1994_13

THE PITFALLS OF DUPLICATE RADON MEASUREMENTS

Raymond Johnson, Douglas Heim, and Lori Faiola
Key Technology, Inc.
Jonestown, PA

ABSTRACT

The EPA radon measurement device protocols recommend duplicate measurements for 10% of devices deployed or a maximum of 50 per month. In addition, one of the three options for real estate testing calls for simultaneous, side-by-side measurements. While these side-by-side measurements are not identified as such, they are actually duplicate tests. Since most testers will choose the side-by-side option (rather than sequential or continuous testing) and since most testing is currently done for real estate transactions, we can expect a large increase in the overall number of duplicate tests conducted in the future. Some radon testers have applauded this approach as a big improvement in the confidence of real estate testing results. We contend that this is not necessarily true. Unfortunately, when duplicate measurements yield similar values, this shows that the devices are able to provide reproducible results, but it does not say that either measurement is accurate with respect to the true radon level. In other words, duplicate measurements can lead to unjustified confidence in results. Furthermore, the interpretation of duplicate measurements represents one of the most difficult, and often most misunderstood, of any radon measurements.

To properly evaluate duplicate results requires some understanding of radiation statistics. Many radon testers, and essentially all homeowners, have very little understanding of statistics. Consequently, the natural approach is simply to compare duplicate results with the anticipation that they should be exactly the same. When they are not the same (and usually they are not), everyone involved is challenged to explain the difference. Homeowners, in particular, are inclined to want to reject duplicate results that are not identical. They, and unfortunately many radon testers also, do not understand that radon testing involves the measurement of radiation, and a fundamental property of radioactivity is that it is a statistically random process. Consequently, if you expose 100 identical devices in a radon chamber at the same radon level, you will get 100 different results.

Therefore, EPA has attempted to provide guidance on the application of statistics to evaluate duplicate test results. The guidance describes methods involving the coefficient of variation, the relative percent difference, and mean-range control charts. This paper will review the pros and cons of each of these methods, as well as how to effectively communicate duplicate test results to homeowners.

We conclude that mean-range control charts provide the best method for interpreting duplicate results. We also conclude that better information for real estate tests would be achieved by measurements in several locations in a house rather than by side-by-side tests. Our main concern is that duplicate tests yield unjustified confidence when the results are similar (and unneeded conflicts when they are not). We believe that the main source of uncertainty of radon measurements is related to location of the test devices (and occupancy factors) rather than reproducibility of devices.

INTRODUCTION

The Environmental Protection Agency (EPA), July 1992, protocols for radon measurement devices specify for duplicate (collocated) detectors that, *"anyone providing measurement services with AC devices should place duplicate detectors in enough houses to test the precision of measurements. The number of duplicate detectors deployed should be approximately 10 percent or the number of detectors deployed each month or 50, whichever is* *smaller"* (EPA 1992). In addition, the EPA July 1993, protocols for radon measurements in homes presents the following option for real estate testing: *"Option 2. Simultaneous Testing. This option involves the use of two tests, conducted simultaneously and side-by-side, made for similar durations, and producing results in the same units"* (EPA 1993). While these simultaneous tests are not identified as such, they meet all the criteria for duplicate tests.

Our laboratory data indicate that currently 80 to 90 percent of all radon testing is done for real estate transactions. Furthermore, we see that most radon testers are choosing the option of simultaneous duplicate detectors, rather than the options for real estate testing that use either sequential testing or active continuous monitoring. When we add together the QA requirement for duplicates and the primary use of duplicates for real estate testing, we are now seeing a large increase in the number of duplicate AC detectors.

Some radon testers have applauded the use of duplicate simultaneous detectors as a big improvement in the confidence of real estate testing results, in comparison to the result of only a single measurement. We contend that the use of duplicate tests for real estate does not necessarily improve the quality of results over a single measurement. In fact, the additional data of a duplicate test may easily lead to misinterpretation of the results.

