



Statistical modeling of some NDE measurement uncertainties

**Albert van Roodselaar, Luc Huyse, CP Hsiao,
Mark Seiwald**

Chevron ETC



Presenter & Contributors

Part of Chevron Energy Technology Company

- In-house specialists
- Statistical modeling & NDE expertise
 - API 510 & 570 certified inspector
 - Vice-chair of ASCE Probabilistic Methods Committee
 - Member of API inspection committee
 - Member of API/ASME FFS committee

Current job

- Reliability-Based Inspection, Fitness-for-Service
- Inspector training, qualifications (Chevron NDE Exams)
- Quantify POD and sizing accuracy for various NDE techniques
- Hands-on advanced inspection services
- Research and development of emerging UT and ET techniques
- Development of inspection strategies



Setting the Tone

■ Purpose of the Presentation

- Quantify **probability of detection** and **sizing accuracy** of various NDE inspection methods
- The resulting statistical model can be used to calculate the confidence in the inspection results
- Focus of this presentation is on UT crack sizing, remote field eddy-current exchanger tubes and manual UT thickness gauging inspections
- Results for some other NDE techniques have also been generated

■ Sources

- Proprietary Chevron inspector testing results



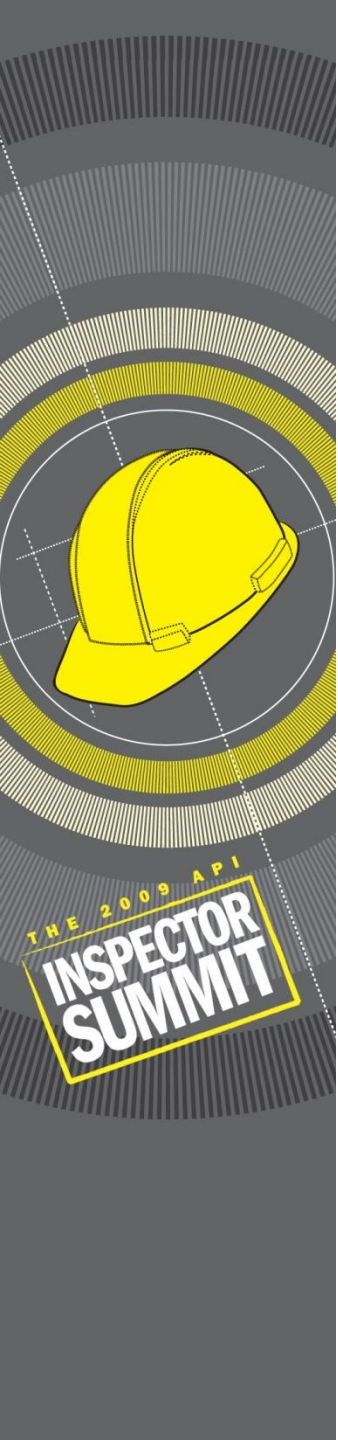
NDE Contractor Qualification Status

- Chevron administers on the order of 100 exams per year
- Results are remaining consistent
 - No major changes in pass rates since 2006
- Ultrasonic Programs
 - Ultrasonic Thickness
 - Ultrasonic Crack Detection
 - Ultrasonic Planar Flaw Through Wall Sizing
 - Ultrasonic Heavy Wall Crack Detection and Sizing
- Electromagnetic/Remote Field/IRIS Programs
 - Electromagnetic for tubing
 - Remote Field for tubing
 - IRIS for tubing
 - Surface Eddy Current for carbon steel welds



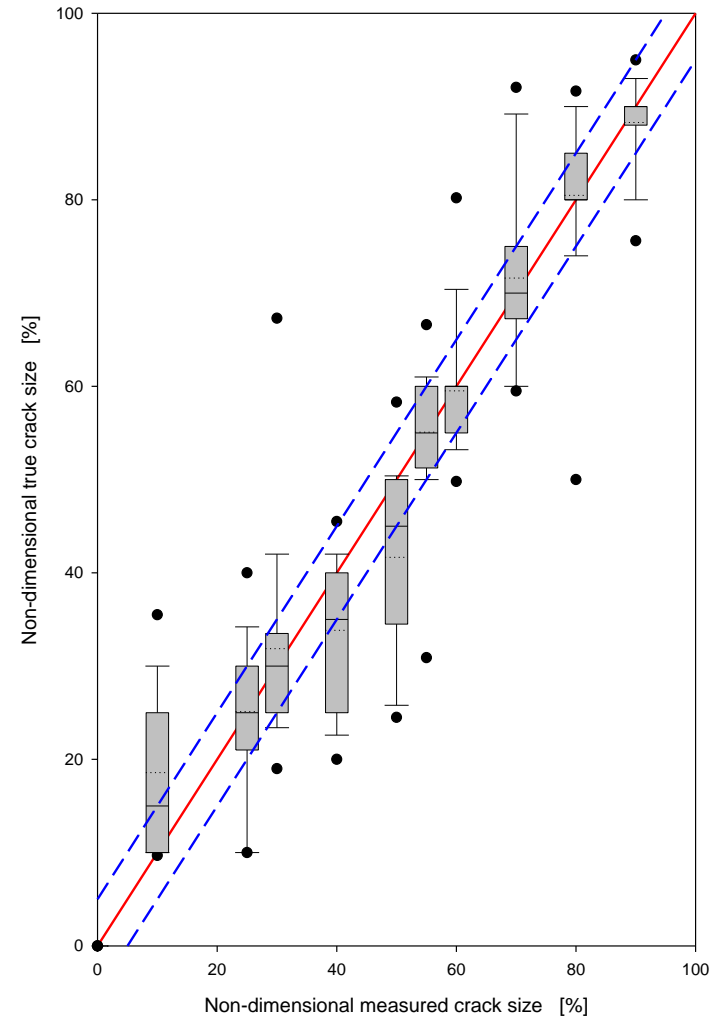
Crack sizing accuracy

- Sizing bars are used in the test
 - Crack sizes reported in 0.05" increments
 - 64 test results generated for different inspectors
- Overcalls: identifying a crack where none exists
 - Exams may contain un-cracked specimens
 - Overcalls trigger an automatic failure of test
 - ▶ More than 10% of inspectors overcall
- Removing these inspectors from the pool does not improve the overall accuracy
 - Summary statistics (mean and standard deviation do not change much due to overcall)



