DE GRUYTER
OPEN

Beiträge zur Tabakforschung International
*Contributions to Tobacco Research*
Volume 27 · No. 6 · April 2017
DOI: 10.1515/cttr-2017-0008

# Analysis of the Effect of Multiple Testing in Assessing Tobacco Product Differences*

*by*

*Thomas Verron, Xavier Cahours, and Stéphane Colard*

*SEITA, Imperial Tobacco Limited - 48, rue Danton, 45404 Fleury-les-Aubrais, France*

SUMMARY

During the last two decades, an increase of tobacco product reporting requirements from regulators was observed, such as Europe, Canada or USA.
However, the capacity to compare and discriminate accurately two products is impacted by the number of constituents used for the comparison. Indeed, performing a large number of simultaneous independent hypothesis tests increases the probability of rejection of the null hypothesis when it should not be rejected. This leads to virtually guarantee the presence of type I errors among the findings. Correction methods have been developed to overcome this issue like the Bonferroni or Benjamini & Hochberg ones. The performance of these methods was assessed by comparing identical tobacco products with different sizes of data sets. Results showed that multiple comparisons lead to erroneous conclusions if the risk of type I error is not corrected. Unfortunately, reducing the type I error impacts the statistical power of the tests. Consequently, strategies for dealing with multiplicity of data should provide a reasonable balance between testing requirement and statistical power of differentiation. Multiple testing for product comparison is less of a problem if studies restrict to the most relevant parameters for comparison. [Beitr. Tabakforsch. Int. 27 (2017) 78–85]

KEYWORDS

Tobacco product, mainstream cigarette smoke, smoke toxicant emissions, statistical comparison, measurement variability

ZUSAMMENFASSUNG

Über die letzten zwei Jahrzehnte lässt sich beispielsweise in Europa, Kanada und den USA ein deutlicher Anstieg des Umfangs der gesetzlich vorgeschriebenen Berichterstattungspflicht zu Inhaltsstoffen von Tabakprodukten beobachten. Die Möglichkeit zwei Produkte miteinander zu vergleichen, ist stark von der Anzahl der betrachteten Faktoren abhängig. Die Durchführung einer Vielzahl von simultanen unabhängigen Hypothesentests erhöht die Wahrscheinlichkeit der Zurückweisung der Nullhypothese, wenn diese nicht zurückgewiesen werden soll. Dies führt zu einer virtuellen Garantie der Existenz von Typ 1 Fehlern in den Ergebnissen. Es wurden Korrekturmethoden wie die von Bonferroni oder Benjamini & Hochberg eingeführt, um das Problem zu umgehen. Die Leistungsfähigkeit dieser Methoden wurde anhand des Vergleichs von äquivalenten Tabakprodukten mit einem unterschiedlich großen Datensatz mit dem Ergebnis untersucht, dass der multiple Vergleich zu fehlerhaften Entscheidungen führt, wenn Typ 1 Fehler nicht korrigiert werden. Die Reduzierung von Typ 1 Fehlern reduziert jedoch die statistische Aussagekraft. Die Strategie, um mit multiplen Daten zu verfahren, sollte resultierend daraus vernünftig und ausgewogen zwischen analytischem Aufwand und statistischer Differenzierung sein. Multiple Analyse als Tool zum Vergleich von Produkten ist weniger problematisch und eindeutiger, wenn Studien sich auf relevante Parameter zur Vergleichbarkeit konzentrieren. [Beitr. Tabakforsch. Int. 27 (2017) 78–85]

* Received: 22nd November 2016 – accepted: 30th March 2017

Unauthenticated
Download Date | 3/16/20 2:40 PM

RESUME

Pendant les deux dernières décennies, les régulateurs n'ont cessé de réclamer de plus en plus d'information à l'industrie du tabac concernant leurs produits. C'est par exemple le cas en Europe, au Canada ou encore aux Etats-Unis.
La difficulté rencontrée est que la comparaison entre deux produits est dépendante du nombre de paramètres (information/constituants) que l'on considère. En effet, il est connu que la réalisation d'un nombre important de tests d'hypothèses indépendants augmente la probabilité de rejet de l'hypothèse nulle alors qu'elle ne devrait pas être rejetée. Ceci conduit à une augmentation virtuelle de la présence d'erreurs de type I dans les conclusions. Afin de contourner ce problème, des méthodes de corrections, telles que Bonferroni ou Benjamini & Hochberg ont été développées dans de nombreux domaines, notamment en génétique.
Les performances de ces méthodes ont été évaluées en comparant des produits équivalents du tabac et en considérant plus ou moins de paramètres. Les résultats ont montré que si le risque de type I n'était pas corrigé, les comparaisons multiples générées un nombre important de conclusions erronées. Malheureusement, la réduction du risque d'erreur de type I impact la puissance statistique des tests de comparaison. Par conséquent, la stratégie pour gérer cette problématique, liée aux données multivariées, consisterait à trouver un équilibre raisonnable entre les exigences en matière de tests et la puissance statistique de différenciation. Autrement dit, les comparaisons de produit seront plus faciles à gérer si on se limite aux paramètres les plus pertinents pour la comparaison. [Beitr. Tabakforsch. Int. 27 (2017) 78–85]

