advanced chemometric tools to transform them in un understandable information. This data analysis phase requires highly skilled staff that is not always available. However, once calibrated, the devices may be easily used by minimally trained users which increases the device's usability (see previous paragraph).

A second aspect of this item is the "maintenance". This term may have several meanings. For example, in the case of reagents-based devices (e.g. Minilab), a maintenance of the reagents/standards stock is mandatory to enable its continuous use. In the case of spectroscopy-based devices, the maintenance is related to the chemometric models' performances that must be monitored and checked over time. Indeed, several changes in samples' composition, device ageing, etc. may impact the model's performance and require a model's maintenance [37].

One must be aware of these challenges before selecting this kind of technology since it will impact its performances and possibly its future routine costs that are non-negligeable.

2.3.4. Items' weights

As previously mentioned, the evaluation of the techniques/methods should be made considering their final purpose. Keeping this in mind it appears that, depending on the final-user's priorities, different weights could be given to the different colours, items and sub-items. Therefore, the final White colour score is the weighted geometric mean of the individual colour scores (CS) and is computed following the notation of Nowak et al. [13]:

$$CS_{white} = \sqrt[(W_{red}+W_{green}+W_{blue})]{CS_{red}^{W_{red}} \times CS_{green}^{W_{green}} \times CS_{blue}^{W_{blue}}} \quad (1)$$

Where W_{red}, W_{green} and W_{blue} are the weights assigned to each CS.

Each individual CS being itself the weighted geometric mean of its elements:

$$CS_{colour} = \sqrt[(w_1+w_2+\ldots+w_n)]{item_1^{w_1} \times item_2^{w_2} \times \ldots \times item_n^{w_n}} \quad (2)$$

Where n is the number of items for the given colour (green or blue). For the *ex-post* analysis, Redness having only one item, therefore no average is computed.

The geometric mean is justified by the fact that it tends to decrease if one element strongly diverges from the other and to tend to zero if one of its elements does so. However, a characteristic of the geometric mean is that none of its element may be exactly zero. Therefore, to avoid this, a baseline score of 10 is given to each item. More details are available in the Excel template and in the previous Sections 2.3.1. to 2.3.3.

For the *ex-ante* analysis, we applied weights of $W_{red} = 1$ and $W_{green} = W_{blue} = 2$. This choice was made to emphasize most on the accurate values of Greenness and Blueness and to lower the importance of Redness that is based on assumptions or on a literature review. These weights are changed to $W_{red} = W_{green} = W_{blue} = 1$ for the *ex-post* analyses.

For the Greenness, we applied weights of 1 for each item.

For the Blueness, we based our weights' choice on the article by Roth et al. [38] who questioned different stakeholders from geographically and economically different countries on the ideal qualities of a medicines' quality screening device. From their findings, we prioritized and gave more weight to items the most frequently cited by LMIC stakeholders. This led to the following weights: Cost effectiveness and portability: $w = 10$; sample throughput, sample destruction, usability, calibration/maintenance: $w = 5$, formation: $w = 1$.

2.4. Medicines' quality screening device comparison

2.4.1. Ex-ante analysis

The *ex-ante* comparison of the different selected devices was made according to the data found in the literature and from the experience of the authors. Several scientific articles, reports and the manufacturer's websites were used as source of information. These sources are listed in the corresponding supplementary data.

Contrarily to Nowak et al. [13] and Jiménez-Carvelo et al. [18], no "acceptable" or "satisfactory" levels were set. However, a boxplot analysis and the interquartile range was used to compare the devices and detect the best (Q1, top 25 %) and the poorest (Q4, least 25 %) ones regarding the different CSs (Red, Green, Blue, White).

2.4.2. Ex-post analysis

The results of the *ex-post* analyses are based on the following studies of the authors: [28,39–42]. The main advantage of these studies is the possibility to compare the same device in different applications or several devices with the same samples in field setting with optimized chemometric models.

2.5. Data analysis

All data analyses have been performed in Excel [43] except the boxplot analysis that was performed in Matlab [44] with the PLSToolbox [45].

3. Results

This section will describe and compare the results of the WAC analysis of the different medicines' quality screening devices following the items developed in the previous section. For each section, the different devices will be compared colour by colour. The detailed scores may be found in the supplementary Excel file S1. Fig. 1 and Table 2 summarize the different devices' scores.

3.1. Ex-ante analysis

3.1.1. Redness

As described above, the *ex-ante* redness score is based on the results of Zambrzycki et al. [25] except for NIR-S-T2 whose results are based on Ciza et al. [28] and HPLC whose score is fixed at 100 %. The CS_{red} of all devices is generally good and ranges from 71.3 % for the PADs to 90.6 % for the Minilab. These results are open to debate, since the performance of spectrophotometers depends on the particular application, the optimisation of the chemometric models developed, etc. However, this is a good basis since all the devices were compared on the same samples which included either no API, the wrong API or too few API. We considered this comparison as fair. Nevertheless, for the *ex-ante* CS_{white}