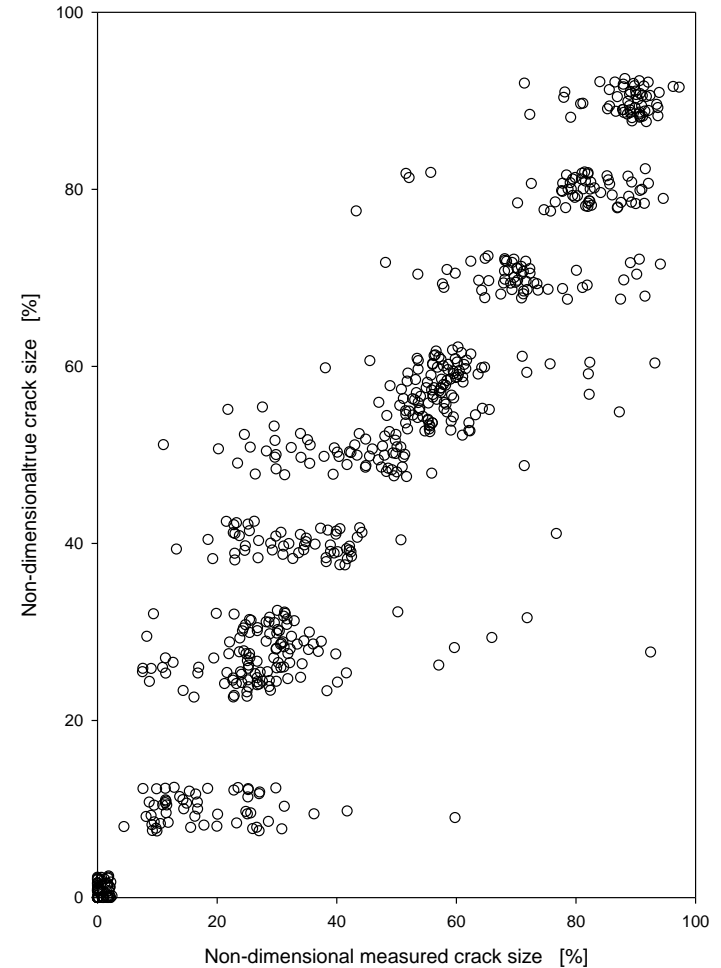
Crack sizing accuracy

- Box and whisker plot
- Median is solid line
 - Median never off by more than 0.05"
- Average is dashed line
 - If above median, skewed distribution with a longer right tail
- 50% of data fall inside the gray box
 - Near middle of crack size range, there is significant sizing inaccuracy, where $\frac{3}{4}$ of the data fall either below or above the correct answer
- Whiskers (error bars extending from the box) indicate the 10th and 90th percentiles
- Individual dots represent the 5th and 95th percentile
 - 90% of data fall inside the dots



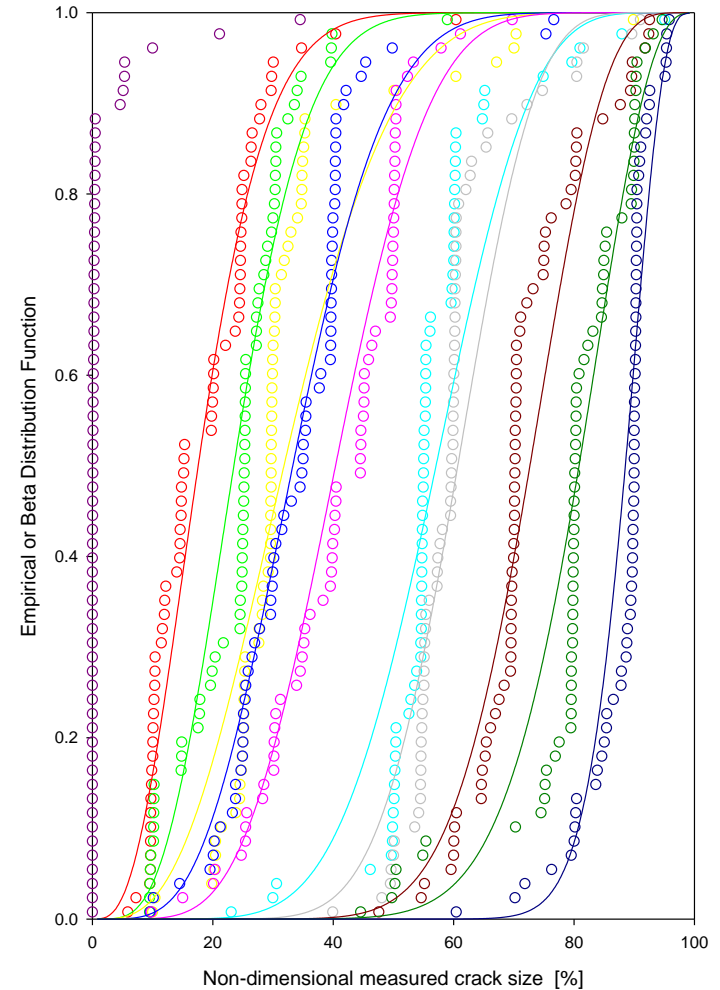
Crack sizing accuracy

- Statistics of crack size measurements:
 - Median value of crack sizes within 0.05" of the true crack size
 - RMS (root mean square) of error is 0.1"
 - 30% has exact size
 - 60% within ± 0.05 "
 - 75% within ± 0.1 "
 - 85% within ± 0.15 "
 - 95% within ± 0.2 "



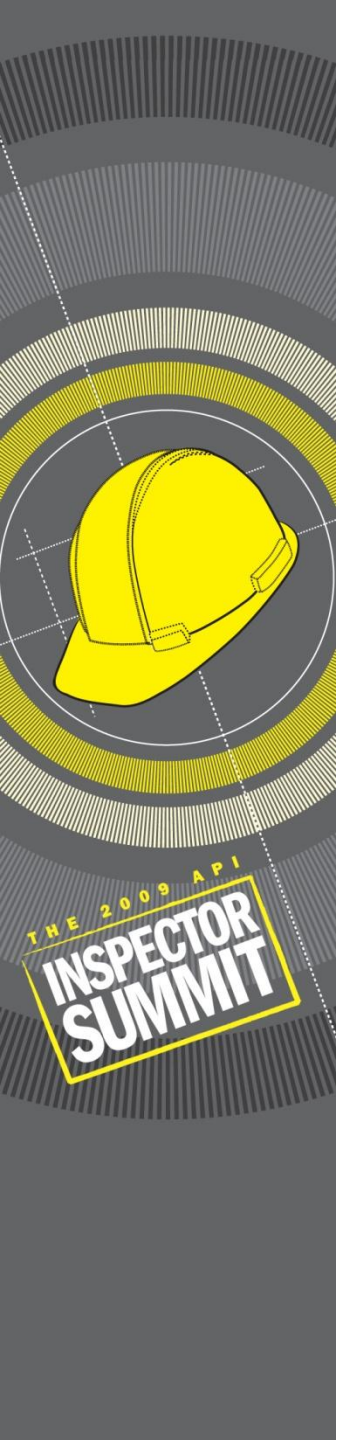
Crack sizing accuracy

- Single statistical distribution model of all data created
- Beta distribution limits max crack size to plate thickness limits
 - Parametric probability distribution model evolves from a long right tail to symmetric to a long left tail as the reported crack size increases from 0 to 100% through-wall



Summary of Crack Sizing Study

- Approximately 10% of inspectors overcall crack where none exists
- Crack sizing accuracy:
 - Results fall within 0.05" 60% of the time
 - Almost Normally distributed for crack sizes near 50% of plate thickness
 - Clearly skewed for crack sizes near 0% and 100% of plate thickness



