INTER LABORATORY STUDIES

What are they, and why do them?

Henk Blom
Director of Technical Services
Rollprint Packaging Products, Inc.

Outline

- The big picture
  - What is a measurement system?
  - What is measurement system analysis?
- E691 – The ILS Standard
- ILS & Gage R&R's
- Pass/fail tests
  - Dye leak test case study

Precision and Bias Statements

10. Precision and Bias

10.1 *Precision*—A research report[5] describes a round robin conducted in 2002 in accordance with Practice E691, involving

[5] A research report is available from ASTM International Headquarters. Request RR:F02-1018.

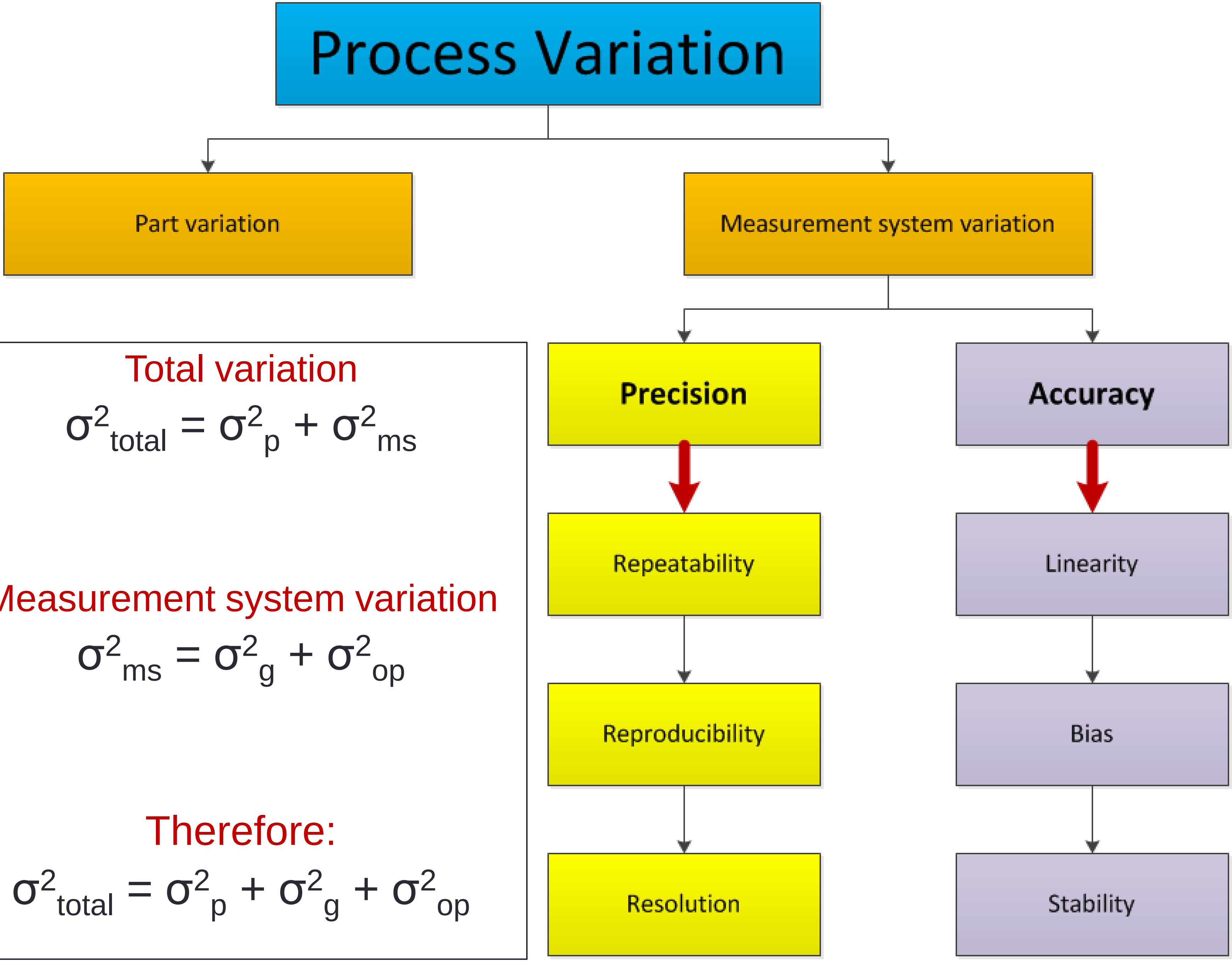
F2251 – 03 (2008)