INTRODUCTION

Tobacco product manufacturers are increasingly being asked by regulatory authorities to report information on their own products. Starting in 2000 with the mandated measurement and reporting of smoke emissions from cigarettes in Canada (1), the requirement to measure and report emissions has spread to other countries, e.g., Venezuela in 2004 (2), Brazil and Taiwan in 2007 (3, 4). The objective of reporting is presumably either to gain a better understanding of the products, to be able to compare tobacco products or to set limits on selected constituents. For instance, since 2009, in the United States of America, the law grants Food and Drug Administration (FDA) new authority to regulate the manufacture, marketing, and distribution of tobacco products. FDA requires each tobacco product manufacturer or importer, or an agent, to report constituents, including smoke constituents. FDA also published draft guidance on the reporting of an abbreviated list of 24 harmful and potentially harmful constituents (HPHCs), 18 in mainstream cigarette smoke and 6 in the cigarette filler blend, for which analytical protocols are assumed to be well established and widely available (5). However, several analytical protocols are not internationally validated and standardized to date.
On a global scale, the World Health Organization (WHO) published through its study group on Tobacco Product Regulation (TobReg) a strategy for tobacco regulation based on product assessments, with the goal of reducing the mainstream smoke levels of selected constituents (6). In 2015, TobReg established a non-exhaustive priority list of 39 contents and emissions of cigarettes (7). TobReg recommends that contents and emissions of tobacco products are measured by the validated methods of its affiliated laboratory network TobLabNet. Methods for measuring nicotine, CO, tobacco-specific nitrosamines (TSNA), benzo[*a*]pyrene in smoke and humectants in blend have already been developed and published, and validation of methods for measuring ammonia content, volatile organic compounds, and aldehyde emissions is under way following WHO recommendations. The proposed regulatory strategy would be implemented in phases, starting with a period of annual reporting of emission levels by cigarette manufacturers to the regulatory authorities. This would be followed by the promulgation of emission levels above which brands cannot be offered for sale.
Collecting smoke constituent data may provide useful information on commercial cigarettes although any comparison between products shall take into account the sources of variability likely to affect the testing results in order to avoid wrong or misleading conclusions. Examples of potential data misinterpretation due to temporal variability within one laboratory (8) and among laboratories (9) have been discussed previously. PURKIS *et al.* reviewed the works undertaken on smoke constituent measurements and related variability, and highlighted the factors influencing this variability (10). ELDRIDGE *et al.* also published on the variation in tobacco and mainstream smoke constituents from selected commercial cigarette products (11), and observed that coefficient of variation (CV) values averaged 20% in a single laboratory for commercial products. This observation was valid for tobacco blend and smoke emissions determined under either smoking regime and when expressed as a ratio to nicotine. However, the CV value reached in excess of 50% for some low-level constituent emissions, for which levels were near the quantitation limits of the analytical method. BELUSHKIN *et al.* have recently discussed the impact of long term variability in product comparison (12). The authors showed that investigation of the phenomena associated with long term variability was critical to enable the implementation of statistical methodologies capable of providing definitive answers with respect to product evaluations. The situation is further complicated when attempting to compare constituent yields obtained in different laboratories. Taken together, the published studies demonstrate the importance of taking into account all sources of variation when comparing testing figures, mandating maximum product levels of constituents and setting tolerance limits. Repeatability and reproducibility derived from the analytical method validation process provide a part of the global variability but not all. The other part corresponding to the inherent variability of manufacturing, e.g., from raw materials or product design features, shall also be considered in the comparison procedure. For that purpose, sources of variability can be aggregated in one formula giving the critical difference (CD) (13), which is the smallest difference between two results so that they can be considered as statistically different. As the usual manufacturing variability of the