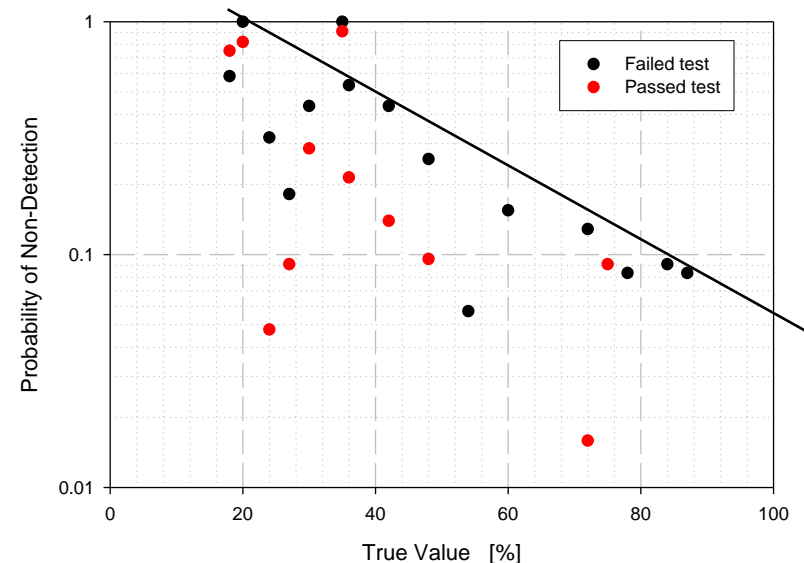
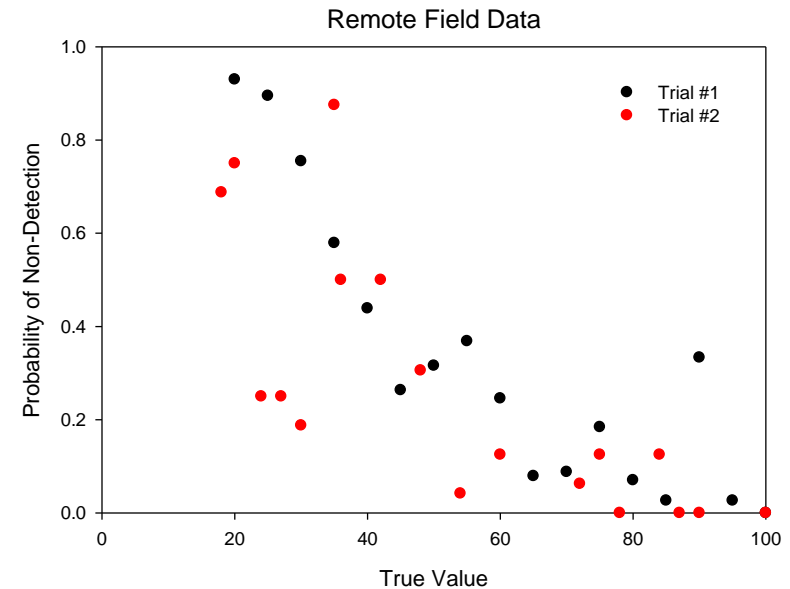
Remote Field Eddy Current Inspection

POD of RFET

- Varies with flaw size
- Better POD for qualified inspectors than for inspectors that failed the qualification test

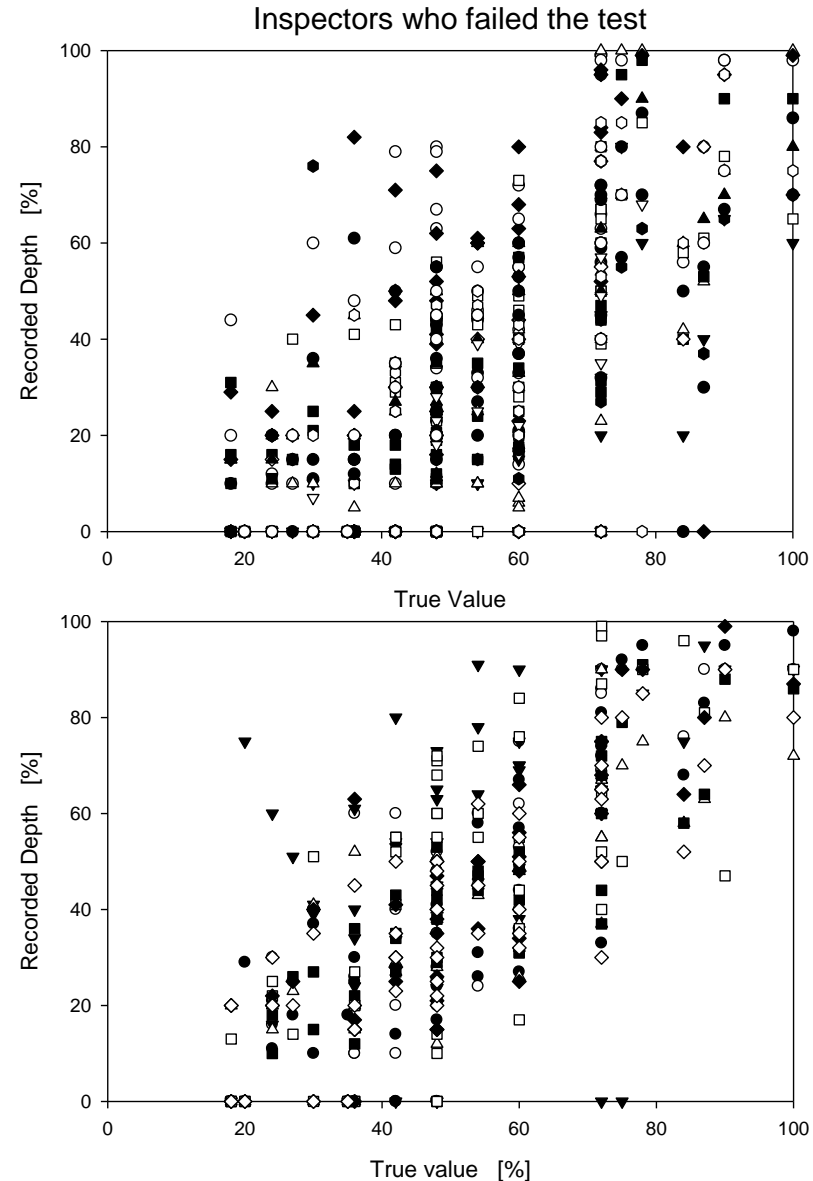
Built a probabilistic model that predicts the probability of detecting a flaw as function of the flaw depth

