oncimmune®

Leading early cancer detection

The use of sensitivity analysis to set acceptance criteria in the validation of biomarker assays

Graham Healey, Chief Statistician,
Oncimmune, UK

oncimmune®

Leading early cancer detection

SECTION 1

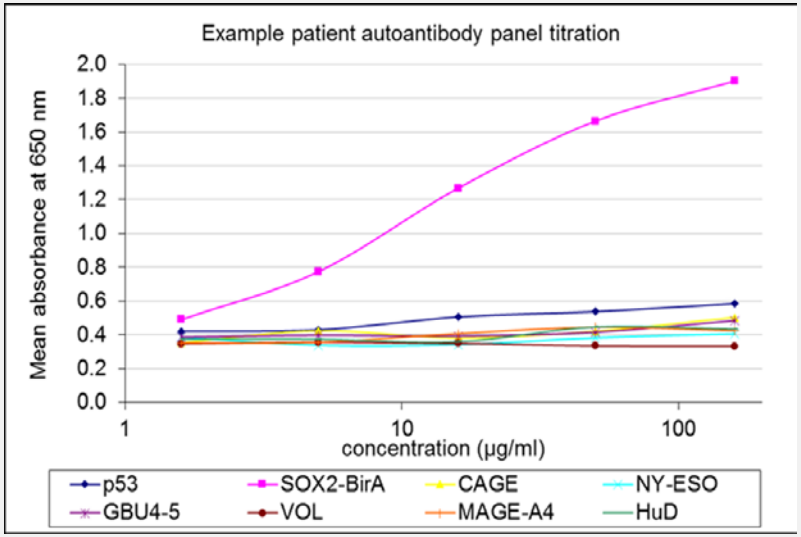
Introduction

Autoantibody panel diagnostic test

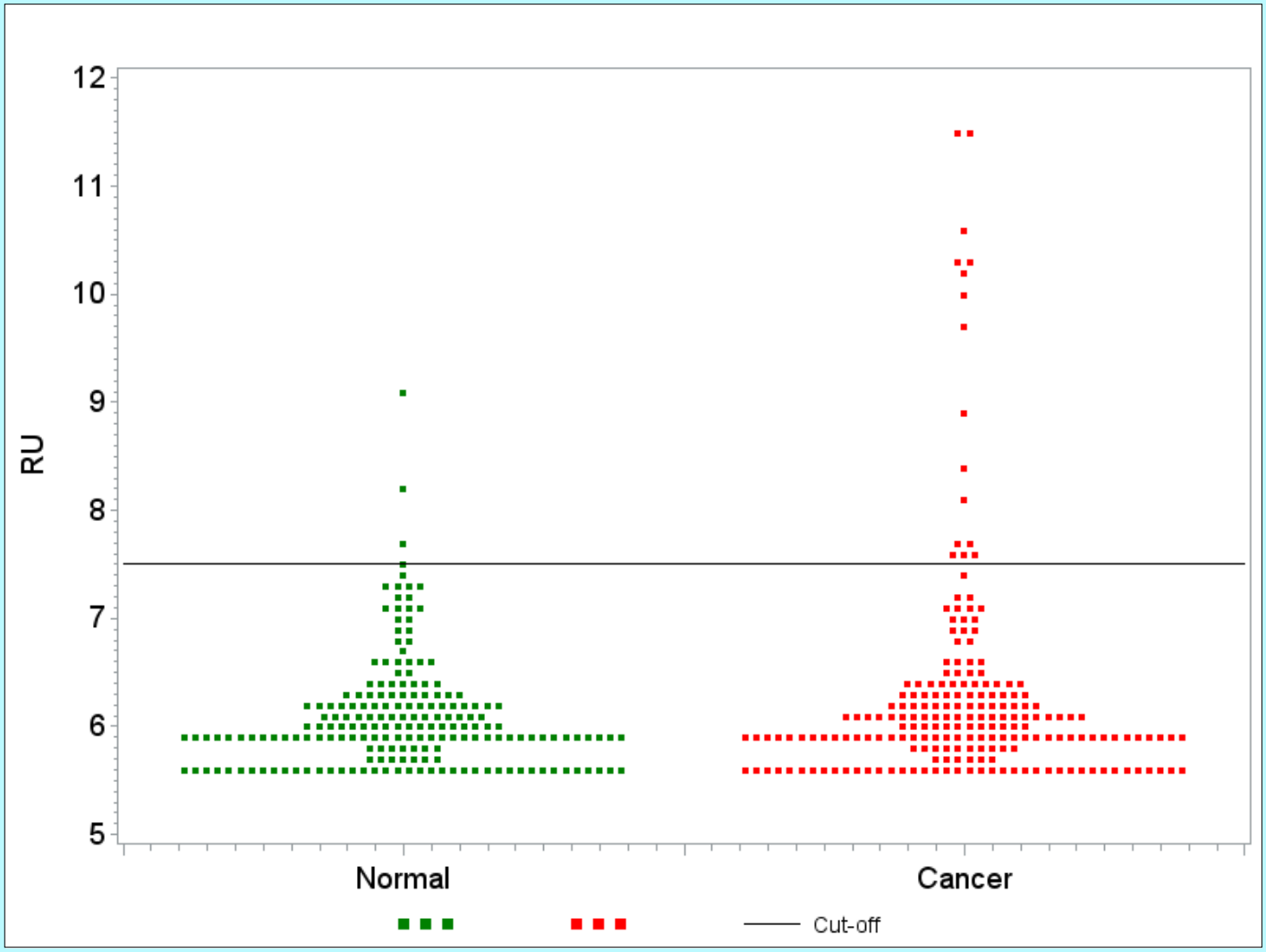
- A panel of autoantibody (AAb) markers measured in serum samples.
  - Early detection of cancer (lung and liver)
  - Companion diagnostic for prognostic or predictive purposes

