



Cost analysis of screening methods to find non-compliant models using the example of tumble dryers

Arno Knieschewski · Rainer Stamminger · Anne Simo

Received: 31 October 2018 / Accepted: 7 June 2019 / Published online: 7 August 2019
© The Author(s) 2019

Abstract The European energy label was established to reduce energy consumption in Europe. All classes and values on the energy label are stated by the supplier. The task of market surveillance is to ensure correct labels and to validate their values through compliance tests. However, this can only be done for a small fraction of all models on the market, since physical tests are expensive. Screening methods can be used to narrow down the number of critical products before compliance tests are done to decrease the costs of finding non-compliant models. This study shows a mathematical approach to analyze the cost benefits of screening methods to find non-compliant models. Furthermore, the analysis presented has been applied to 18 screening methods for tumble dryers. The performance test for tumble dryers consists of seven test runs. All screening methods evaluated are composed of a reduced number of test runs. The most efficient screening method reduces the surveillance costs by 40%.

Keywords Screening method · Market surveillance · Tumble dryer · Energy label

Introduction

The first framework directive for the European energy label has been established in 1979 with the goal to stimulate improvements in the energy efficiency of selected appliances (Council Directive 79/530/EEC 1979). An increased energy efficiency has multiple benefits, such as protecting the climate, minimizing the dependency on fossil fuel imports, and increasing the living standard of the people. The energy label supports the consumer in choosing more energy-efficient products (Granda et al. 2013; Stadelmann and Schubert 2018; Wüstenhagen and Sammer 2006). A large majority of European consumers recognize and trust the energy label and use it in their purchasing decisions (Molenbroek et al. 2014). One of the main incentives for consumers to invest in energy efficiency is the cost savings from the reduced energy consumption (Bull 2012; Camilleri and Larrick 2014; Cole et al. 2018). All information on the energy label is declared by the supplier. The accuracy of the declared information is randomly monitored by the market surveillance of each country. Thus, market surveillance shall ensure that the suppliers label their products truthfully and that the public maintains confidence in the energy label.

Tumble dryers have had a European energy label since 1996 (Commission Directive 95/13/EC 1995).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12053-019-09805-y>) contains supplementary material, which is available to authorized users.

A. Knieschewski · A. Simo (✉)
Federal Institute for Materials Research and Testing, Rudower
Chaussee 11, 12489 Berlin, Germany
e-mail: anne.simo@bam.de

R. Stamminger
Institute for Agricultural Engineering, University of Bonn,
Nußallee 5, 53115 Bonn, Germany

All values on the energy label for tumble dryers are calculated according to regulation (EU) no. 392/2012. The harmonized test standard EN 61121:2013 describes the standardized measurement procedure to obtain all necessary data. This standard requires that the standard performance test consists of three runs in treatment full and four runs in treatment half. A treatment full is a test run with the rated capacity of laundry which is called a full load. A treatment half is a test run with half of the rated capacity of laundry, which is called a partial load. From these seven measurements, weighted values, such as energy consumption, are calculated. In addition to these seven test runs, conditioning runs are demanded whenever the dryer was idle for more than 36 h. Hence, the conditioning runs are carried out at the beginning of the performance test and after weekends. A performance test usually includes two conditioning runs.

The seven measured test runs can only be evaluated if each of the runs achieves a certain final moisture content that is defined as “cupboard dry” laundry. If the required final moisture content is not achieved in a run, that run is declared invalid and can be repeated once. If a second run does not achieve the final moisture content, the whole performance test is declared invalid.

