Product comparison: the risk associated with multiple testing

Thomas Verron[1], Xavier Cahours[1], Stéphane Colard[1,2]

[1]*SEITA, Imperial Tobacco Group - France*
[2]*Imperial Tobacco Limited - U.K.*

Context

- Manufacturers are increasingly being asked by regulatory authorities to report data on their products

- The objective of these requests is to use data for
  - Public communication
  - Product comparison
  - Introduction of smoke yield ceilings

Context

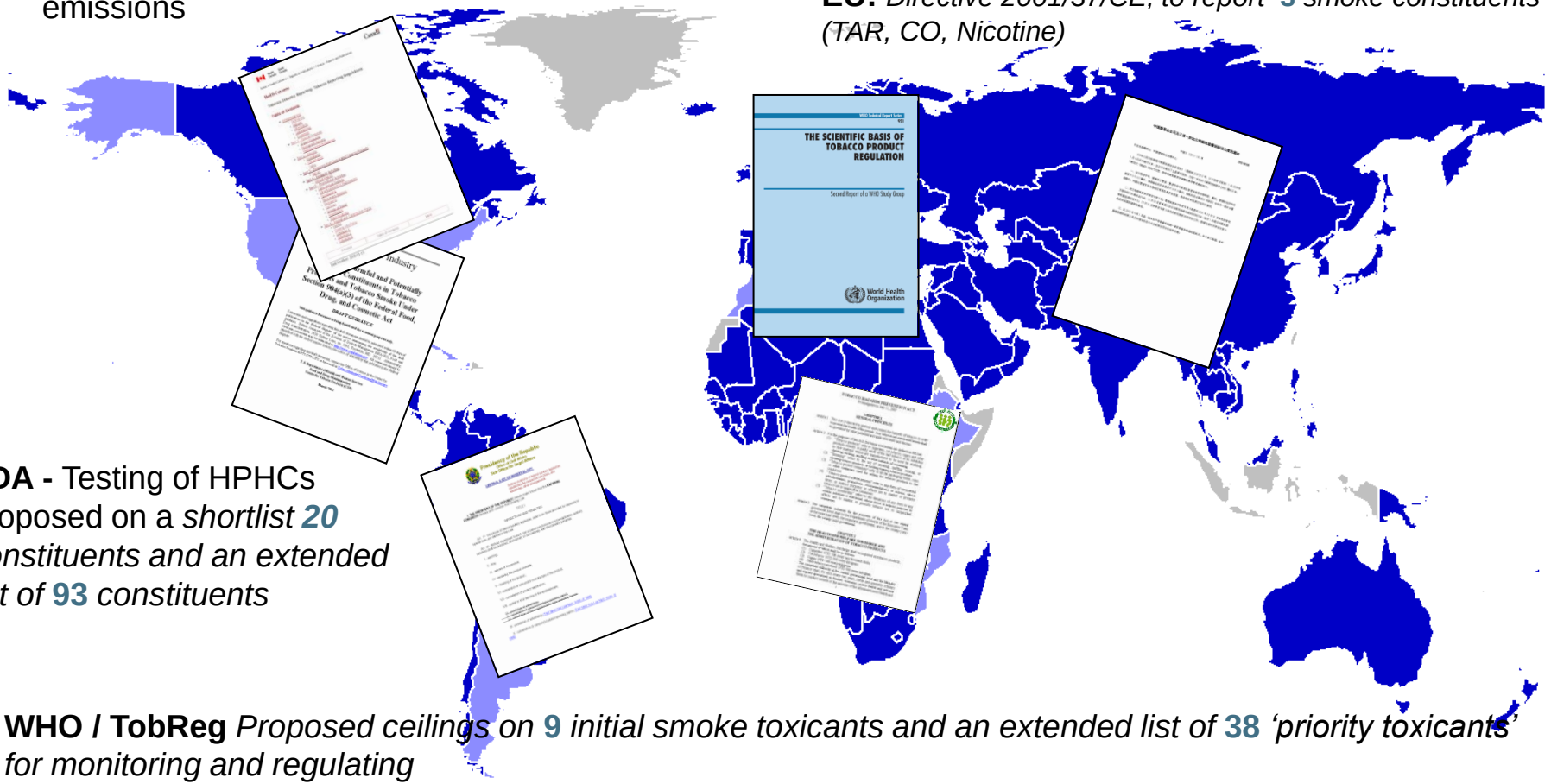
- Manufacturers are increasingly being asked by regulatory authorities to report data on their products

- The objective of these requests is to use data for
  - Public communication
  - Product comparison
  - Introduction of smoke yield ceilings

Health Canada - Testing of **44** smoke emissions

Other Regulatory Authorities
e.g. in Taiwan, Brazil, China, etc –

EU: *Directive 2001/37/CE, to report* **3** *smoke constituents (TAR, CO, Nicotine)*



FDA - Testing of HPHCs proposed on a *shortlist* ***20*** *constituents and an extended list of* **93** *constituents*

WHO / TobReg *Proposed ceilings on* **9** *initial smoke toxicants and an extended list of* **38** *'priority toxicants' for monitoring and regulating*

Product comparison

Context:

Product comparison

Objective:

Make a decision

Product comparison

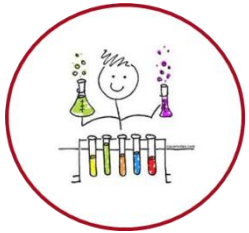
Context:

Product comparison

2013, we showed*:

ISO, CRM…

validated and standardised methods



Objective:

Make a decision

* B. Teillet, X. Cahours, T. Verron, S. Colard, S. Purkis. Comparison of Smoke Yield Data Collected from Different Laboratories. Beitr. Tabakforsch. Int. 25 (2013) 663-670.

Product comparison

Context:

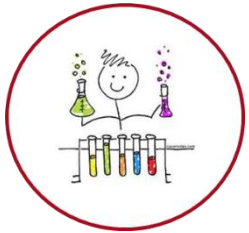
Product comparison

2013, we showed*:

ISO, CRM…

ISO 5725

validated and standardised methods



Uncertainties

Objective:

Make a decision

* B. Teillet, X. Cahours, T. Verron, S. Colard, S. Purkis. Comparison of Smoke Yield Data Collected from Different Laboratories. Beitr. Tabakforsch. Int. 25 (2013) 663-670.

Product comparison

Context:

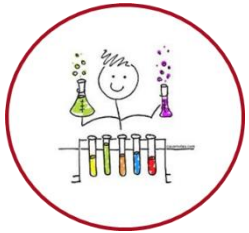
Product comparison

2013, we showed*:

Objective:
Make a decision

ISO, CRM…

validated and standardised methods



ISO 5725

Uncertainties

ISO 5725

Critical Difference

the best way to compare data from <u>different laboratories</u>

* B. Teillet, X. Cahours, T. Verron, S. Colard, S. Purkis. Comparison of Smoke Yield Data Collected from Different Laboratories. Beitr. Tabakforsch. Int. 25 (2013) 663-670.

