Test design for consumer & producer risk

How to save time & money and still approve equipment with confidence

David Anderson
European Commission, Joint Research Centre

(Presentation by video conference)

ADSA-20, 16 May 2019, Northeastern University, Boston.

European Commission

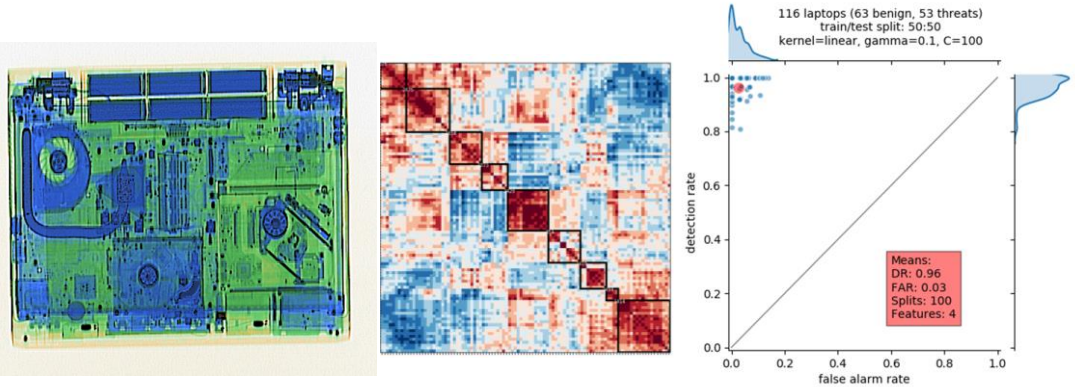
About the speaker

David Anderson is a scientist at the European Commission's Joint Research Centre (JRC). The European Commission is the civil service of the European Union (EU), and the JRC is its internal research arm.

Other current research interests:

Simulants for field testing explosive detection systems.

(2019 H2: available in 28 EU Member States)



AI algorithms for checkpoint X-ray images.

(e.g. sheets in laptops: DR>95% FAR<5%.

More info: David.Anderson@ec.europa.eu

European Commission

Testing equipment: "So what, who cares?"

- **Space**: Do you test detection equipment?
- **Challenge 1**: Can you answer this question:

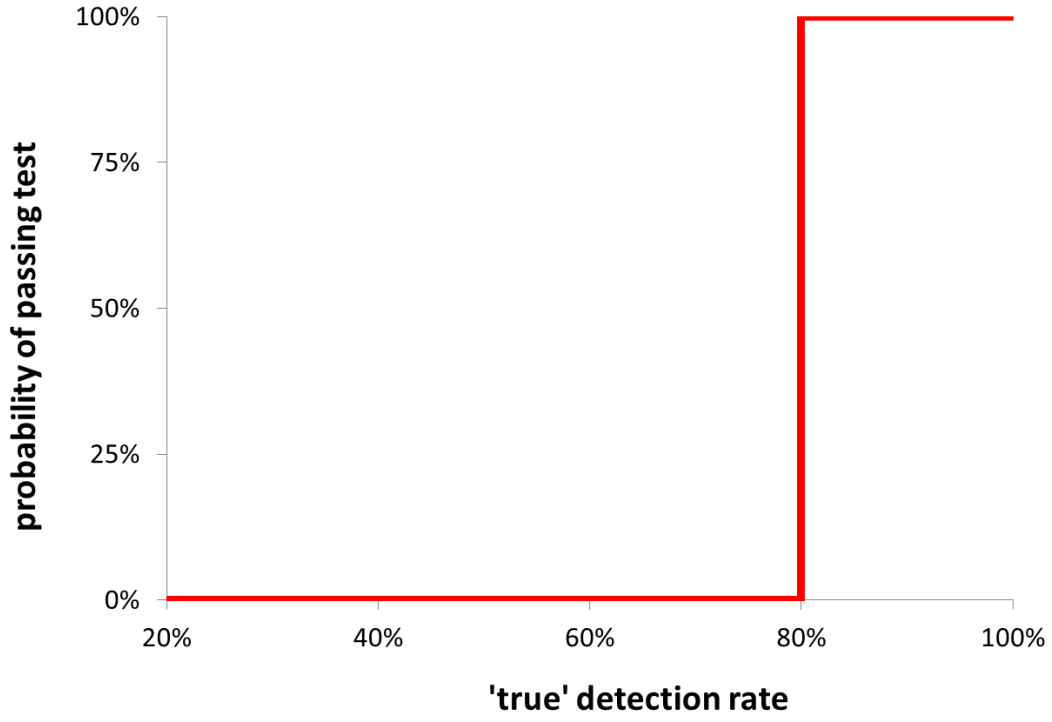
  "What is the risk that your test will lead to the wrong decision?"
- **Challenge 2**: Do you want to optimise the test duration (e.g. to save money, reduce disruption, or free-up more testing resources)…. and still preserve statistical confidence?
- **Solution**: This presentation explains how to do it …

  (and a free Excel software tool is available).