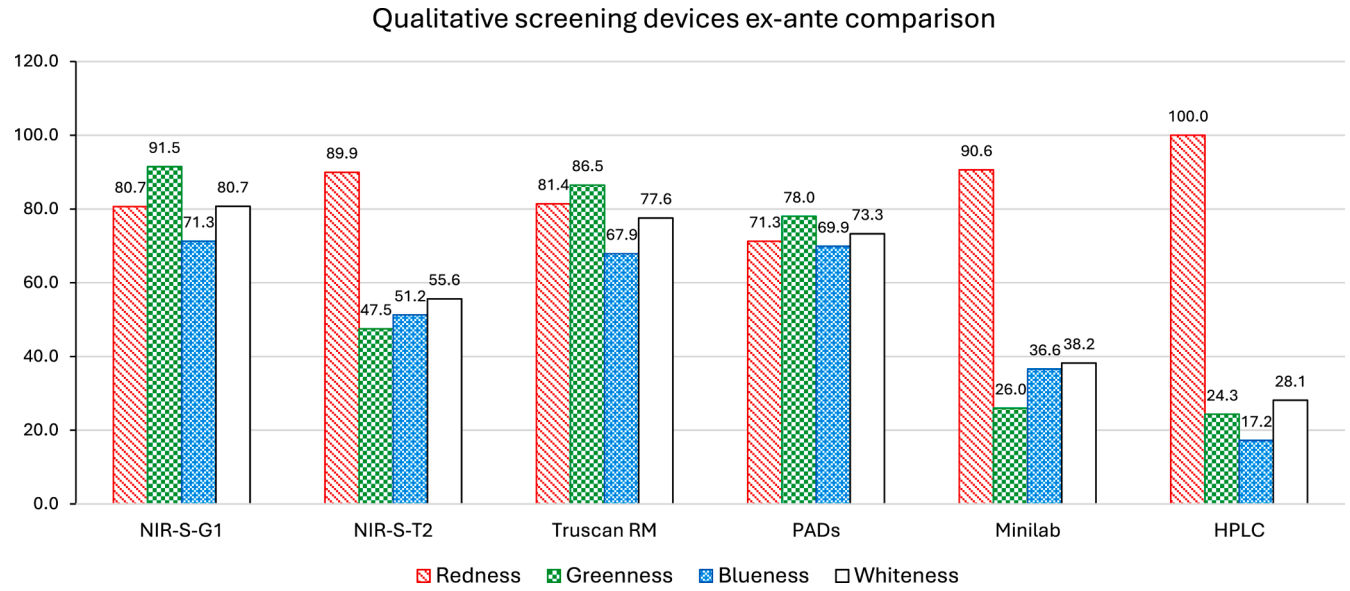


Fig. 1. Medicines' Quality screening device white analytical chemistry ex-ante evaluation for qualitative application (i.e. detection of poor-quality medicines).

Table 2
Scores attributed to the different studied devices during the Ex-ante and the qualitative and quantitative Ex-post analyses.

		Device	Redness	Greenness	Blueness	Whiteness
Ex-ante evaluation		NIR-S-G1	80.7	91.5	71.3	80.7
		NIR-S-T2	89.9	47.5	51.2	55.6
		Truscan RM	81.4	86.5	67.9	77.6
		PADs	71.3	78.0	69.9	73.3
		Minilab	90.6	26.0	36.6	38.2
		HPLC	100.0	24.3	17.2	28.1
Qualitative Ex-post evaluation	**ciprofloxacine**	NIR-S-G1	78.3	91.5	68.2	78.7
		PADs	62.6	78.0	71.6	70.5
	coartem	NIR-S-G1	100.0	91.5	73.3	87.5
		Truscan	42.1	86.5	55.1	58.5
Quanitative Ex-post evaluation		NIR-S-T2 ciprofloxacine	82.6	56.2	48.1	60.7
		NIR-S-T2 quinine	90.4	47.5	48.1	59.1
		Truscan ibuprofene	65.6	86.5	50.6	70.6

evaluation, CS_{red} was given a lower weight compared to the CS_{green} and the CS_{blue}.

Fig. 2 displays the CS_{colour} of the *ex-ante* evaluation of all devices. By analysing boxplots, one can visually compare devices and easily distinguish between those that perform better and those that perform worse. Looking at the CS_{red} boxplot, the two best performing techniques are the HPLC (by convention) and the Minilab which is very close to the NIR-S-T2. The Truscan RM and the NIR-S-G1 have comparable results. The fact that the two technologies (NIR and Raman) are close may be surprising. However, as discussed in other papers [40,46] and will be discussed in the *ex-post* analysis, the analytical performances of the two technologies are very different depending on the application (specific brand or API identification). PADs were the least performing in the considered study.

3.1.2. Greenness

NIR-S-G1, as a NIR reflexion spectrophotometer has the highest CS_{green} of 91.5%. Indeed, there is no hazard, no generation of waste etc. The only sub-item that is not fulfilled is the consumption of electricity when in use although this may be balanced by the use of rechargeable batteries. The Truscan RM is also a reflexion mode spectrophotometer and therefore has almost the same Greenness than the NIR-S-G1. However, as it uses a class 3B laser, there is a physical hazard for the user and its final CS_{green} is of 86.5 %.

The PADs are the third best performing device with a Greenness score above 50 %. Their lower score compared to spectrophotometers is due to the consumption of water (although limited), they generate wastes and are consumables *per se*. They also generate chemical reactions (colorimetric tests) but since the byproducts are not toxic and the amount of reagents is very low, they have a score of 0.75 for the "no chemical reaction" sub-item. Their final CS_{green} is of 78 %.

The last devices, Minilab, NIR-S-T2 and HPLC are based on "wet chemistry" and therefore several items of Greenness are not met. Some sub-items require explanations on their quotation: the Minilab does not necessary need electricity when in use, but some TLC revelations are done under UV light excitation requiring electricity and some reactions need a hot plate which explain the 0.5 score for this sub-item. NIR-S-T2 has also a 0.5 score for the "chemical hazard" and the "hazardous waste" because, depending on the method used, some hazardous reagents may be involved (e.g. hydrochloride acid). The HPLC has a score of zero to almost all sub-items excepting the use of electricity when not in use and the biological hazard. Their final CS_{green} is of 47.5 %, 26 % and 24.3 % for the NIR-S-T2, the Minilab and the HPLC respectively.

When looking at Fig. 2, the CS_{green} interquartile range is quite large which is explained by the presence of two separate groups: a "greener" group: NIR-S-G1, Truscan and PADs and a "least-greener" group:

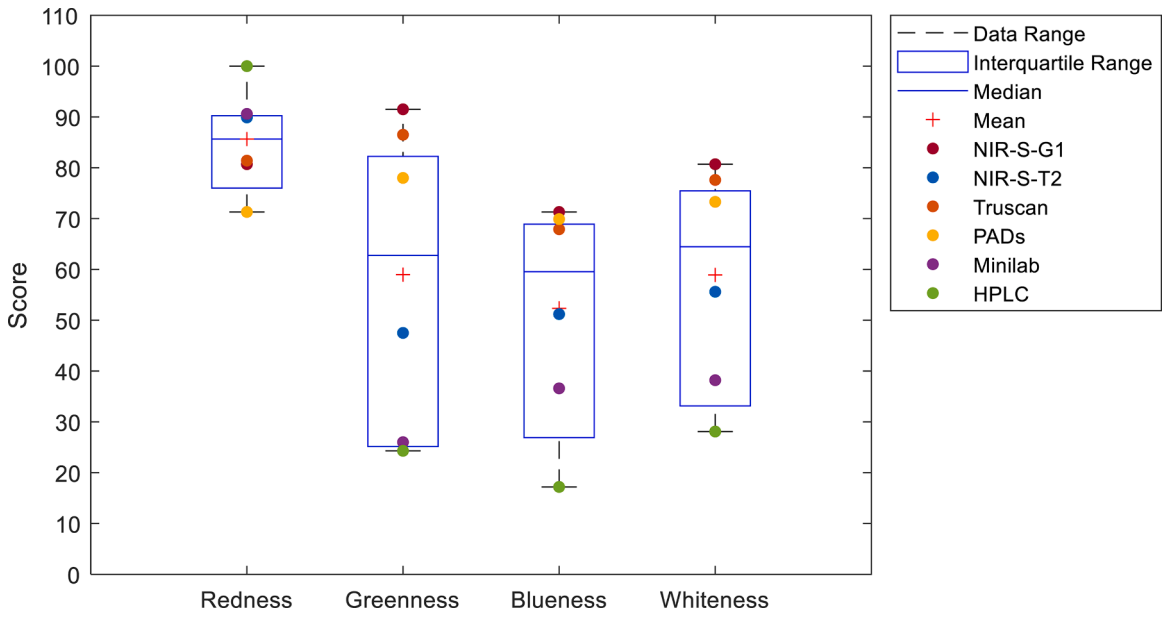


Fig. 2. Boxplot of the ex-ante evaluation for qualitative application.