When market surveillance in the European Economic Area (EEA) controls the energy label of appliances such as tumble dryers, it examines one appliance in a standard performance test. If all measured parameters are within their verification tolerance, the model is declared as compliant. The verification tolerances are defined in each appliance-specific regulation and state for chosen parameters how much the measured value is allowed to deviate from the one declared by the supplier. The allowed deviation is usually based on the measurement uncertainty. If at least one of the measured values does not lie within the verification tolerance, three more units of the same model are tested. If the arithmetic mean of the three additional units tested does not lie within the verification tolerance either, the model is declared as non-compliant. The rules on the organization and operation of the market surveillance are defined in a framework regulation (Regulation (EC) No 765/2008 2008). The power and funding of the market surveillance is regulated according to the national law of each EU member state and may differ. In many countries, the appliances for these tests can be borrowed from the supplier free of charge and picked up without prior notice because free testing appliances delivered by the supplier might be preselected units that will meet the

standard. According to regulation (EU) 2017/1369, market surveillance has the right to recover the testing costs from the supplier and impose penalties, e.g. fines, if a model is declared as non-compliant. The amount of a fine can differ within the EU member states. The expenses for the tests of compliant products are covered by market surveillance.

When the market offers many different models or often replaces older models with new ones, market surveillance does not have the funding to examine an appropriate share of models in a standard performance test (Pahal et al. 2013). This results in undetected non-compliant models. Precise data on compliance rates is hard to obtain. The most-cited estimate for the EEA is an average of non-compliant products across all product groups in the range of 10 to 25% (Molenbroek et al. 2013). Assuming this rate, Krivosik estimated that 100 TWh of final energy¹ is wasted annually in the EU due to non-compliant products (Krivosik 2015). Therefore, a screening method which filters the models on the market and determines the ones that are most likely to be non-compliant might be helpful to increase the efficiency of compliance tests.

Function of screening tests

In the context of this study, the screening tests are not meant to shorten the compliance verification procedure. Screening tests are performed before the compliance verifications start and select models which have a high chance to be non-compliant. Hence, if a screening test selected a model, it is first tested in one standard performance test and when it has any value out of verification tolerance, three more units are tested. It is possible to complement a part of the standard performance test with the result of the screening test if the test procedures are identical.

A screening test is only useful if it is cheaper than the standard performance test. It can be applied on a market with too many models to test each of them properly. The preselection with screening tests has the advantage that it reduces the number of more expensive standard performance tests on compliant models and thus increases the share of testing costs which are recovered by the supplier. However, not every screening method is

¹ Final energy is the energy consumed by end users, such as households.

economically beneficial. This study describes how a cost analysis is carried out and applies it to 18 screening methods for tumble dryers. These screening methods are based on not performing the full set of measurements as requested by the standard performance test procedure.

Cost analysis of screening methods

To evaluate a screening method, different models from the same product group are chosen randomly and examined with a standard performance test. Afterwards, the screening method is used for the very same models. Consequently, for each model, the result of the standard performance test can be compared with the result of the screening test. For example, if the standard performance test for one model detected that at least one of the values is not within the verification tolerance and the screening test selected the same model for further testing, the screening test was successful. The number of models to which this applies is registered as true positive. The same counting is done for false positive, false negative, and true negative results. The terms are used according to Table 1.

The subsequent cost analysis for the market surveillance can be used for any screening method that selects models for compliance verifications.

The benefit of a certain screening method is determined by comparing its costs with the costs of using other potential screening methods or none. The expenses for the market surveillance are calculated for each case with the following parameters:

a: The share of models on the market that have values out of verification tolerance in a standard performance test.

b: The share of models that the screening method selects for further testing (true positive) from all models that have values out of verification tolerance in a standard performance test.

c: The share of models that the screening method selects for further testing (false positive) from all models that do not have values out of verification tolerance in a standard performance test.

d: The share of testing expenses that must be covered by market surveillance from all standard performance tests with models that have values out of verification tolerance. In the EEA, this equals the share of compliant models from all models that

have values out of verification tolerance in a standard performance test.

y: The costs of a screening test.

z: The costs of a standard performance test.