Unauthenticated
Download Date | 3/16/20 2:40 PM

commercial cigarettes still need to be investigated, its contribution has been assumed to be null for the calculation of the CD in this study. Since 2015, a CORESTA task force is conducting a long term collaborative study to improve the understanding of overall tobacco and smoke constituent variability relevant to commercial cigarette design features. In 2013, TEILLET *et al*. (14) showed the relevance of using the CD to compare results from different laboratories in order to avoid erroneous conclusions. However, if the comparison of products involves several constituents simultaneously, then the use of CD independently from each constituent, using the same level of statistical significance ($p < 0.05$), could lead to erroneous conclusions as well. This issue well known in statistics is still a very active topic of research with many challenges to be taken into account such as the notion of independence and statistical power. The aim of this paper is to define the multiple testing issue and concepts, to present the risks, when tobacco products data are compared, and to introduce some methods for addressing these risks.

## EXPERIMENTAL

The absolute difference between two testing results originating from two laboratories is compared to the CD in order to determine whether data are statistically different or not. If the difference is lower than the CD, there is no statistically significant difference between the testing results produced by the two laboratories. By contrast, if the difference is higher than the CD, then there is a statistically significant difference. However, problems arise when we do not perform a single hypothesis test but several in parallel. The probability of making a type I error is bounded in hypothesis tests by α, an 'acceptable' risk of type I errors, conventionally set at 0.05. Since each test again has a probability of producing a type I error, performing a large number of simultaneous hypothesis tests virtually guarantees the presence of type I errors among the findings. This problem is called the "inflation" of the α level.

For a single statistical test, the null hypothesis, $H_0$, is rejected when the p-value is lower than $\alpha = 0.05$. In that case, the probability of making an uncontrolled type I error, i.e., the rate of rejecting the null hypothesis when it should not be rejected, while performing the test is 5%. For a set of n independent tests, the probability of not making a type I error is $(1 - \alpha)^n$ and the probability of making at least one type I error on the set of tests is $1 - (1 - \alpha)^n$.

Therefore, corrections when making multiple comparisons of large sets of data are recommended to avoid a too readily null hypothesis rejection. The key goal of multiple statistical testing method corrections is to control the type I errors that arise when several hypothesis tests are performed simultaneously. Multiple-testing correction refers to adjusting the α level associated with each comparison. In order to retain a desired group-wise error rate α in an analysis involving more than one comparison, the error rate for each comparison must be more stringent than α.

The simplest multiple test procedure is the Bonferroni method (15). Specifically, the p-value is calculated by testing each individual hypothesis at a significance level of α divided by the number of hypotheses.

The Bonferroni method is simple and applicable in most of hypothesis tests situation. However, the cost for this simplicity and universality is a low power. While this reduces the number of false rejections, it also increases the probability that the null hypothesis is not rejected when in fact it should have been, this is what is called the type II error.

Observing the weaknesses of the Bonferroni correction, other researchers proposed more sophisticated procedures. Typically, the goal of these methods is to reduce the probability of making one or more false type I errors with less loss of power excessively (16–18). Despite being more powerful than the simple Bonferroni method, the modified Bonferroni methods still tend to result in a loss of power. A more recent class of approaches to this problem focuses not on reducing the type I errors but instead on controlling the expected proportion of false positives among all rejected null hypotheses (19, 20). Controlling the proportion of false positives rather than the type I error rate leads to more powerful and less conservative testing procedure. These methods make particular sense in scientific fields like genetics where it is important to identify the maximum number of significant differences.

In the context of product comparison, the objective is to assess if there is no difference or at least one whatever the number of constituents used for the comparison. Therefore, all the methods of correction will lead to the same conclusion. Given this fact, in our study, only the Bonferroni method was used.

Comparisons of simulated tobacco products were performed without correction and using Bonferroni correction using different number of constituents, K. Figure 1 illustrates the process of simulation. The n replicates for each smoke constituent were simulated by using a normal