Critical Difference (CD)

Maximum difference expected between two final values with a specified probability

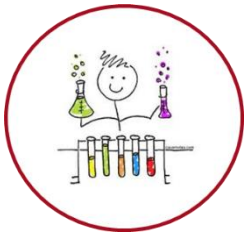
$$|\overline{X_A} - \overline{X_B}| \leq \quad CD = \sqrt{\left(2\sqrt{2}\sigma\right)^2 - \left(2\sqrt{2}\sigma_r\right)^2 \left(1 - \frac{1}{2n_A} - \frac{1}{2n_B}\right)}$$

Critical Difference (CD)

Maximum difference expected between two final values with a specified probability

Notion of risk

$$\underbrace{|\overline{X_A} - \overline{X_B}|}_{\text{Measurements}} \leq CD = \sqrt{\left(2\sqrt{2}\sigma\right)^2 - \left(2\sqrt{2}\sigma_r\right)^2 \left(1 - \frac{1}{2n_A} - \frac{1}{2n_B}\right)}$$



Take into account uncertainties

Critical Difference (CD)

Maximum difference expected between two final values with a specified probability

$$|\overline{X_A} - \overline{X_B}| \leq \quad CD = \sqrt{\left(2\sqrt{2}\sigma\right)^2 - \left(2\sqrt{2}\sigma_r\right)^2 \left(1 - \frac{1}{2n_A} - \frac{1}{2n_B}\right)}$$

Number of *replicates*

Critical Difference (CD)

Maximum difference expected between two final values with a specified probability

Standard deviation of repeatability

$$|\overline{X_A} - \overline{X_B}| \leq \quad CD = \sqrt{\left(2\sqrt{2}\sigma\right)^2 - \left(2\sqrt{2}\sigma_r\right)^2 \left(1 - \frac{1}{2n_A} - \frac{1}{2n_B}\right)}$$

Number of *replicates*

Critical Difference (CD)

Maximum difference expected between two final values with a specified probability

$$|\overline{X_A} - \overline{X_B}| \leq \quad CD = \sqrt{(2\sqrt{2}\sigma)^2 - (2\sqrt{2}\sigma_r)^2 \left(1 - \frac{1}{2n_A} - \frac{1}{2n_B}\right)}$$

Standard deviation of repeatability

Standard deviation

Number of *replicates*

repeatability

Intermediate precision

Reproducibility

Measured in the same lab (short period of time)		Measured in the same lab (long period of time*)		Measured in two labs*
$\sigma = \sigma_r$	$\leq$	$\sigma = \sigma_{IP}$	$\leq$	$\sigma = \sigma_R$

** Assumed that the same method and protocol are used*

Critical Difference (CD)

Maximum difference expected between two final values with a specified probability

Risk to have a false positive: 5%

*Risk to conclude that the yields of the two products are **not equivalent** whereas actually they are.*

Standard deviation of repeatability

$$|\overline{X_A} - \overline{X_B}| \leq \quad CD = \sqrt{\left(2\sqrt{2}\sigma\right)^2 - \left(2\sqrt{2}\sigma_r\right)^2 \left(1 - \frac{1}{2n_A} - \frac{1}{2n_B}\right)}$$

Standard deviation

Number of *replicates*

repeatability

Intermediate precision

Reproducibility

Measured in the same lab (short period of time)		Measured in the same lab (long period of time*)		Measured in two labs*
$\sigma = \sigma_r$	$\leq$	$\sigma = \sigma_{IP}$	$\leq$	$\sigma = \sigma_R$

** Assumed that the same method and protocol are used*

Product comparison

Two products are equivalent for several smoke analytes, if for each smoke analyte there are not significant differences

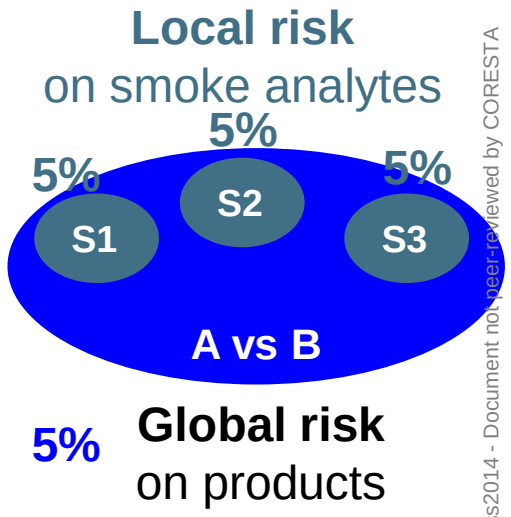
CD controls **"Local"** risk of concluding **for one smoke analyte** that the two products are not equivalent whereas actually they are.

Product comparison

Two products are equivalent for several smoke analytes, if for each smoke analyte there are not significant differences

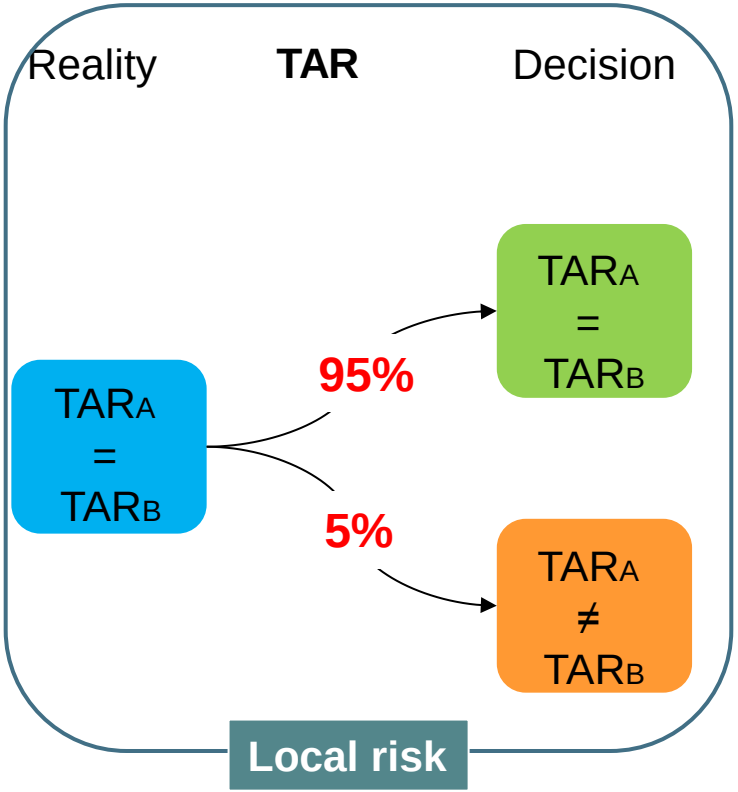
CD controls **"Local"** risk of concluding **for one smoke analyte** that the two products are not equivalent whereas actually they are.

We want to control **"Global"** risk of concluding when **considering all smoke analytes**, that the two products are not equivalent whereas actually they are.