Using these parameters, the costs of finding models with values out of verification tolerance can be calculated as follows:

No screening

If *n* represents the number of models tested, *n · a* models with values out of verification tolerance are found. The expenses for market surveillance when no screening (*e_{ns}*) is used can be calculated with Eq. (1).

$$e_{ns} = n \cdot z (1 - a(1 - d)) \tag{1}$$

where:

n · z are the costs of all the standard performance tests.

(1 - *a*(1 - *d*)) is the share of the testing costs that must be covered by market surveillance.

With screening

In order to find the same number of models with values out of verification tolerance as in “No screening,” $\frac{n}{b}$ models need to be tested, because $\frac{n}{b} \cdot a \cdot b = n \cdot a$. The expenses for market surveillance with screening (*e_{ws}*) can be calculated with Eq. (2).

$$e_{ws} = \frac{n}{b} \cdot y + \frac{n}{b} \cdot z (a \cdot b \cdot d + (1 - a)c) \tag{2}$$

where:

$\frac{n}{b} \cdot y$ are the costs of all screening tests.

$\frac{n}{b} \cdot z (a \cdot b \cdot d)$ are the testing costs of all models that have values out of verification tolerance in the standard performance test, but must be covered by market surveillance anyway, for example, because the models are subsequently declared as compliant.

$\frac{n}{b} \cdot z ((1 - a)c)$ are the testing costs of all false positive models.

Equation (1) can also be considered a special case of Eq. (2). If no screening is done, the costs for the screening test are zero (*y* = 0); all models that have values out

Table 1 Definitions of the relations between a screening outcome and the result of a standard performance test

	Standard performance test— out of verification tolerance	Standard performance test— within verification tolerance
Screening—further testing	True positive	False positive
Screening—no further testing	False negative	True negative

of verification tolerance in a standard performance test are chosen ($b = 1$) and all models that do not have values out of verification tolerance in a standard performance test are chosen ($c = 1$).

Results

We applied the cost analysis presented to the case of tumble dryers and market surveillance in the EEA with access to free appliances and the recovery of testing costs in the case of non-compliance. A total of 48 condenser tumble dryers were purchased from the market and tested once according to the standard EN 61121:2013. All data of our study is available in the [supplementary material](#). The weighted energy consumption (E_t), weighted condensation efficiency (C_t) and weighted program time (T_t) were calculated for all tumble dryers. Nine tumble dryers had at least one of the three parameters out of verification tolerance and four tumble dryers did not reach the final moisture content. The other 35 tumble dryers had all values tested within the verification tolerance.

The data from each single test run was available. Therefore, it could be calculated how E_t , C_t and T_t would have changed if fewer partial load and fewer full load runs were made. This gave us 18 hypothetical screening methods with less test runs. The missing runs to calculate the parameters E_t , C_t and T_t were derived from the runs that were used in the hypothetical screening test. For example, if the screening test consists of four runs at partial load and only one run at full load, the one run at full load was multiplied by three to obtain the seven runs which are needed to calculate the weighted parameters.

If the screening did not contain any full load run, the missing values were calculated from the partial load runs with conversion factors and vice versa. Conversion factors have been derived from the 48 measurements from the average of the three full load runs divided by the average of the four partial load runs. The conversion factors that were used for all screening methods that had

either no full load or no partial load runs are given in Table 2.

If the screening had at least one of the parameters out of verification tolerance, the model was selected for further testing.

Four tumble dryers did not reach the required final moisture content in two test runs, which made it impossible to calculate E_t , C_t and T_t for the standard performance test. In these four cases, the screening method was counted as true positive if it contained a test run that did not reach the required final moisture content. If all test runs from the screening reached the required final moisture content, E_t , C_t and T_t could be calculated. If at least one of the parameters was out of verification tolerance, the case was counted as true positive as well. Otherwise, it was counted as false negative.