MISINTERPRETATION OF DUPLICATE ANALYSES

Misinterpretation of duplicate results will often occur in two ways. When the two results are very close, or identical, this is an indication of good precision or reproducibility of results. Unfortunately, while the two tests may demonstrate good precision, they do not provide any measure of accuracy relative to the true radon level in the house, as noted in Figure 1. In fact, a single measurement may very well be more accurate than two simultaneous measurements. Misinterpretation occurs when the closeness of the two results is interpreted as an indication of good accuracy which could lead to unjustified confidence in the results. The other misinterpretation of duplicates occurs when the results of the two tests are not close together and this is interpreted as poor accuracy and a lack of confidence in the results. We are reminded of the saying,

" A person with one watch always knows what time it is.
But, a person with two watches is never quite sure."

We predict that the interpretation of duplicate measurements will become the source of some of our greatest difficulties in providing radon measurement services. The credibility of radon measurement businesses may very well be judged on the basis of duplicate analysis results. Lack of understanding on how to interpret duplicate analyses will cause some companies to lose credibility in the eyes of clients and others to gain credibility, when in fact, neither judgement is warranted by the data.

INTERPRETATION OF DUPLICATE ANALYSES

To properly evaluate duplicate analyses requires some understanding of the fundamental properties of radiation and radiation statistics. Many radon testers, and essentially all home buyers and sellers, have very little understanding of either radiation or statistics. Without such understanding, the normal approach to evaluating duplicate measurements is simply to compare the reported results as absolute values. Most clients assume that duplicate measurements should be exactly the same. When they are not the same, and usually they are not, then the radon tester and the radon laboratory are challenged by the client to explain the difference. The natural inclination of clients, home owners, home buyers and real estate agents, is to want to reject duplicate results that are not exactly the same. When clients pay $100 to $200 for a professional radon test, they expect a high quality measurement service. They also expect that duplicate results should be exactly the same concentrations, or at least very close together.

Most people do not understand that radon testing involves the measurement of radiation, and a fundamental property of radiation is that it is a statistically random process. This means that if you expose 2 radon detectors, or 10, or even 100, to the same radon concentration, such as in a radon chamber, each of the individual results will be different because of the randomness of radiation. These random variations have nothing to do with the quality of the detectors or the detection system. No matter how good a radon detection system may be, it cannot improve on the natural randomness of the radiation signal which is used to quantify radon. Variations in radon instrument response at a given radon level can only add to the natural variations expected due to the randomness of radiation.

If a "perfect" radon detector could be designed, without any source of variation in detector response, then duplicates would still differ as a result of randomness in radiation. Since, despite marketing claims, no perfect device has yet been invented for radon detection, differences in duplicates will reflect both the randomness of radiation and other sources of variability in the detection system due to design, manufacture, handling, or analyses. Therefore, duplicate results which come out the same, or even close together, are mainly a matter of statistical chance. The chances are greater statistically that duplicate results will always be different.

DEALING WITH RADIATION STATISTICS

EPA has attempted to provide guidance on how to evaluate duplicate tests, which might be called; *"how to evaluate duplicates without using statistics", or "how to apply statistics without being a statistician."* Rather than attempting to teach radon testers the rudiments of simple statistics, EPA has devised ways of interpreting duplicate analyses without having to understand basic statistics. Unfortunately, we believe the proposed EPA methodology is drastically oversimplified (at the same time it is also very confusing) and will result in erroneous interpretations by anyone not understanding the normal randomness of radiation. We are referring to the protocol for interpreting simultaneous testing as presented in the June 1993 homes protocol for real estate testing (EPA 1993).

EPA's protocol for duplicate analyses involves evaluating the data in three categories: 1) both results are less than 4 pCi/L, 2) one result is greater and one is less than 4 pCi/L, and 3) both results are greater than 4 pCi/L. The reason for the three categories is to account for variations due to the randomness of radiation which is a function of the size of the radon signal, i.e. the radon concentration. The randomness of radiation decreases as the signal or the concentration increases. Or vice versus, as the concentration decreases the randomness increases until you reach the lower limits of detection, which is often estimated as the smallest amount of radon which can be reported as a real value with an uncertainty of +/- 100 percent.