TABLE 2 Summary of Interlaboratory Results by Measurement Set

Part	Avg Thickness		Repeatability, Std Dev, s_r		Reproducibility, Std Dev, s_R		95 % Repeatability, Limit, r		95 % Reproducibility, Limit, R	
	mm	mils	mm	mils	mm	mils	mm	mils	mm	mils
G: 0.48 mil PET	0.012954	0.5100	0.001008	0.0397	0.002197	0.0865	0.00282	0.1111	0.006152	0.2422
A: 2 mil PE	0.053679	2.1133	0.000881	0.0347	0.002691	0.1059	0.00247	0.0971	0.007534	0.2966
B: 2.5 mil PET/PE	0.06405	2.5217	0.000489	0.0192	0.001915	0.0754	0.00137	0.0539	0.005361	0.2111
H: 2.94 mil Calibration Feeler Gage	0.074083	2.9167	0.001316	0.0518	0.00152	0.0598	0.00369	0.1451	0.004255	0.1675
F: 42# Uncoated Medical Paper	0.090678	3.5700	0.000733	0.0289	0.001943	0.0765	0.00205	0.0808	0.00544	0.2142
D: 4.2 mil PET/PE/FOIL/PE	0.10994	4.3283	0.000811	0.0319	0.001966	0.0774	0.00227	0.0894	0.005505	0.2167
E: Uncoated 1073B Tyvek	0.143722	5.6583	0.00164	0.0645	0.009781	0.3851	0.00459	0.1807	0.027387	1.0782
C: 10 mil EVA/SURLYN/EVA	0.260181	10.2433	0.001222	0.0481	0.002596	0.1022	0.00342	0.1347	0.007268	0.2862
Average			0.001069	0.0421	0.004001	0.1575	0.00299	0.1178	0.011204	0.4411

TABLE 3 Summary of Interlaboratory Averaged Results

Repeatability (within a laboratory) standard deviation (s_r)	0.001069 mm (0.0421 mils)
95 % repeatability limit (r)	0.00299 mm (0.1178 mils)
Reproducibility (between laboratories) standard deviation (s_R)	0.004001 mm (0.1575 mils)
95 % reproducibility limit (R)	0.011204 mm (0.4411 mils)



