"You can... never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant."\(^1\)

Sherlock Holmes was describing a situation with which Medical Directors must deal on a daily basis. Ultimately, the job of the Medical Director is to assist in the risk classification process; i.e., "foretelling" into what risk group or "average number" an individual should be assigned for mortality. In order to do this, the Medical Director must understand the statistical principles underlying standard mortality and then develop a consistent approach to determining an individual's increase in mortality.

Once a medical impairment has been identified in an individual, to classify the risk, the impairment can be conceptually categorized as one that is correlated with a single pathological entity such as cystic fibrosis or one that represents an abnormality such as elevated liver enzymes which may be caused by a number of underlying pathological conditions.

Determining appropriate ratings for the first type of impairment is generally straightforward (though not always easy). If studies of insured populations with the impairment exist, then the mortality of those with the impairment can be compared to the mortality of a standard insured population to determine the degree of substandard mortality. If no insurance mortality studies exist, then studies from the medical literature may be found and the mortality methodology can be applied.\(^2\) Once the assumed mortality is adjusted for an appropriate mortality benchmark, ratings can be determined. If there are no studies because of the rarity of the disease, then experience with a more common impairment with a similar pathophysiology may be chosen to make a reasonable "guesstimate" by analogy. For many impairments of this type, the background research has been done and is presented in underwriting manuals. The task then becomes matching the proposed insured's characteristics to those of the appropriate section of the manual to determine the ratings.

With the second type of impairment, the situation is more complex because the underlying pathological condition causing the abnormality is not known. In some cases, such as certain electrocardiographic findings (which may be normal variants or caused by cardiac disease), there are adequate insurance studies on those isolated abnormalities (e.g., minor T wave changes) to be able to determine a rating.\(^3\) If no studies of insured lives exist, it is unusual to find a study from the medical literature to analyze because most medical studies look at specific diseases and not isolated abnormalities.

If there are no clinical or insurance mortality studies available on a given abnormality, such as an isolated elevated Gamma-Glutamyl Transpeptidase (GGTP) result, then Medical Directors must have an approach to analyze these abnormalities. Although this analysis may be used for physical exam findings (e.g., rhonchi) or findings on paraclinical data (e.g., x-ray abnormalities), the evaluation of abnormal laboratory results will be the main topic of this discussion.

The approach requires answers to several questions:

1. Is the result abnormal?
2. What is the significance of the abnormality?
3. What should be done about the abnormality?

Although "abnormal" can be defined in many ways, the initial consideration is if the result lies outside the "normal range" for the laboratory performing the test. The reference value of normal that accompanies most laboratory tests is the range of test results within two standard deviations around the mean test value in a population without disease.\(^3\)

This definition has limitations because it assumes that laboratory values follow a "normal" or Gaussian distribution. Although this assumption is generally not valid, using the "normal range" does serve as a reasonable starting point.\(^4\) If the result falls outside these limits then the Medical Director must proceed to answer the next two questions.

This involves examining the specific test on multiple levels:

1. Analytic level
2. Diagnostic level
3. Operational level
4. Decision Making level.\(^5\)

The analytic level is concerned with the technical factors of the test's performance. On the analytic level, two variables are important, precision and accuracy. They determine the reliability of the number. Precision refers to the agreement between repeated measurements performed in the same way. It measures test reproducibility so that precise tests have little variation if measurement of the same sample is repeated. Accuracy of a test represents agreement between the measured quantity and the true value.\(^6\)
Imprecision is caused by random error. This may arise from small changes in procedure; innate diurnal changes in metabolism; or age, sex, and race differences. Inaccuracy is caused by systematic error, i.e., error that can be attributed to a specific etiology such as the laboratory instrumentation, reagents, specimen labeling, or interfering substances. A laboratory test must be precise to be accurate. However, accuracy is not required for precision, that is, the same result may be obtained on repeated testing, but have no relationship to the true value. These types of errors are additive and produce the overall analytic variability.\(^7\)