Testing: idealised vs. reality

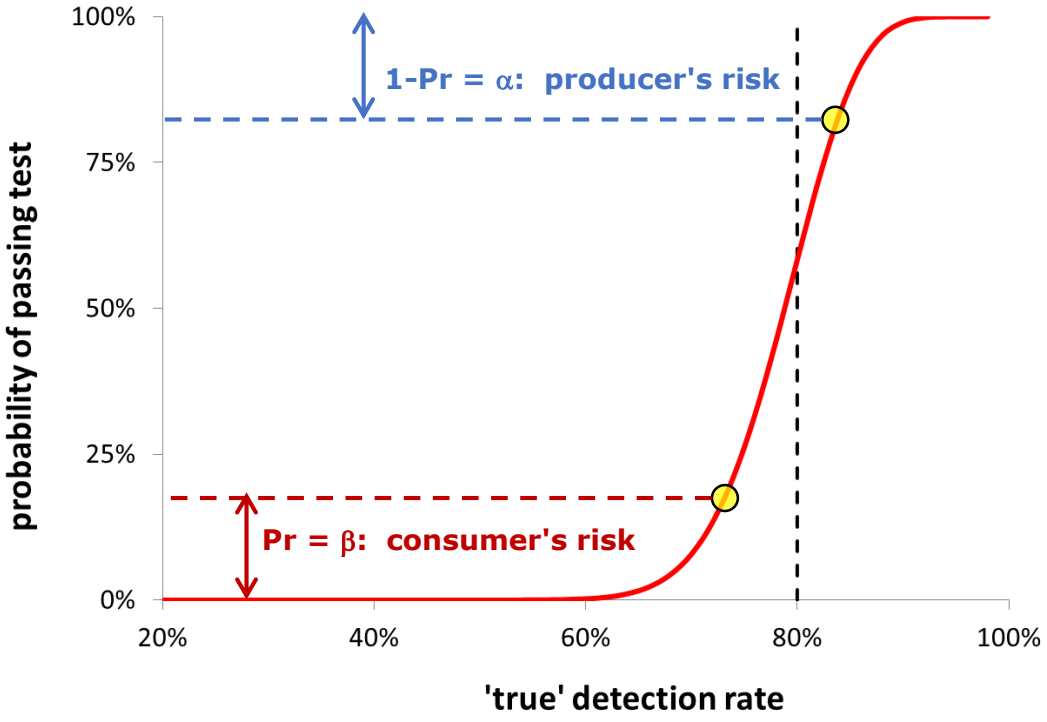
Idealised:

risks ignored



Reality:

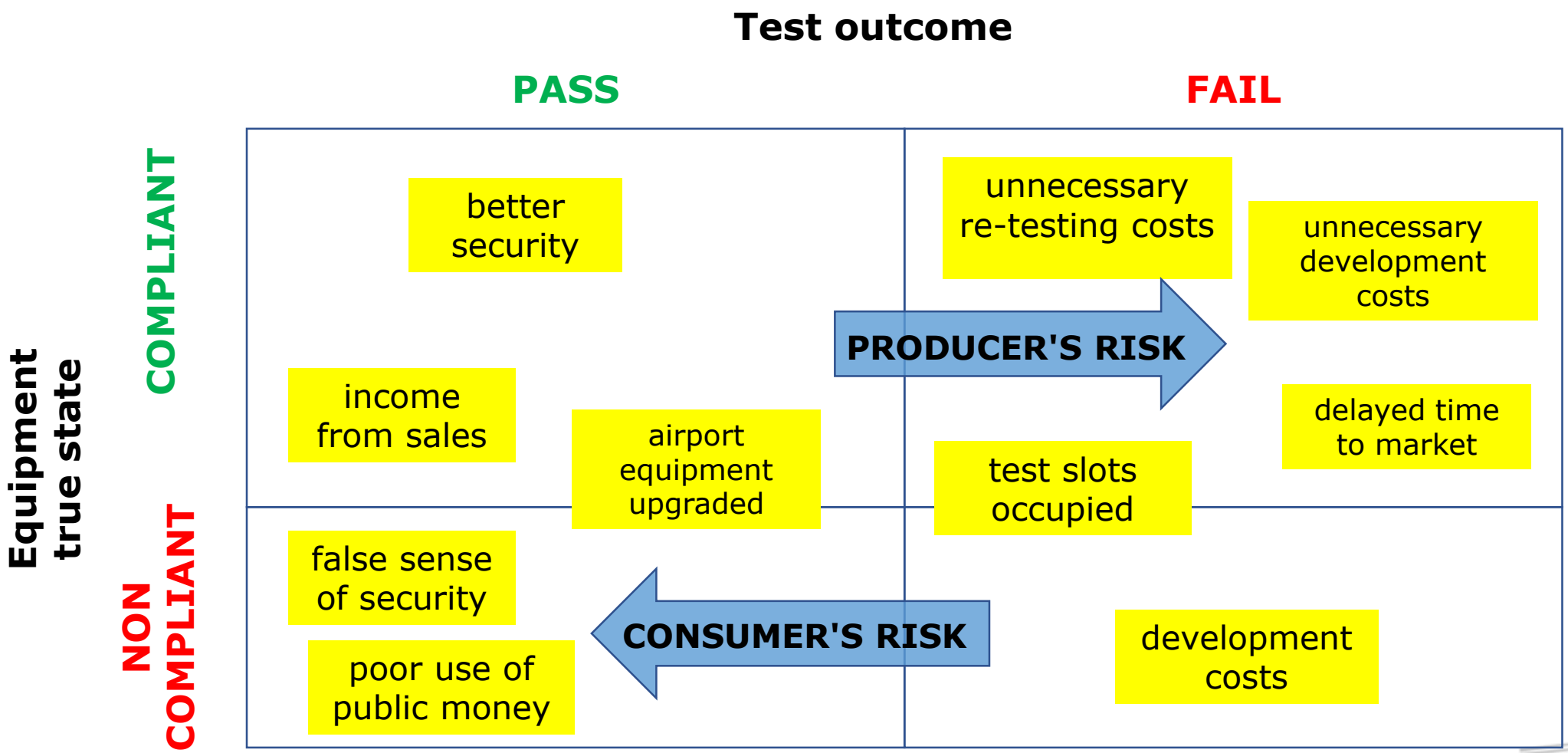
risks managed



Producer's risk: risk that compliant equipment is rejected
Consumer's risk: risk that non-compliant equipment is accepted

European Commission

The big picture: testing risk and economic impact



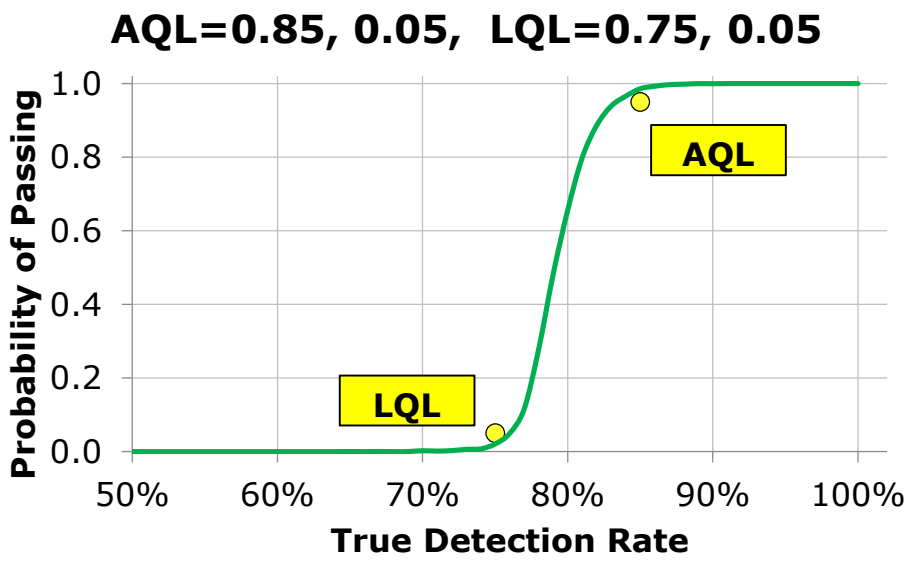
First things first!