The measurements of each test run are affected by random errors. Their influence on the result of the standard performance test is reduced by measuring multiple test runs and calculating weighted values. The screening methods are not using all test runs. To improve the significance of the calculated cost savings, we analyzed the screening methods with different combinations of test runs from the standard performance test. Afterwards, the arithmetic mean of all combinations was used as the result for each screening method. For example, the cost savings of the “1 half” screening was calculated four times, once for each partial load run. For screenings with one full load run, either the first or second full load run of the standard performance test could be used. The third full load run could not be used with our data set because some tumble driers did not reach the demanded final moisture content in the first two test runs with full load. Consequently, the

Table 2 Conversion factors with standard deviation: averaged values of full load runs divided by partial load runs

Energy consumption	Condensation efficiency	Program time
1.80 ± 0.08	0.99 ± 0.03	1.68 ± 0.08

Table 3 Average value and standard deviation of parameters b (share of true positive) and c (share of false positive) of the screening methods with 1 partial load and 1 full load (“1 half + 1 full”) and with 1 full load (“1 full”)

	1 half + 1 full	1 full
b	0.84 ± 0.10	0.77 ± 0.00
c	0.09 ± 0.04	0.14 ± 0.08

laboratory did not measure a third full load run because the tumble dryer already failed the test. For example, for the screening “1 half + 1 full,” we combined each of the four partial load runs with each of the first two full load runs. All eight combinations were used to calculate the cost savings of the screening “1 half + 1 full.”

Parameter a was chosen to be $\frac{13}{48} \approx 0.27$ for our calculations, since 13 out of 48 devices had values out of verification tolerance or could not reach the final moisture content. Parameters b and c were calculated for each screening method. Parameters b and c for the two most beneficial screening methods are shown in Table 3.

Parameter d could not be determined by our data, because the study did not include any compliance tests with three additional units. The share of compliant models in a verification test with three additional units was estimated to be about 25%, as obtained by an interview of a German market surveillance authority: $d = 0.25$.

Instead of assigning average prices to parameters y and z , we used the ratio of the two parameters (y/z). z represents the costs of a standard performance test; thus, it counts for the costs of seven test runs and two conditioning runs. Since the partial load, full load, and conditioning runs cost approximately the same, the screening costs² can be compared with the costs of a standard performance test by the fraction of test and conditioning runs that are needed (Table 4).

Using the parameters a , b , c and d and the known ratio of y/z , we can determine which screening method offers the most cost savings.

The screening methods analyzed can be applied in two ways. They can be used to select tumble dryers which are subsequently tested in a full standard performance test (option A), or the test runs from the screening can be used as the first test runs of the standard performance test (option B).

² According to three interviewed testing institutes that have already been commissioned by the market surveillance, the setup costs for tests on tumble dryers are negligible. Thus, the testing institutes charge according to the number of test runs.

1. *Option A:* The screening test is carried out separately from the standard performance test. The proportion of costs are determined by the ratio of $\frac{e_{ws}}{e_{ns}}$, which is shown for the two most beneficial screening methods in Table 5.

2. *Option B:* The costs can be reduced even further if the test runs of the screening are used as the first runs of the standard performance test. When the screening is part of the standard performance test, its cost will be recovered by the supplier in case of non-compliance. Furthermore, the costs of the screening runs y are subtracted from the costs of the standard performance test z . Equation (2) needs to be modified to represent the expenses for market surveillance for option B (e_{Bws}). The modifications of Eq. (2) are highlighted in Eq. (3) with a bold font.

$$e_{Bws} = \frac{n}{b} \cdot y(1 - a \cdot b(1 - d)) + \frac{n}{b} \cdot (z - y)(a \cdot b \cdot d + (1 - a)c) \tag{3}$$

The proportion of costs for the two most beneficial screening methods with option B are shown in Table 6.

So far, we have compared the surveillance costs up to the point where the same number of models with values out of verification tolerance was found from one standard performance test. If we want to know the costs of the entire compliance tests, the testing costs from three additional units of the models with values out of verification tolerance in the standard performance test need to be added. Therefore, $3n \cdot a \cdot d \cdot z$ is added to e_{ns} , e_{ws} , and e_{Bws} . This addition does not change whether a screening method is cost-efficient or not. It only changes the relative cost savings. The proportion of costs for the entire compliance tests are shown in Table 7 for option A and in Table 8 for option B.

