# Is That Flaw Really There?

Robert M. Ryan

Twining, Inc.
2883 E. Spring Street, Suite 300, Long Beach, CA 90806
(562) 426-3355; fax (562) 426-6424; e-mail rryan@twininginc.com

## Abstract

Bayesian inference is a statistical methodology wherein probabilities are calculated and modified based on "priors," that is, the current state of knowledge. We utilize Bayesian inference methodology to assess the reliability of ultrasonic testing of weldments in accordance with the California Building Code and American Welding Society (AWS) D1.8 – Structural Welding Code-Seismic Supplement (a supplement to AWS D1.1 – Structural Welding Code-Steel). Utilizing reasonable estimates of the likelihood that a skilled welder will install a weld with a rejectable UT indication and the requirements for qualification to perform ultrasonic testing as specified in Annex F (Normative) of AWS D1.8, we show that a rejectable indication reported by a technician qualified at the minimum level of Annex F has approximately a 75% likelihood of being a Type 1 error. That is, under this set of assumptions (i.e., this set of priors), there is approximately a 75% probability that no rejectable defect exists at the location reported. We then analyze some of the factors that lead to this high probability of so-called "false calls" and the economic consequences of such reports. We briefly describe possible remedies to this situation.

## Introduction

Since 1969, the American Welding Society has recognized ultrasonic inspection as one component to determine acceptance or rejection of welds in steel structures[1]. For welding in steel building structures governed by the American Welding Society AWS D1.1-Structural Welding Code, Steel[2], the criteria for acceptance and rejection utilize the measured length of detected ultrasonic reflectors and a comparison between the amplitude of the reflector and that of a standard reference reflector (the former being determined by the so-called "6 dB drop" method and the latter being modified for attenuation due to length of sound path). Then, based on the location of the reflector in the weld, the proximity of other reflectors, and the thickness of the weld, the ultrasonic technician determines from a table whether the reflector will be accepted or rejected.

In the aftermath of the 1994 Northridge Earthquake, the Structural Engineers Association of California (SEAOC), the Applied Technology Council (ATC), and the California Universities for Research in Earthquake Engineering (CUREe) formed the FEMA funded SAC Joint Venture partnership[3] to investigate all aspects of steel moment frame construction in light of the multiple failures of welded connections discovered by our firm and others.

Pursuant to this investigation, in 2000, FEMA published FEMA-353[4], "Recommended Specifications and Quality Assurance Guidelines for Steel Moment-Frame Construction for Seismic Applications." In this document, Part I Appendix E, "Supplementary Ultrasonic Technician Testing," describes a required (that is, to the extent that the designer requires compliance with the document) examination for determining the ability of UT technicians to accurately locate and characterize reflectors in welds. That requirement has been slightly modified and carried over into AWS D1.8 – Structural Welding Code Seismic Supplement in Annex F of that document[5]. This examination requires that the technician evaluate a series of welds of differing configurations with known flaws (that is, known to the examiner, not to the technician undergoing qualification testing) of various types. He or she is scored based on successfully locating and characterizing actual flaws and on not reporting "false indications," that is, not reporting flaws at locations where no flaw exists. The technician must demonstrate the ability to detect a minimum of 87% of the actual flaws in the series of test pieces. This number, D, is reported as the ratio of detected flaws to total flaws. The ratio must be at least 0.87. The ratio of false indications to total indications, F, cannot exceed 0.15 or 15%. Finally, combining the detection rate and the false indication rate, the final rating score, calculated as $½ \cdot (1 + D - F)$ must be at least 0.9 or 90%. This score is termed the "overall rating" and is reported as R. Thus, a technician cannot perform minimally on both the D and the F scores, this would result in an R score of 0.86.

Then, in AISC 360, "Specifications for Structural Steel Buildings" Specification N, Quality Control and Quality Assurance, Section N5.5, Nondestructive Testing of Welded Joints, Subsection 5e., Reduction of Ultrasonic Testing Rate,[6] for welded joints that otherwise require an ultrasonic testing rate of 100%, the specification permits the testing rate to be reduced to 25% for those welders who demonstrate a rate of ultrasonic rejects not exceeding 5%.

Thus, we have the following useful information to construct a model: 1) A welder that the applicable codes will, arguably, characterize as skilled (based on the implication of allowing a reduction in testing rate) may have a rate of installing welds with rejectable defects as high as 5%; 2) An ultrasonic technician must be able to detect actual rejectable flaws at a rate of 87%; 3) An ultrasonic technician may produce type 1 errors, that is, false positives or "false indications" at a rate as high as 15%; and 4) the technician's R score as calculated above must be at least 0.9 for the technician to be characterized as sufficiently skillful to evaluate welds in demand critical seismic connections.