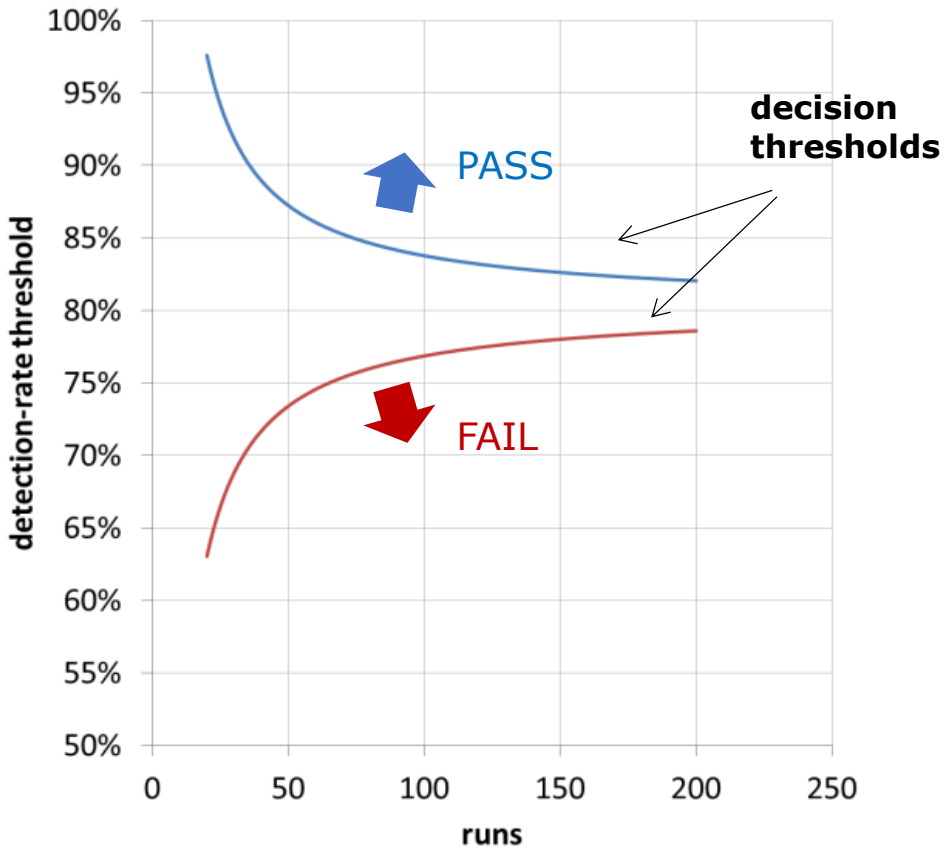
- Producer and consumer risks should **not** be an after-thought.
- On the contrary, **start** by defining the maximum tolerable risks, **then** work out the number of runs.
- Deciding the acceptable risk levels is primarily a **policy** and **economic** consideration.
- Risks can be defined using 'acceptance sampling' terminology:

Acceptable quality level (AQL): the rate at which there should be a very low probability (e.g. producer risk, α = 5%) of incorrectly rejecting.

Limit quality level (LQL): a lower rate, for which there should be a very low probability (e.g. consumer risk, β = 5%) of incorrectly accepting.

From risks to decision thresholds vs. runs





The math - useful resources:

- NIST Engineering Handbook, [6.2.6 What is a Sequential Sampling Plan?]
- JRC Technical Report, [Application of Sequential Sampling to Field Testing of Aviation Security Equipment], (2016).
- RAC, [Understanding Binomial Sequential Testing].