Discussion

The cost calculation of option A does not consider the two following aspects:

- i. We assumed by giving Eqs. (1) and (2) the same parameter d that the same number of models found with values out of verification tolerance in one

Table 4 The columns show the number of partial load (“Half”), full load (“Full”), and conditioning runs (“Cond.”) and the ratio of costs of a screening test to costs of a standard performance test (y/z) for all 18 screening methods

Half	4	4	4	3	3	3	3	2	2	2	2	1	1	1	1	0	0	0
Full	2	1	0	3	2	1	0	3	2	1	0	3	2	1	0	3	2	1
Cond.	2	2	1	2	2	1	1	2	1	1	1	2	1	1	1	1	1	1
y/z	8/9	7/9	5/9	8/9	7/9	5/9	4/9	7/9	5/9	4/9	3/9	5/9	4/9	3/9	2/9	4/9	3/9	2/9

standard performance test will also lead to the same number of non-compliant models detected. However, if the screening runs are additionally carried out, the models were tested more often than the models without screening. If a model has values out of verification tolerance in more test runs, it is more likely to be found non-compliant after three additional units are tested. Thus, parameter d should have been lower for Eq. (2) and the cost savings from the screening methods are slightly higher than evaluated.

- ii. The screening was always a part of the standard performance test it was compared with. Consequently, the result of the standard performance test was influenced by the screening test, and thus, the screening had a higher “true” rate than it would have had if the screening test and the standard performance test had been two individual tests. This influence is especially important for screenings with many test runs. In that case, the test reproducibility and repeatability define whether the cost benefits calculated stay the same. We determined the relative standard deviation for the parameters E_t , C_t , and T_t in a round robin test with one tumble dryer that was sent to five laboratories and underwent four standard performance tests in each laboratory (Table 9).

The parameters in Table 9 show moderate relative standard deviations for the repeatability and reproducibility if compared with the verification tolerance of 6% in the regulation (EU) no. 392/2012. A screening with a lot of test runs has a good repeatability, and therefore, the “true” rate should only be slightly lower than evaluated if

Table 5 Average ratio of $\frac{e_{ws}}{e_{ns}}$ with standard deviation for the screening methods “1 half + 1 full” and “1 full”

	1 half + 1 full	1 full
$\frac{e_{ws}}{e_{ns}}$	0.69 ± 0.07	0.62 ± 0.10

the screening is carried out separately. If a screening involves fewer test runs, the repeatability worsens; however, the influence it had on the result of the standard performance test also decreases. Consequently, this effect only leads to slightly fewer cost savings than evaluated.

The two aspects (i) and (ii) roughly compensate for each other. Hence, the calculated cost benefits of option A are also significant without considering them.

Two further effects that influence the cost benefits of both options and can occur in the practical application are the following:

- iii. We assumed in Eq. (1) that the chances of finding a model with values out of verification tolerance are equal to the share of models on the market with values out of verification tolerance if no screening is used. This assumption is correct. However, the market surveillance often uses simple screening methods to increase the chances of finding non-compliant models. Therefore, the costs of the screening method presented here could also have been compared with the costs of simple screening methods applied already. However, these simple screening methods, similar to examining models in energy classes that were more likely to fail in the past or certain brands which had non-compliant models, can differ greatly between market surveillances from different countries. Additionally, the models for the screening method presented in this study can also be chosen by a simple screening beforehand. Hence, the simple screening reduces the costs of both e_{ns} and e_{ws} and has little influence on our cost benefit evaluation.