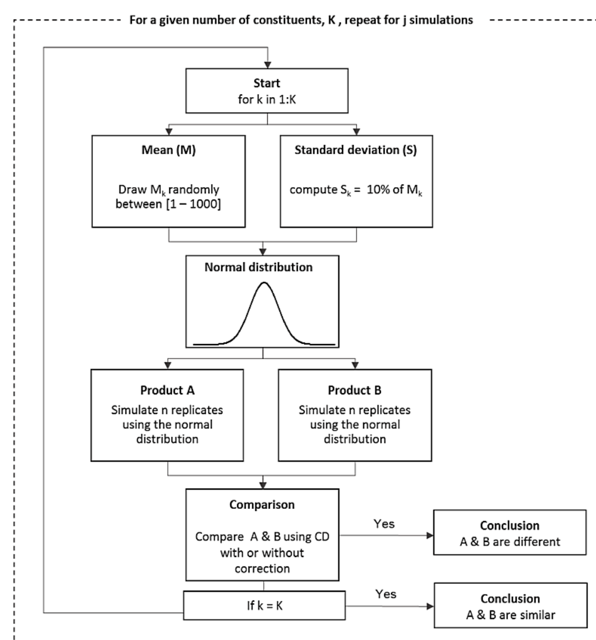


**Figure 1. Process of simulation**



Unauthenticated
Download Date | 3/16/20 2:40 PM

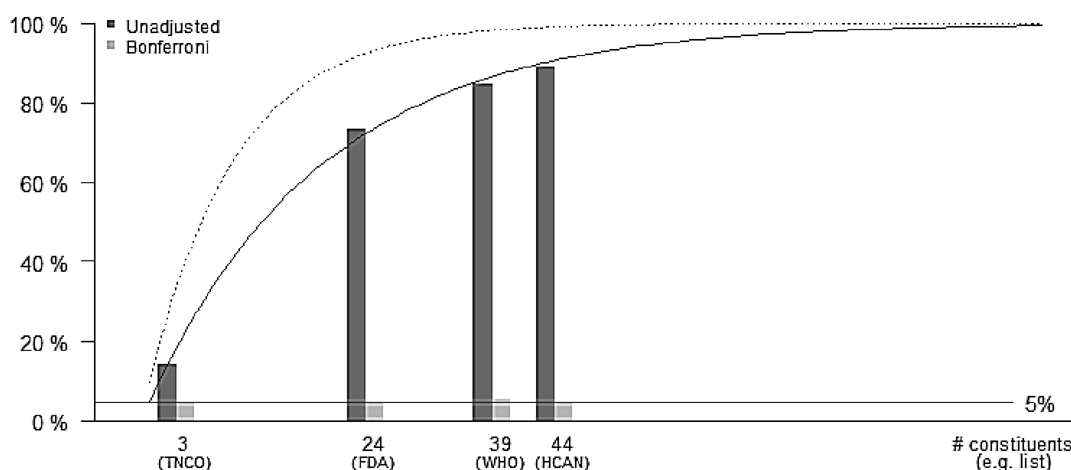


**Figure 2. Percentage of chance to make a wrong decision, i.e., false positive rate, when comparing two identical products *versus* the number of smoke constituents.** Critical differences were corrected using Bonferroni correction. The horizontal line represents a risk of 5% (type I error set to control) and the curves represent the theoretical percentage to make wrong decision with one smoking regime (solid line) and with two smoking regimes (dotted line).

distribution centered on mean value, $M_k$, randomly sampled between 1 and 1000, and with a standard deviation, $S_k$, assumed to be 10% of the average for all constituents. For each smoke constituent, the two products were compared using the CD (13) based on the simulated standard deviations without and with correction.

RESULTS AND DISCUSSION

*Application to comparison of identical products*

*- Simulated data*

The process described above was used to simulate the testing results of products with the same constituent yields. The cases of different number K of smoke constituents was considered with K equal to 3, 24, 39, and 44, corresponding to proposed list described previously. A total of 500 simulations were performed to assess and illustrate trends. Figure 2 shows the percentage of chance to make a wrong decision, i.e., to detect differences when actually two products with the same constituent yields are compared, *versus* the number of constituents analysed. These comparisons were made without and with correction.
Comparison of data sets based on the FDA's abbreviated list of 24 HPHCs leads to a percentage of chance to make a wrong decision equal to 71% when no correction is made. The percentage reaches 90% with the Health Canada list of 44 HPHCs. As a result, corrections need to be made.
By contrast, the type I errors are below the significance level of $\alpha = 5\%$ when correction is made, and this whatever the number of constituents included in the comparison.
In the simulations made, the modification of the level of statistical significance controls the rate of false positives. Therefore, without correction the higher number of constituents is used for the comparison, the more likely it is to reject the null hypothesis when it should not.