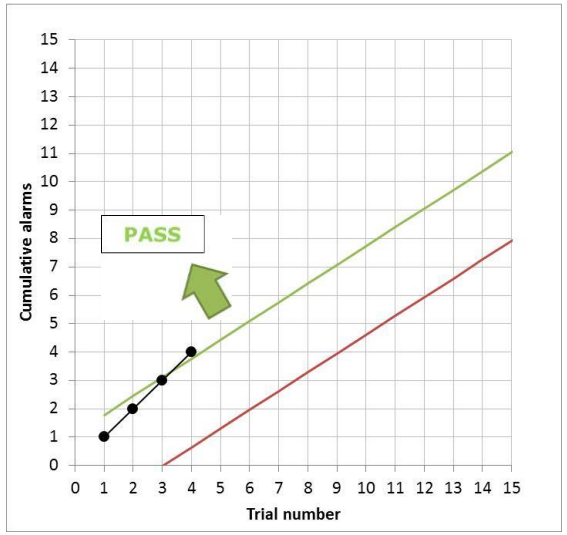
European Commission

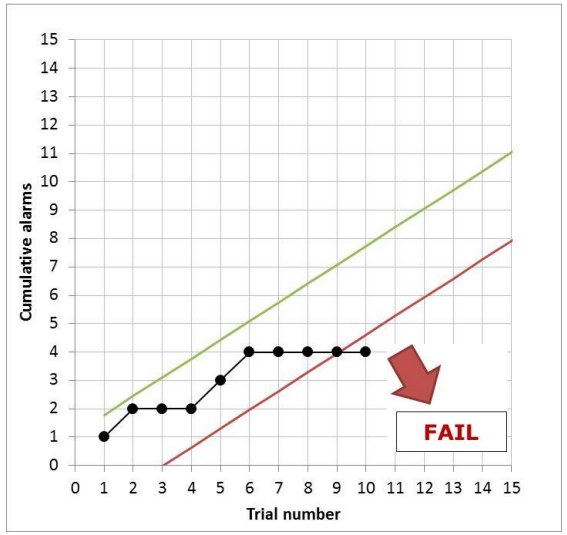
The next step: sequential sampling

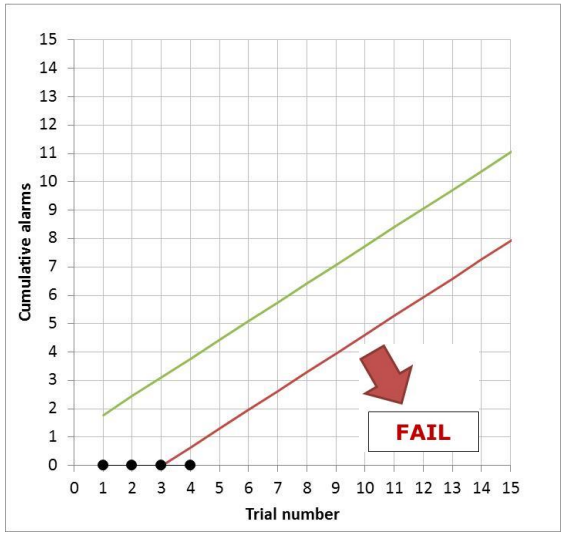
Sequential Sampling: a version of acceptance testing, but the number of observations is not pre-determined.

Observations are made **one at a time**; after each observation the **cumulative** results are assessed. When the desired degree of confidence is reached, the process is stopped.

Sequential sampling allows for **early stopping** and a saving of time and resources, compared to a pre-determined, fixed sampling plan.