Further, there is a "multiplier effect" if many tests are done at one time. Independently, each test has a 95% chance of being normal (this is the range of the mean and 2 standard deviations). For \(n\) number of tests the probability that a completely normal person will have all normal results on all \(n\) tests is the probability of one test being normal times the probability of the next test being normal, i.e., 0.95 times 0.95 times or 0.95 to the \(n\)th power. This means that if there are 20 tests there is a 36% probability that all tests will be normal.\(^8\)

To help decide if an abnormal result is due to analytic variability, the test must be examined on the diagnostic and operational levels. The diagnostic aspects of a test are its sensitivity and specificity. The sensitivity is the likelihood of a positive test result in a person with a disease. The specificity of a test is the likelihood of a negative test result in a person without disease.\(^9\) These can only be determined if the test has been subjected to an independent "blind" comparison with a "gold standard" of diagnosis.\(^4\) Further, the population in which the test has been evaluated must be explicitly defined to avoid problems of spectrum bias in which the reference population is different (e.g., hospitalized patients) than the population in which the test is being performed (e.g., ambulatory population).\(^10\)

Sensitivity and specificity are excellent indices of the diagnostic aspects of a test. Likelihood ratios (sensitivity/1-specificity) are another useful index. They express the odds (probability of event/1-probability of event) that a given level of diagnostic test result would be expected in a person with the underlying disease. For example, the sensitivity of an abnormal stress test (one with greater than 1 mm of ST depression) indicating greater than 70% narrowing of at least one coronary artery is often given as about 65%. The likelihood ratio of an abnormal stress test with 1-1.49 mm of ST depression is 2.1 to 1, which means that these results are 2.1 times more likely to come from people with coronary artery disease than from people without coronary artery disease. The likelihood ratio of a test with greater than 2.5 mm ST depression is 39 to 1.\(^11\)

Although the diagnostic analysis of test results adds to the understanding of interpreting abnormalities because it tells how well the test performs in the presence or absence of disease (i.e., how often the person with disease will have a positive test and how often the person without disease will have a negative test), it really only deals with the situation in which there is already knowledge of the presence or absence of the underlying disease. Because the Medical Director has only an abnormal test result with which to deal, the test must next be examined on the operational level which tells how often a person with a positive test has disease and how often a person with a negative test does not have disease.

On the operational level, two indices are important, the predictive value positive and the predictive value negative. The predictive value positive is the probability of a disease being present if a test result is positive. The predictive value negative is the probability of a disease being absent if a test result is negative.\(^9\) These are calculated from the sensitivity and specificity (or the likelihood ratios) of the test, the prevalence of the disease, and the prevalence of the test abnormality. This can be determined by using Bayes' theorem.\(^12\)

Bayes' theorem states that the probability of a disease, given an abnormal test result, is equal to the probability of the test being abnormal in the diseased population, multiplied by the prevalence of the disease divided by the prevalence of abnormal test results in the normal and diseased populations. One formula for the predictive value positive is:

\[
\text{Predictive Value Positive} = \frac{(\text{prevalence}) \times (\text{sensitivity})}{(\text{prevalence}) \times (\text{sensitivity}) + (1-\text{prevalence}) \times (1-\text{specificity})}
\]

The predictive value negative has a corresponding formula. There are several other ways of expressing formulas for Bayes' theorem including calculations which can be done with likelihood ratios.\(^4\)

Once the test has been examined on the operational level and the predictive value positive and negative have been calculated using Bayes' theorem, the Medical Director has an idea of the significance of the abnormality, i.e., the probability that the test result indicates disease and is not due to analytic variability or is otherwise falsely positive. However, for risk classification, the final question, "What should be done about the abnormality?" must be answered by working on the decision making level. This involves the Medical Director determining and explicitly weighing the consequences of true positive and negative results versus false positive and negative results in terms of mortality and cost. An explicit cost-benefit analysis needs to be done. Decision Analysis, a quantitative technique for making decisions about complex problems in situations of uncertainty, is an effective way to do this.\(^14\)