Minilab and HPLC. NIR-S-T2 has an intermediate score but below the median. Considering the "greener" group, NIR-S-G1 and Truscan RM have a significatively better score and a in the 1st quartile making them the two best performing techniques for CS_{green}.

3.1.3. Blueness

The evaluation of the Blueness criterion will be detailed below. However, before analysing each device, a general constatation is that HPLC performs the worst in almost every item. Indeed, HPLC is a costly destructive technique, not portable, with a low sample throughput, and that needs a highly skilled staff to be run. However, the HPLC is the best device regarding the versatility. These aspects make the HPLC the less "blue" device with a final CS_{blue} score of 17.2 %. Nevertheless, it was included in the *ex-ante* comparison as a benchmark technique.

The CS_{blue} is composed of 8 items having different weights. Fig. 3 summarizes the scores (not yet weighted) of each screening device for each of these items.

The most important items ($w = 10$) are the cost effectiveness and the portability. The least expensive techniques are the NIR-S-G1 and PADs with an annual cost of $287.8 and $612 respectively. The Minilab and NIR-S-T2 have an intermediate annual cost of $2546 and $1120 respectively. Cost is currently one of the major barriers to implementing Raman spectroscopy in LMICs. This is clearly visible here since it has the lowest score (apart from HPLC) with an annual cost of $12,564.

Considering the portability, the most portable devices are again the NIR-S-G1, NIR-S-T2 and PADs that are considered as "miniature". The second most portable device is the Truscan RM ("handheld") and the Minilab ("transportable"). Obviously, HPLC has a baseline score since it is a benchtop device.

Another item is the sample throughput ($w = 5$) that considers the analysis time per sample. We based our evaluation on the "median total time per sample" reported in Caillet et al. [24] except for HPLC whose analysis time per sample was evaluated as >60 min (considering a screening method with a gradient time of 60 min). The time per sample for NIR-S-T2 was evaluated at 10 min per sample based on our experience. The fastest techniques, as expected, were the spectroscopies

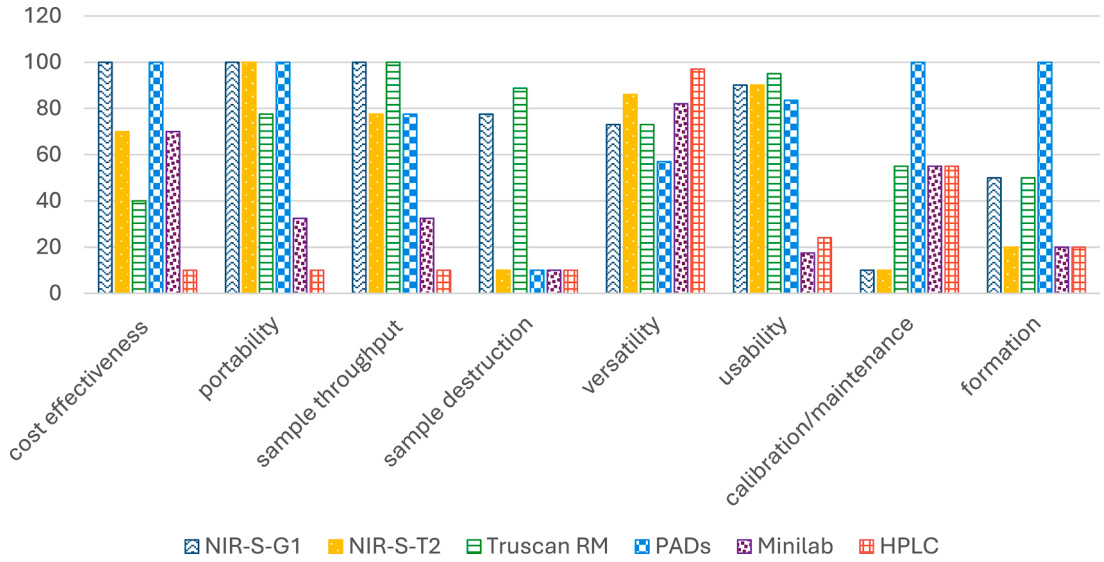


Fig. 3. Medicines' Quality screening device Blueness items ex-ante evaluation for qualitative application (detection of poor-quality medicines).

performed in reflexion mode (1.5 min and 2 min for NIR-S-G1 and Truscan RM respectively). Then come the PADs and the NIR-S-T2 (10 min for both techniques) and finally the Minilab (35 min per sample).

The sample destruction is another item considered for Blueness. We gave it a weight of 5 ($w = 5$) although it may be of paramount importance to keep samples non destroyed for further forensic analyses or, during pharmacy routine inspection, to allow their consumption by patients after analysis. Clearly, two categories of devices appear: reflexion mode spectroscopies and the others. Indeed, PADs, Minilab, NIR-S-T2 and obviously HPLC require the destruction of the sample. Although a fraction of it may be kept for further confirmatory analysis (as an aliquot of the powdered sample), its integrity and physical state is lost. In contrast, NIR-S-G1 and Truscan RM allow the non-invasive analysis of the sample. Indeed, both techniques allow the analysis of samples through most thin translucid containers (e.g. PVC blisters) but are incapable of analysing through most opaque containers (e.g. aluminium blisters). Other issues may also appear regarding the fluorescence of coloured blisters or the ageing of the blister and its impact on the spectral response. Therefore, we gave a 0.5 score for the non-invasive character of NIR-S-G1 and 0.75 score for Raman spectroscopy since the latter may analyse through a more diverse set of containers (e.g. thick glass or thin plastic bottles). We didn't considered the possibility to use different Raman technologies such as SORS to analyse through containers [47,48].