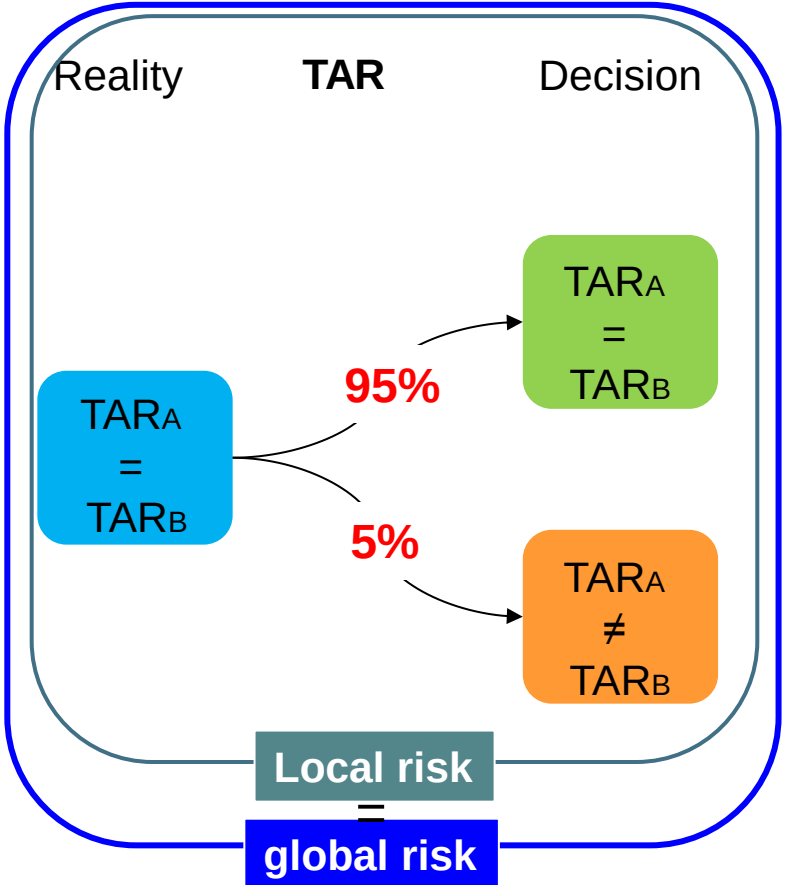
Product comparison

Let A and B be two equivalent products for **one smoke analyte i.e.TAR**

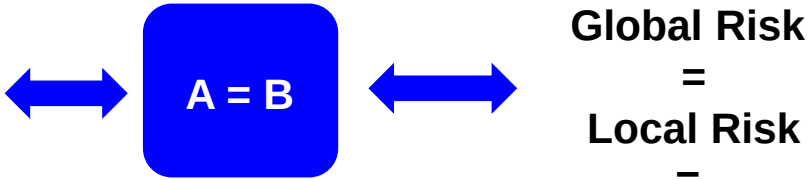


Product comparison

Let A and B be two equivalent products for **one smoke analyte i.e.TAR**



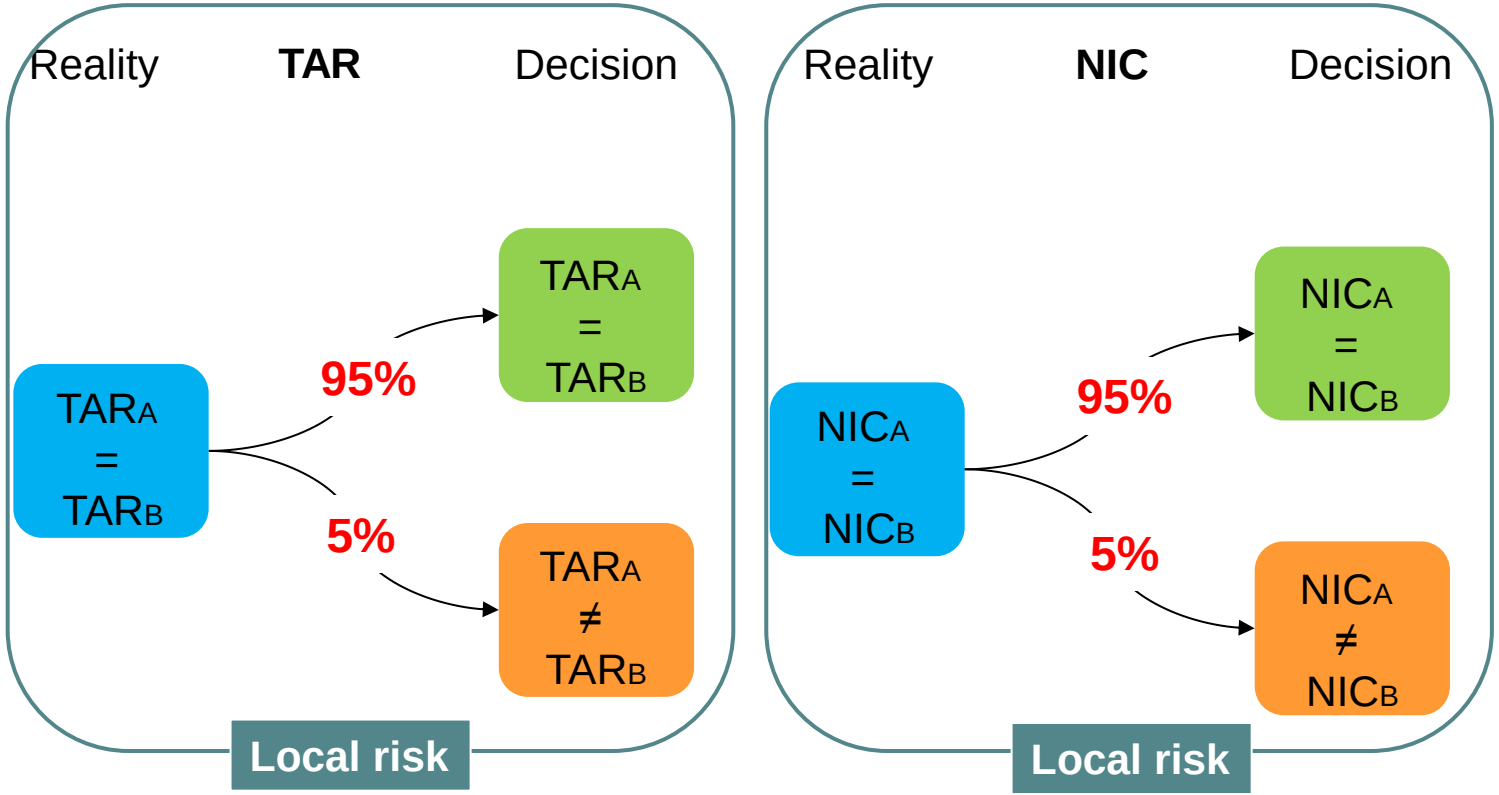
Product decision



Global Risk = Local Risk = 5%