*- Experimental data*

Three different commercial products coded A, B, and C covering a range of ISO "tar" yields from 2 to 10 mg/cig approximatively, were analyzed in two different ISO 17025 accredited laboratories coded Lab1 and Lab2. The yields of abbreviated list of HPHC smoke compounds and the "tar" yields under ISO smoking regimes were quantified in two replicates by the two labs for the three products (see Table Annex 1). Three product comparisons were performed using critical difference corrected or not. Adjustment of critical difference was made by applying the Bonferroni correction (see Table Annex 2). Comparison was only possible with the constituents for which the repeatability and reproducibility values were available. Therefore the comparisons were made on a list of 15 constituents because the repeatability and reproducibility of aromatic amines and ammonia were not available at the time of the publication. In such situation, the three following product comparisons A-Lab1 *vs.* A-Lab2, B-Lab1 *vs.* B-Lab2 and C-Lab1 *vs.* C-Lab2 were made.
Table 1 shows the conclusions of comparisons using or not a correction, two options were noted: I for Identical or NI for Non-Identical.
When no correction was used the two comparisons out of three have one or more statically significant differences, even though the products are identical. By contrast the comparison of the identical product analysed in two different laboratories concluded on the equivalence when the correction was made. It confirms the need to apply a correction for product comparison when multiple constituents are considered.

*Issues related to correction*

As explained previously, the correction is required to control the risk to conclude there are differences between products that are actually identical. However, this correction is performed at the potential expense of many



Unauthenticated
Download Date | 3/16/20 2:40 PM

**Table 1. Effect of correction on the comparison of identical products analysed in two different laboratories when considering 15 smoke constituents under ISO smoking regime.**

| | Correction | A-Lab1 *vs.* A-Lab2 | B-Lab1 *vs.* B-Lab2 | C-Lab1 *vs.* C-Lab2 | Percentage of well-classified |
|---|---|---|---|---|---|
| Decision | without | NI | NI | I | 33% |
| | with | I | I | I | 100% |

more false negatives. A false negative corresponds to not finding a difference between products when a difference exists. In order to assess such cases, comparisons with simulations using the same approach as described before but with a change of one constituent level were considered. Two levels of changes were investigated. In the first situation S1, one smoke constituent was significantly changed in such a way that the difference is statistically detectable with 80% of chance, at a risk of $\alpha = 5\%$ (Figure 3). In the second situation S2, one smoke constituent was moderately changed in such a way that the difference is statistically detectable with only 50% of chance, at a risk of $\alpha = 5\%$ (Figure 4). For both situations S1 and S2, the probability to differentiate correctly the product decreases with the increase of the number of constituents simultaneously compared. For instance, for S1 and a list of 44 HPHCs the probability to see a difference between the two products is equal to 49% instead of the expected 80% with one constituent.

The effect on power is sensitive to the number of constituents which are compared. Therefore, when correction methods are used in product comparisons it becomes more difficult to identify a "real" difference when the number of constituents increases.

CONCLUSION

Multiple comparisons can lead to erroneous conclusions if the risk of type I error for individual tests is not corrected downward. Therefore, adjustments for making multiple comparisons in large sets of data are recommended to avoid too readily null hypothesis rejection. Unfortunately, reducing the type I error impacts the statistical power of the tests. Multiple testing is less of a problem if studies limit the number of comparisons. Sharply focusing research questions on few constituents reduces the chance of finding unexpected observations, and this is one of the best ways to address the multiple comparisons problem. Therefore, the multiple comparisons testing strategy should be first based on a process that reduces and prioritizes parameters for comparison.

CONFLICT OF INTEREST

The authors are employees of Imperial Tobacco Limited.