Versatility has been compared on four levels: the possibility to perform quantitative and qualitative analyses, the impact of the dosage form on the implementation of the device and the possibility to analyse several API with the same device. All devices allowed the qualitative analysis of medicines (detection of API). NIR-S-T2 and HPLC allow for the quantitative analysis of samples whereas NIR-S-G1, Truscan RM, PADs and Minilab have semi-quantitative capabilities and were attributed a score of 0.5. The difference between dosage forms has a greater impact on spectroscopic techniques (particularly in reflection mode) and a lesser impact on Minilab and HPLC techniques. Finally, PADS are designed for a specific API and are not yet available for many of them (20 APIs as listed in the 2022 PAD manual [49]) and NIR-S-T2 is also limited to highly dosed and easy to dissolve APIs which explains their score of 0 and 0.5 respectively for this sub-item.

As can be seen on Fig. 3, most devices have a good usability with scores above 80 % at the notable exceptions of Minilab and HPLC (17.5 % and 24.1% respectively). This item is divided in three sub-items:

- The analysis of a reference sample, that is required for HPLC, Minilab (score: 0). Non mandatory but recommended is the testing of a chemical reference standard when using new PADs batches as quality control check (score: 0.5).
- The automatic interpretation of the results is also a key difference between the devices. Indeed, PADs have the highest score because a mobile application is now available for the automatic interpretation of the results. For Truscan RM, there is an integrated software for basic analyses (correlation to reference spectra) and the integration of chemometric models is available with an add-on (which explains that the sub-item's score is a bit lower than 1). Regarding NIR-S-G1 and NIR-S-T2, mobile applications are available to pilot the devices, but the interpretation of results remains basic. However, it is straightforward to implement automatic results analysis on personal computer. Therefore, the score is of 0.7 ("automatic interpretation not available but possible at short term"). For the Minilab, there is no easy to implement automatic interpretation of the results at the notable exception of a mobile app whose use and source code are available at [50,51]. Nevertheless, its implementation is not as straightforward as for the other devices which explains its final score of 0.5 for the automatic interpretation of results. Finally, most commercial software allows for the automatic interpretation of chromatograms for routine analyses However, this requires development of the interpretation method that explains the 0.7 score (automatic interpretation not available but possible at short term).
- The possibility to pre-calibrate the device. This clearly constitutes a plus for the "usability" item. Indeed, pre-calibrated devices may be directly used by minimally trained users. This is the case for spectroscopic devices and PADs (that are pre-calibrated by nature). However, this is not the case for Minilab nor HPLC which explains their low score.

Even if devices may be pre-calibrated, a major limitation to the implementation of NIR spectrophotometers is the complexity of this pre-calibration phase. This is shown in the "calibration/maintenance" item. Currently, everyone that wants to use a NIR device for medicines quality check must perform the calibration itself. This has a large cost in manpower that must be highly skilled, in time for the collection of a representative dataset and in money for the collection and reference analysis of samples. In addition, once built, the chemometric models must be maintained over time which complicates again the routine use of these devices. This explains the baseline score of 10 for NIR-S-G1 and NIR-S-T2. Truscan RM is in a different situation since, for the identification of APIs, the user may rely on commercial databases of pure substances. However, for the identification of specific brands, the problem is the same as for NIR spectroscopy. Therefore, the score for Truscan RM is 55 % for this item (10 + 45). The Minilab and HPLC do not require previous calibration since it is performed by the analysis of reference samples concomitantly to the analysis of the samples (already accounted in the usability item). Nevertheless, Minilab and HPLC require a maintenance (buying of reagents, maintenance of equipment etc.) limiting their score to 55 %%. Finally, PADs are the best for this item since no maintenance nor pre-calibration is needed.

The eighth item ($w = 1$) is the formation time required before a basic user can use the device. HPLC is the worst one since a basic user cannot be rapidly trained to its use. Minilab and NIR-S-T2 also require a basic laboratory formation which requires more than one day. On the other side, because of its simplicity, PADs are the best performing devices in this category since a small formation of less than a half day may be sufficient to provide good results even for a basic user. NIR-S-G1 and Truscan RM are in an intermediary position since a basic formation of a single day may be sufficient.

The weighted geometric mean of these eight items resulted in CS_{blue} scores of 71.3 %, 69.9 %, 67.9 %, 51.2 % 36.6 % and 17.2 % for NIR-S-G1, PADs, Truscan RM, NIR-S-T2, Minilab and HPLC respectively. Looking at Fig. 2, the "bluest devices" are NIR-S-G1 and PADs. Truscan RM is very close to the top, but its cost efficiency impacts its overall score. NIR-S-T2, Minilab and HPLC are the least performing devices below the median score of 59.6 %.

3.1.4. Whiteness

After evaluating the devices for each aspect (analytical, environmental, practical), it is time to gather the information in a single criterion. As described above, for the *ex-ante* comparison, less weight was given to the analytical performances since these are evaluated based on the literature and is likely to change according to the specific application. *Ex-ante* Whiteness was evaluated as the weighted geometric mean of the different *CS* with the following weights: $W_{red} = 1$; $W_{green} = W_{blue} = 2$.

Looking at Fig. 2, the median CS_{white} is of 65.3. The best performing devices are, in descending order, NIR-S-G1, Truscan and PADs with CS_{white} of 80.7 %, 77.6 % and 73.3 % respectively. NIR-S-T2 is in an intermediate position, but below the median, with CS_{white} of 55.6 % The least performing devices are the Minilab and HPLC with CS_{white} of 38.2 % and 28.1 % respectively.

3.2. Ex-post analysis

Ex-post analysis has been performed on two different applications:

qualitative and quantitative analysis of medicines. Two major differences are to be noted between the *ex-ante* and the diverse *ex-post* analyses:

- "versatility" item of the CS_{blue} is removed from the ex-post analyses since the specific application is defined. In addition, the quantitative or semi-quantitative character is now directly reflected in the CS_{red} for the ex-post analysis of quantitative applications.
- Each CS_{colour} has the same weight for the CS_{white} computation.

3.2.1. Qualitative ex-post analysis

The qualitative *ex-post* analysis was performed on two different common applications of medicines quality screening device: API identification and brand identification. For the API identification, NIR-S-G1 and PADs were used to confirm the presence of ciprofloxacin in different medicines collected in the field in Cameroon [39]. The brand identification targeted Coartem® as reference versus different generics and active ingredients [40]. Coartem® is an antimalarial combination therapy composed of artemether and lumefantrine. What is interesting in this analysis is the presence of the same device (NIR-S-G1) in two different applications and how it impacts its score (see Fig. 4).