- Only partial data shown



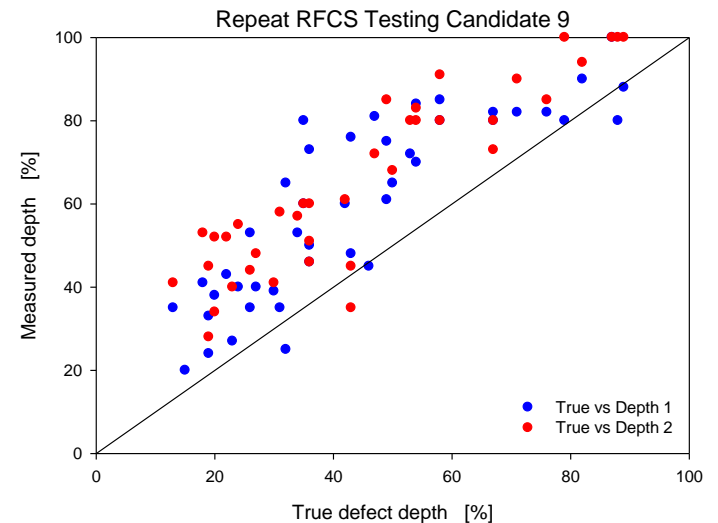
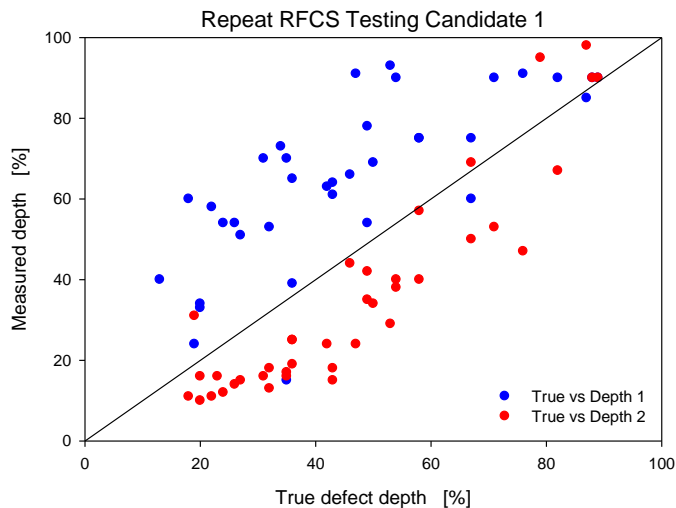
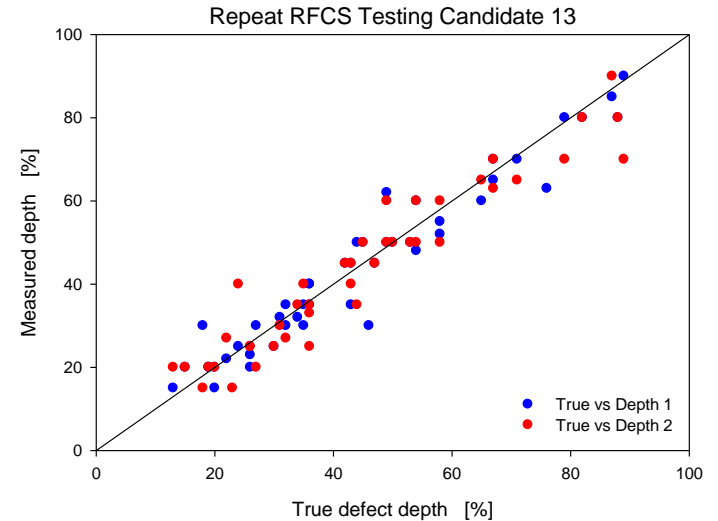
Remote Field Eddy Current Inspection

- Sizing accuracy is virtually identical for inspectors who pass or fail the test
- 80% confidence interval for this configuration almost $[-20\%, +20\%]$
 - A defect reported as 60% through has a 95% confidence interval of $[30\%, 90\%]$
- There is also a depth-dependent bias in the results
 - Deep defects likely overcalled, defects called as shallow defect are often undercalls of deeper defects
 - 95% confidence interval of a defect reported as 40% loss is $[30\%, 70\%]$



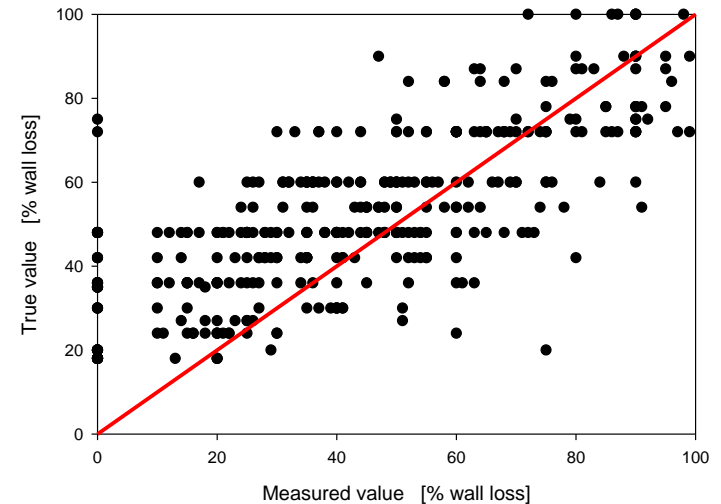
Analysis of Some Repeat RFEC Exams

- Large fluctuation in sizing accuracy between different inspectors
 - Some inspectors are much more accurate than others
- Variation in sizing accuracy within repeat exams of same inspector
 - Accuracy
 - ▶ Both bias and scatter
 - Repeatability



Statistical Modeling of RFET data

- Practical application is “inverse problem”
 - Estimate true value from measured value
- Raw data (black dots) are not centered around red one-to-one line
 - Raw data do not provide an unbiased estimate; clear trend in error values
 - Bias correction (green line) accounts for this
 - ▶ Raw data must be shifted
 - ▶ Bias-corrected data are still subject to uncertainty
- Bias correction improves estimates and reduces overall RMS error from 21% to 14%



Summary of RFET tests

■ Probability of Detection

- Depends on the flaw size
- Qualified inspectors clearly have improved POD

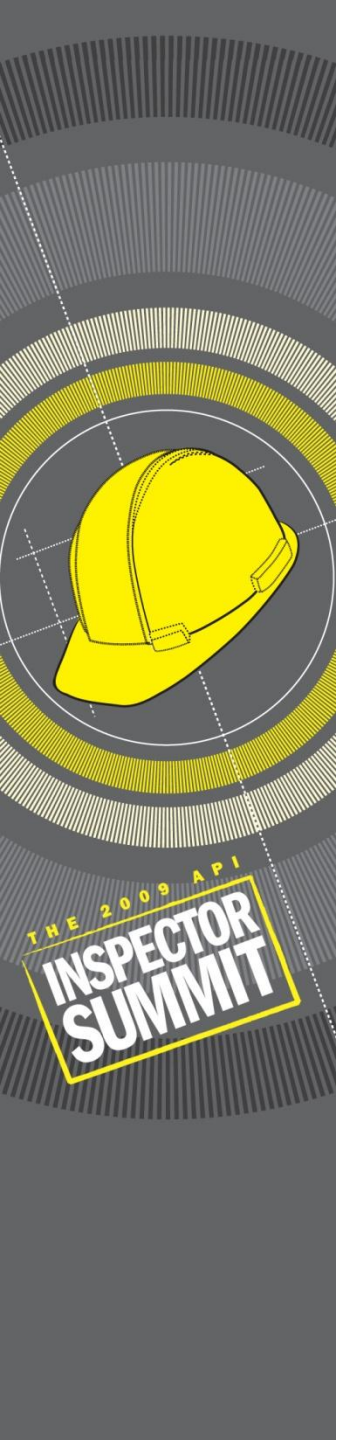
■ Sizing accuracy

- Bias in the measurements:
 - ▶ For large reported values, true defects often smaller than reported
 - ▶ For small reported values, true defects often larger than reported
- Overall significant amount of scatter in the data
 - ▶ A general rule of thumb for 80% confidence bounds is: [-20%, 20%]
- Huge variation of these statistics between inspectors
 - ▶ Some inspectors have no bias and almost no sizing scatter