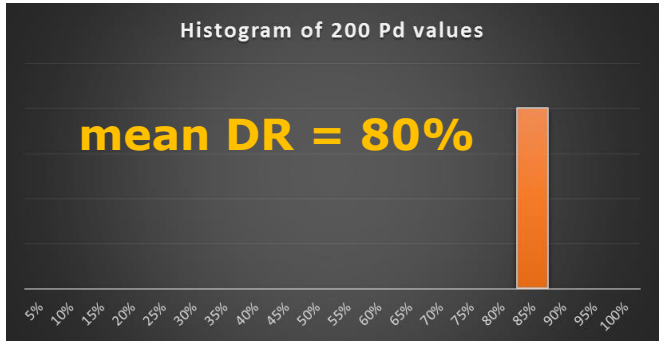


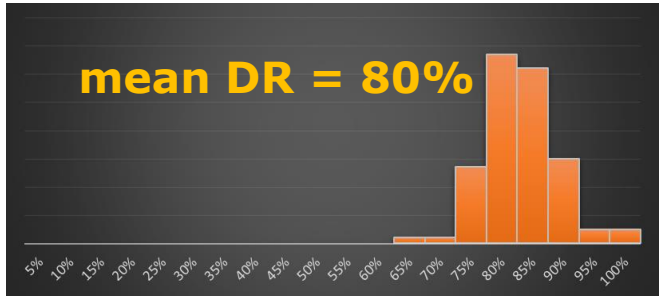


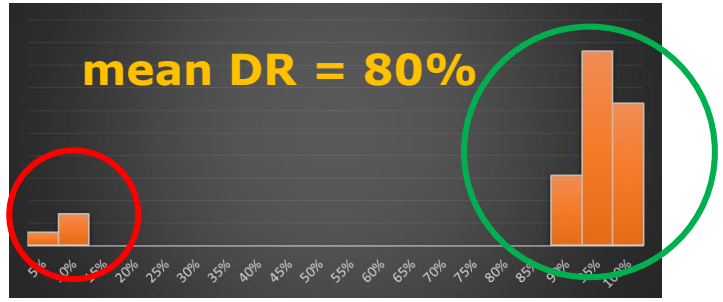


European Commission

Pd distributions across the test set







'challenging' test items

'easy' test items

The distribution of Pd values across test items is normally unknown. However, all distributions can be represented by the **ordinary binomial distribution*.**

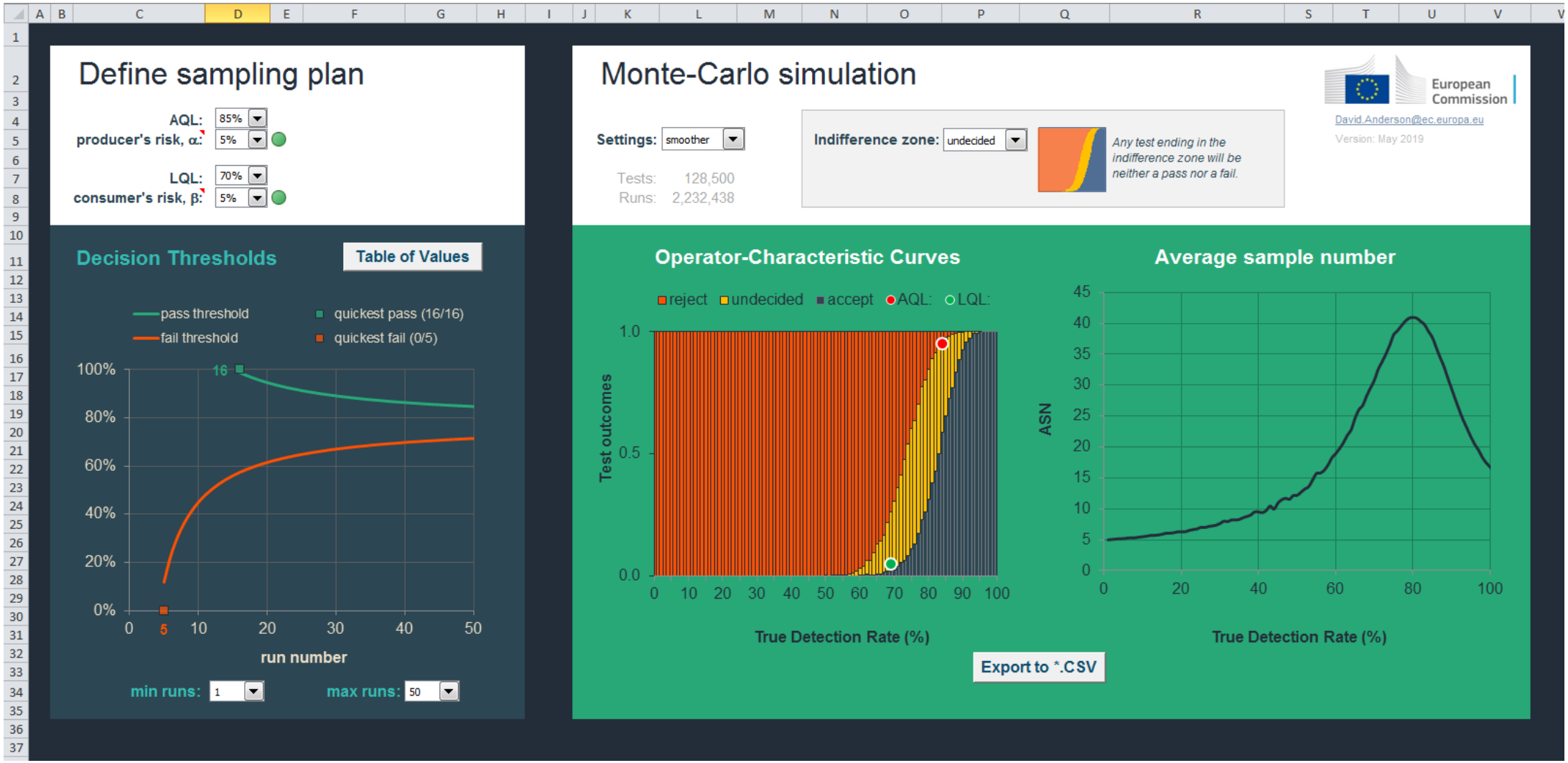
However, any **subset** of test items **must be chosen randomly**, to avoid introducing bias.