- The assay is an ELISA (a type of LBA = Ligand-binding assay)
  - 96-well plates are coated with 7 tumour-associated antigens TAA.
  - Incubate with serum samples, add a colour-labelled secondary antibody.
  - Read optical densities (OD).
  - Separate standard curves for each antigen.
  - Calibrated reference units (RU)

- Negative/Positive cut-off for each marker
- LDT for our central lab (CLIA/COLA)
- Kit for approved labs (CE, ISO 13485)



Typical data for a single marker



Clinical requirement and MAPV

- For clinical practice it is necessary to maintain the performance within limits.
- We publish a test performance claim, namely specificity and sensitivity.
  - To control the false-positive and false-negative rates.
- Define MAPV = maximum allowable performance variation around the claim.
  - Part of a Product specification
- We need to control the assay to keep the spec/sens within MAPV=5%.
  - Performance claim is a specificity of 91% and a sensitivity of 41%.
  - So an allowed specificity range of 86% to 96%, and sensitivity of 36% to 46%.
  - Derived from discussion with KOLs and other stake-holders.
- How do we maintain this claim ?
- The spec/sens are a function of the individual markers, so what level of change in the individual markers can be tolerated?
- We focus on multi-analyte biomarker assays.

Robustness and Ruggedness

- Robustness: Deliberate alteration of experimental conditions
- Ruggedness: Natural variation in experimental conditions (eg different labs)
  - A-F Aubry & N Weng 2015 “So you think your assay is robust?” Bioanalysis 7(23) 2969-71
  - An ICH definition, not a requirement, makes no distinction between robustness and ruggedness.
- A key aim is to identify factors which must be more strictly controlled.
  - Such as change of batches, instruments, operators, etc.
- Our approach can be called “Indirect Outcome Studies”
  - Sandberg *et al* 2015 “Defining analytical perf. spec.” Clin Chem Lab Med 53(6) 833-5
  - Model 1: The effect of analytical performance on clinical outcomes.

Qualitative Assay

- Run a set of test samples under the various conditions.
- Assess positivity versus the MAPV.
- e.g. for a Standard condition with sens=40%, compare to 35%/45% (TABLE 1)
- Look at concordance (TABLE 2 for New vs Standard conditions [Made-up data])

TABLE 1		Temperature		
		18C	20C	22C
Incubation time	10min	3/10	3/10	4/10
	15min	4/10	4/10	5/10
	20min	4/10	4/10	5/10

TABLE 2		New		
		Neg	Pos	Total
Standard	Neg	13	2	15
	Pos	4	11	15
	Total	17	13	30

- Calculate standard errors. e.g. for P=40%:-
  - se(P) = √(P*(100-P)/n) = +/-15%, needs n=10
  - se(P) = √(P*(100-P)/n) = +/-5%, needs n=90

- Probably not sensitive enough. Too many samples needed.
  - Look at James Westgard's literature
  - Design for within-sample changes

Multi-analyte options

- Better to look at signal levels (OD/RU).
- So for a panel of k markers:-

- OPTION 1: k individual markers > k cut-offs > a positivity rule > Neg/Pos
  - Set acceptance limits on the individual markers.
  - **Sample is positive if any marker is positive ("One-marker-high")**
  - This is what we currently do.

- OPTION 2: k individual markers > 1 score > 1 cut-off > Neg/Pos
  - Form a single composite score, e.g. logistic regression, pooled over markers.
  - **Sample is positive if score > cut-off ("Logistic-score")**
  - This is what most people do!