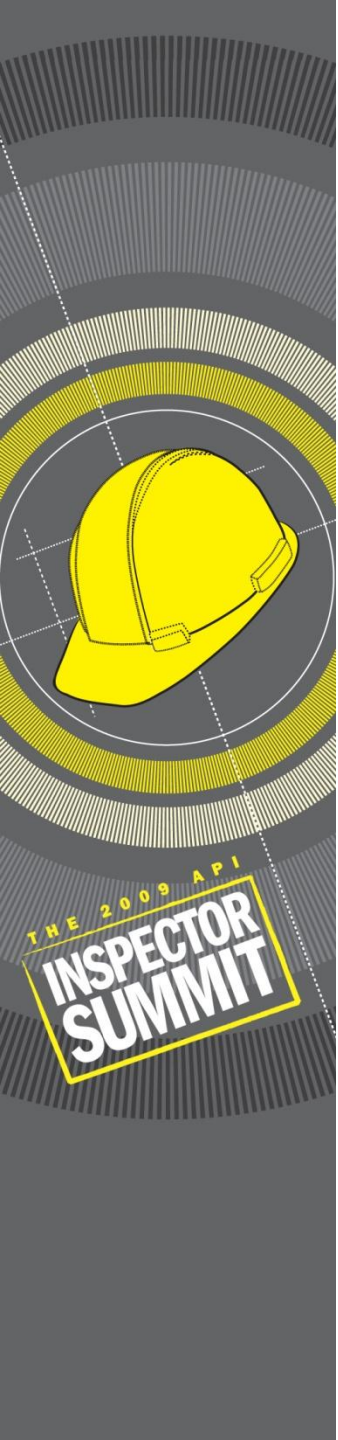
Manual UT inspections

- Single Chevron refinery location
 - 5 plates must be sized
 - Sizing must be called within set bounds
 - ▶ No relative thickness error beyond 10%
 - ▶ Absolute thickness error within 1% of plate thickness
 - Laminations must be identified properly



Manual UT inspections

- 24 inspectors, 5 plates, no repeat exams
 - 120 individual thickness readings
 - 42 readings are within the acceptable bounds and are considered a pass (35%)
 - A little more than 50% of the tested inspectors fail to correctly identify laminations
 - Of all inspectors who correctly identify the laminations, 70% of them have all their thickness sizings outside the acceptable bounds and therefore failed the inspector exam
 - Only 4 out of 24 inspectors passed the complete exam of 5 plates
 - ▶ Passing rate for subgroup that correctly indentifies laminations is identical to passing rate for all inspectors
 - ▶ Therefore, there is no statistical evidence that ability to identify laminations increases likelihood of accurate sizing of thickness



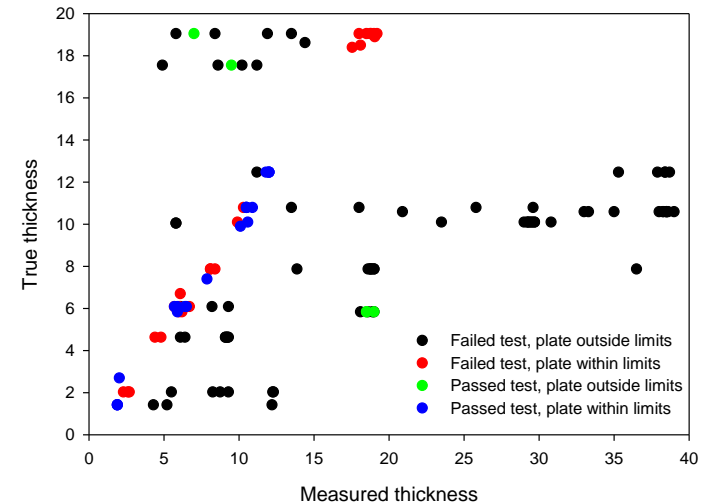
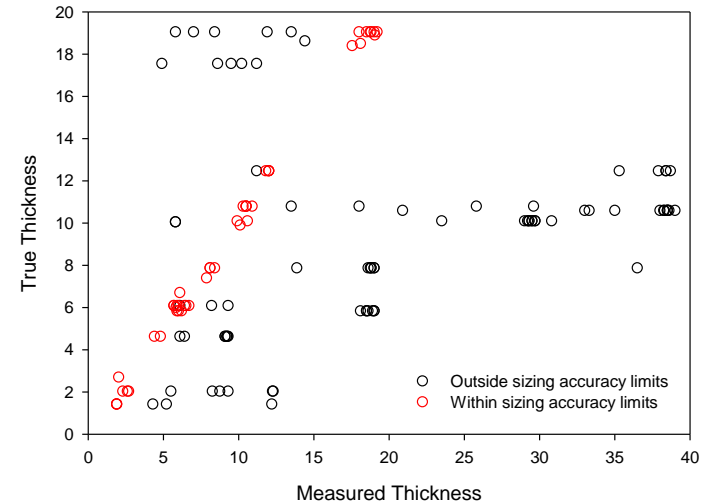
Presentation of Manual UT results

■ Pass-Fail charts

- Data within sizing accuracy limits closely follows one-to-one line

■ Compare correct/incorrect sizing of single plate to overall pass/fail of test

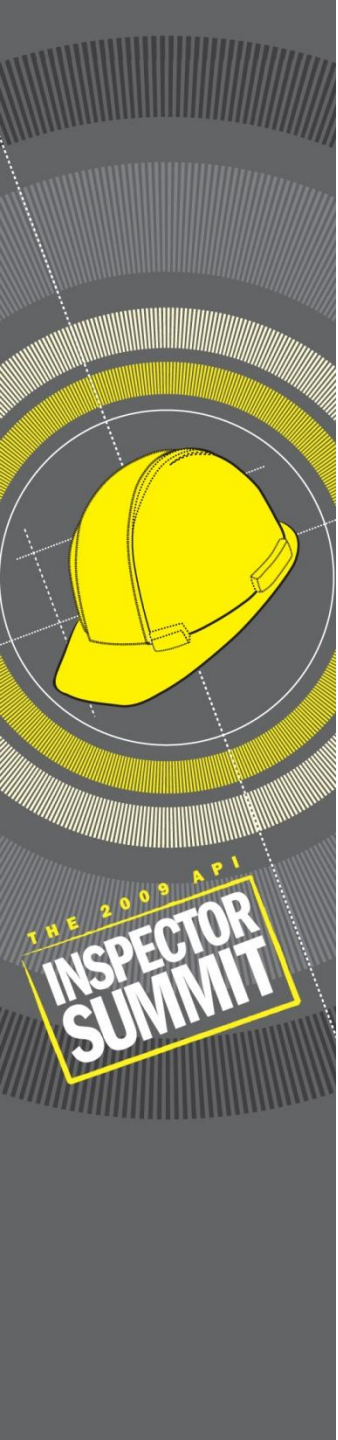
- Green dots indicate failed plates for inspectors who pass overall test
- Red dots indicate correct sizings by inspectors who failed overall test due to either missing laminations or incorrect sizing of too many plates
- Black dots indicate some readings that are grossly erroneous



Manual UT inspections

■ Summary of test results

- Many inspectors fail to recognize laminations
- Most inspectors who passed overall exam still fail one individual plate thickness sizing
 - ▶ There are more accuracy issues with the thicker plates than with the thinner plates
 - ▶ This means that, even if a certified inspector is used, it is still possible that at least some thickness readings are grossly erroneous



More Manual UT thickness tests

- Different test setup

- Previous test consisted of 5 plates with corrosion and/or lamination
- This test setup requires
 - ▶ Measuring thicknesses for up to 10 plates with corrosion and/or lamination
 - ▶ Various plate thicknesses