A random subset can also **prevent over-training**.

** The binomial sum variance inequality states that the variance of the sum of binomially distributed random variables will always be less than or equal to the variance of a binomial variable with the same n and p parameters.*

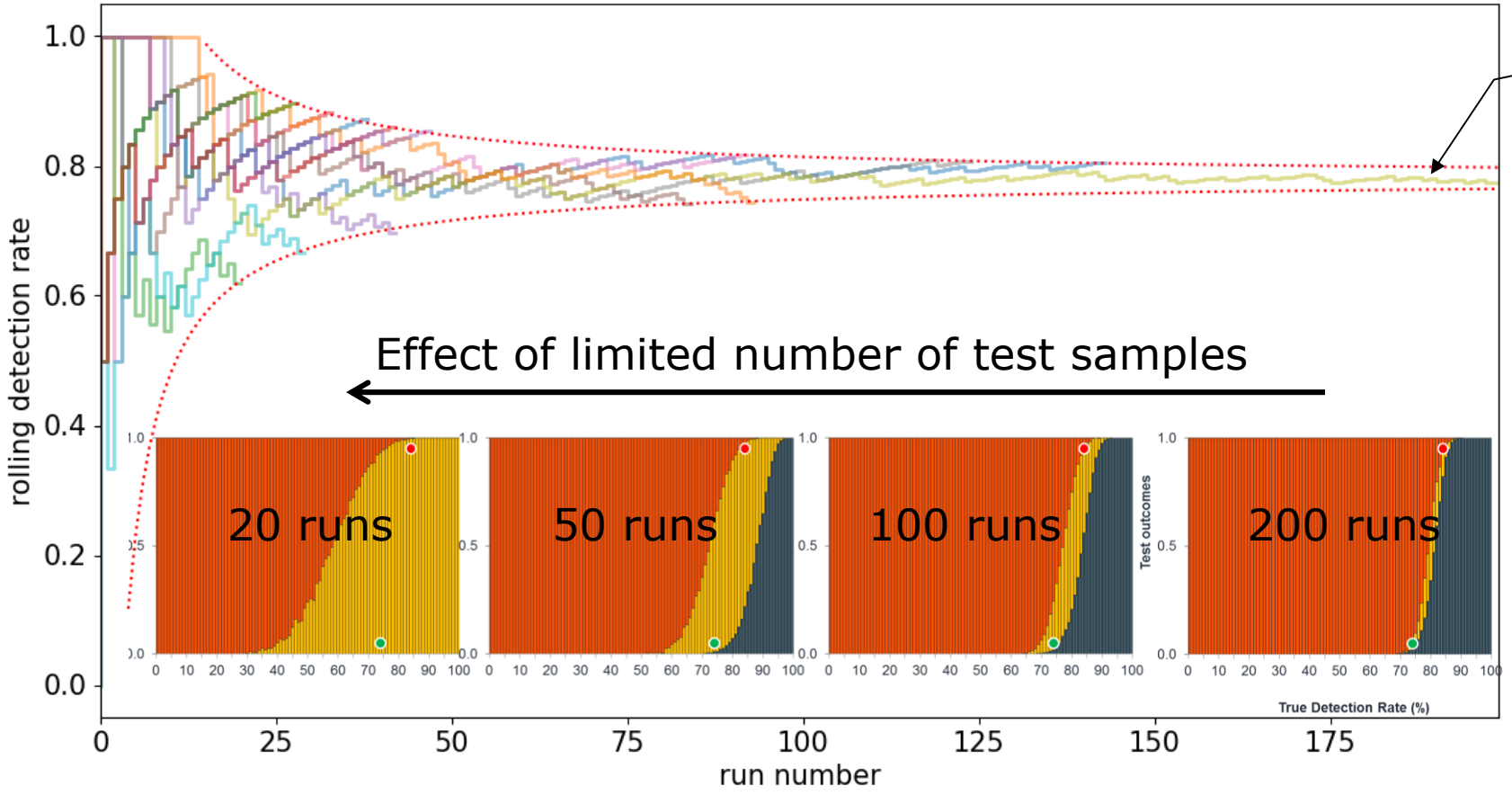
European Commission

Excel software: define test & simulate outcome



A look under the hood…

15 examples of simulated sequential tests.
(all from distribution with mean DR=80 %)