oncimmune®

Leading early cancer detection

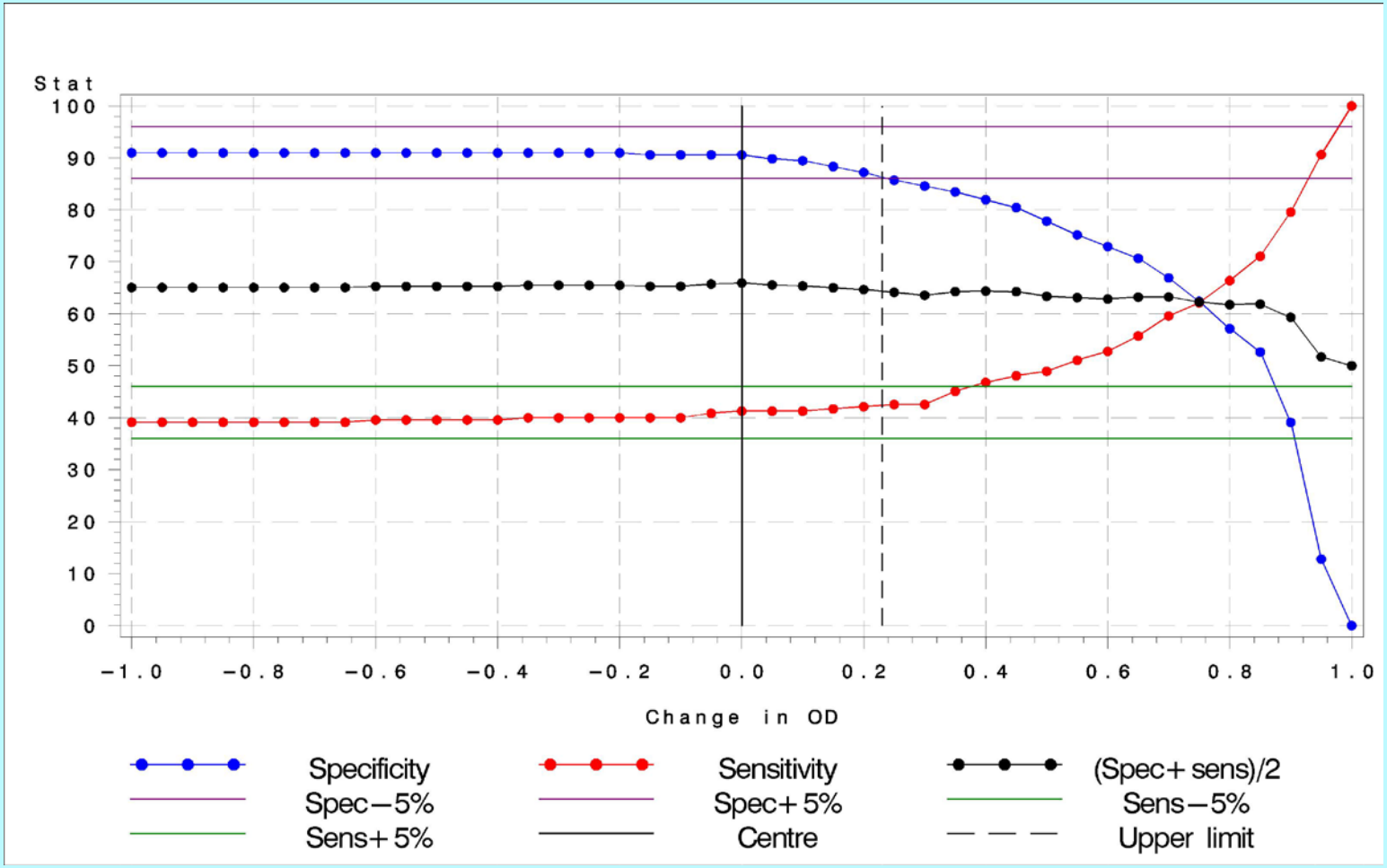
SECTION 2

Setting acceptance limits

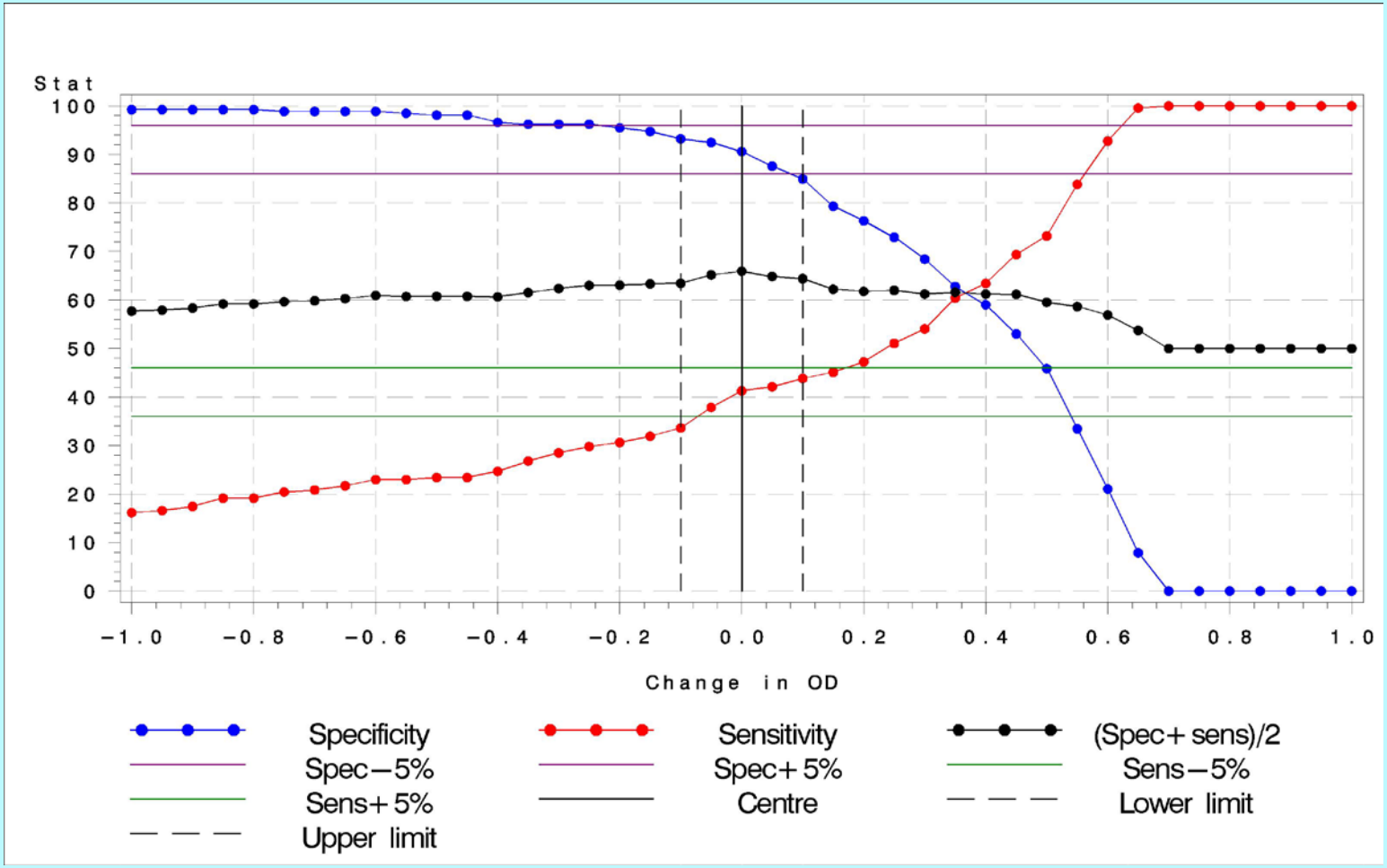
Sensitivity Plot

- Took data from a previous case-control study (approx. 250C/250N).
- Varied the observed signals in small increments
  - For each increment calculate spec/sens using current cutoffs
- "Sensitivity plot" of %age change in spec/sens versus increment.
- Define MASD as the maximum allowable signal deviation in the individual markers to maintain the MAPV
  - MASDs are equivalent to engineering tolerance
- Read this directly off the plot.
  - Vertical lines indicate the limits of signal variation allowable.
- Applied at two levels:
- 1) MASD for a Single-marker: Use marker giving largest effect
- 2) Same MASD applied to All-markers: Use average over all markers