Table 6 Average ratio of $\frac{e_{Bws}}{e_{ns}}$ with standard deviation for the screening methods “1 half + 1 full” and “1 full”

	1 half + 1 full	1 full
$\frac{e_{Bws}}{e_{ns}}$	0.54 ± 0.06	0.50 ± 0.07

Table 7 Average ratio of $\frac{e_{ws}+3n \cdot a \cdot d \cdot z}{e_{ns}+3n \cdot a \cdot d \cdot z}$ with standard deviation for the screening methods “1 half + 1 full” and “1 full”

	1 half + 1 full	1 full
$\frac{e_{ws}+3n \cdot a \cdot d \cdot z}{e_{ns}+3n \cdot a \cdot d \cdot z}$	0.75 ± 0.06	0.70 ± 0.08

iv. When the screening method presented here is applied, market surveillance does not strictly have to stick to the rule that every appliance which fails the screening test must be checked in a standard performance test. Market surveillance can look at the results from the screening test and, in the case where the verification tolerance was exceeded by just one parameter by just a little bit, it can refrain from further testing. Thus, the false positive cases (parameter *c*) and, with it, the costs of the screening method can be reduced.

If aspects (iii) and (iv) were considered, this study would become much more complex without improving much in its significance, since these aspects only represent minor influences which also compensate for each other.

Parameters *b*, *c* and *d* and the quotient *y/z* are values that originate from the characteristic of the screening method and the test standard. Therefore, they can be applied as long as the standard remains the same. Parameter *a*, however, depends on the way suppliers label their tumble dryers. Thus, parameter *a* can change significantly in the future. Figure 1 shows the proportion of costs of the screening methods “1 half + 1 full” and “1 full” to find models with values out of verification tolerance in one standard performance test for option B as a function of parameter *a*. A dotted line at *a* = 0.27 highlights the value of parameter *a* that was used in our calculations. Its intersection with the function $\frac{e_{Bws}}{e_{ns}}(a)$ gives the result of $\frac{e_{Bws}}{e_{ns}}(a = 0.27) = 0.54$ for “1 half + 1

Table 8 Average ratio of $\frac{e_{Bws}+3n \cdot a \cdot d \cdot z}{e_{ns}+3n \cdot a \cdot d \cdot z}$ with standard deviation for the screening methods “1 half + 1 full” and “1 full”

	1 half + 1 full	1 full
$\frac{e_{Bws}+3n \cdot a \cdot d \cdot z}{e_{ns}+3n \cdot a \cdot d \cdot z}$	0.64 ± 0.05	0.60 ± 0.06

Table 9 Relative standard deviation for *E_t*, *C_t*, and *T_t* measured for one tumble dryer in five laboratories four times each. The term repeatability refers to the degree of agreement between the results of experiments performed at the same laboratory. The term reproducibility refers to the degree of agreement between the results of experiments performed at different laboratories

	<i>E_t</i>	<i>C_t</i>	<i>T_t</i>
Repeatability in %	0.7	1.0	1.3
Reproducibility in %	1.0	1.7	2.9

full” at *P*₁ and $\frac{e_{Bws}}{e_{ns}}(a = 0.27) = 0.50$ for “1 full” at *P*₃ shown already in Table 6. The break-even point at which the surveillance with screening costs as much as the surveillance without screening is highlighted with the dashed line at $\frac{e_{Bws}}{e_{ns}} = 1$. $\frac{e_{Bws}}{e_{ns}}(a) = 1$ solved for *a* gives *P*₂ at *a* = 0.89 and *P*₄ at *a* = 0.90. Consequently, if less than 89% of all tumble dryers on the market have parameters out of verification tolerance in one standard performance test, the screening method “1 half + 1 full” reduces the costs of finding non-compliant dryers. The “1 full” screening is cost-efficient if less than 90% of all tumble dryers on the market have parameters out of verification tolerance in one standard performance test.

For option A, we have the break-even point for “1 half + 1 full” at *a* = 0.55 and for “1 full” at *a* = 0.65. If the entire compliance test is considered and *3n · a · d · z* is added to the expenses, the break-even points remain the same.