Total variation

$$\sigma^2_{total} = \sigma^2_p + \sigma^2_{ms}$$

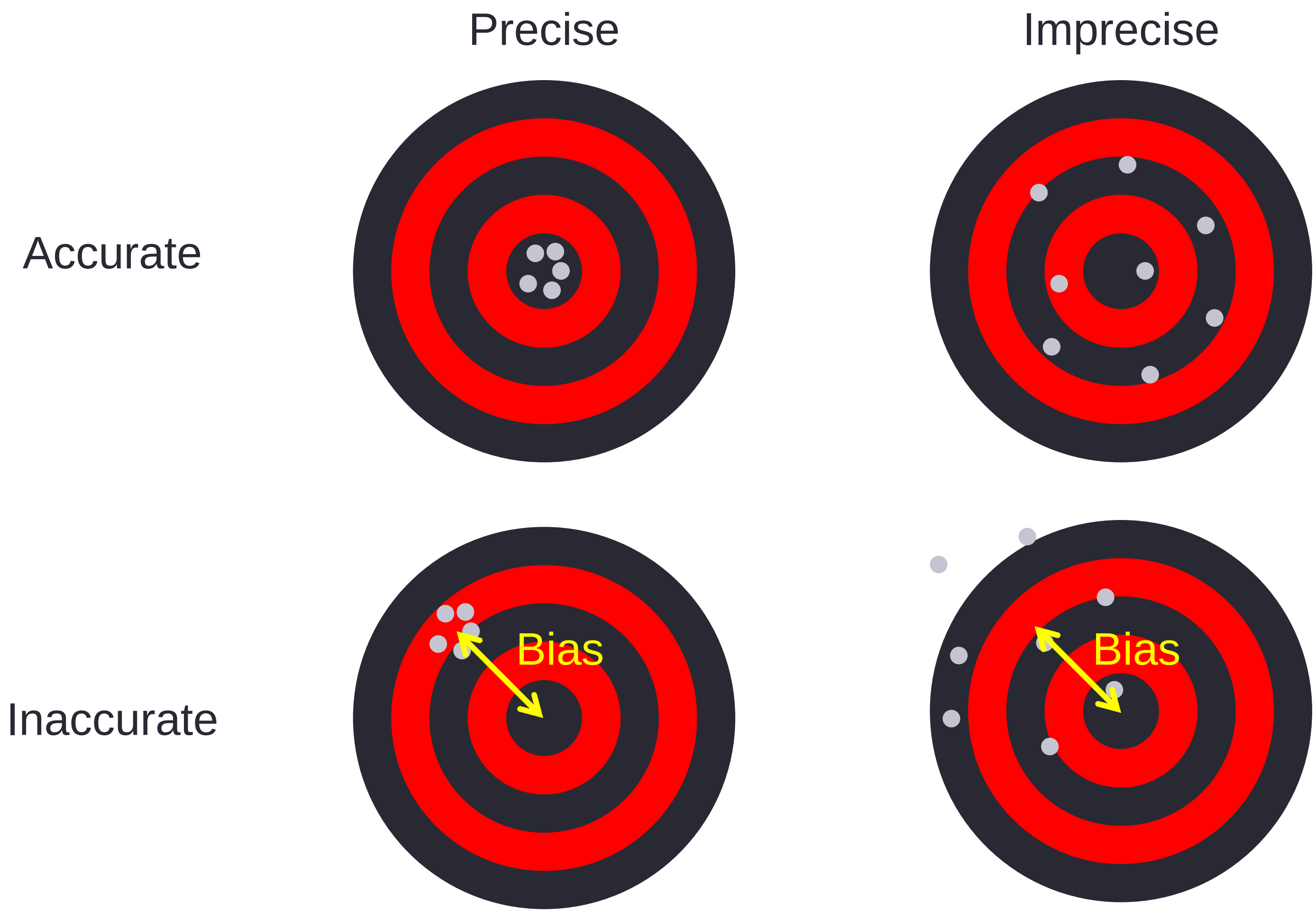
Measurement system variation

$$\sigma^2_{ms} = \sigma^2_g + \sigma^2_{op}$$

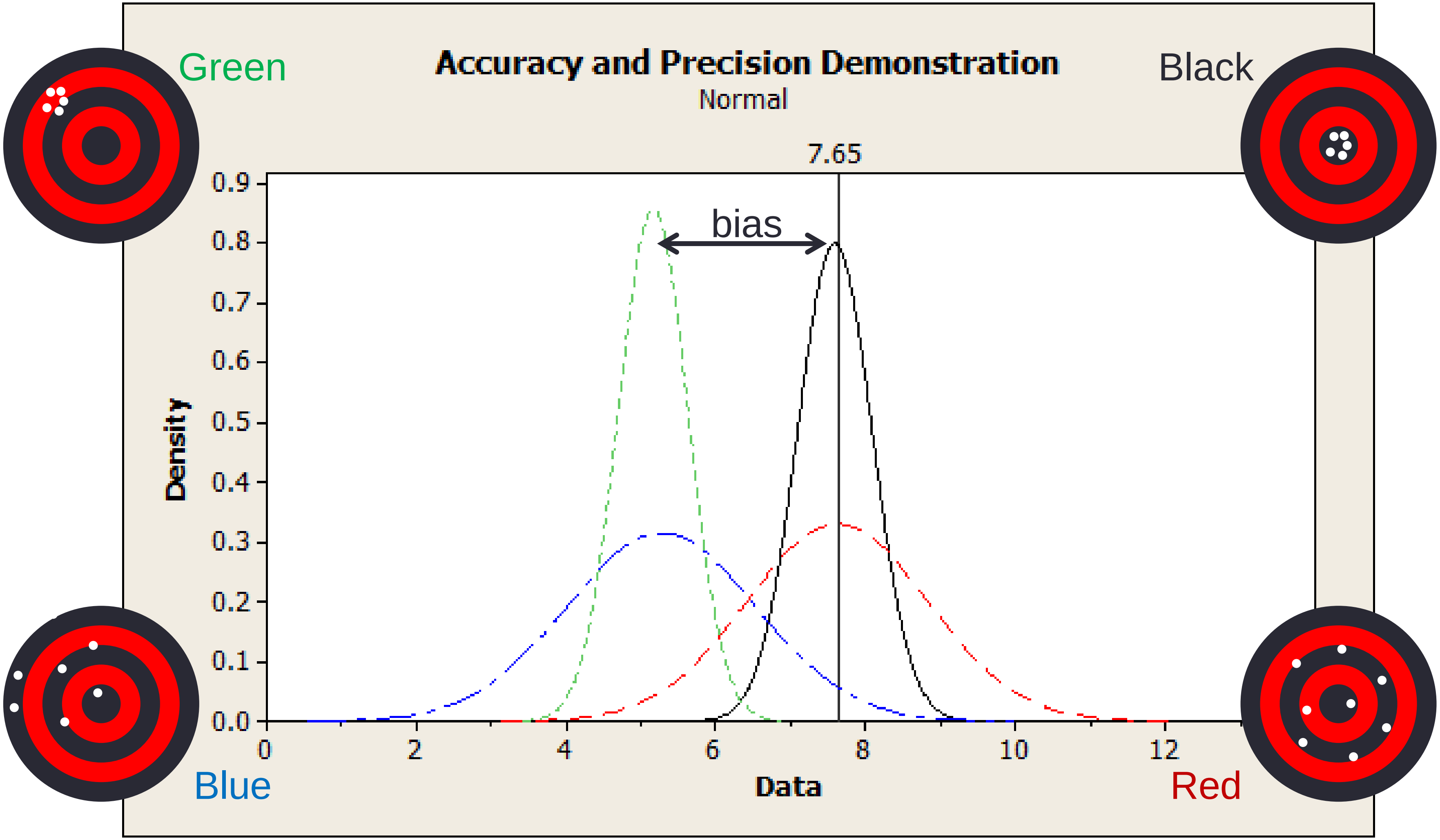
Therefore:

$$\sigma^2_{total} = \sigma^2_p + \sigma^2_g + \sigma^2_{op}$$

Accuracy and precision



Accuracy and precision



Measurement System Analysis

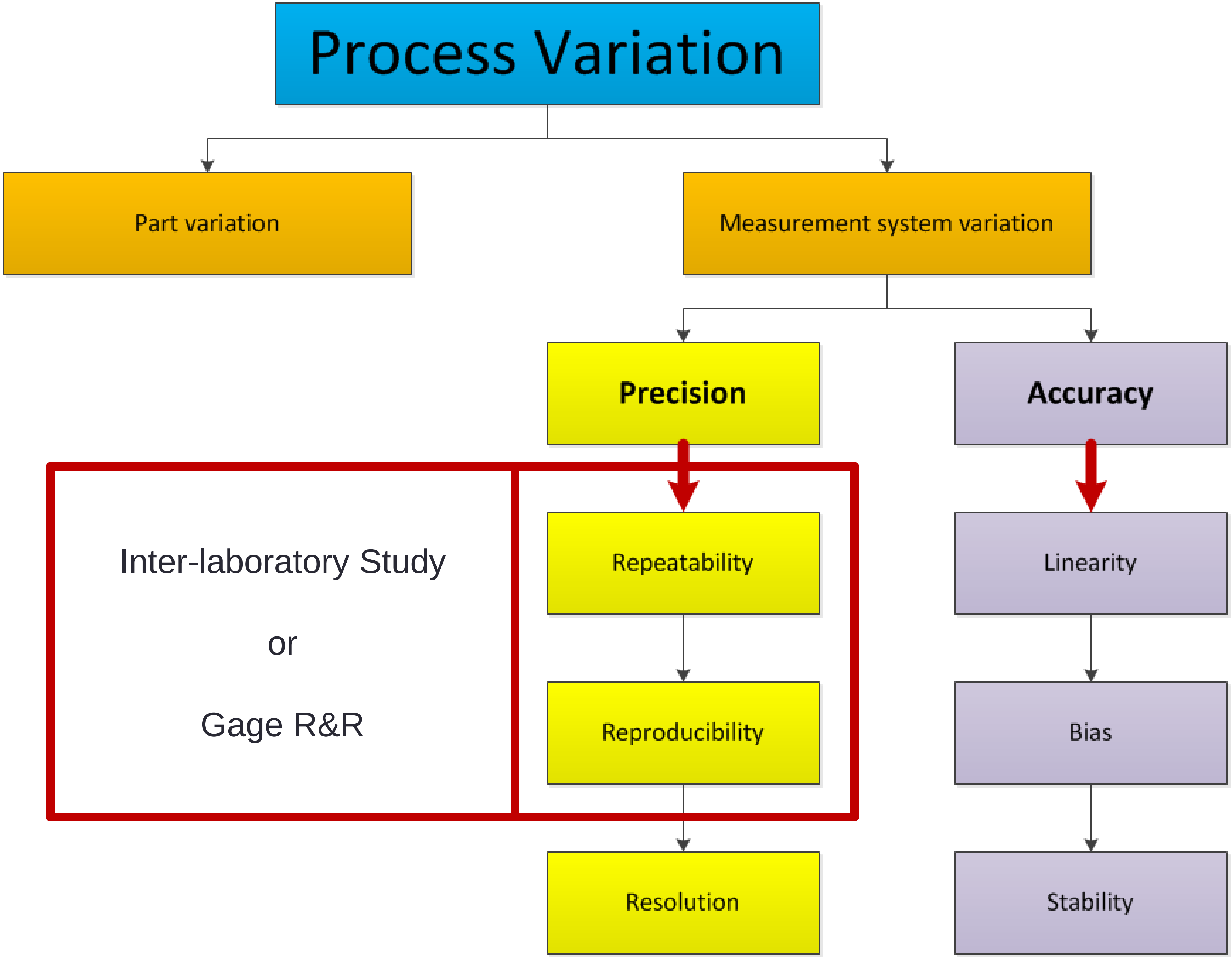
An analysis of the collection of

equipment, operations, procedures, software

and personnel

that affects the assignment of a reading to a

measurement characteristic.


Process Variation
Part variation
Measurement system variation
Precision
Accuracy
Inter-laboratory Study
or
Gage R&R
Repeatability
Reproducibility
Resolution
Linearity
Bias
Stability

Repeatability and reproducibility

- Repeatability
  - Repeatability is the variation in measurements observed when one operator repeatedly measures the same characteristic in the same place on the same part with the same gauge. It is related to the standard deviation of the measured values.
- Reproducibility
  - Reproducibility is the variation in average measurements due to factors other than gauge variation.
    - Accounts for operator-to-operator (including lab-to-lab) and gauge-to-gauge variability, as well as changes in temperature and humidity.