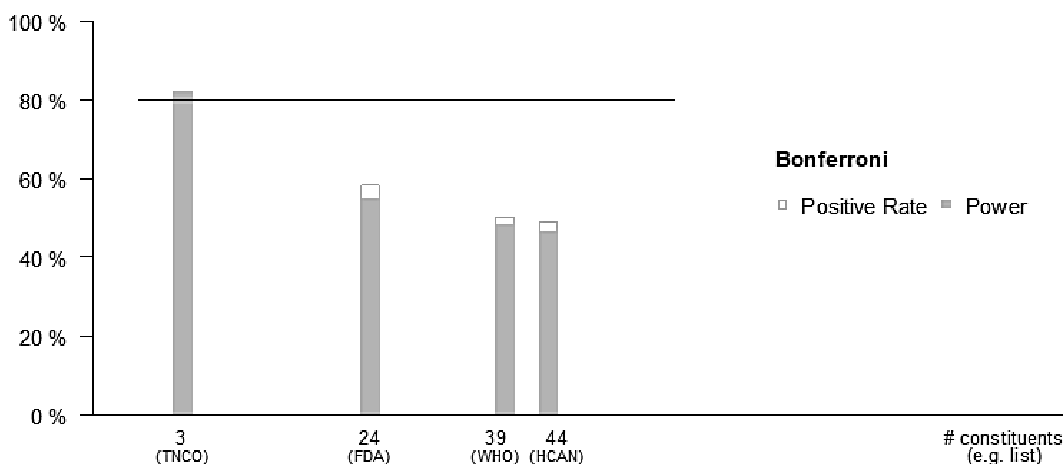


**Figure 3. Probability to detect a difference between two products significantly differentiated on one constituent (detectable in 80% of chance) versus the number of constituents using critical difference corrected using Bonferroni method**



Unauthenticated
Download Date | 3/16/20 2:40 PM

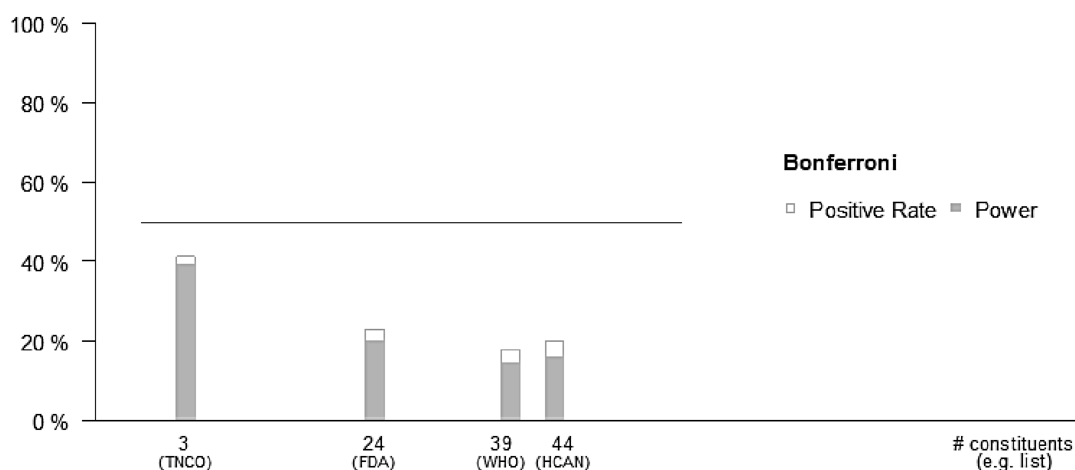


**Figure 4. Probability to detect a difference between two products moderately differentiated on one constituent (detectable in 50% of chance) *versus* the number of smoke constituents using critical difference corrected using Bonferroni method.**