Although a screening is often used to test a large number of specimens, it should be noted that the screening method does not have to be applied many times to be beneficial. Already one screening test has the expected value of the calculated cost savings. Only the probability of reaching this expected value increases with the number of screening tests.

This study focuses on the cost savings regarding the number of non-compliant models found. However, even if the screening methods are not cost-efficient in this regard, they can still be useful in terms of energy savings. The false negative results of the screening only occur with models which are slightly out of verification tolerance. All models that have measured values which are much worse than the declaration have a high probability of being selected by the screening. Hence, the

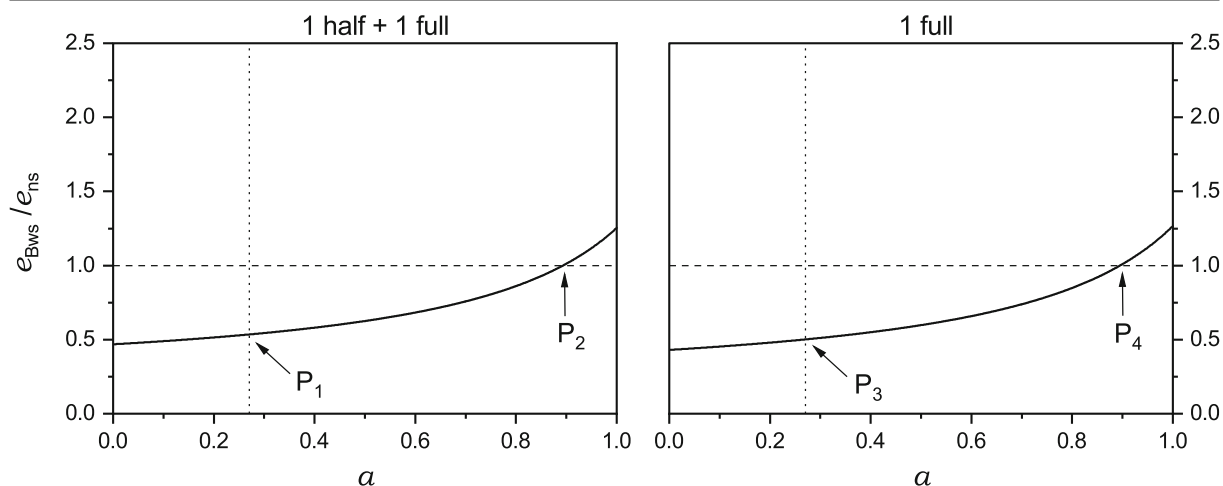


Fig. 1 Quotient of the expenses with and without screening for option B as a function of parameter a for the screening “1 half + 1 full” and “1 full”

screening helps to clear the market from the worst wrongly labeled models.

The same concept to evaluate screening methods can be used for other tumble dryer standards, such as the IEC 61121:2012, or for screening methods of other appliances that use multiple test runs, such as washing machines and dishwashers. Moreover, this cost analysis can be applied to evaluate any other kind of screening method, such as a shorter test method for fridges that has been suggested by Hermes et al. (2013).

Conclusion

The performance test for tumble dryers consists of many measurements from repeated test runs. The purpose of measuring multiple runs and taking the average is to obtain results with a low uncertainty, which is important for meaningful compliance verification tests. However, each additional run also increases the costs of the test. Therefore, a screening test is useful to identify products with a high probability of non-compliance which are subsequently checked properly in a full compliance verification.

Applying the cost calculation concept of screening methods to the test standard EN 61121:2013 for tumble dryers showed that a screening can help market surveillance to find non-compliant dryers less expensively. The “1 half + 1 full” and the “1 full” screening showed the highest cost savings of

all 18 screening methods. With option B, the “1 half + 1 full” screening reduces the costs of finding non-compliant models by $(36 \pm 5) \%$ and the “1 full” screening reduces the costs by $(40 \pm 6) \%$. Although the “1 full” screening has higher cost savings, we suggest the market surveillance to use the “1 half + 1 full” screening because it is independent of the conversion factors which might change over time. The calculated cost savings are only applicable for countries in which the market surveillance can take free samples and have the testing costs recovered by the supplier in the case of non-compliance. If a market surveillance does not have this power, most screenings will be inefficient. To reinforce the important role of market surveillance, we recommend equipping them with this power.