E691 – The ILS Standard Practice

Designation: E691 – 13

An American National Standard

Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[1]

This standard is issued under the fixed designation E691; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

This standard has been approved for use by agencies of the Department of Defense.

"This practice provides a standard procedure for determining the precision of a test method. Precision, when evaluating test methods, is expressed in terms of two measurement concepts, repeatability and reproducibility."

(from the Introduction)

The ILS Team

- ASTM has an in-house ILS group, led by Phil Godorov
- They are available to:
  - Design an interlaboratory study,
  - Identify potential samples,
  - Solicit volunteer laboratories,
  - Find available suppliers,
  - Contract with distributors,
  - Review laboratory instructions,
  - Collect and analyze data,
  - Produce a draft precision statement,
  - Compile information for and generate the research report, and
  - Give recognition to participating laboratories.

ILS Basics

- **Scope**
  - § 1.1 – …techniques for planning, conducting, analyzing, and treating the results of an ILS…
  - § 1.4 – This practice is concerned exclusively with test methods which yield a single numerical figure as a test result
- **Test Method**
  - § 8.1 – Of prime importance is the existence of a valid, well-written test method that has been developed in one or more competent laboratories…
- **Laboratories**
  - § 9.1.1 – An ILS should include 30 or more laboratories…
  - § 9.1.2 – Under no circumstances should the final statement of precision …be based on…test results…from fewer than 6 laboratories.

ILS Basics

- **Materials**
  - § 10.2.2 – An ILS of a test method should include at least three materials representing different test levels…
- **Number of test results**
  - § 11.1 – It is generally sound to limit the number of test results on each material in each laboratory to a small number, such as three or four.
    - “Generally, the time and effort invested in an ILS is better spent on examining more materials across more laboratories than on recording a large number of test results per material within a few laboratories.”

Data analysis

calculations for that material on the same worksheet, as illustrated in Table 2. Work on only one material at a time.

15.4 *Cell Statistics:*

15.4.1 *Cell Average, $\bar{x}$*—Calculate the cell average for each laboratory using the following equation:

$$\bar{x} = \sum_{1}^{n} x/n \quad (1)$$

where:

$\bar{x}$ = the average of the test results in one cell,
x = the individual test results in one cell, and
n = the number of test results in one cell.

Thus from Table 2 for Material A, Laboratory 1 (that is, for Cell A1):

$$\bar{x} = (41.03+41.45+41.37)/3 = 41.2833$$

15.4.2 *Cell Standard Deviation, s*—Calculate the standard deviation of the test results in each cell using the following equation:

$$s = \sqrt{\sum_{1}^{n} (x - \bar{x})^2/(n-1)} \quad (2)$$

The symbols have the same meaning as for Eq 1. Thus for Cell A1:

$$s = \sqrt{(41.03 - 41.2833)^2 + (41.45 - 41.2833)^2 + (41.37 - 41.2833)^2]/(3-1)} = 0.2230$$

While Eq 2 shows the underlying calculation of the cell standard deviation, inexpensive pocket calculators are available that calculate both the average and the standard deviation directly. Check to be sure the calculator uses (n – 1) as the divisor in Eq 2, not n, and has adequate precision of calcula-

15.5.3 *Standard Deviation of the Cell Averages, $s_{\bar{x}}$*—Calculate this statistic using the following equation:

$$s_{\bar{x}} = \sqrt{\sum_{1}^{p} d^2/(p-1)} \quad (5)$$

Thus for Material A:

$$s_{\bar{x}} = \sqrt{[(-0.2350)^2 + (-0.0783)^2 + (-0.0683)^2 + (-0.0616)^2 + (-0.0550)^2 + (0.5017)^2 + (-1.0616)^2 + (1.0584)^2]/(8-1)} = 0.6061$$

15.6 *Precision Statistics*—While there are other precision statistics, introduced later in this practice, the fundamental precision statistics of the ILS are the repeatability standard deviation and the reproducibility standard deviation. The other statistics are calculated from these standard deviations.

15.6.1 *Repeatability Standard Deviation, s_r*—Calculate this statistic using the following equation:

$$s_r = \sqrt{\sum_{1}^{p} s^2/p} \quad (6)$$

where:

s_r = the repeatability standard deviation,
s = the cell standard deviation (p of them from Eq 2), and
p = the number of laboratories.