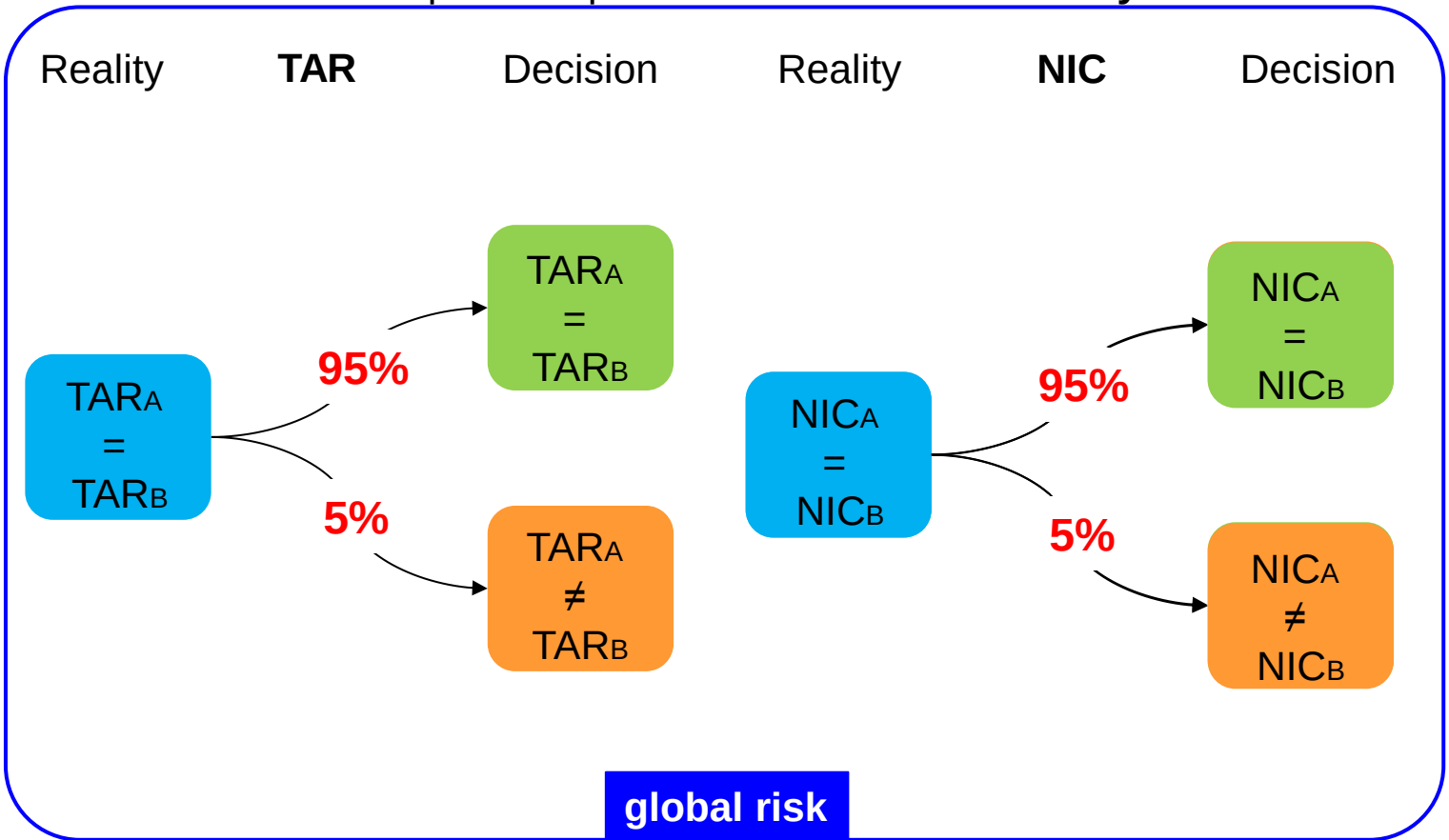
Product comparison

Let A and B be two equivalent products for **two smoke analytes i.e.TAR & NIC**



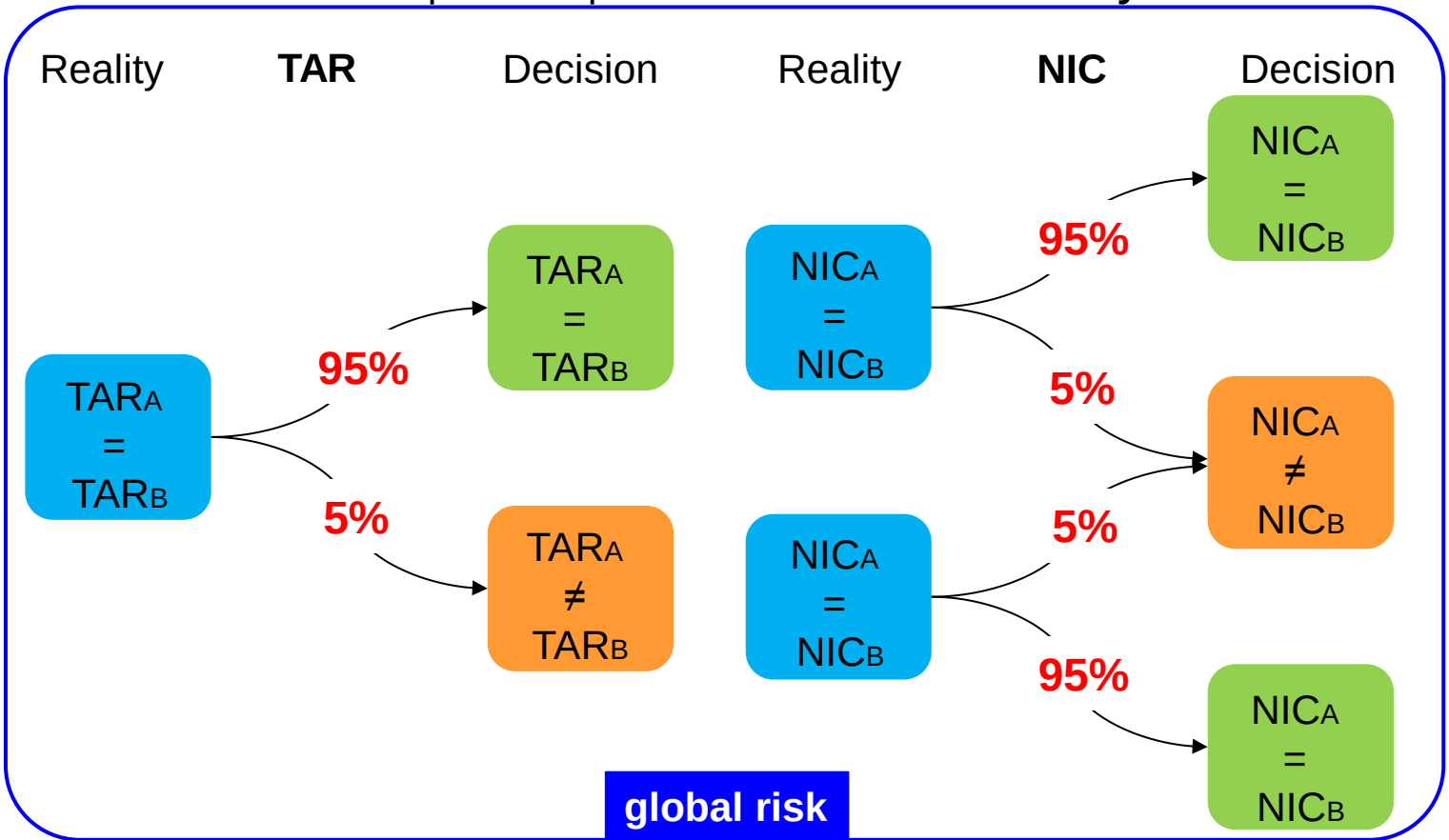
Product comparison

Let A and B be two equivalent products for **two smoke analytes i.e.TAR & NIC**



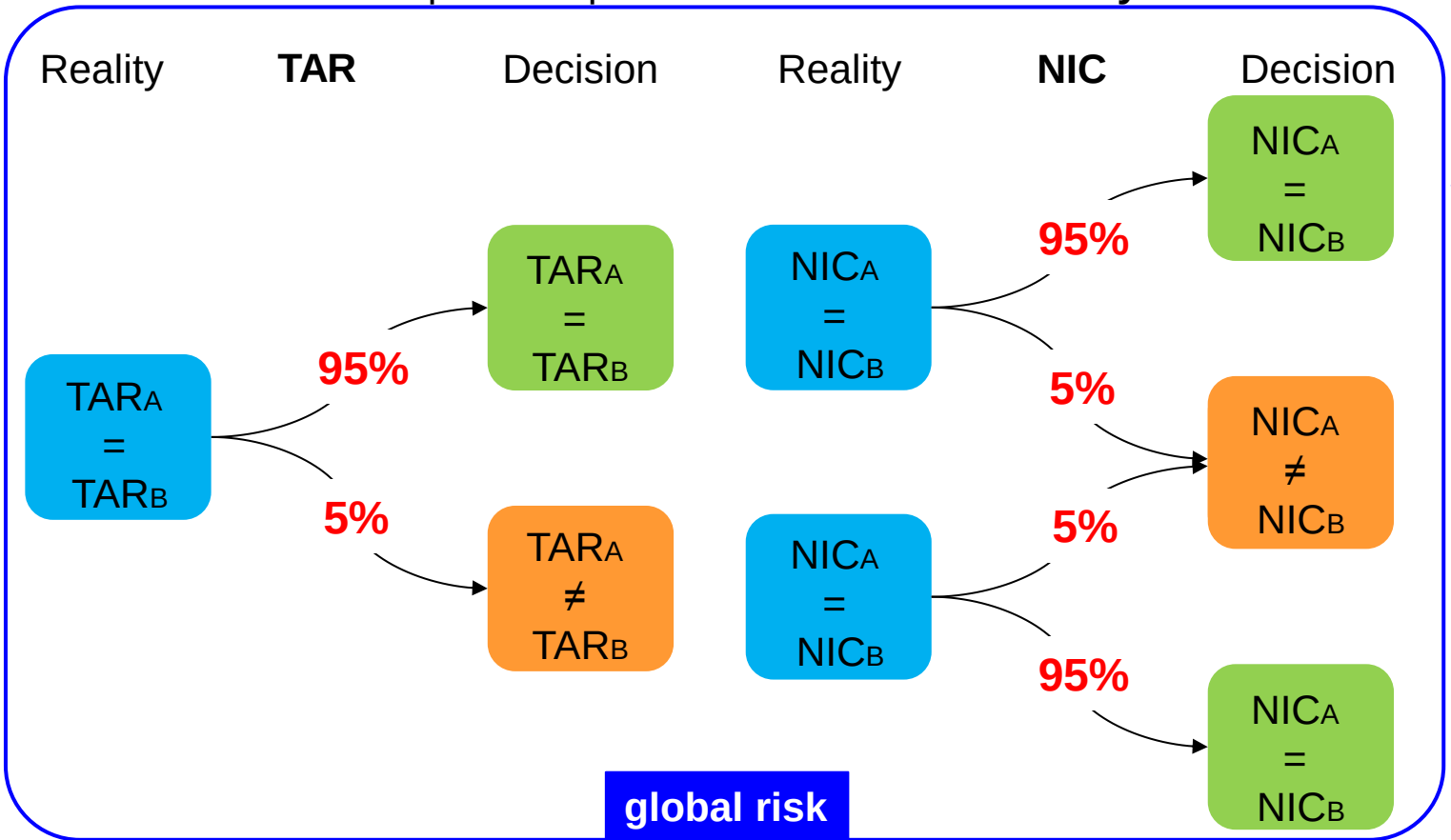
Product comparison