1: One-marker-high: Single-marker change



1: One-marker-high: All-marker change



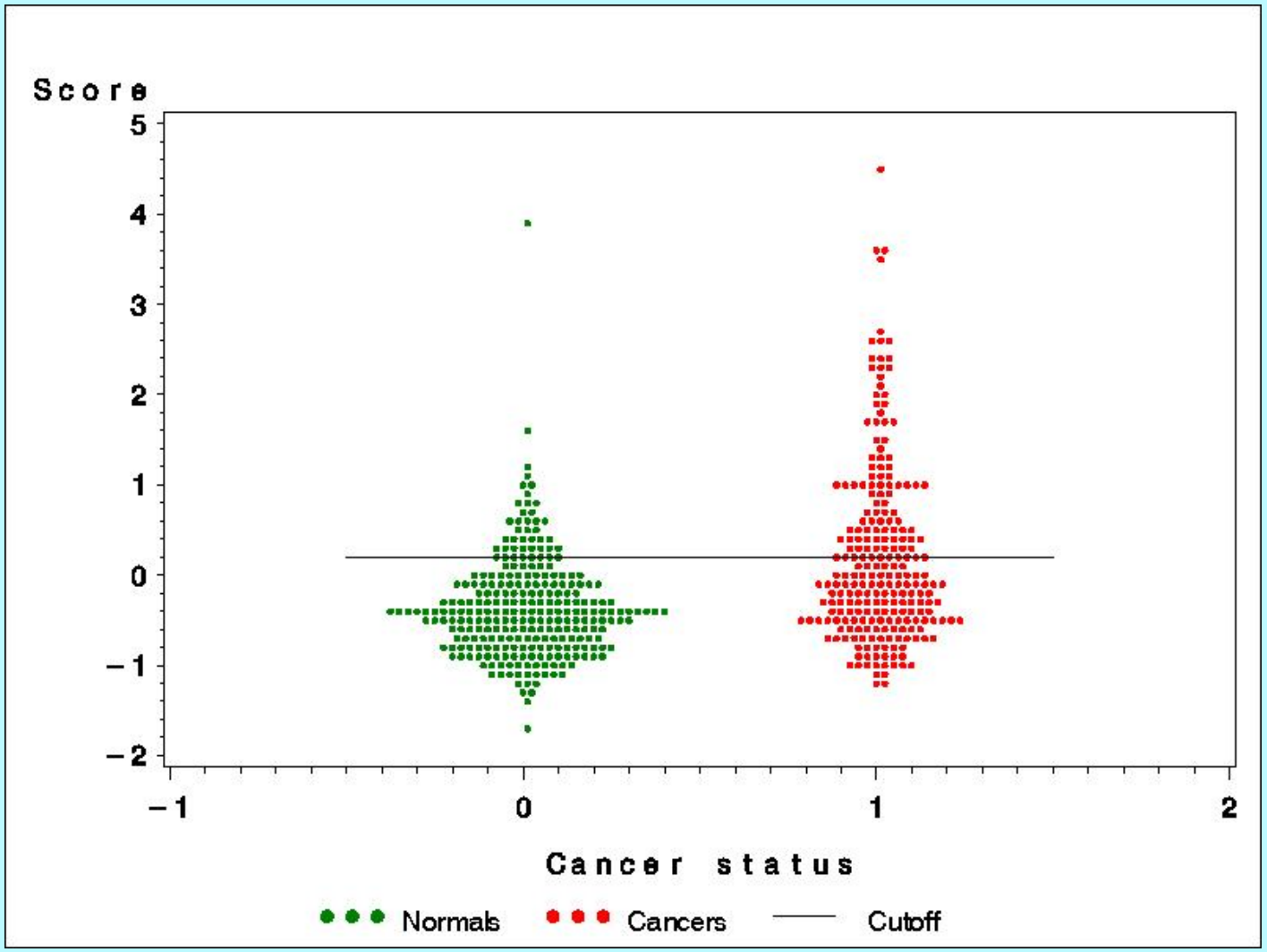
1: One-marker-high: Limits

- Increases in signal have much greater effect on overall spec/sens than decreases.
- Reduction in signal for a single antigen has virtually no overall effect.
- In our system the MASDs are asymmetric.

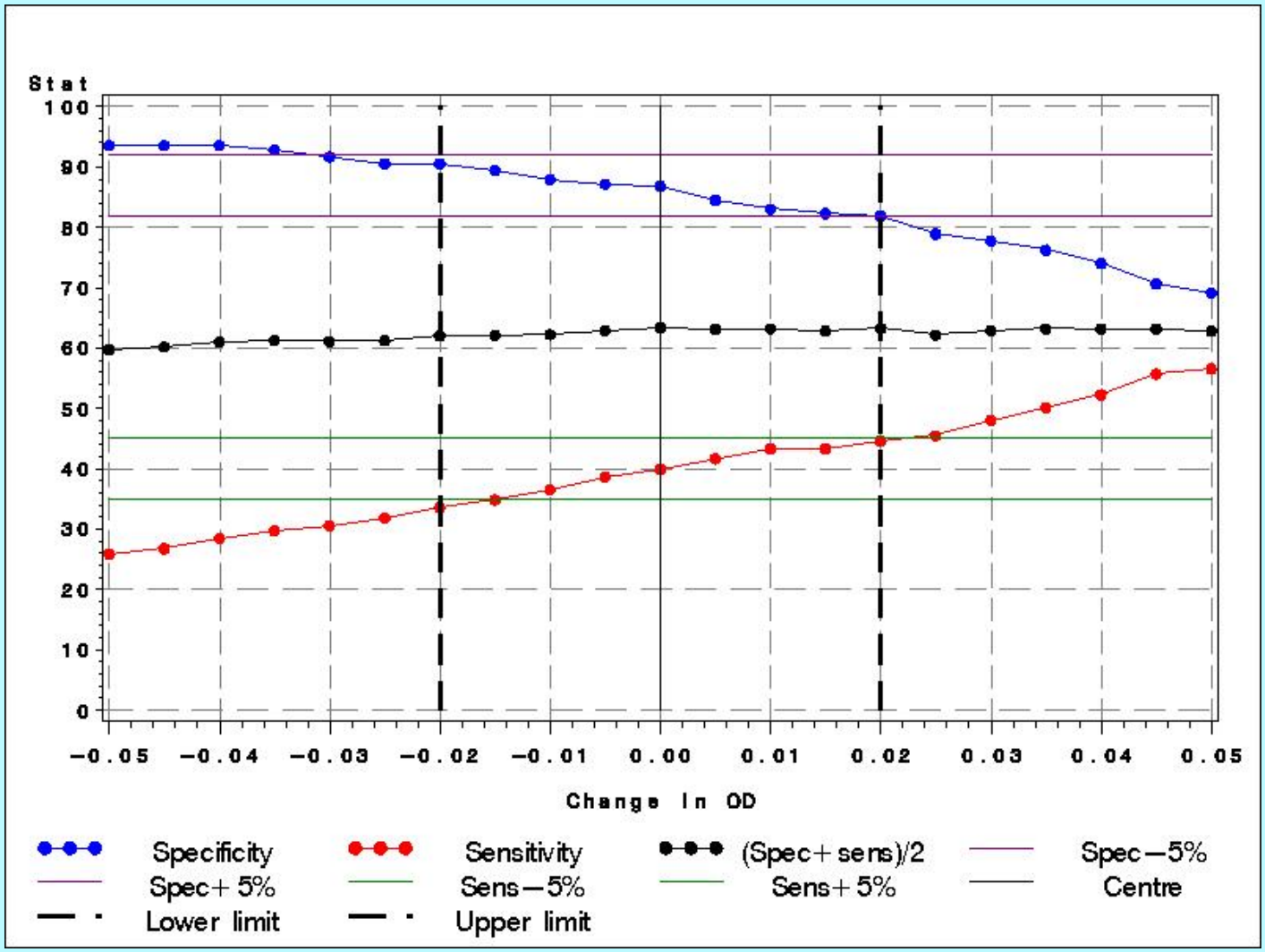
Unit	Level	Kit	
		Lower	Higher
OD	Single-marker	-1.50	+0.30
	All-marker	-0.15	+0.15