Thus for Material A:

$$s_r = \sqrt{[(0.2230)^2 + (0.4851)^2 + (1.0608)^2 + (1.8118)^2 + (0.3667)^2 + (1.4081)^2 + (1.2478)^2 + (0.8225)^2]/8} = 1.0632$$

15.6.2 *Reproducibility Standard Deviation, s_R*—Calculate a provisional value of this statistic using the following equation:

Output

TABLE 11 Glucose in Serum—Precision Statistics

NOTE 1—This table (with the column for $s_{\bar{x}}$ omitted) is a useful format for the presentation of the precision of a test method as required by Section A21 of the *Form and Style of ASTM Standards*.

Material	$\bar{x}$	$s_{\bar{x}}$	s_r	s_R	r	R
A	41.5183	0.6061	1.0632	1.0632	2.98	2.98
B	79.6796	1.0027	1.4949	1.5796	4.19	4.42
C	134.7264	1.7397	1.5434	2.1482	4.33	6.02
D	194.7170	2.5950	2.6251	3.3657	7.35	9.42
E	294.4920	2.6931	3.9350	4.1923	11.02	11.74

$\bar{x} = cell\ average$

$S_{\bar{x}} = standard\ deviation\ of\ cell\ averages$

$S_r = repeatability\ standard\ deviation$

$S_R = reproducibility\ standard\ deviation$

$r = 2.8 S_r$ (95% repeatability limit)

$R = 2.8 S_R$ (95% reproducibility limit)

$$S_r = \sqrt{\sum_{1}^{p} s^2/p}$$

$$S_R = \sqrt{{S_{\bar{x}}}^2 + \frac{(n-1){S_r}^2}{n}}$$

The $1,000,000 question(s)…

TABLE 11 Glucose in Serum—Precision Statistics

NOTE 1—This table (with the column for $s_{\bar{x}}$ omitted) is a useful format for the presentation of the precision of a test method as required by Section A21 of the *Form and Style of ASTM Standards.*

Material	$\bar{x}$	$s_{\bar{x}}$	s_r	s_R	r	R
A	41.5183	0.6061	1.0632	1.0632	2.98	2.98
B	79.6796	1.0027	1.4949	1.5796	4.19	4.42
C	134.7264	1.7397	1.5434	2.1482	4.33	6.02
D	194.7170	2.5950	2.6251	3.3657	7.35	9.42
E	294.4920	2.6931	3.9350	4.1923	11.02	11.74

…how do I interpret the data in the table above?

…is this a good measurement system?

Gage R&R's

- Very similar to an ILS – although usually done within one company across several operators
  - Methodology is equally applicable to an ILS situation
- Industry-accepted guidelines to assess gage suitability
- Many companies in the medical device industry (OEM's and suppliers) use Gage R&R's routinely

Output comparison - ILS

F2251 – 03 (2008)

TABLE 2 Summary of Interlaboratory Results by Measurement Set

Part	Avg Thickness		Repeatability, Std Dev, s_r		Reproducibility, Std Dev, s_R		95 % Repeatability, Limit, r		95 % Reproducibility, Limit, R	
	mm	mils	mm	mils	mm	mils	mm	mils	mm	mils
G: 0.48 mil PET	0.012954	0.5100	0.001008	0.0397	0.002197	0.0865	0.00282	0.1111	0.006152	0.2422
A: 2 mil PE	0.053679	2.1133	0.000881	0.0347	0.002691	0.1059	0.00247	0.0971	0.007534	0.2966
B: 2.5 mil PET/PE	0.06405	2.5217	0.000489	0.0192	0.001915	0.0754	0.00137	0.0539	0.005361	0.2111
H: 2.94 mil Calibration Feeler Gage	0.074083	2.9167	0.001316	0.0518	0.00152	0.0598	0.00369	0.1451	0.004255	0.1675
F: 42# Uncoated Medical Paper	0.090678	3.5700	0.000733	0.0289	0.001943	0.0765	0.00205	0.0808	0.00544	0.2142
D: 4.2 mil PET/PE/FOIL/PE	0.10994	4.3283	0.000811	0.0319	0.001966	0.0774	0.00227	0.0894	0.005505	0.2167
E: Uncoated 1073B Tyvek	0.143722	5.6583	0.00164	0.0645	0.009781	0.3851	0.00459	0.1807	0.027387	1.0782
C: 10 mil EVA/SURLYN/EVA	0.260181	10.2433	0.001222	0.0481	0.002596	0.1022	0.00342	0.1347	0.007268	0.2862
Average			0.001069	0.0421	0.004001	0.1575	0.00299	0.1178	0.011204	0.4411

TABLE 3 Summary of Interlaboratory Averaged Results

Repeatability (within a laboratory) standard deviation (s_r)	0.001069 mm (0.0421 mils)
95 % repeatability limit (r)	0.00299 mm (0.1178 mils)
Reproducibility (between laboratories) standard deviation (s_R)	0.004001 mm (0.1575 mils)
95 % reproducibility limit (R)	0.011204 mm (0.4411 mils)