Let A and B be two equivalent products for **two smoke analytes i.e.TAR & NIC**



Product comparison

Let A and B be two equivalent products for **two smoke analytes i.e.TAR & NIC**



Product decision

A = B

Global Risk
=
1 - (.95 x .95)
=
9.75%

Assuming independence among smoke analytes

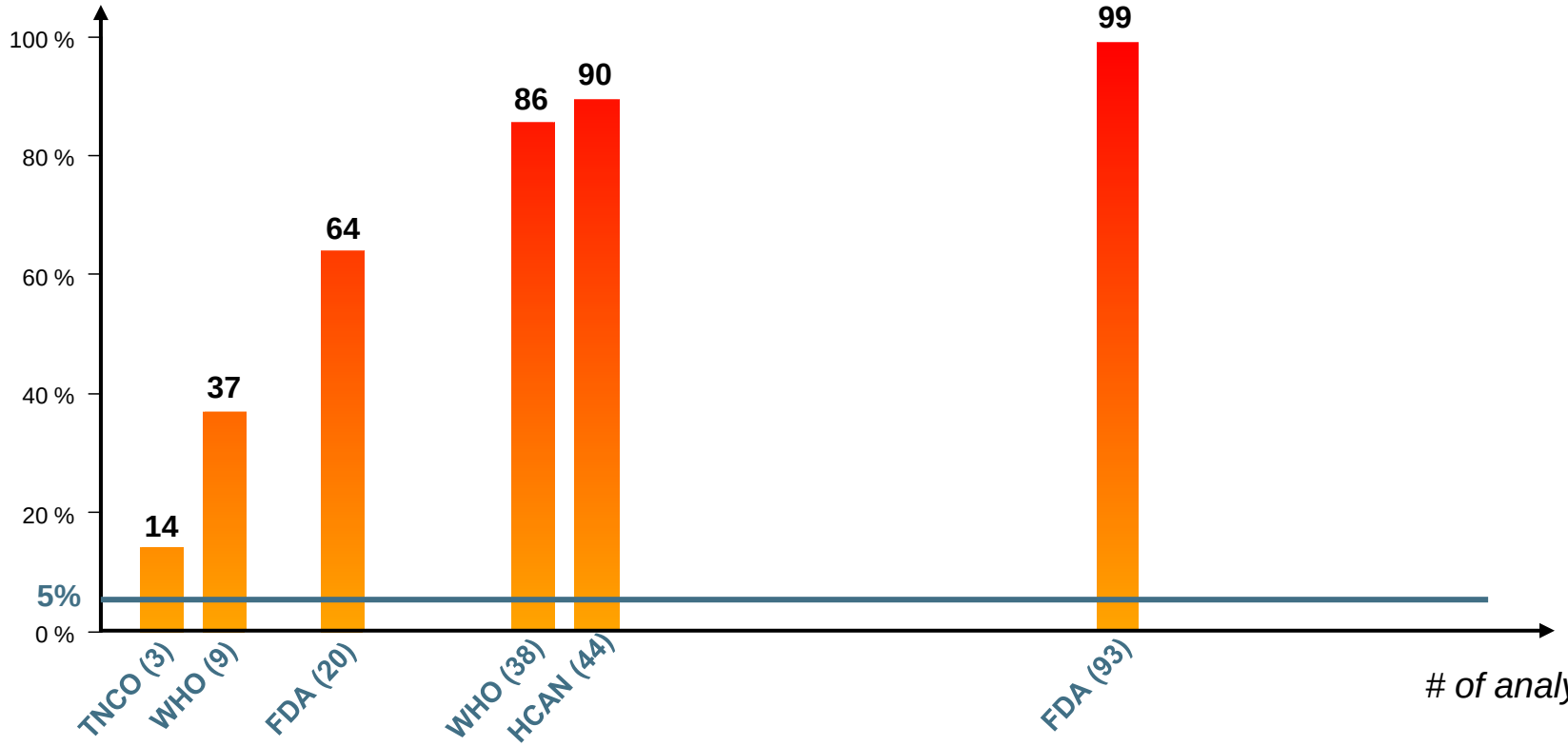
Product comparison vs number of analytes