OD on a scale of 0.00 to 2.00
Lower = Observed signal lower than Control
Higher = Observed signal higher than Control

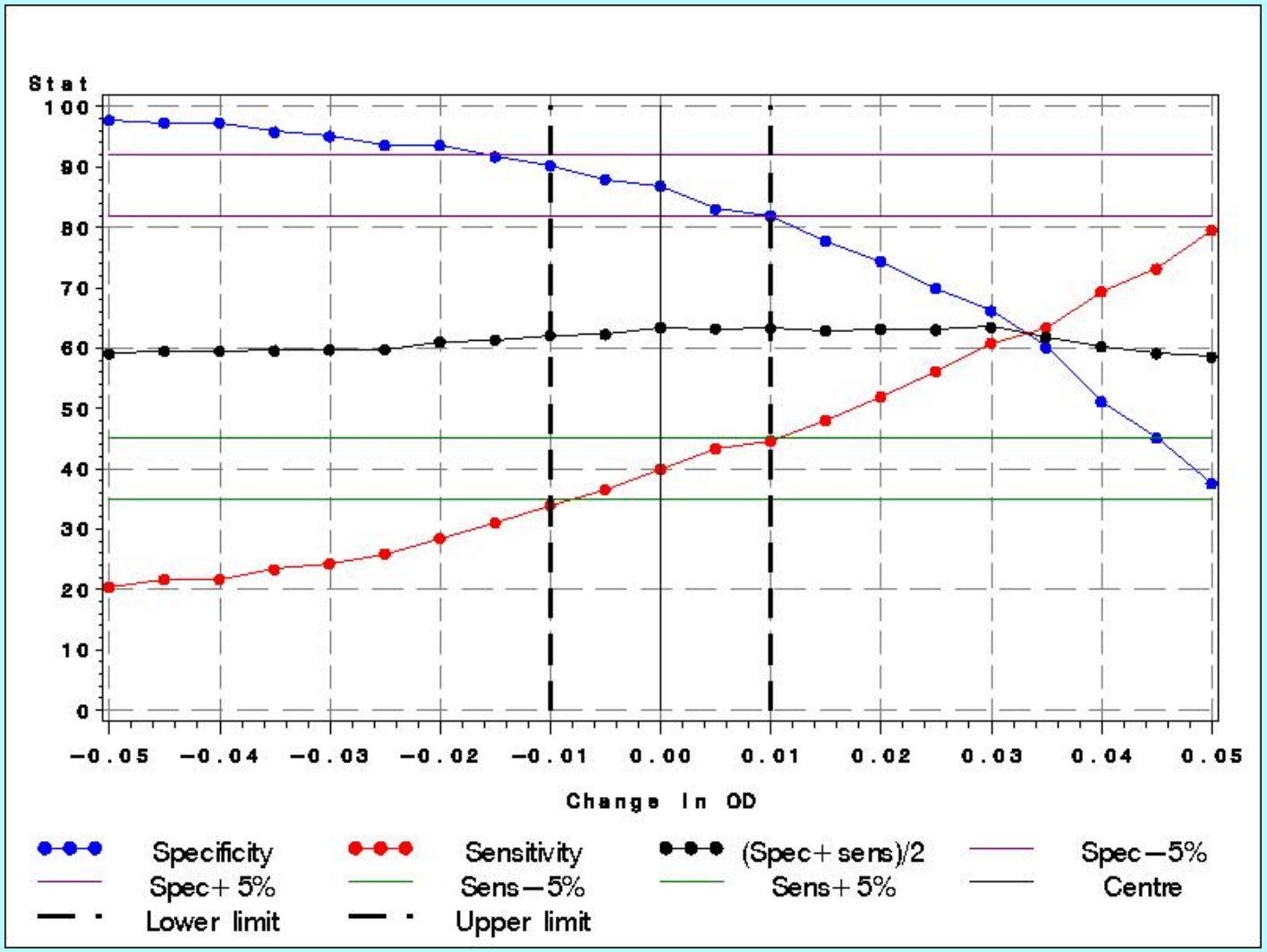
2: Logistic-score: Score distribution



2: Logistic-score: Single-marker change



2: Logistic-score: All-marker change



2: Logistic-score: Limits

- Derive the logistic regression equation. For 7 markers for example:
- Linear predictor $Z = b0 + b1*y1 + b2*y2 \ldots\ldots b7*y7 = b0 + \sum(bi*yi)$
- Apply the deviation D to all marker signals yi:
- Shifted predictor: $Z' = b0 + \sum(bi*(yi+D)) = Z + D.\sum(bi)$
- Seems very sensitive !