Output comparison – Gage R&R

Gage R&R (Nested) for C4

Source	DF	SS	MS	F	P
Operators	7	0.0000003	0.0000000	0.00	1.000
Parts (Operators)	56	0.0014843	0.0000265	8174.94	0.000
Repeatability	128	0.0000004	0.0000000		
Total	191	0.0014849			

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.0000000	0.04
Repeatability	0.0000000	0.04
Reproducibility	0.0000000	0.00
Part-To-Part	0.0000088	99.96
Total Variation	0.0000088	100.00

% Contribution – Total Gage R&R
Acceptable: < 1%
Possibly acceptable: 1-9%
Needs improvement: > 9%

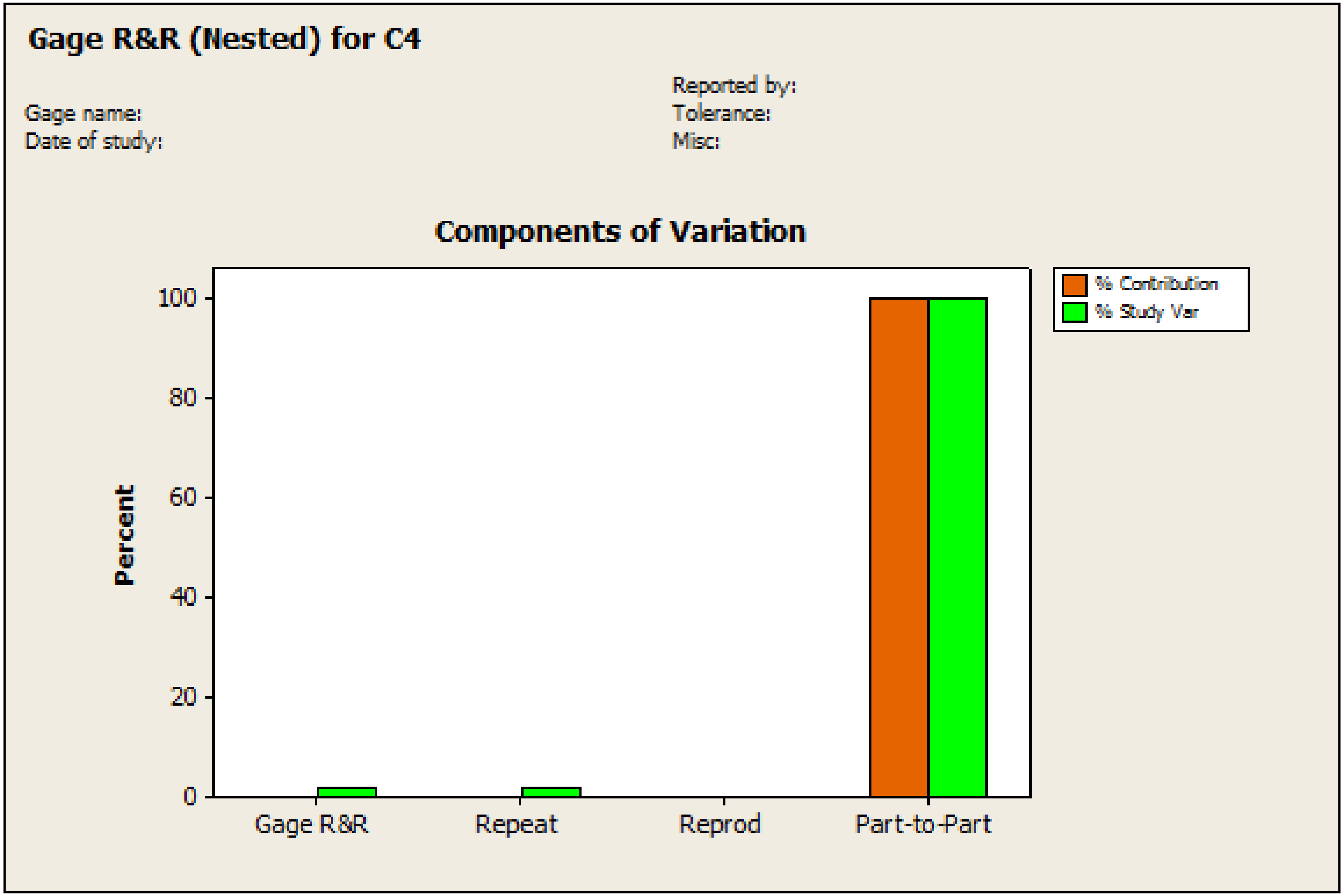
Source	StdDev (SD)	Study Var (6 * SD)	%Study Var (%SV)
Total Gage R&R	0.0000569	0.0003416	1.92
Repeatability	0.0000569	0.0003416	1.92
Reproducibility	0.0000000	0.0000000	0.00
Part-To-Part	0.0029722	0.0178330	99.98
Total Variation	0.0029727	0.0178363	100.00

%Study Var – Total Gage R&R
Acceptable: < 10%
Possibly acceptable: 10-30%
Needs improvement: > 30%

Number of Distinct Categories = 73

Should be >5

Gage R&R Graphical Summary



What about binary data?