*% of chance to make a wrong decision**

***wrong decision** = when it is concluded that two products are not equivalent when actually they are*

Product comparison vs number of analytes

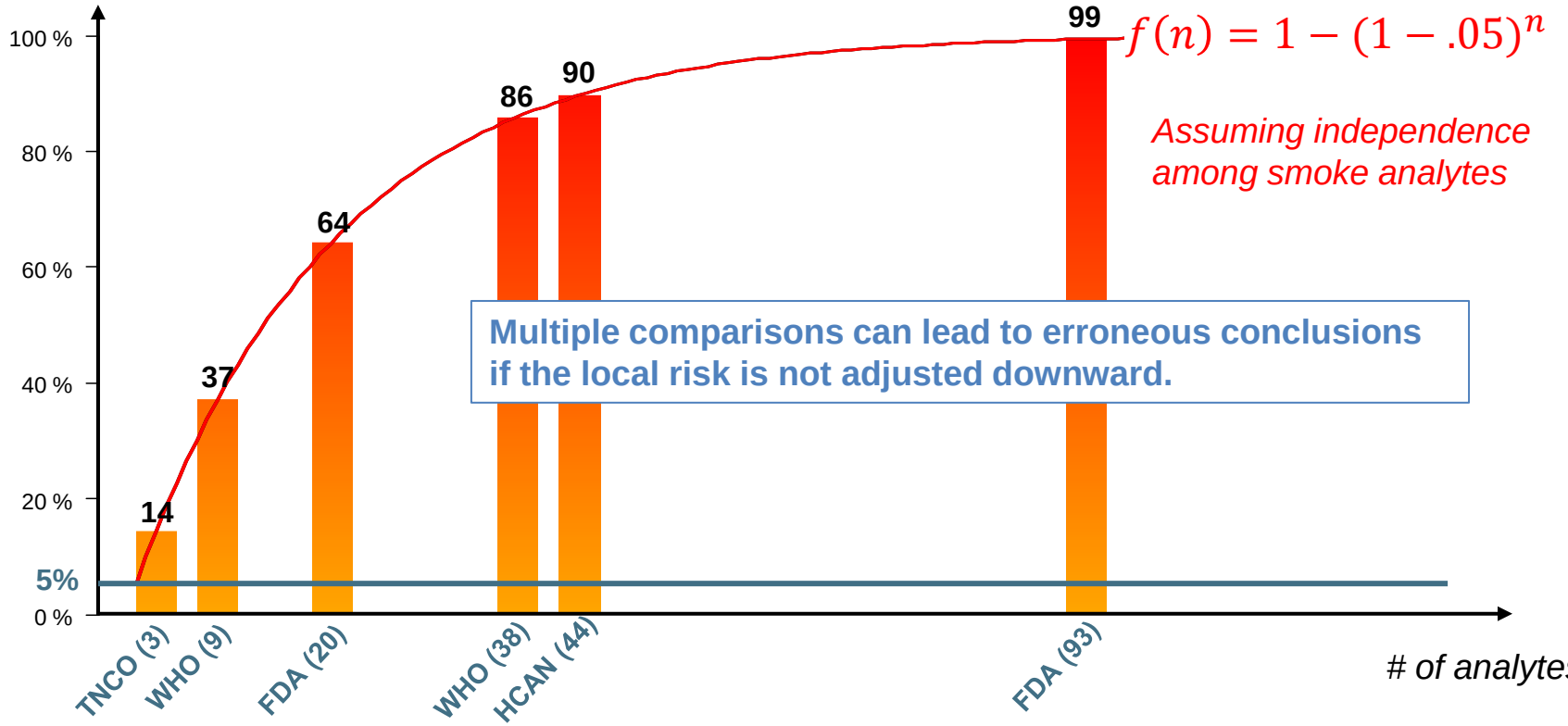
*% of chance to make a wrong decision**



***wrong decision** = *when it is concluded that two products are not equivalent when actually they are*

Product comparison vs number of analytes

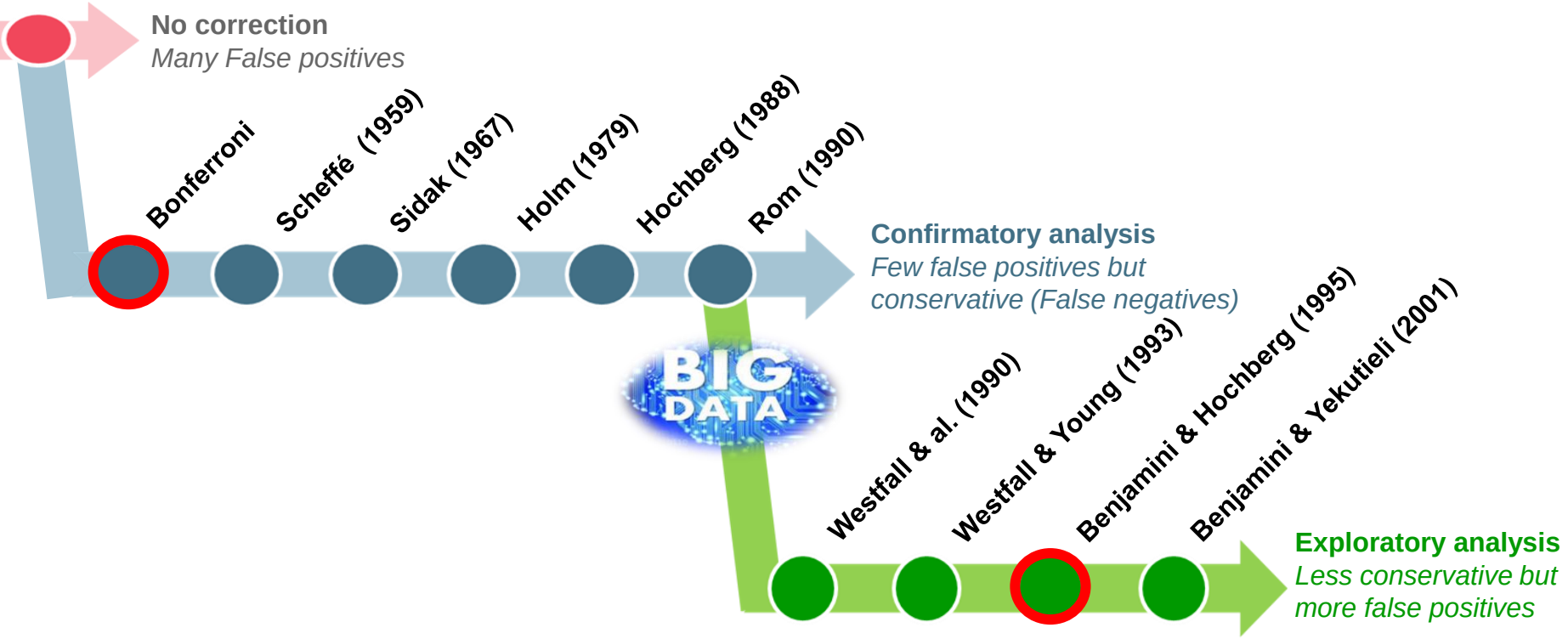
*% of chance to make a wrong decision**



***wrong decision** = *when it is concluded that two products are not equivalent when actually they are*

How to address the multiplicity problem

Objective:
Minimize the chance of obtaining falsely significant findings



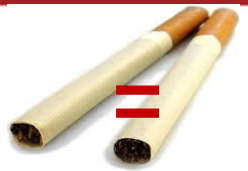
Two simulations

- Comparison of two equivalent products considering several analytes with and without correction

  % of wrong decision in function of number of analytes