REFERENCES


1. Health Canada: Tobacco Reporting Regulations SOR/2000-273; 2000. Available at: http://laws-lois.justice.gc.ca/eng/regulations/SOR-2000-273/ (accessed April 2017).
2. Venezuela Ministry of Health & Social Development: Regulation and Control of Cigarette & Other Derivative Tobacco for Human Consumption; 2004.
3. Brazilian National Health Surveillance Agency: RDC No. 90 and its Amendment; 2007. Available at: http://www.tobaccocontrollaws.org/legislation/country/brazil/laws (accessed April 2017).
4. Taiwan Department of Health: Regulations Governing Reporting of Tobacco Product Information under the Tobacco Hazards Prevention & Control Act (THPCA); 2007.
5. FDA: Draft Guidance; Reporting Harmful and Potentially Harmful Constituents in Tobacco Products and Tobacco Smoke Under Section 904(a)(3) of the Federal Food, Drug and Cosmetic Act; 2012. Available at: http://www.fda.gov/downloads/TobaccoProducts/GuidanceComplianceRegulatoryInformation/UCM297828.pdf (accessed April 2017).
6. World Health Organization (WHO): The Scientific Basis of Tobacco Product Regulation, Second Report of a WHO Study Group (TobReg); WHO technical report series No. 951, WHO Press, Geneva, Switzerland, 2008. Available at: http://www.who.int/tobacco/global_interaction/tobreg/publications/9789241209519.pdf (accessed April 2017).
7. World Health Organization - WHO Study Group on Tobacco Product Regulation: The Scientific Basis of Tobacco Product Regulation, Fifth Report of a WHO Study Group (TobReg); WHO Technical report series; 989, WHO Press, Geneva, Switzerland, 2015. Available at: http://apps.who.int/iris/bitstream/10665/161512/1/9789241209892.pdf?ua=1&ua=1 (accessed April 2017).
8. Purkis, S.: Letter to the Editor: Analysis of the Data Variability in the Australian Benchmark Study 2000 - 2001; 2009. Available at https://www.degruyter.com/view/j/cttr.2009.23.issue-5/cttr-2013-0867/cttr-2013-0867.xml (accessed April 2017). DOI: 10.2478/cttr-2013-0867
9. Hyodo, T., K. Minagawa, T. Inoue, J. Fujimoto, N. Minami, R. Bito, and A. Mikita: Estimation of Mouth Level Exposure to Smoke Constituents of Cigarettes with Different Tar Levels Using Filter Analysis; Regul. Toxicol. Pharmacol. 67 (2013) 486–498. DOI: 10.1016/j.yrtph.2013.09.009
10. Purkis, S.W., M. Meger, and R. Wuttke: A Review of Current Smoke Constituent Measurement Activities and Aspects of Yield Variability; Regul. Toxicol. Pharmacol. 62 (2012) 202–213. DOI: 10.1016/j.yrtph.2011.10.006
11. Eldridge, A., T.R. Betson, M.V. Gama, and K. McAdam: Variation in Tobacco and Mainstream Smoke Toxicant Yields from Selected Commercial Cigarette Products; Regul. Toxicol. Pharmacol. 71 (2015) 409–427. DOI: 10.1016/j.yrtph.2015.01.006
12. Belushkin, M., G. Jaccard, and A. Kondylis: Considerations for Comparative Tobacco Product Assessments Based on Smoke Constituent Yields; Regul. Toxicol. Pharmacol. 73 (2015) 105–113.
13. International Organization for Standardization (ISO): ISO 5725-6:1994/Cor 1:2001 - Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 6: Use in Practice of Accuracy Values – Technical Corrigendum 1; ISO, Geneva, Switzerland, 2001.
14. Teillet, B., X. Cahours, T. Verron, S. Colard, and S.W. Purkis: Comparison of Smoke Yield Data Collected from Different Laboratories; Beitr.Tabakforsch. Int. 25 (2013) 662–670. DOI: 10.2478/cttr-2013-0943
15. Bland, J. M. and D. G. Altman: Multiple Significance Tests: The Bonferroni Method; Brit. Med. J. 310 (1995) 170. DOI: 10.1136/bmj.310.6973.170
16. Holm, S.: A Simple Sequentially Rejective Multiple Test Procedure; Scand. J. Statist. 6 (1979) 65– 70.
17. Levin, B.: On the Holm, Simes, and Hochberg Multiple




Unauthenticated
Download Date | 3/16/20 2:40 PM

Test Procedures; Am. J. Public Health 86 (1996) 628–629.
18. Aickin, M. and H. Gensler: Adjusting for Multiple Testing When Reporting Research Results: The Bonferroni *vs*. Holm Methods; Am. J. Public Health 86 (1996) 726–728.
19. Benjamini, Y. and Y. Hochberg: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing; J. Royal Statist. Soc. B 57 (1995) 289–300.
20. Benjamini, Y. and D. Yekutieli: The Control of the False Discovery Rate in Multiple Testing Under Dependency; Ann. Statist. 29 (2001) 1165–1188.

*Corresponding author:*

*Thomas Verron*
*SEITA*
*Imperial Tobacco Limited*
*48, Rue Danton*
*45404 Fleury-les-Aubrais*
*France*
*E-mail: thomas.verron@fr.imptob.com*



Unauthenticated
Download Date | 3/16/20 2:40 PM

**Annex 1. Raw data obtained in two laboratories on the three commercial products A, B, and C with the list of published standards for the repeatability and reproducibility.**