- Remember – an ILS only applies to variable data, not to pass/fail tests
- But…
  - Some test methods only provide pass/fail results
    - Dye leak tests
    - Visual seal integrity tests
    - Particulate inspections
- What do we do in these cases?
- Can a statement be made about the suitability of a test method if the output is pass/fail?

Case study – Dye penetration leak test

F1929 – porous packages

- Determines the presence of a 50 μm or greater channel leak in the seal of a pouch comprised of porous materials

F3039 – non-porous packages

- Method A – determines the presence of a 50 μm or greater channel leak in the seal of a pouch comprised of non-porous materials
- Method B – determines the presence of a 10 μm or greater through-pinhole in a sheet sample of non-porous material

Channel leaks

- 50 µm tungsten wire used to create channels in seals
- Set of 50 pouches created for each of the labs
  - ~25 of these samples had the channel leak, and the balance did not
  - Defect samples randomly distributed
- Each lab tested all 50 samples and recorded Y or N for detection of channel leak
- Results analyzed with Minitab
  - 95% confidence interval for probability of detection of 50 µm channel

Minitab analysis

13.2.2 When combining the edge dip data population of all labs, the results show that this method provides the correct response of detecting channels created by a 50 µm wire in seals 95 % of the time. The 95 % confidence interval is 93.8 % to 96.0 %. The results of correctly identified, false positives and false negatives are shown in Tables 2-4.

13.2.2.1 When comparing the edge dip labs through ANOVA, labs 6520 and 6521 were significantly different than the other five. This would suggest that there is potentially some other cause of their greater error rate. If they were excluded, the rate improves to 98 % (96.8 % to 98.5 % at a 95 % confidence level).

TABLE 1 Summary of Results

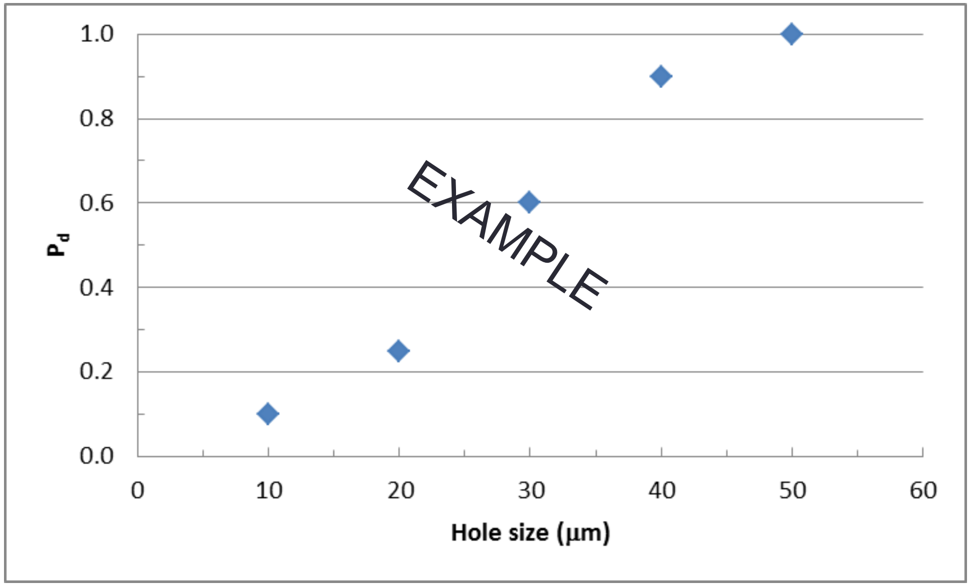
Total specimens tested	False positives	False negatives	95% confidence interval	Error rate (%)
2000	2	4	99.13–99.94	0.3

What about “Method B”?

- Initial attempt to detect 10 μm holes yielded poor and highly variable results
- Investigation led to key areas where method could be improved
  - Saw some improvement, but not enough
- Tough questions…
  - Do we really have a test method here?
  - Was the hole size we targeted too small?
- Discussions with E11 (Statistics) led to an alternate approach

Method B – Limit of detection

- Six distinct samples
  - 10 μm
  - 20 μm
  - 30 μm
  - 40 μm
  - 50 μm
  - Control (no hole)
- 3 samples of each per set
- Random distribution



Conclusions

- A measurement system analysis is a critical part of every test method
  - ILS
  - Gage R&R
- The ILS team at ASTM is eager to help ILS study directors
- Precision statements <u>can</u> be made for pass/fail data
- Working with E11 group to formalize a “limit of detection” approach to R&R studies for pass/fail test methods