On the one hand, for the API identification, the analysis was performed on tablets removed from the blister. On the other hand, the analysis of tablets for Coartem® identification was performed through the blister. This impacts the CS_{blue} of NIR-S-G1 that changes from 68.2 % for API identification to 73.3 % for brand identification. This confirms that the *ex-ante* CS_{blue} of 71.3 % was appropriate as first guess. NIR-S-G1 also exhibited very different performances for the Redness criterion. Indeed, NIR spectroscopy is very sensitive to both API, excipients and the physical properties of the sample. Therefore, it may exploit its full potential for brand identification with a homogeneous target class spectral variability. However, it is a disadvantage for API identification since the various API/excipient ratio and various formulations increase the spectral variability of the target class. As expected, CS_{green} was not impacted by the specific application. PADs had lower CS_{red} compared to the *ex-ante* analysis because of specificity issues between ciprofloxacin and other fluoroquinolones. All-in-all, both NIR-S-G1 and PADs had lower but consistent whiteness scores compared to the *ex-ante* analysis.

Regarding the brand identification, Truscan RM exhibit a smaller CS_{blue} value since the automated results were not available for the specific chemometric model used (i.e. data-driven SIMCA, [52]). In addition, as discussed earlier, the specific brand identification application requires a pre-calibration of the chemometric model and its maintenance. These two aspects also impact CS_{blue} that decreases from 67.9 % for the ex-ante analysis to 55.1 % for the ex-post analysis. Nevertheless, the biggest difference between *ex-ante* and *ex-post* analysis for Truscan RM concerns the CS_{red}.Indeed, Raman spectroscopy is less sensitive to physical properties of the sample especially when high scatterers are present in large amount. This was the case for Coartem® where lumefantrine largely dominated the measured signal masking the excipients and artemether signals. Therefore, the distinction between Coartem® and the generics was not complete which explains the low CS_{red} of 42.1 %. Unfortunately, Truscan RM was not used in the ciprofloxacin study where it could have outperformed the two competing devices for the CS_{red}.

Interestingly, the *ex-post* analysis of the different devices, despite some adjustments in the different CS_{white} scores exhibit the same tendency as the *ex-ante* analysis. This reinforces the fact that this preliminary investigation prior to any technical or analytical development may greatly assist in choosing the most appropriate device for the specific application envisaged.

3.2.2. Quantitative ex-post analysis

In this paragraph, two different quantitative methods were developed with the same device (NIR-S-T2): the assay of ciprofloxacin in tablets [41] and the assay of quinine in different dosages forms [28]. The main differences between the two methods are the dilution medium (water for ciprofloxacin and HCl 1 M for quinine) and their respective analytical performances. Therefore, $CS_{blue} = 48.1$ % remains the same for both methods, CS_{green} is higher for ciprofloxacin ($CS_{green} = 56.2$ %) since water is less hazardous than HCl 1 M used for the assay of quinine ($CS_{green} = 47.5$ %). CS_{red} is different for the two methods, however it is less relevant for the comparison since the purpose is different and, since both methods were validated, that they were considered fit-for-purpose for CS_{red}.

The last case study is the use of Truscan RM for the quantitation of ibuprofen in tablets [42]. This quantitation has been performed on intact tablets outside their packaging with the analysis of results directly available in the device. The final scores are shown in Fig. 5.

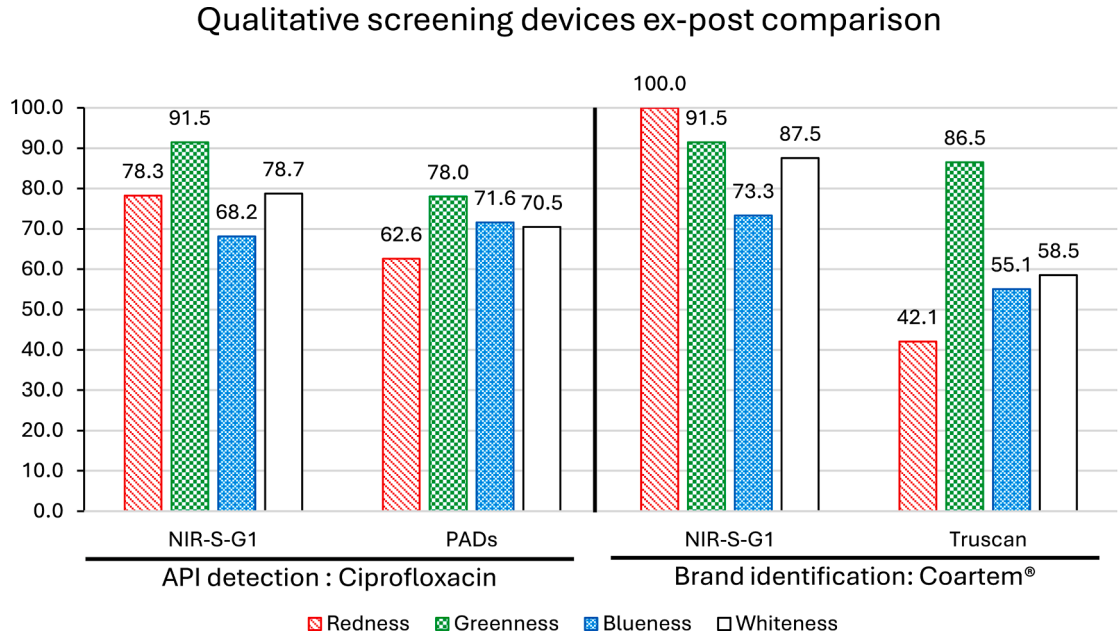


Fig. 4. qualitative screening device ex-post comparison.