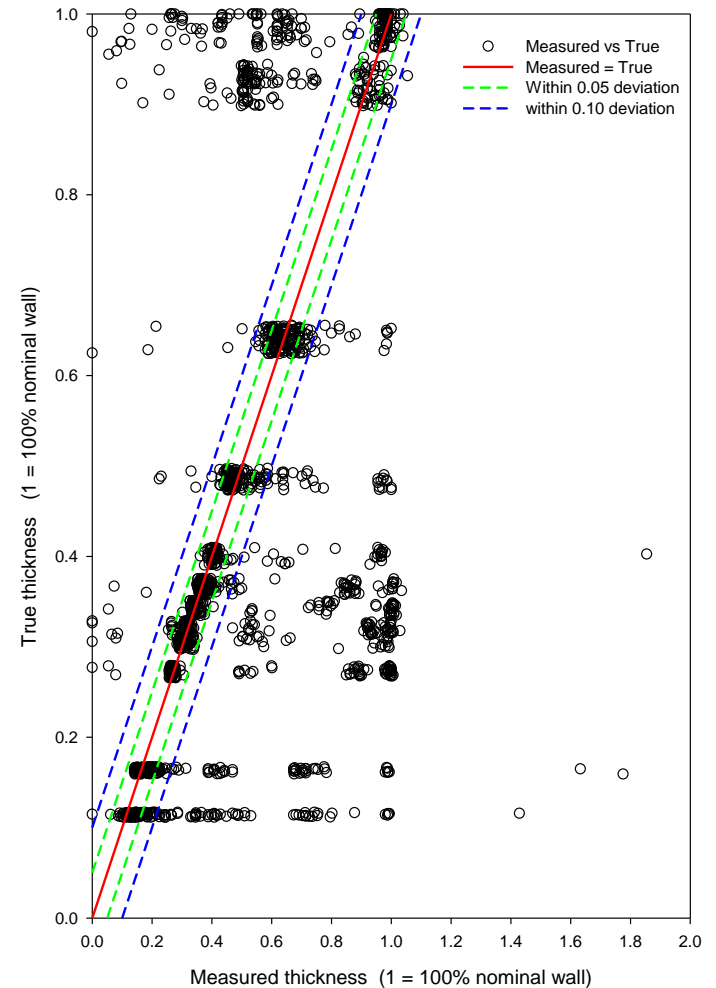
- Multiple inspectors were tested

- Some repeat testing of inspectors



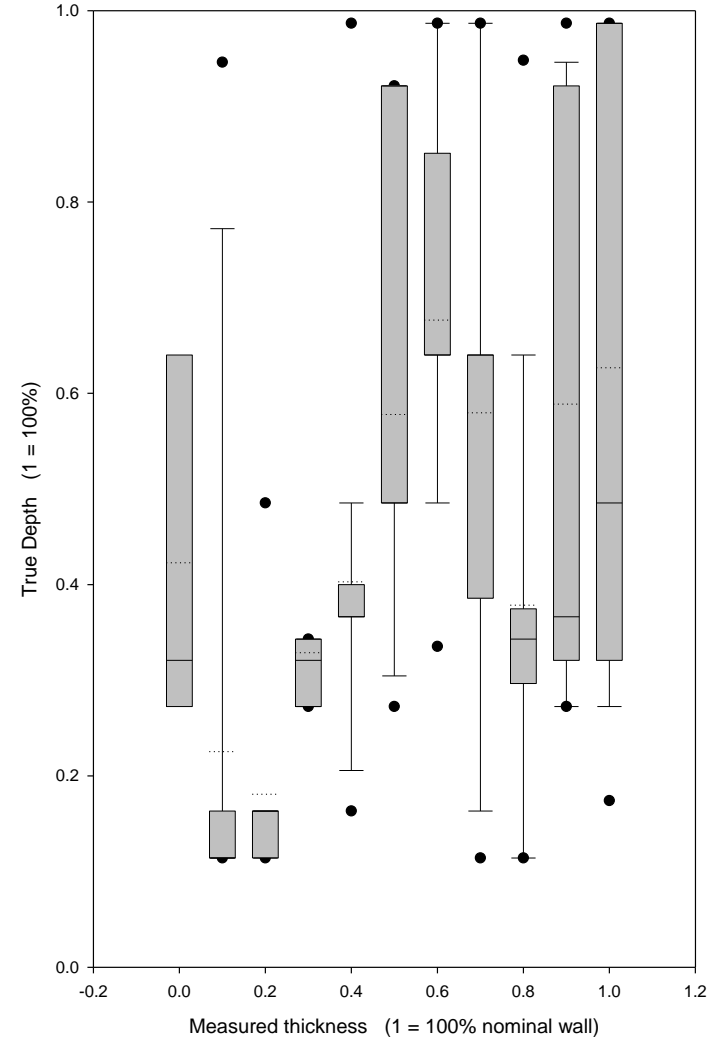
Manual UT inspections

- Many data fall within the 5% or 10% deviation of the nominal thickness
 - The green and blue bands do not represent the actual pass-fail criterion, which is in part based on relative sizing accuracy
- The data outside of these limits can literally be anywhere
 - Measured thickness values range between 0 and 100% remaining wall
 - A few of the reported thicknesses even exceed nominal wall thickness by as much as 85%



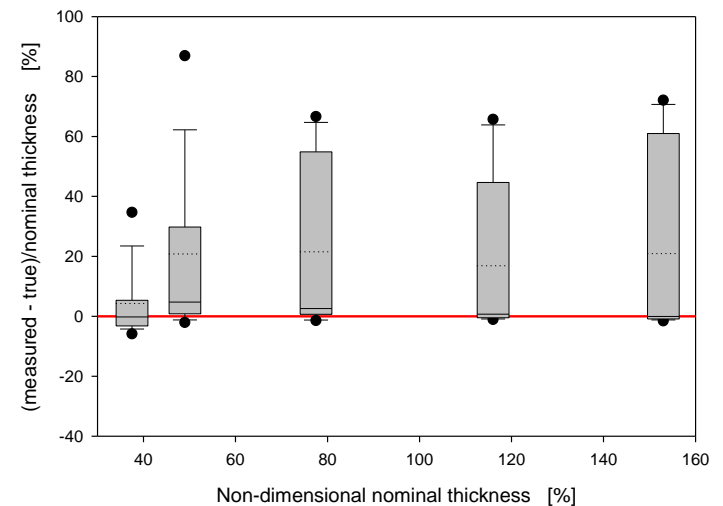
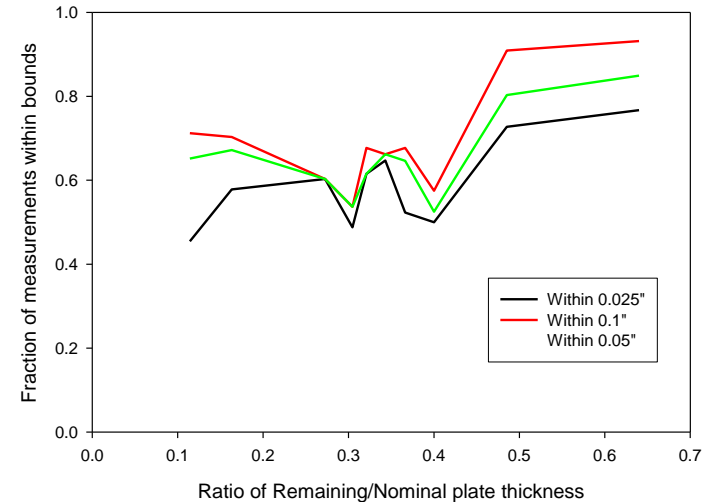
Can we use this for prediction?