- Comparison of two different products considering several analytes with corrections

  % to detect a difference in function of the number of analytes



Simulation 1 :
Two equivalent products

Method & product variability

Repeat 500X

A

B

No. analyte

1

3

9

20

44

93

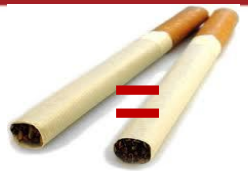
Comparison of A and B using

- No correction
- Bonferroni
- BH

Two products are equivalent?

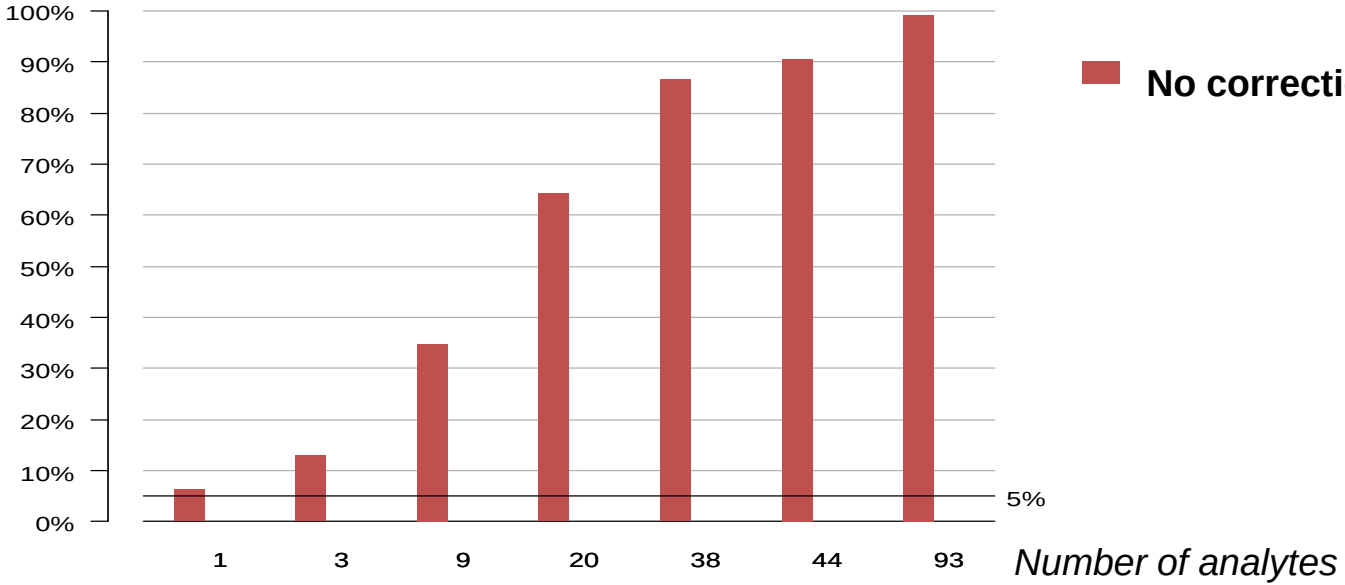
% of wrong decision* in function of number of analytes

***wrong decision** = *when it is concluded that two products are not equivalent when actually they are*

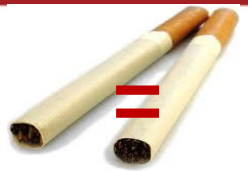


Simulation 1: two equivalent products

% of wrong decision in function of the number of analytes

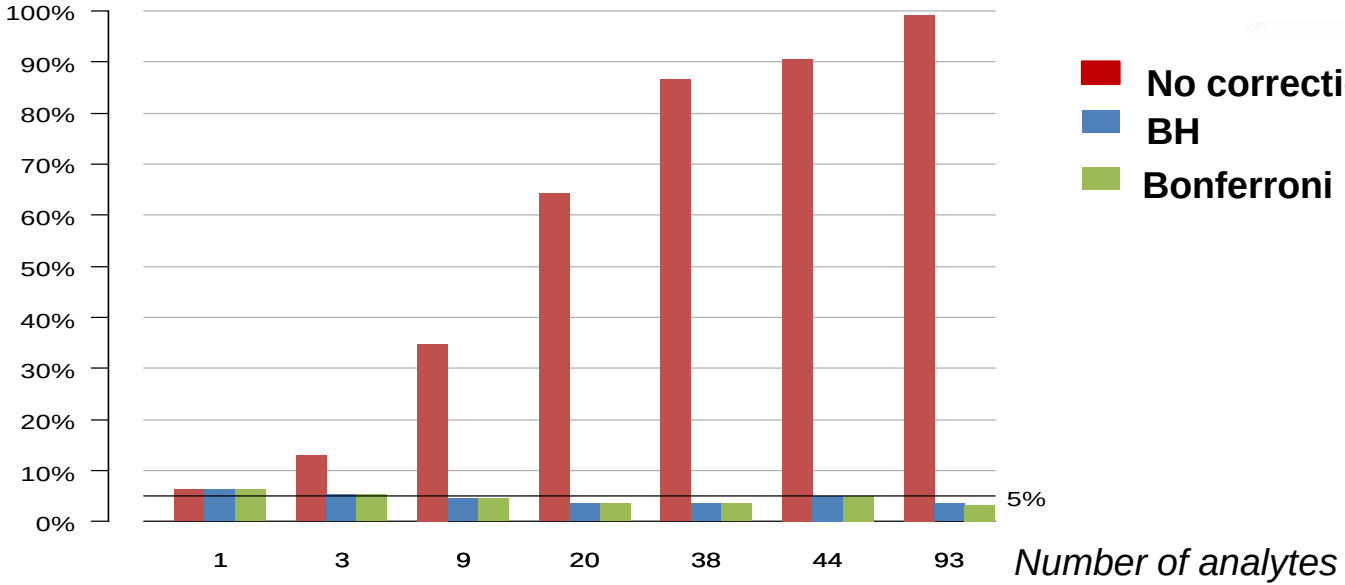


Without adjustment the risk to make false conclusion increase with the number of analytes