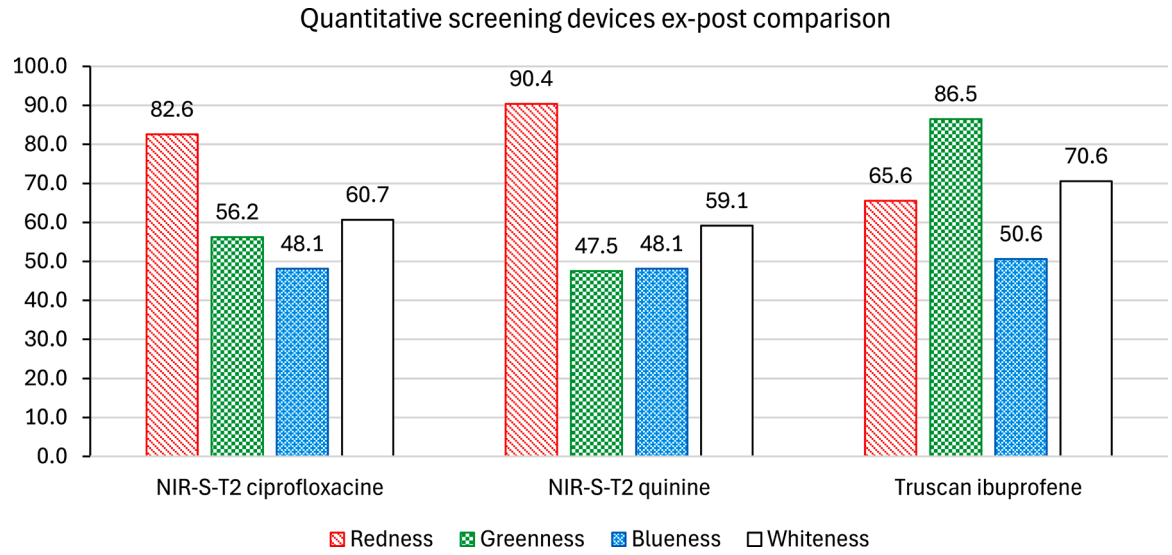


Fig. 5. quantitative screening device ex-post comparison.

4. Conclusion and discussion

During this study, we investigated the opportunity of using the White Analytical Chemistry concept for the evaluation of medicines' quality screening device in LMIC field settings. A generic template covering the Red (analytical), Green (environmental) and Blue (practical/economical) aspects has been developed with specifically selected items. The results showed that wet chemistry-based techniques such as the Minilab, the HPLC and NIR spectroscopy in transmission mode performed less well overall than vibrational spectroscopies in reflexion mode or paper analytical devices. Indeed, the practical and environmental aspects of these techniques counterbalance their better analytical performances. These results are obtained with two major constraints that are the portability and the cost effectiveness. These items may be less important in different settings (e.g. in a fixed laboratory) or depending on the place of each device in the supply chain. Therefore, to allow a certain flexibility on the use and application of this strategy, each colour score and item per colour is given a weight that must be chosen considering the final purpose of the technique. This weight is directly reflected in the CS_{colour} score computed as a weighted geometric mean of its components.

Nevertheless, when it comes to on-site analysis of drug quality, inspectors need highly portable, robust, inexpensive, accurate and versatile techniques. The most promising techniques for this application are the vibrational spectroscopy-based techniques which is also reflected in the CS_{white} scores. This general trend is also present in many other fields where the non-destructive and fast character of portable spectroscopy are needed (e.g. food and agriculture, archaeology, forensics etc. [53, 54]). The main obstacle to their routine implementation is their price (for Raman spectroscopy), the absence of public or commercial databases or pre-trained models and the difficulties in calibration transfer between equipment [55]. These two limitations are considered in the CS_{blue} item. However, this is an active research field, and one may expect that these barriers will soon fall [55,56].

Regarding the Minilab, it is based on rather old principles and techniques that have demonstrated their applicability and interest over time. However, its design implies a low portability and requires laboratory infrastructure. It may be thought as a second line screening device together with the NIR-S-T2 and should be regarded as a technique that may be used to reduce the workload of confirmatory laboratories rather than a first line tool that intends to analyse many samples in a limited time and detect the poorest quality medicines in a non-destructive fashion.

In a general manner, the WAC approach invites the user to investigate the different facets of a screening device and enables to highlight its forces and weaknesses allowing an informed choice to be made on the best device for the intended application. Once one or two devices are selected, the method development may start, and WAC may be used in a more traditional way to compare the *ex-post* results of the different methods developed. This approach is also perfectly aligned with the UN SDGs since it helps selecting the tools that best improve health (Redness), supports local institutions with a limited economic impact (Blueness) while preserving the environment (Greenness).

In view of these results, we suggest that a WAC assessment be carried out on new medicines' quality control devices or their application, to enable a transparent and comparable assessment of performance.

CRediT authorship contribution statement

Pierre-Yves Sacre: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Christelle A. Waffo Tchounga:** Writing – original draft, Formal analysis, Conceptualization. **Charlotte De Bleye:** Writing – review & editing, Conceptualization. **Philippe Hubert:** Writing – review & editing, Conceptualization. **Roland D. Marini:** Writing – review & editing, Conceptualization. **Eric Ziemons:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.greeac.2024.100158.

References

[1] United Nations, Transforming our world: the 2030 agenda for sustainable development, 2015. https://sdgs.un.org/2030agenda.
[2] World Health Organization, Substandard and falsified medical products, (2018). https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products (accessed June 18, 2024).
[3] R.K. Salami, S.V. de Almeida, A. Gheorghe, S. Njenga, W. Silva, K. Hauck, Health, economic, and social impacts of substandard and falsified medicines in low- and