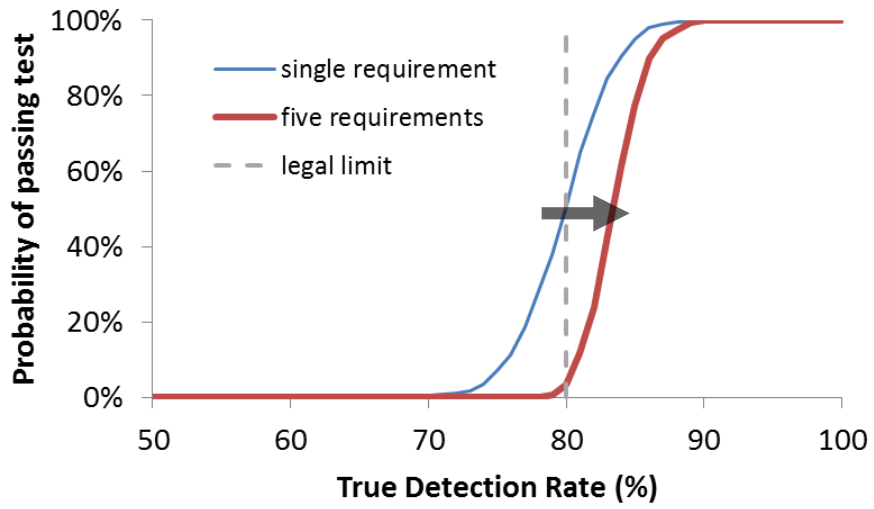
What about these?

Options:

1. "I don't know"
2. pass all (increases β)
3. fail all (increases α)
4. "shared risk" (use limit)

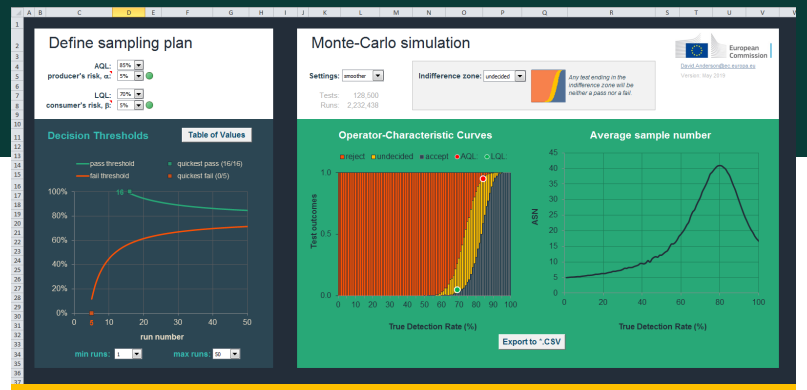
Multiple Requirements

- Detection equipment typically has to meet **more than one** requirement.
- The probability of passing the overall test is the **product** of the probabilities of passing each requirement (i.e. if fail one requirement = fail entire test).
- If each criterion uses a 'shared risk' decision rule (i.e. use the limit value as a strict cut-off), then the operator-characteristic curve for the overall test will **shift to the right**:



- Asymmetric (**higher producer' risk**)
- Something for regulators and testers to bear in mind when adding requirements.
- If desired, can be redressed by modifying AQL and LQL points on individual criteria.

Summary



Excel software available:

David.Anderson@ec.europa.eu

- Risks of incorrect decisions **at test level** lead to undesirable consequences **at the macro level**.
- **Sequential testing**:
  - ✓ Manage the risks
  - ✓ Reduce test duration whilst satisfying risk limits
- **Software available**:
  - ✓ Design tests
  - ✓ Simulate outcome
  - ✓ Visually support our understanding
- **Drivers**: reduce cost, improve time to market, minimise operational disruption, avoid unnecessary re-testing, cope with ever-increasing demands, free-up precious testing resources, spend taxpayers money wisely, …

European Commission