The EPA protocol, which looks at three radon levels in categories, is fundamentally wrong for two reasons. First, the overall approach is based on the assumption that each radon measurement is an absolute value. Therefore, when a result is reported as 4.1 pCi/L, it means the result is exactly as reported. In fact, anyone who has experience with radiation measurements knows that there is no such thing as an absolute measurement for radiation, or for radon in particular. **All radon measurements are only best estimates.** Furthermore, any radon result only has meaning when accompanied by a measure of the uncertainty of the result which includes an estimate of expected variations due to the randomness of radiation and variations due to accuracy and precision of the measurement system. Therefore, any radon measurement that is reported without its associated uncertainty is meaningless and does not warrant careful interpretation. From this perspective, the EPA protocol is flawed because it goes on to use absolute values (which we define as meaningless numbers) to calculate a measure of precision called the *"Relative Percent Difference (RPD)."* The RPD is defined as the difference between the duplicate results divided by the mean of the two results and multiplied by 100 to convert to a percentage.

The second reason the EPA protocol is fundamentally wrong is that it attempts to evaluate RPDs as a measure of precision error in three categories of radon levels as if precision error varies as a step function in discrete categories. In fact, precision error (which includes the randomness of radiation) varies as a continuous function over all radon levels. As noted above, differences due to randomness of radiation alone would be expected to range in a smooth transition from a few percent at high radon levels (greater than 10 to 20 pCi/L) to a 100 percent or greater at levels below 1 pCi/L. Since the randomness of radiation increases rapidly at lower levels of radon measurements, the chances for misinterpreting duplicate measurements also increases at the lower radon levels (below 4 pCi/L).

The data in Table 1 illustrate how duplicate results can be misinterpreted at low radon levels by the Relative Percent Difference method. These data were selected from about 100 duplicate measurements to illustrate the flaws in the RPD method at low radon levels. All of these duplicates fall outside of the warning or control levels for evaluating RPDs (when the average value is less than 4 pCi/L) and were the basis for letters of complaint to Key Technology. Basically we were challenged to explain the high RPD values as an indication of poor quality in the performance of our activated charcoal detection systems. According to the EPA protocol, measurements should cease anytime a single RPD value exceeds 67%. In this data set of 15 duplicates, six sets exceed the control limits indicating that measurements should be stopped until the problem is identified and corrected. Furthermore, all 15 sets are at or above the warning level that would be expected only 5% of the time by chance. Again, the conclusion would be to stop measurements until the problem is corrected.

Since the RPD method for evaluating duplicate tests is simple to perform, i.e. the calculations are easy, most radon testers will use this method for judging the quality of duplicate data. However, according to the EPA protocols this approach will often lead to conclusions that the duplicate data are unacceptable. Unfortunately, such conclusions are erroneous because they fail to recognize the inherent uncertainty of individual measurements and the large increase in these uncertainties at low radon levels.

Since the average radon level in homes in the United States is about 1.3 pCi/L, we should expect most radon measurements (and duplicates) to give results below 4 pCi/L. When radon testers apply the RPD method in this range of low radon levels, they will also likely find a high percentage failing the EPA evaluation criteria. When radon testers or their clients see such results, the implications are bad for the radon measurement industry. A purely mathematical interpretation of the data in Table 1 would lead to conclusions that these measurements represent very poor quality and this leads to questions about the competence and credibility of the radon tester and the primary laboratory.

MEAN RANGE CONTROL CHARTS ARE A BETTER WAY

Mean range control charts, as described by Goldin, represent a method for interpreting duplicate data which is both simple and statistically rigorous at the same time (Goldin 1984). This method is based on the inherent uncertainty of radon measurements at each radon level as determined by precision variations due to counting error. The construction of a mean range control chart is shown in Figure 2, which includes the results of duplicate measurements from Table 1. The mean of duplicate measurements is plotted on the horizontal scale and the range, or difference, of the two measurements is plotted on the vertical scale. The plotted data points are then interpreted on the control chart in relation to three lines representing: 1) expected mean range results, 2) a warning level, and 3) a control level.

The control chart lines show the range limits of duplicates using the mean range and the standard deviation of the range. The standard deviation (counting error) is represented by s for a single measurement. Since the standard deviation will vary with increasing radon levels, then the mean range will also vary with concentration. For duplicate measurements, the mean range (R) is given by

$$R = 1.13\ s$$

The warning level is given by

$$R = 2.53\ s$$

And the control level is given by

$$R = 3.69 s.$$

The warning level is set where only 5% of duplicates would be expected to exceed this level. The control level is set for less than 1% probability that duplicates would exceed this level by random variations in counting error.