- In practical application, the true thickness is not known
- Practical question is how well the true thickness can be inferred from the reported (i.e. measured) thickness, and what the possible errors are on a single measurement
- Box and whisker plot
 - Median (solid line) and Average (dashed line) are typically well off
 - 50% of data fall inside the gray box
 - ▶ There is significant sizing inaccuracy,
 - ▶ $\frac{3}{4}$ of the data fall either below or above the correct answer
 - One particular thickness reading really does not give a strong indication of what the true remaining thickness really is
- One would need at least 3 or 4 independent, repeat measurements to conclusively say what the "correct" value would be
 - Need to assure independence of assessments → human factor



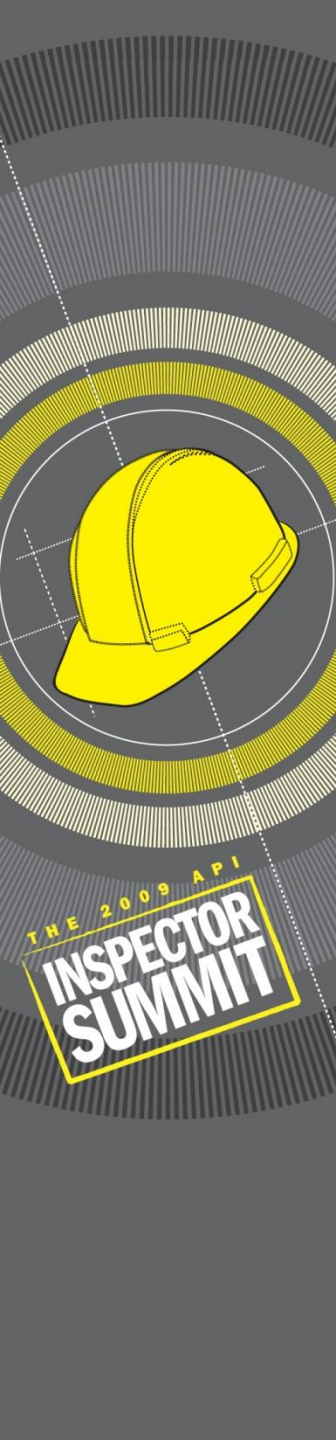
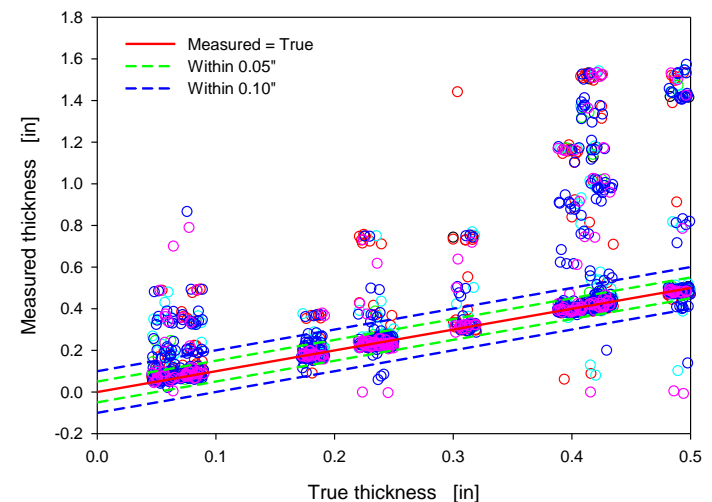
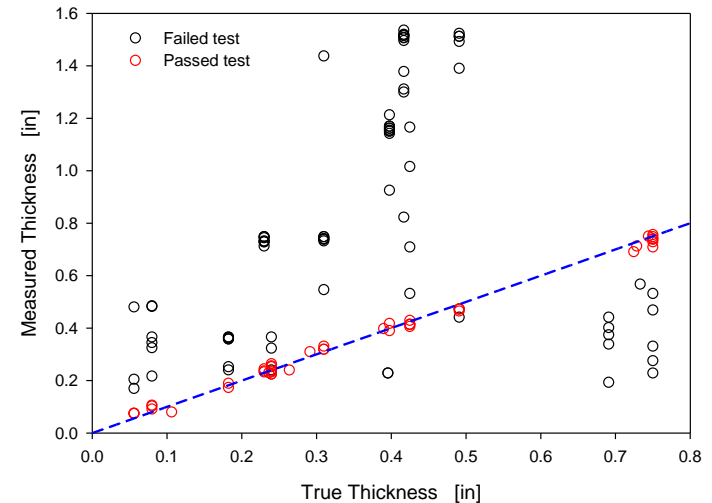
Can we use this for prediction?

- Accuracy of manual UT measurements
 - For small remaining plate thickness, accurate in approximately 60% of all cases
 - When remaining plate thickness is more than half of nominal, accurate in approximately 80% of all cases
- Average value of relative error is approximately 20% of nominal thickness
 - The average error (dotted line) is MUCH larger than median error (solid line) and is heavily influenced by the presence of grossly erroneous results



Manual UT inspections: comparison of two test batches

- Remarkable consistency in the statistics between 2 batches of tests with different test setup
 - Top chart: 1st setup
 - Bottom chart: 2nd setup
- Similar behavior:
 - Many readings are very accurate
 - Grossly erroneous readings exists



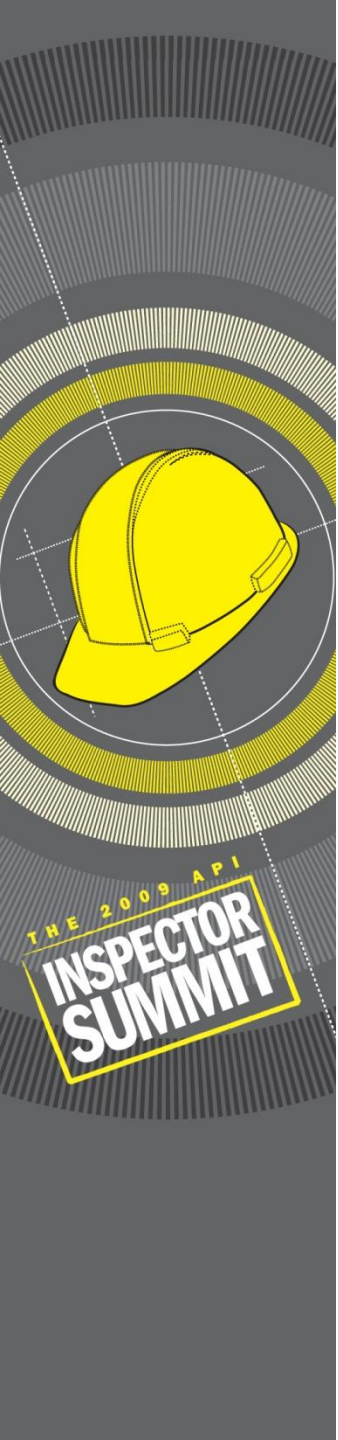
Manual UT: Summary of Observations

- Total number of exams administered in last three years – **313**
- Total number of successful exams – **140 (45%)**
 - Not necessary to correctly size all flaws to pass exam
 - Total number of inspectors that correctly size all flaws – **21 (or 7%)**
- Total number of failed exams – **173 (or 55%)**