Here, an example case will be developed to demonstrate the entire process by determining the best course of action to take in a case with an isolated laboratory abnormality. The underwriter presents to the Medical Director a case for $100,000 on a 55-year-old male. There are no problems with the financial underwriting, inspection report, or examination. He is without complaints, and has no known history of alcohol abuse, alcoholism, or liver disease. The only abnormality is a Gamma-Glutamyl Transpeptidase (GGTP) result of 210 U/L on a routine blood chemistry profile.
As discussed above, the first question, "Is the result abnormal?" must be answered. A recent reference text for laboratory medicine lists the "normal range" for GGTP for adult males to be 2-65 U/L.\(^7\) Further, the reference laboratory at which the test was performed (Home Office Reference Laboratory) lists 2-65 U/L as their "usual clinical range." Since the result falls outside that range, it can be considered "abnormal" and the other questions need to be answered.

Next, "What is the significance of the abnormality?" Reviewing the test on the various levels, it can be seen that the analytical aspects of the GGTP have been well studied.\(^7\) It is a test with a well-defined conventional methodology which can be readily performed in most laboratories.

On a diagnostic level, what is its sensitivity and specificity? To answer this, the Medical Director must decide which disease is being considered. In this example, it will be assumed that liver disease due to alcohol is the important underlying disorder. The sensitivity of GGTP for alcoholic liver disease (defined against the "gold standard" of liver biopsy) was 87.7%.\(^7\) In that study the population was hospitalized patients which may have led to some spectrum bias. However, in another study from an ambulatory population with alcoholic liver disease (which also used liver biopsy as the "gold standard") the sensitivity was 52% using the cutoff value of 200 U/L. The specificity of GGTP was found to be 85%.\(^8\)

Operationally, as stated above, to calculate the predictive value positive with Bayes' theorem, the prevalence of alcoholic liver disease needs to be known. The National Institute on Alcohol Abuse and Alcoholism estimates that 13% of the adult population (21% of the males) is composed of heavy drinkers—defined as drinking more than 60 drinks a month.\(^9\) However, only about 10% of the population reach the stage of alcohol abuse—defined as medical, social, and occupational complications caused by high (greater than seven drinks a day) alcohol intake.\(^9\) Liver biopsy studies correlated with alcohol consumption has demonstrated that 60% of people ingesting this much alcohol will have alcoholic liver disease.\(^7\) For the insured population who drinks enough to be at risk for liver disease, this estimate will be decreased to 5%. (Since the prevalence of alcoholic liver disease in the insured population is not known, it is estimated to be 5%. The reasonableness of this estimate will be tested later.)

Using these numbers in the formula for Bayes' theorem given above, the predictive value of the GGTP for alcoholic liver disease is 15.4% \([\{(1.0)(.15)/(.85)\} \times (.95) \times (.15)\]. That is, 15.4% of those with an elevated GGTP will have alcoholic liver disease (which conversely means that 84.6% of those with an elevated GGTP will not have alcoholic liver disease).

Now, since the significance of the abnormality is known, as asked above, "What should be done about the abnormality?" According to the previous analysis, there is only a 15.4% chance that the proposed insured has alcoholic liver disease. How can that information be used to decide what should be done with this application? The test must be examined on the decision making level using Decision Analysis. This involves five steps:

1. Determine the possible options and their consequences for decision making
2. Determine the chance events and their probabilities within each option
3. Quantify the results of each option
4. Calculate the value of each option
5. Determine the best option and the changes in the options with changes in the probabilities.\(^8\)

To do this, the problem is dissected and then structured as a decision tree in which each branch looks at the choices and chance events which may occur. Each branch is then evaluated to determine the best option.