After plotting duplicate data on the mean range control chart, one should examine the chart to evaluate the quality of the duplicate measurement and take action as indicated. When duplicate results fall outside of the warning level, it is necessary to consider how many results have occurred above the warning level in relation to the total number of duplicates according to the criteria in Table 2 (Goldin 1984). If the total number of duplicate measurements that have been entered on the chart by that time is given in column 4, stop making the measurements and identify and correct the trouble. If the number of measurements that have been entered on the chart by that time is given in column 3, investigate the situation but continue making measurements. If the number of the measurements that have been made by that time is given in column 2, no action is needed but the analyst should be aware of possible problems with the measurement system.

If a measurement exceeds the control limit, the measurement should be repeated for verification. For AC measurements, this means the canisters should be recounted. If the repeat measurement is also outside the control limit, then the entire measurement should be repeated. This means going back to the place of measurement and placing the devices again to repeat the original measurement.

Dealing With Discrepancies

Since the discrepancy in results which occur outside of the control limit implies some defect in the device, or analysis of the device, at Key Technology we provide the repeat measurement without additional charge. While we have been using the mean range control chart method for evaluating duplicates over the past four years, we have had very few duplicates fall above the control limit. In each case, a repeat of the measurement with new charcoal canisters has resulted in acceptable results.

Figure 2 shows the results of plotting the data from Table 1 which failed the Relative Percent Difference Method. As you can see, all of these data fall comfortably within the expected level for duplicates at low radon concentrations. The reason that these data pass the duplicate precision test by the mean range control chart is that this method includes the known standard deviation for random counting errors associated with each radon concentration. Therefore, each measurement is not treated as an absolute value, but as a best estimate which recognizes the normal random variation of radiation measurements.

In at least 90 to 95 percent of the cases where duplicates have fallen outside of control limits when plotted on a mean range control chart, the problem has been related to placement of the devices. For example, this problem occurs most often in school measurement programs. When we ask where were the duplicates placed, we get an answer such as; one of the canisters was placed on the teacher's desk and the other was placed on the speaker box of the public address system. Most cases of failures in duplicate tracking are the result of not placing the duplicates side-by-side, or 4 inches apart, if we are to follow the EPA protocol which says devices should not be placed closer than 4 inches to other objects.

One exception to the question about device placement was recently observed by one of the authors (Faiola). Duplicates were carefully placed according to EPA protocols in the basement of a home and the results were 123 and 188 pCi/L. These measurements were obviously far outside of the control chart limits. A repeat count verified the original results, which meant that the only answer was to repeat the entire measurement with new devices. The repeat duplicates measured in the same location gave results of 151 and 148 pCi/L. Interestingly, the average results for both sets of measurements were 156 and 150 pCi/L, which is about the same for both sets. We cannot explain the discrepancy in the first set of duplicates. It does seem very strange that the average of both set were so close. One factor would seem to be the possibility of homeowner tampering, but no particular evidence was noted. The homeowner expressed great concern for the initial results and demanded a repeat measurement.

The two main reasons for recommending the interpretation of duplicates by a mean range control chart is 1) simplicity and 2) statistical rigor. The chart is very simple because it allows the interpretation of duplicates based only the calculation of the difference between the measurements (range) and the average of the two results (mean). This method does not require knowledge or calculations involving the uncertainty or counting error of individual measurements. Anyone using charcoal canisters from Key Technology will be given a mean range control chart which is drawn on the basis of known standard deviations for each radon level. Thus, users of these devices have only to plot the mean and range, and visually observe the appropriate interpretation. The important statistics have already been built into the mean range control chart. The technical advantage of this method is that one can use the normally reported radon concentrations, without the associated uncertainty terms, and evaluate duplicate data on the assumption that each concentration is an absolute value. The associate uncertainty is accounted for by the slope and placement of the lines on the control chart.

If you wish to plot a control chart for your particular device, you will need to determine the uncertainty or standard deviation (variation in precision) at each level of radon concentrations that you expect to find in duplicate measurements. This can best be done with 20 sets of duplicates at each radon level included on the chart. Typical radon concentrations for establishing standard deviations of duplicates for mean range control charts would be 5, 10, 25, 50, and 100 pCi/L. It is necessary to verify standard deviations at several radon concentrations, because as the concentrations increase the acceptable range values go up faster than the means, i..e, the control lines are not linear, but tend to slope upward with increasing radon concentration.