middle-income countries: a systematic review of methodological approaches, Am. J. Trop. Med. Hyg. 109 (2023) 228–240, https://doi.org/10.4269/ajtmh.22-0525.
[4] World Health Organization, Quality control laboratories for medicines, (2024). https://extranet.who.int/prequal/medicines/prequalified/quality-control-labs (accessed June 20, 2024).
[5] S. Kovacs, S.E. Hawes, S.N. Maley, E. Mosites, L. Wong, A. Stergachis, Technologies for detecting falsified and substandard drugs in low and middle-income countries, PLoS. One 9 (2014) e90601, https://doi.org/10.1371/journal.pone.0090601.
[6] J.M. Chisholm, R. Zamani, A.M. Negm, N. Said, M.M. Abdel daiem, M. Dibaj, M. Akrami, Sustainable waste management of medical waste in African developing countries: a narrative review, Waste Manage Res. 39 (2021) 1149–1163, https://doi.org/10.1177/0734242X211029175.
[7] P.T. Anastas, Green chemistry and the role of analytical methodology development, Crit. Rev. Anal. Chem. 29 (1999) 167–175, https://doi.org/10.1080/10408349891199356.
[8] S. Armenta, S. Garrigues, M. de la Guardia, Green analytical chemistry, TrAC, Trends Anal. Chem. 27 (2008) 497–511, https://doi.org/10.1016/j.trac.2008.05.003.
[9] A. Gałuszka, Z. Migaszewski, J. Namieśnik, The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices, TrAC, Trends Anal. Chem. 50 (2013) 78–84, https://doi.org/10.1016/j.trac.2013.04.010.
[10] M. Locatelli, A. Kabir, M. Perrucci, S. Ulusoy, H.I. Ulusoy, I. Ali, Green profile tools: current status and future perspectives, Adv. Sample Preparation 6 (2023) 100068, https://doi.org/10.1016/j.sampre.2023.100068.
[11] M. Branković, Green chemical analysis: main principles and current efforts towards greener analytical methodologies, Anal. Methods 15 (2023) 6631–6642, https://doi.org/10.1039/D3AY01644G.
[12] F.A. Semysim, B.K. Hussain, M.A. Hussien, E.A. Azooz, D. Snigur, Assessing the greenness and environmental friendliness of analytical methods: modern approaches and recent computational programs, Crit. Rev. Anal. Chem. (2024) 1–14, https://doi.org/10.1080/10408347.2024.2304552.
[13] P.M. Nowak, P. Kościelniak, What color is your method? Adaptation of the RGB additive color model to analytical method evaluation, Anal. Chem. 91 (2019) 10343–10352, https://doi.org/10.1021/acs.analchem.9b01872.
[14] P.M. Nowak, R. Wietecha-Posłuszny, J. Pawliszyn, White analytical chemistry: an approach to reconcile the principles of green analytical chemistry and functionality, TrAC, Trends Anal. Chem. 138 (2021) 116223, https://doi.org/10.1016/j.trac.2021.116223.
[15] P.M. Nowak, What does it mean that “something is green”? The fundamentals of a unified greenness theory, Green Chem. 25 (2023) 4625–4640, https://doi.org/10.1039/D3GC00800B.
[16] B. Jain, R. Jain, P.M. Nowak, N. Ali, M.N. Ansari, A. Kabir, L.P. Chandravanshi, S. Sharma, Comparison of various sample preparation methods for benzodiazepines in terms of the principles of white analytical chemistry, TrAC Trends Anal. Chem. 171 (2024) 117524, https://doi.org/10.1016/j.trac.2024.117524.
[17] N. Manousi, W. Wojnowski, J. Płotka-Wasylka, V. Samanidou, Blue applicability grade index (BAGI) and software: a new tool for the evaluation of method practicality, Green. Chem. 25 (2023) 7598–7604, https://doi.org/10.1039/D3GC02347H.
[18] A.M. Jiménez-Carvelo, A. Arroyo-Cerezo, L. Cuadros-Rodríguez, Evaluating the whiteness of spectroscopy-based non-destructive analytical methods – application to food analytical control, TrAC, Trends Anal. Chem. 170 (2024) 117463, https://doi.org/10.1016/j.trac.2023.117463.
[19] S. Vickers, M. Bernier, S. Zambrzycki, F.M. Fernandez, P.N. Newton, C. Caillet, Field detection devices for screening the quality of medicines: a systematic review, BMJ Glob. Health 3 (2018) e000725, https://doi.org/10.1136/bmjgh-2018-000725.
[20] C. Caillet, S. Vickers, S. Zambrzycki, N. Luangasanatip, V. Vidhamaly, K. Boutsamay, P. Boupha, Y. Lubell, F.M. Fernández, P.N. Newton, Multiphase evaluation of portable medicines quality screening devices, PLoS. Negl. Trop. Dis. 15 (2021) e0009287, https://doi.org/10.1371/journal.pntd.0009287.
[21] L. Roth, K.B. Biggs, D.K. Bempong, Substandard and falsified medicine screening technologies, AAPS. Open. 5 (2019) 2, https://doi.org/10.1186/s41120-019-0031-y.
[22] General Chapter, ⟨2⟩Oral Drug Products—Product Quality Tests, USP-NF, United States Pharmacopeia, Rockville, MD, 2024.
[23] General Chapter, ⟨1850⟩ Evaluation of Screening Technologies For Assessing Medicine Quality, USP-NF, United States Pharmacopeia, Rockville, MD, 2023.
[24] C. Caillet, S. Vickers, S. Zambrzycki, F.M. Fernández, V. Vidhamaly, K. Boutsamay, P. Boupha, P. Peerawaranun, M. Mukaka, P.N. Newton, A comparative field evaluation of six medicine quality screening devices in Laos, PLoS. Negl. Trop. Dis. 15 (2021) e0009674, https://doi.org/10.1371/journal.pntd.0009674.
[25] S.C. Zambrzycki, C. Caillet, S. Vickers, M. Bouza, D.V. Donndelinger, L.C. Geben, M.C. Bernier, P.N. Newton, F.M. Fernández, Laboratory evaluation of twelve portable devices for medicine quality screening, PLoS. Negl. Trop. Dis. 15 (2021) e0009360, https://doi.org/10.1371/journal.pntd.0009360.
[26] C. Caillet, S. Vickers, V. Vidhamaly, K. Boutsamay, P. Boupha, S. Zambrzycki, N. Luangasanatip, Y. Lubell, F.M. Fernández, P.N. Newton, Evaluation of portable devices for medicine quality screening: lessons learnt, recommendations for implementation, and future priorities, PLoS. Med. 18 (2021) e1003747, https://doi.org/10.1371/journal.pmed.1003747.
[27] N. Luangasanatip, P. Khonputsa, C. Caillet, S. Vickers, S. Zambrzycki, F. M. Fernández, P.N. Newton, Y. Lubell, Implementation of field detection devices for antimalarial quality screening in Lao PDR—a cost-effectiveness analysis, PLoS. Negl. Trop. Dis. 15 (2021) e0009539, https://doi.org/10.1371/journal.pntd.0009539.
[28] P.H. Ciza, P.-Y. Sacre, M.R. Kanyonyo, C.T. Waffo, M.A. Borive, L. Coïc, J. K. Mbinze, Ph. Hubert, E. Ziemons, R.D. Marini, Application of NIR handheld transmission spectroscopy and chemometrics to assess the quality of locally produced antimalarial medicines in the Democratic Republic of Congo, Talanta Open. 3 (2021) 100025, https://doi.org/10.1016/j.talo.2020.100025.