Here, there are three options: take the case standard (i.e., ignore the GGTP result assuming it is an analytic error or falsely positive because the predictive value is too low); rate the case because of the possibility of alcoholic liver disease; or decline the case because of the possibility of liver disease. These are the main branches of the tree. Each branch has different possible uncertain events which can be represented as further branches. If the case is taken standard, then the chance event is that the elevated GGTP is due to alcoholic liver disease which will lead to early mortality. If the case is rated, then the chance events are the same as before, as well as the chance that a rated case will not be placed in an asymptomatic person. All this can be depicted in the decision tree. (See Fig. 1)

![Figure 1](https://example.com/figure1.png)
Further, what is the consequence of having alcoholic liver disease; i.e., what is the mortality (all causes) of drinking enough alcohol to cause liver disease? The 1983 Medical Impairment Study showed that for the impairment of alcohol abuse, the standard mortality was 199%, the highest mortality ratio for any standard risk category in the study! For the minimally and moderately substandard groups the mortality ratios were 207% and 284% respectively. Therefore, in the analysis, the mortality ratio for alcohol abuse will be estimated as 200%. (It could be argued that the mortality ratio of 200% is too low for those with alcohol abuse and liver disease. The reasonableness of this assumption will be examined later in the analysis.)

The last assumption for the analysis is that if the case is rated (table D) for the elevated GGTP, in an asymptomatic individual without an admitted history of alcohol abuse, there is only a one in ten chance that the case will be placed. (Once again, the reasonableness of this assumption will be examined also.)

Next, the explicit consequences of each option need to be determined. What is the consequence of early mortality on a $100,000 policy if the proposed insured does have liver disease? This can be estimated as a loss of $10,000 (the mortality cost “unanticipated” by premium revenue when a 55-year-old male with a mortality ratio of 200% is misclassified as a standard risk) (personal communication, J. Mast). The expected profit of a standard case would be about $490 for a standard case and $890 for a case rated table D (personal communication, D. Becker).

Each of these figures can be inserted into the tree and then by following a formula, \[ V = b \times (u_3) + (1 - b) \times [a \times (u_2) + (1 - a) \times (u_1)] \], on each major branch, the value of each option can be calculated. (See Fig. 2) In Decision Analysis, this is called “folding back the tree.” The calculations show that the first option, taking the case standard, has a value of $1083 (i.e., a loss of $1083). The second option, declining the case has a value of $50 (i.e., the loss of $50 to review the case). The third option, rating the case, has a value of $397. There could also be a greater chance of antiselection with an increased number of “asymptomatic” people accepting rated cases. So, declining the case would still be the best option.

Hence, for a wide range of values, the same decision holds, that is, the option which maximizes profitability (the best decision) for the case is to decline. In a competitive environment, there may be other considerations for the underwriter to make a “business decision” to “take the case” Standard (i.e., take a loss of $1083). But the role of the Medical Director, as stated above, is to assist in the risk classification and not necessarily make the final decision.

In theory, several more complex analyses could be done. The best cutoff to define an “elevated” GGTP level could be examined by using Receiver Operating Characteristic (ROC) curves which graph the tradeoffs of sensitivity and specificity for different cutoff levels. A family of decision trees using a series of cutoffs with different likelihood ratios could be created. Other causes of elevated liver function tests (such as hepatitis or tumors) could be included as well as analyses for other liver function tests (e.g., SGOT). However, for the above case, they would not add substantially to the above analysis.

In summary, by answering several key questions and by approaching abnormal laboratory test results on the analytic, diagnostic, operational, and decision making levels using probability theory, Bayes' theorem, and Decision Analysis, Medical Directors can further assist underwriters in the evaluation of cases with isolated laboratory abnormalities. Although this methodology appears complex, with the improvements in information Technology, especially artificial intelligence, it can be programmed and distributed to underwriters and Medical Directors. [I am indebted to the members of the Medical, Underwriting Research and Development, and Actuarial Departments of Lincoln National Life Insurance Company, especially Jess Mast, for assistance.]
References