## Analysis

Bayesian inference utilizes Bayes' theorem to update the probability of a hypothesis as evidence accumulates. Such evidence is referred to as "priors" and represents the state of knowledge prior to the evaluation. The iconic examples of the usefulness of this method of reasoning about probabilities are in medical screening of various types (drug screening for athletes, cancer screening, etc.[7]) Multiple examples are available wherein groups misinterpret positive test results, overestimating the probability of an actual positive result. This may be characterized as failing to distinguish a positive test result vs. the event of the [flaw, cancer, banned substance, etc.] actually being present.

Bayes' Theorem takes prior knowledge into account in evaluating the probability that a positive test result means an actual positive event. In the case of an ultrasonic technician who reports a rejectable indication in a weld, we'll look at the prior knowledge indicated above, i.e., the likelihood that a given weld installed by a skilled welder has an actual rejectable flaw is 5% and the likelihood that a report by the technician of a positive test result indicates that an actual rejectable defect exists is 85%.

Bayes' theorem is, mathematically, $\Pr(F|I) = \frac{\Pr(I|F) * \Pr(F)}{\Pr(I|F) * \Pr(F) + \Pr(I|{\sim}F) * \Pr({\sim}F)}$ (1)

Where:

- $Pr(F|I)$ is the probability that an actual flaw exists, given a rejectable indication. *Read as "the probability of F given I."*
- $\Pr(I|F)$ is the probability that a rejectable indication will be found, given that a rejectable flaw actually exists. *Read as "the probability of I given F."*
- $Pr(I|{\sim}F)$ is the probability that a rejectable indication will be reported where no rejectable flaw exists. *Read as "the probability of I given not F."*
- $Pr(F)$ is the probability that a rejectable flaw exists in a given weld. Read as *"the probability of F."*
- $\Pr({\sim}F)$ is the probability that no rejectable flaw exists in a given weld. Read as *"the probability of not F."*

For our purposes, let $Pr(I|F) = 0.95$; $Pr(I|{\sim}F) = 0.15$; $Pr(F) = 0.05$; and $Pr({\sim}F) = 0.95$ (a technician with these scores will pass the requirements of AWS D1.8, Annex F). Suppose a UT technician who achieved these scores, examining a weld, reports a rejectable indication. Plugging these numbers into the above equation, we find that $Pr(F|I) = 0.25$, that is, there is a 25% probability that there is actually a rejectable flaw at the location reported.

The table below shows how this non-intuitive result arises. For a sample of 10,000 welds using the probabilities above, the table entries show the number of welds in each applicable category. A reported flaw will fall into the left column and, because the number of actual flaws in the 10,000 welds (actual events, using the terminology above) is such a small fraction of the total, the actual flaws reported are overwhelmed by the false indications.

| | Technician Reports Flaw | Technician Reports No Flaw |
|---|---|---|
| Weld Has No Flaw | 1425 (Type 1 Error) | 8075 |
| Weld Has Flaw | 475 | 25 (Type 2 Error) |

Now this is a welder who minimally achieves a sufficiently low rejection rate to qualify for a reduction in testing frequency and a technician who minimally passes the supplemental UT examination. But, manipulating the equation, we find that a technician who achieves a D, detection, score of 0.95 and an F, false indication score, of 0.05 and reports a flaw in a weld produced by a welder with an overall reject rate of 10% (and on projects in our experience, this welder will likely be removed by his or her foreperson) suffers a 32.1% likelihood that there is no rejectable defect at the reported location, i.e., there is a 67.9% probability that the reported indication is an actual rejectable defect. If this welder had a reject rate of 5%, the probability would be 50% that the rejectable defect actually exists – a coin toss!

For the ultrasonic technicians employed by our firm and qualified in accordance with AWS D1.8 Annex F, the mean scores in each category were: D = 0.917; F = 0.034; R = 0.947. Using a reasonable probability of installation of a weld with a rejectable defect of 2.5% as performed by a highly skilled welder, the mean score for a qualified technician at our firm would imply a 40.8% probability that a reported rejectable indication actually exists, a 59.2% probability that there is no actual rejectable defect. A probability of a flawed weld of 5% implies a probability of 58.7% that a reported rejectable indication actually exists.

## Economic Implications

It's clear that a Type 1 error (that is, reporting a false indication) is preferable to a Type 2 error (failure to locate an actual flaw) in a seismic or demand critical weld. Nevertheless, there are significant economic implications of such false indications. Typically, the welder is required to gouge out weld metal at the location of any reported rejectable indication. This involves completely different equipment than that used to install the weld.