[29] D. Chicco, G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC. Genomics. 21 (2020) 6, https://doi.org/10.1186/s12864-019-6413-7.
[30] A.L. Pomerantsev, O.Ye. Rodionova, New trends in qualitative analysis: performance, optimization, and validation of multi-class and soft models, TrAC, Trends Anal. Chem. 143 (2021) 116372, https://doi.org/10.1016/j.trac.2021.116372.
[31] General Chapter, ⟨1210⟩ Statistical Tools For Procedure Validation, USP-NF, United States Pharmacopeia, Rockville, MD, 2023.
[32] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), ICH Q2 (R2): validation of Analytical Procedures, 2023.
[33] E. Rozet, V. Wascotte, N. Lecouturier, V. Préat, W. Dewé, B. Boulanger, Ph. Hubert, Improvement of the decision efficiency of the accuracy profile by means of a desirability function for analytical methods validation, Anal. Chim. Acta 591 (2007) 239–247, https://doi.org/10.1016/j.aca.2007.04.002.
[34] E. Psillakis, F. Pena-Pereira, The twelve goals of circular analytical chemistry, TrAC, Trends Anal. Chem. (2024) 117686, https://doi.org/10.1016/j.trac.2024.117686.
[35] U.S. Pharmacopeial Convention, USP Technology Review: global Pharma Health Fund (GPHF) - Minilab., Rockville, MD, 2020.
[36] P.E. Leary, R.A. Crocombe, B.W. Kammrath, Introduction to Portable Spectroscopy, in: R. Crocombe, P. Leary, B. Kammrath (Eds.), Portable Spectroscopy and Spectrometry, 1st ed., Wiley, 2021, pp. 1–13, https://doi.org/10.1002/9781119636489.ch1.
[37] B.M. Wise, R.T. Roginski, A calibration model maintenance roadmap, IFAC-PapersOnLine 48 (2015) 260–265, https://doi.org/10.1016/j.ifacol.2015.08.191.
[38] L. Roth, A. Nalim, B. Turesson, L. Krech, Global landscape assessment of screening technologies for medicine quality assurance: stakeholder perceptions and practices from ten countries, Global. Health 14 (2018) 43, https://doi.org/10.1186/s12992-018-0360-y.
[39] C.A. Waffo Tchounga, P.-Y. Sacré, R. Ravinetto, M. Lieberman, P. Hamuli Ciza, R. Ngono Mballa, E. Ziemons, P. Hubert, R. Djang’eing’a Marini, Usefulness of medicine screening tools in the frame of pharmaceutical post-marketing surveillance, PLoS. One 18 (2023) e0289865, https://doi.org/10.1371/journal.pone.0289865.
[40] P.H. Ciza, P.-Y. Sacre, C. Waffo, L. Coïc, H. Avohou, J.K. Mbinze, R. Ngono, R. D. Marini, P. Hubert, E. Ziemons, Comparing the qualitative performances of handheld NIR and Raman spectrophotometers for the detection of falsified pharmaceutical products, Talanta 202 (2019) 469–478, https://doi.org/10.1016/j.talanta.2019.04.049.
[41] C.A. Waffo Tchounga, R.D. Marini, E. Nnanga Nga, P. Ciza Hamuli, R. Ngono Mballa, P. Hubert, E. Ziemons, P.-Y. Sacré, In-field implementation of near-infrared quantitative methods for analysis of medicines in tropical environments, Appl. Spectrosc. 77 (2023) 1264–1279, https://doi.org/10.1177/00037028231201653.
[42] L. Coic, P.-Y. Sacré, A. Dispas, E. Dumont, J. Horne, C. De Bleye, M. Fillet, P. Hubert, E. Ziemons, Evaluation of the analytical performances of two Raman handheld spectrophotometers for pharmaceutical solid dosage form quantitation, Talanta 214 (2020) 120888, https://doi.org/10.1016/j.talanta.2020.120888.
[43] Microsoft® Excel® for Microsoft 365 MSO (Version 2405 Build 16.0.17628.20006), (2024). https://www.microsoft.com/microsoft-365.
[44] Matlab version 9.13.0.2553342 (R2022b) Update 9, (2022). https://www.mathworks.com.
[45] PLS_Toolbox 9.3, (2023). http://www.eigenvector.com.
[46] R. Deidda, P.-Y. Sacre, M. Clavaud, L. Coïc, H. Avohou, P. Hubert, E. Ziemons, Vibrational spectroscopy in analysis of pharmaceuticals: critical review of innovative portable and handheld NIR and Raman spectrophotometers, TrAC, Trends Anal. Chem. 114 (2019) 251–259, https://doi.org/10.1016/j.trac.2019.02.035.
[47] S. Mosca, Q. Lin, R. Stokes, T. Bharucha, B. Gangadharan, R. Clarke, L. G. Fernandez, M. Deats, J. Walsby-Tickle, B.Y. Arman, S.R. Chunekar, K.D. Patil, S. Gairola, K. Van Assche, S. Dunachie, H.A. Merchant, R. Kuwana, A. Maes, J. McCullagh, C. Caillet, N. Zitzmann, P.N. Newton, P. Matousek, Innovative method for rapid detection of falsified COVID-19 vaccines through unopened vials using handheld Spatially Offset Raman Spectroscopy (SORS), Vaccine 41 (2023) 6960–6968, https://doi.org/10.1016/j.vaccine.2023.10.012.
[48] M.A. Mansouri, P.-Y. Sacré, L. Coïc, C. De Bleye, E. Dumont, A. Bouklouze, P. Hubert, R.D. Marini, E. Ziemons, Quantitation of active pharmaceutical ingredient through the packaging using Raman handheld spectrophotometers: a comparison study, Talanta 207 (2020) 120306, https://doi.org/10.1016/j.talanta.2019.120306.
[49] University of Notre Dame, Paper Analytical Device, Paper Analytical Device Project (2023). https://padproject.nd.edu/(accessed June 13, 2024).
[50] TLCyzer, n.d. https://tlcyzer.github.io/(accessed June 11, 2024).
[51] C. Hauk, M. Boss, J. Gabel, S. Schäfermann, H.P.A. Lensch, L. Heide, An open-source smartphone app for the quantitative evaluation of thin-layer chromatographic analyses in medicine quality screening, Sci. Rep. 12 (2022) 13433, https://doi.org/10.1038/s41598-022-17527-y.

[52] S. Kucheryavskiy, O. Rodionova, A. Pomerantsev, A comprehensive tutorial on Data-Driven SIMCA: theory and implementation in web, J. Chemom. (2024) e3556, https://doi.org/10.1002/cem.3556, n/a.
[53] R.A. Crocombe, Portable spectroscopy, Appl. Spectrosc. 72 (2018) 1701–1751, https://doi.org/10.1177/0003702818809719.
[54] R.A. Crocombe, P. Leary, B. Kammrath, Portable Spectroscopy and Spectrometry, 1st ed., Wiley, 2021 https://doi.org/10.1002/9781119636489.
[55] R. Nikzad-Langerodi, E. Andries, A chemometrician's guide to transfer learning, J. Chemom. 35 (2021) e3373, https://doi.org/10.1002/cem.3373.
[56] P. Mishra, R. Nikzad-Langerodi, F. Marini, J.M. Roger, A. Biancolillo, D. N. Rutledge, S. Lohumi, Are standard sample measurements still needed to transfer multivariate calibration models between near-infrared spectrometers? The answer is not always, TrAC, Trends Anal. Chem. 143 (2021) 116331, https://doi.org/10.1016/j.trac.2021.116331.