The screening methods become more useful if the share of compliant models on the market increases and, on the other hand, can become inefficient if the share of non-compliance becomes too high. In our case, the share of tumble dryers with values out of verification tolerance in one standard performance test would have to more than triple to make the two screenings mentioned no longer cost-effective. Therefore, it is likely that the screening methods will also reduce the costs of finding non-compliant models in the future.

Funding information This project is financed as part of the National Action Plan on Energy Efficiency (NAPE) of the Federal Ministry for Economic Affairs and Energy.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bull, J. (2012). Loads of green washing – can behavioural economics increase willingness-to-pay for efficient washing machines in the UK? *Energy Policy*, 50, 242–252.
- Camilleri, A. R., & Larrick, R. P. (2014). Metric and scale design as choice architecture tools. *Journal of Public Policy & Marketing*, 33(1), 108–125.
- Cole, J. C., McDonald, J. B., Wen, X., & Kramer, R. A. (2018). Marketing energy efficiency: perceived benefits and barriers to home energy efficiency. *Energy Efficiency*, 1–14. <https://doi.org/10.1007/s12053-018-9614-z>.
- Commission Directive 95/13/EC of 23 May 1995 implementing Council Directive 92/75/EEC with regard to energy labelling of household electric tumble driers (OJ L 136, 21.6.1995, p. 28.). ELI: <http://data.europa.eu/eli/dir/1995/13/oj>.
- Council Directive 79/530/EEC of 14 May 1979 on the indication by labelling of the energy consumption of household appliances.
- Granda, C., Bush, E., & Wold, C. (2013). Market transformation for clothes dryers: lessons learned from the European experience. In *ECEEE Summer Study Proceedings* (pp. 523–527).
- Hermes, C. J. L., Melo, C., & Knabben, F. T. (2013). Alternative test method to assess the energy performance of frost-free refrigerating appliances. *Applied Thermal Engineering*, 50, 1029–1034.
- Krivosik, J. (2015). Energy label display trends in shops – second round results. Market Watch. http://www.market-watch.org.pt/wp-content/uploads/2015/07/Market_Watch_ShopVisit_2ndROUND.pdf Accessed 25 February 2019.
- Moltenbroek, E., Smith, M., Groenenberg, H., Waide, P., Attali, S., Fischer, C., et al. (2013). Background report I: Literature review, evaluation of the Energy Labelling Directive and specific aspects of the Ecodesign Directive ENER/C3/2012–523.
- Moltenbroek, E., Groenenberg, H., Cuijpers, M., Janeiro, L., Smith, M., Surmeli, N., et al. (2014). Evaluation of the Energy Labelling Directive and specific aspects of the Ecodesign Directive ENER/C3/2012–523.
- Pahal, S., Pilenko, D., Cohen, F., Tinetti, B., & Mudgal, S. (2013). Implications of the new Energy Labelling Directive (2010/30/EU) and the Ecodesign of energy-related products (Ecodesign) Directive (2009/125/EC) on market surveillance activities, ATLETE II: second work package.
- Regulation (EC) No 765/2008 of the European parliament and of the council of 9 2008 Setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93. ELI: <http://data.europa.eu/eli/reg/2008/765/oj>.
- Stadelmann, M., & Schubert, R. (2018). How do different designs of energy labels influence purchases of household appliances? A field study in Switzerland. *Ecological Economics*, 144, 112–123.
- Wüstenhagen, R., & Sammer, K. (2006). The influence of eco-labelling on consumer behaviour – results of a discrete choice analysis for washing machines. *Business Strategy and the Environment*, 15, 185–199.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.