We've estimated that, for a weld connecting a W36X150 beam with flange thickness of 0.94" where a rejectable indication with a length of 1.5" is found at a depth of 0.5", that it takes approximately 3.5 hours to redeploy equipment, gouge out weld metal to that depth and length, to replace the removed weld metal, and to retest the affected area. Based on this, the estimate below includes the costs of time to redeploy equipment, provide preheat, and other logistical issues as well as the actual gouging, welding, and testing. It does not include second order costs such as welding and arc gouging consumables, fuel for power supplies for the welding equipment and the compressor (as hinted at above, the weld metal will likely be removed using "air carbon arc gouging" which requires a compressor), and others. It also does not include any costs of delays for those welded joints that are reported to have rejectable indications and are on the critical path for construction of the structure. Finally, it does it include intangible costs such as the welder's aggravation and the loss of credibility of the technician when the welder is unable to find visual evidence of the reported defect and the foreperson's evaluation of the welder.

In California, a welder either in the applicable Labor Union or working for his or her craft's "prevailing wage" will cost his or her employer approximately $90/hour and so the total cost to remove and replace the reported flaw will be on the order of $315. Some will cost less, but many will cost far more, depending on project logistics, defect size, beam weld thickness, column flange thickness (which typically determines preheat requirements), etc. – the cost distribution will have a "long tail." Roughly, each rejected weld may cost about ½ of a person day!

## Corrective Measures?

What can be done? Mathematically speaking, we can increase $Pr(F|I)$, the probability that an actual flaw exists when an indication is reported, either by increasing the numerator of the fraction in equation (1), decreasing the denominator, or some combination of these. If we evaluate the fraction, we find that one way to increase $Pr(F|I)$ is to increase $Pr(F)$, the probability that a flaw actually exists in a weld by utilizing less highly skilled welders. This is, perhaps, not an optimal solution.

$Pr(F|I)$is relatively insensitive to $Pr(D)$, that is, the probability that a reported indication represents a real flaw is relatively insensitive to the probability of detecting an existing flaw. This leaves reducing $Pr(I|{\sim}F)$, the probability of reporting a flaw where none exists, as the most effective method of reducing the economic impact of the reporting of false indications. And, if we agree that type 1 errors are preferable to type 2 errors, it would be a mistake to modify procedures or requirements so as to uniformly reduce reported rejectable indications. Rather, we must provide equipment, education, training, and procedures that assure a high probability of detection (POD) for actual rejectable flaws and yet a reduction in false indications.

And yet, if we look at a realistic situation where a welder has a rate of rejected welds of 2.5% and a technician has a POD of 95%, to reduce the probability of a false indication below 50% in the field, we must have a technician whose rate of false indications is 2.4%, about one sixth of the maximum rate allowed for passing the requirement for certification to Annex F of AWS D1.8. And, as stated above, the mean F (false indication) score of our technicians is 3.4%, far better than the requirement of Annex F of AWS D1.8.

We are currently engaged in research to determine what factors lead to a false indication. We are looking at equipment, technique, joint configuration, and procedures. This work is continuing, and we hope to have actionable information in the intermediate future.

## Conclusions

We have shown the practical result of the current state of the art and practice of ultrasonic inspection of welded joints in steel structures installed, inspected, and ultrasonically tested in accordance with current codes and specifications leads to a strong likelihood that a report of a rejectable indication by a qualified ultrasonic technician represents a "false indication," that is, that no true rejectable flaw exists at that location. We have also shown that the impact of such "false calls" (to use the trade vernacular) can have a significant economic impact for subcontractors, contractors, developers, and owners. Finally, we have described research efforts underway at our firm to find effective ways to minimize the incidence and impact of false calls.

---

[1] American Welding Society, D1.0:1969 Code for Welding in Building Construction, Appendix C.

[2] Clause 6, Inspection. (2015). In AWS D1.1/D1.1M:2015 Structural Welding Code-Steel (2015 ed.). American Welding Society.

[3] SAC Mission Statement. (n.d.). Retrieved August 15, 2015, from http://www.sacsteel.org/project/index.html

[4] Recommended Specifications and Quality Assurance Guidelines for Steel Moment-Frame Construction for Seismic Applications. (2000). Federal Emergency Management Agency (FEMA).

[5] Annex F (Normative) Supplemental Ultrasonic Technician Testing. (2016). In AWS D1.8/D1.8M:2016 Structural Welding Code-Seismic Supplement (2016 ed.). American Welding Society

[6] Specification for Structural Steel Buildings, American Institute of Steel Construction (2016 ed.).

[7] Manrai AK, Bhatia G, Strymish J, Kohane IS, Jain SH. Medicine's Uncomfortable Relationship With Math - Calculating Positive Predictive Value. JAMA Intern Med. 2014;174(6):991-993. doi:10.1001/jamainternmed.2014.1059