Unit	Level	LDT	
		Lower	Higher
OD	Single Ag	-0.02	+0.02
	All Ags	-0.01	+0.01

OD on a scale of 0.00 to 2.00
Lower = Observed signal lower than Control
Higher = Observed signal higher than Control

Single-markers vs All-markers

- The situation will determine if it is likely to be Single-marker or All-marker change.
- A stability problem might only affect a single antigen.
- In our assay the same secondary antibody (secAb) is used for all antigens.
- So a between-run signal deviation in secAb will cause an All-marker change.
- We tend to look at both Single-marker and All-marker rules.

oncimmune®

Leading early cancer detection

SECTION 3

Application of the limits

Validation studies

- Use the MASDs to set acceptance limits around standard conditions.
  - Shipping, storage and freeze-thaw of samples, QC materials or assay reagents.
  - Dilution pipetting, Incubation times, Operating temperature
  - Plate shaker times, Plate reader wavelength, Plate washer cycles
- We ran designed studies varying assay conditions from the standard set-up.
  - e.g. If the optimum temperature is 20°C, run a study at 18°C, 20°C and 22°C.
- Looked at the mean signal change from standard conditions.
- (One-marker-high option from now on)

Example 1. Incubation times

- Two Kit expts: SecAb and TMB Substrate
- Means over 6 subjects and 7 markers (2 reps for the Standard).
- MASD: All-marker limits = -0.15, +0.15 (See earlier Table)

SecAb Incubation time	N	OD mean	Difference vs standard
50 min	42	0.539	-0.051
60 min (Standard)	84	0.590	x
70 min	42	0.651	0.061
90 min	42	0.715	0.125

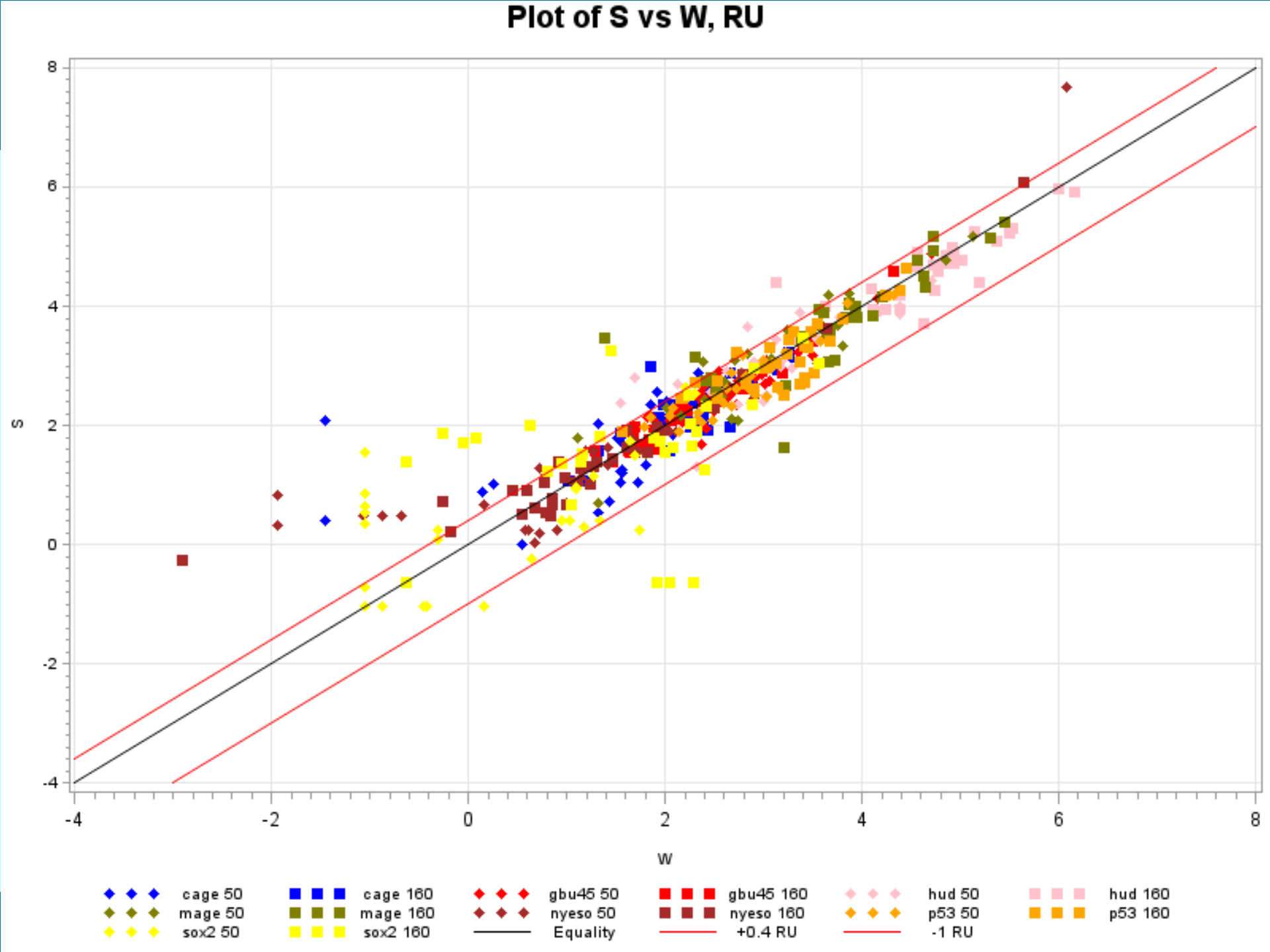
TMB Substrate Incubation Time	n	OD mean	Difference vs standard
10 min	42	0.446	-0.122
15 min (Standard)	84	0.568	x
20 min	42	0.647	0.079
25 min	42	0.751	0.184