| Constituent | Unit | Method | Product A | | Product B | | Product C | |
|---|---|---|---|---|---|---|---|---|
| | | | Lab1 | Lab2 | Lab1 | Lab2 | Lab1 | Lab2 |
| "TAR" | mg/cig | ISO 4387 (2000) | 1.8 | 1.7 | 7.5 | 7.5 | 10.2 | 9.8 |
| Carbon monoxide | mg/cig | ISO 8454 (2007) | 2.2 | 2 | 8.8 | 8.6 | 11.3 | 10.5 |
| Nicotine | mg/cig | ISO 10315 (2000) | 0.18 | 0.19 | 0.55 | 0.56 | 0.81 | 0.81 |
| Benzo[*a*]pyrene | ng/cig | CRM 58 (2013) | 2.6 | 3.2 | 5 | 5.6 | 7.9 | 5.3 |
| 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) | ng/cig | CRM 75 (2012) | 11 | 11.6 | 36.7 | 32.5 | 33.2 | 36.8 |
| *N*-Nitrosonornicotine (NNN) | ng/cig | CRM 75 (2012) | 23.1 | 30.9 | 72.8 | 71.3 | 100 | 77.2 |
| Acetaldehyde | µg/cig | CRM 74 (2013) | 124 | 104 | 458 | 376 | 716 | 552 |
| Acrolein | µg/cig | CRM 74 (2013) | 8.27 | 7.3 | 36.5 | 40.2 | 65 | 59.3 |
| Crotonaldehyde | µg/cig | CRM 74 (2013) | 0.409 | 2 | 7.61 | 9.9 | 13.3 | 15.3 |
| Formaldehyde | µg/cig | CRM 74 (2013) | 2.11 | 3 | 19.7 | 18.6 | 35.8 | 26.1 |
| Acrylonitrile | µg/cig | CRM 70 (2010) | 1.91 | 2.2 | 6.62 | 10.3 | 9.65 | 15.5 |
| Benzene | µg/cig | CRM 70 (2010) | 13 | 8.8 | 35.3 | 31.4 | 44.7 | 42.2 |
| 1,3-Butadiene | µg/cig | CRM 70 (2010) | 10.2 | 13.2 | 40.2 | 45 | 55.6 | 60.9 |
| Isoprene | µg/cig | CRM 70 (2010) | 98.2 | 66.2 | 327 | 192.8 | 345 | 256.2 |
| Toluene | µg/cig | CRM 70 (2010) | 19.7 | 11.6 | 54.4 | 48.2 | 66.5 | 67.4 |

**Annex 2. Critical Differences (CD) obtained for the three commercial products A, B, and C based on the repeatability and reproducibility mentioned in published standards.**

| Constituent | Unit | Method | CD Product A | | CD Product B | | CD Product C | |
|---|---|---|---|---|---|---|---|---|
| | | | Correction | | Correction | | Correction | |
| | | | No | Bonferroni | No | Bonferroni | No | Bonferroni |
| "TAR" | mg/cig | ISO 4387 (2000) | 0.96 | 1.41 | 1.29 | 1.89 | 1.48 | 217 |
| Carbon monoxide | mg/cig | ISO 8454 (2007) | 96 | 1.40 | 1.68 | 2.46 | 1.97 | 289 |
| Nicotine | mg/cig | ISO 10315 (2000) | 0.07 | 11 | 0.11 | 0.16 | 0.13 | 19 |
| Benzo[*a*]pyrene | ng/cig | CRM 58 (2013) | 1.50 | 219 | 217 | 3.19 | 2.56 | 375 |
| 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) | ng/cig | CRM 75 (2012) | 4.97 | 7.30 | 1091 | 1602 | 11.01 | 1616 |
| *N*-Nitrosonornicotine (NNN) | ng/cig | CRM 75 (2012) | 10.15 | 14.90 | 19.82 | 2909 | 23.59 | 3463 |
| Acetaldehyde | µg/cig | CRM 74 (2013) | 76.47 | 112.23 | 134.33 | 19714 | 184.91 | 27137 |
| Acrolein | µg/cig | CRM 74 (2013) | 8.07 | 11.85 | 14.92 | 2189 | 21.57 | 3165 |
| Crotonaldehyde | µg/cig | CRM 74 (2013) | 1.50 | 2.21 | 5.04 | 740 | 7.79 | 1143 |
| Formaldehyde | µg/cig | CRM 74 (2013) | 4.00 | 5.87 | 12.23 | 17.95 | 1886 | 2767 |
| Acrylonitrile | µg/cig | CRM 70 (2010) | 1.73 | 2.54 | 4.30 | 6.31 | 610 | 895 |
| Benzene | µg/cig | CRM 70 (2010) | 7.20 | 10.56 | 15.08 | 22.14 | 1877 | 2754 |
| 1,3-Butadiene | µg/cig | CRM 70 (2010) | 6.75 | 9.90 | 22.11 | 32.46 | 3091 | 4536 |
| Isoprene | µg/cig | CRM 70 (2010) | 79.23 | 116.28 | 145.78 | 213.95 | 16072 | 23587 |
| Toluene | µg/cig | CRM 70 (2010) | 1191 | 1747 | 2418 | 3548 | 3010 | 4417 |



Unauthenticated
Download Date | 3/16/20 2:40 PM