Simulation 1: two equivalent products

% of wrong decision in function of the number of analytes



Without adjustment the risk to make false conclusion increase with the number of analytes

Modification of the level of statistical significance minimises false positives

(risk to conclude that the two products are not equivalent whereas actually they are)

Bonferroni adjustment = **BH adjustment**

Simulation 2: Capacity to detect a difference between two different products on one analyte

Method & product variability

Repeat 500X

A

B

No. analyte

- 3
- 9
- 20
- 38
- 44
- 93

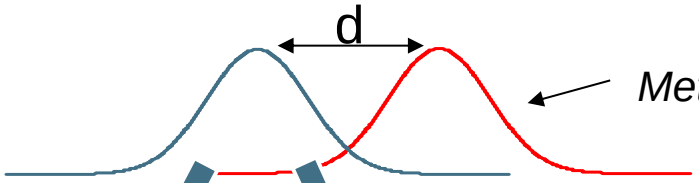
Comparison of A and B using

- Bonferroni
- BH

Two products are equivalent?

% to detect a difference (d) in function of the number of analytes

Simulation 2: Capacity to detect a difference between two different products on one analyte



Repeat 500X

A

B

No. analyte

3

9

20

38

44

93

Comparison of A and B using

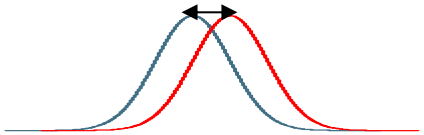
- Bonferroni
- BH

Two products are equivalent?

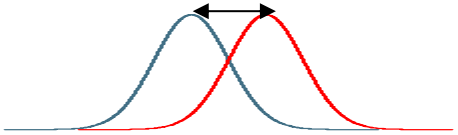
% to detect a difference (d) in function of the number of analytes

Simulation 2: Capacity to detect a difference between two different products on one analyte

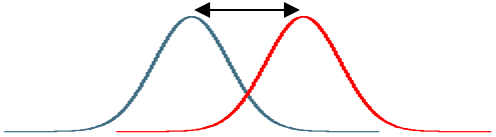
d = 2*s



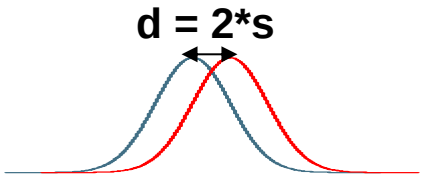
d = 3*s

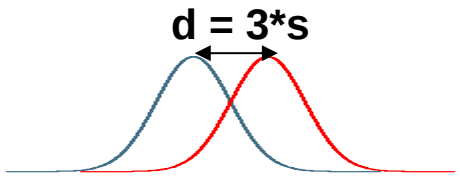


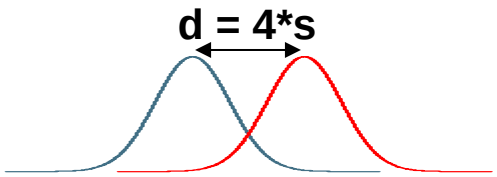
d = 4*s

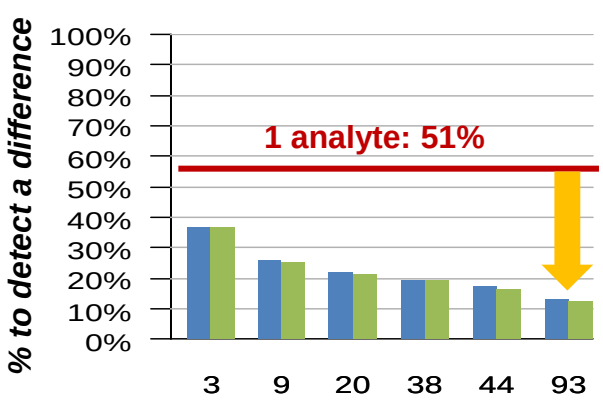


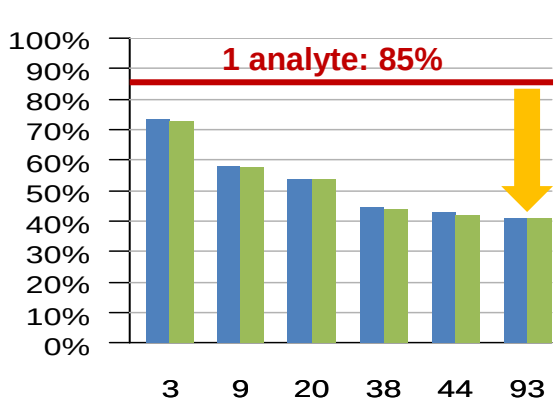
Simulation 2: Capacity to detect a difference between two different products on one analyte

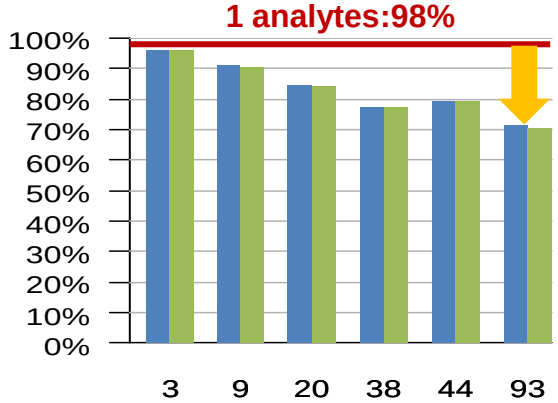












Sensitivity for detecting a difference between the two products decrease with the number of smoke analytes for a given effect size “d”.

The more analytes considered in the comparison, the more difficult to detect a difference between the two products

Bonferroni adjustment = BH adjustment

Conclusions

- **Critical difference (ISO 5725) is the best way** to consider the variability of measurements and laboratories in the comparison of two products for one analyte.
- In multiple testing the critical difference can be used but the **risk must be adjusted** in order to avoid false conclusion about equivalence.

Reduce the number of analytes and select carefully analytes to assure a good statistical power level in order to detect non-equivalent products.

Thank you for your attention

Some references:

- Bland JM, Altman DG. *Multiple significance tests: The Bonferroni method*. British Medical Journal 1995;310:170
- Benjamini Y, Hochberg Y. *Controlling the false discovery rate: A practical and powerful approach to multiple testing*. J R Stat Soc 1995;57:289-300
- Benjamini Y, Yekutieli D. *The control of the false discovery rate in multiple testing under dependency*. Annals of Statistics 2001;29:1165-88

www.imperialtobaccoscience.com