Example 2: Whole Blood Shipping

- Shipping ***serum*** at ambient temperature had no detrimental effect on signals.
- Needed to check freshly drawn ***whole blood***.

- For a set of subjects plotted RU signal for serum vs whole blood.
- Plot all markers on the same correlation "method comparison" plot.
- Display the MASD limits.

		Low	High
RU	Single-Ag	-1.00	+0.40
	All-Ags	-0.20	+0.20


Plot of S vs W, RU
S
W
8
6
4
2
0
-2
-4
-4
-2
0
2
4
6
8
cage 50
mage 50
sox2 50
cage 160
mage 160
sox2 160
gbu45 50
nyeso 50
Equality
gbu45 160
nyeso 160
+0.4 RU
hud 50
p53 50
-1 RU
hud 160
p53 160

oncimmune®

Leading early cancer detection

SECTION 4

Design of validation studies

Operating limits

- For each assay factor estimate the range outside of which the signal deviations are too great. Term this the procedural operating range (POR).
  - Control the assay factors within the stated POR (engineering tolerance).
  - Then the mean marker deviations will be within the MASD limits.
  - Thus diagnostic performance will also be within the MAPV limits (product specification).
- But are these limits achievable ?
- We would like high confidence (say 90%) that the "true" mean is within the limits around the standard condition, given the observed mean.
  - The greater the sample size, the higher the confidence.
- So we need to consider variation, not just means
  - Intra-run and inter-run (process variability)

Confidence intervals

- Suppose we have set all-marker signal limits at +/-0.15.
- We run a validation study at temperature levels (T_{X-2}, T_X, T_{X+2}).
- We run n reps per mean and observe means (Y_{X-2}, Y_X, Y_{X+2}) pooled over all markers.

- EXAMPLE:
- At T_{X-2} a mean difference from standard (Y_{X-2} - Y_X) = (0.483 - 0.603) = -0.120.
- Pooled standard deviation (sd) was 0.052 with n=6 per group.
- Calculate a 95%Confidence Interval for the deviation.
  - 95% CI for the true difference is (Y_{X-2} - Y_X) +/- 1.96.sd.√(1/n+1/n).
- Giving: -0.120 +/- 1.96*0.052*√(2/6) = -0.120+/-0.059
- This is (-0.179, -0.061), so the lower limit is out-of-spec.

- Need to determine optimum sample size.

Effect of sample size

- Simulate observed deviations for a range of true differences D1 and sample sizes.
- Tabulate %-age of simulated runs for which a 90%CI was within +/-0.15 limits.

D1	N=6	N=12	N=24	N=36
-0.17	1%	1%	0%	0%
-0.15	5%	5%	6%	4%
-0.13	15%	22%	36%	50%
-0.11	37%	61%	87%	97%
-0.09	62%	89%	100%	100%
-0.07	81%	98%	100%	100%

- So, for example, if the true mean was -0.11, and we ran N=6, the CI would be within limits only 37% of the time.
- We will only be able to accept narrower POR.

oncimmune®

Leading early cancer detection

SECTION 5

Concluding comments

Concluding comments

- Limited guidance is provided in regulatory documents for acceptance criteria.
  - Authorities need to adapt to multi-marker tests.
- The more markers there are, the less resource available for validation.
  - Lab time, Sample volume, Protein production
- Different markers may have different variability
  - Cannot have a single optimum design for every antigen
  - Get all markers on the same numerical scale and pool, eg log scale
  - Maybe use maxCV across Ags.
  - Focus on high-signal signals.
- The MASD allows you to determine PORs for maintaining performance.
  - Gives a basis on which to assess changes seen in validation studies.
- To statisticians these methods might seem very simple.
- However, for many lab scientists they are not simple !
  - Particularly pooling data over many markers.