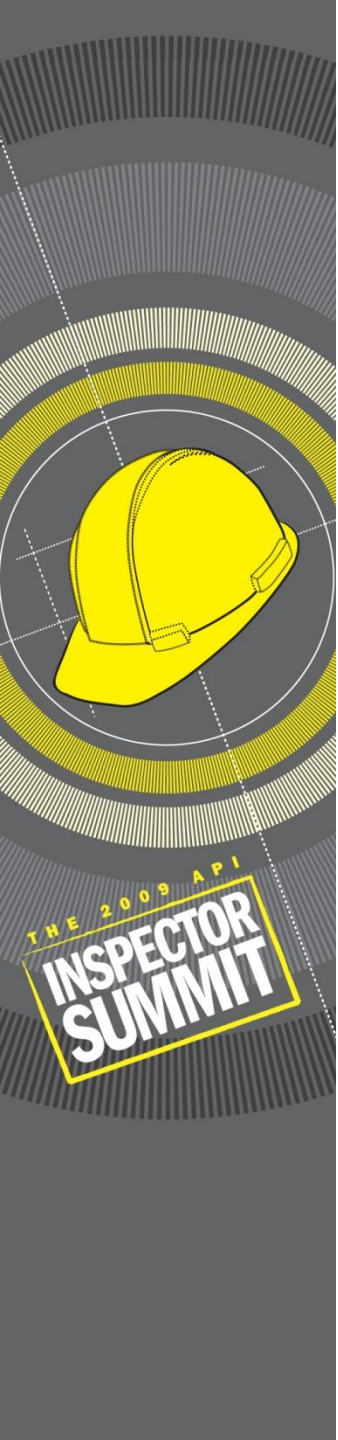
Manual UT: Summary of Observations

- Similar observations for both test setups give credence to the generality of the observations
- UT thickness readings have potential to be very accurate
- About 25% chance of UT reading being completely bogus
- Ability to identify laminations does not improve sizing accuracy



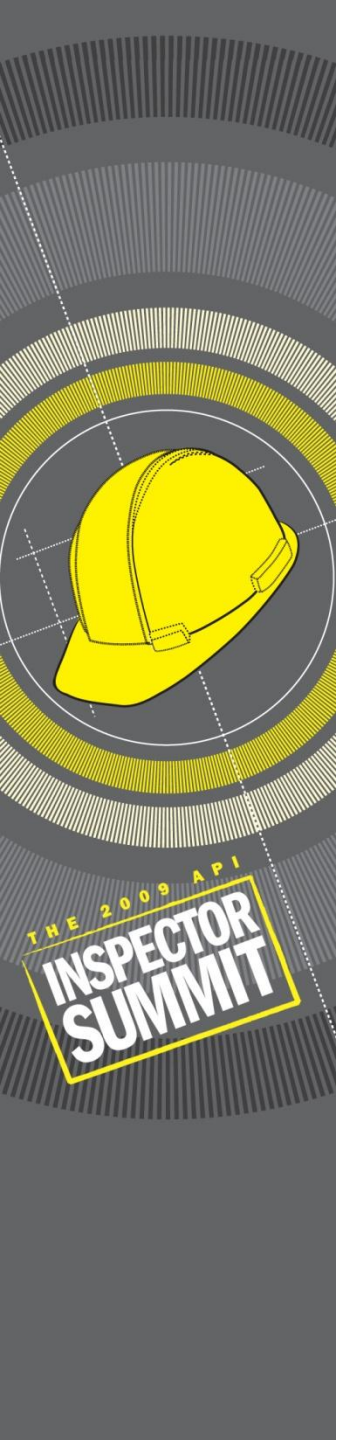
Can we use this for prediction?

- In all tests, the true thickness is known
 - Real application of NDE is to infer true thickness from measured values
- Consider the practical case where we are interested in whether the true remaining thickness falls below 20% of nominal
- Test results indicate that even when inspector reports 70-80% remaining wall thickness, there is actually a 1-in-5 chance that less than 20% of wall thickness remains
 - If you receive 4 or 5 UT readings from a generic UT inspector, statistically speaking there is better-than-even odds that at least one of them is grossly incorrect
 - Chevron tests screen out many of these inspectors



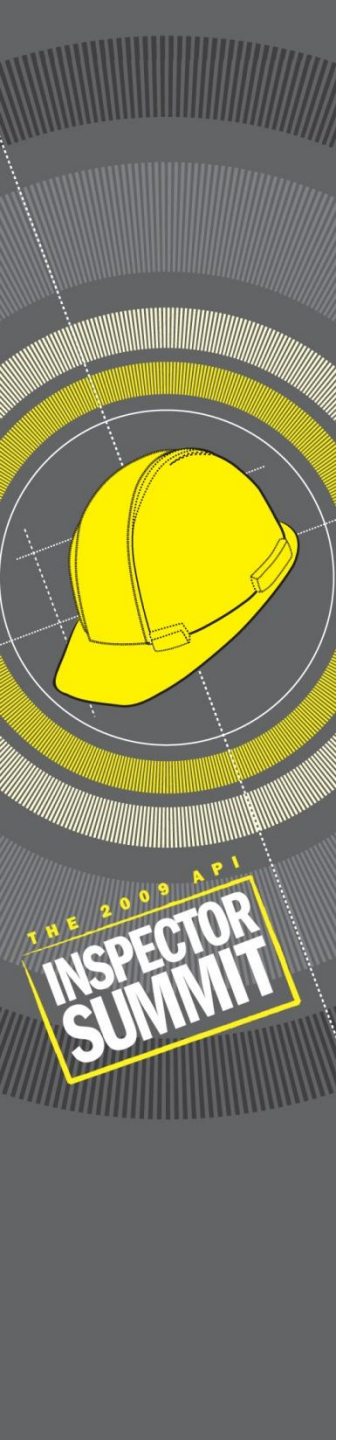
Test and On-Site Performance

- From many points of view, the test results represent a best-case scenario
 - Inspectors know this is a test and that they will be graded
 - Well lit and comfortable environments with small specimens
 - Ample time to perform inspection and interpret data
- On-site performance is likely to be even worse



Going forward

- This presentation quantified the accuracy of several NDE techniques, as measured by inspector performance demonstration exams
 - For given true size
 - Quantify variability of measured values
- Immediate application consists of “inverting error-bands” to infer best estimate and confidence bounds for true value when a single measured value is reported
- This opens up new avenues for applications in asset integrity management



Further use of the statistical models

- Beyond the immediate use of quantifying the confidence intervals of NDE inspections, the results can also be used as an integral part of asset integrity management:
 - Compute the likely effectiveness of repair or mitigation actions and strategies
 - ▶ Did we really find/repair all flaws of concern?
 - ▶ What is the remaining likelihood of failure?
 - Set appropriate action level targets for asset integrity management
 - ▶ What flaw size causes concern and how do POD and sizing uncertainties impact this “red flag” level?
 - ▶ Selection and trade-off between inspection techniques



Summary and Conclusions

- Crack sizing accuracy
 - Median crack size is within 0.05", the RMS of the error is 0.1"
 - 7 out of 64 (11%) inspectors made false calls in non-cracked specimens and were excluded from this analysis.
 - 60% of the remaining inspectors are within 0.05"
 - 75% of the remaining inspectors are within 0.10"
- Remote field measurements
 - Significant scatter on the measurements, 80% confidence bounds are $\pm 20\%$ wall thickness
 - Large fluctuations in sizing accuracy between inspectors
 - Some inspectors have consistent sizing performance, others are highly variable
- Manual UT:
 - Almost 60% of the inspectors fail to properly identify laminations
 - The sizing accuracy of inspectors is very good in 70% of the cases, irrespective whether they pass the lamination test or not.
 - In approximately 25% of the cases the measured remaining wall thickness is grossly over- or underestimated.





Questions ???

Contact information

Chevron ETC

RBI & Inspection Team

Technical team leader:

Albert van Roodselaar

avan@chevron.com

(713) 754-4284