DUPLICATES ARE NOT THE BEST WAY FOR REAL ESTATE MEASUREMENTS

While duplicate measurements can best be interpreted by mean range control charts, the possibility of confusing good precision for accuracy still exists. We believe that more useful information could be obtained by taking measurements in two, or more, locations rather than side-by-side. The use of duplicates for real estate testing assumes that the main source of uncertainty in radon levels is due to lack of precision in devices. We do not believe this is the case. After performing duplicate measurements on 50 sets of duplicates a month for over eight years (over 5000 duplicate sets) we conclude that precision errors with our activated charcoal canisters are exceedingly small and represent an insignificant source of uncertainty in any decisions for retesting or mitigation.

Much greater sources of uncertainty for interpreting radon concentrations include: 1) device placement location, 2) occupancy factors, and 3) other variables, such as weather conditions. One of the authors (Johnson) has found in his own home, that radon levels may be 25 to 50% greater when measured near the outside wall of his living room when compared to measurements near the inside wall. Both sets of measurements were made according to EPA protocols. We assume that the measurements near the outside wall are closer to the radon source in this house which has a slab on grade foundation. The conclusion here is, that with only one measurement (or even duplicates) near the inside wall or near the outside wall, we would not know that simply moving the devices across the room (12 - 14 feet) would cause a change of 25 to 50%. By the way, in the winter months the reading near the outside wall is typically 4 to 5 pCi/L and the readings near the inside wall are normally from 2 to 3 pCi/L.

For real estate measurements, Key Technology has for many years provided a three-pack kit to encourage measurements in several areas of a home. Rather than making possible expensive decisions on the basis of measurements in a single location, we believe that measurements in several locations on a floor suitable for occupancy (without renovations) provide better data than two measurements side-by-side. We do not discount the value of duplicates (at the rate of 10 percent of all measurements, or 50 a month) for verifying precision as a necessary part of a comprehensive QA program. However, we do not believe that using Option 2. "Simultaneous Measurements" for all real estate testing represents the best use of two or more devices in a home.

We believe that occupancy factors, i.e. tampering or interference by parties concerned with the radon measurement, represent a much greater source of uncertainty, than either variations due to precision error or even the specific placement location within the home. In addition, variables which affect radon levels naturally, such as weather conditions, could have even greater effect. With all of these sources of uncertainty of radon measurements, reproducibility or precision error is not very significant. Consequently, we conclude that more useful and reliable information could be gathered for real estate testing by placing devices in two or more locations in the home rather than side-by-side.

DISCUSSION AND CONCLUSIONS

The use of duplicate measurements for real estate testing is fraught with potential for erroneous decisions. First of all, many people still confuse measures of precision by duplicates as a measure of accuracy. Consequently, when duplicate results are close together this is interpreted to mean the results are of high quality and represent a good basis for decisions on retesting or mitigation. Many people fail to understand that good precision does not say that the results are any good, in terms of closeness to the true radon level (accuracy). Therefore, very precise radon measurements may be completely wrong as far as representing a good measure of actual radon levels.

On the other hand, when duplicates do not meet criteria for acceptability, the results are discounted as poor quality. Again, this may be done without understanding that the average of the results may represent good accuracy. Unfortunately, the recommended EPA procedure for interpreting duplicates by the Relative Percent Difference method may add to misinterpretation of duplicate results. This method treats each radon measurement as an absolute value, without consideration of the normal random variation of radiation. Users of this method are then instructed on interpreting the results according to arbitrary categories that attempt to deal with the measurements as statistical variables in three different radon levels. Not only is this an inappropriate use of radon measurements as absolute values, but the three categories treat uncertainties as step functions, when in fact, random precision errors are a smooth continuous function of increasing radon levels.

The main area of difficulty is in the evaluation of duplicate by the Relative Percent Difference method when both results are below 4 pCi/L. The method fails to account for the increasing variability of radon measurements due to the randomness of radiation at decreasing radon levels. Consequently, many duplicates at radon levels below 4 pCi/L appear to fail by this method. Since a large proportion of all radon measurements are likely to have results below 4 pCi/L and since this method is commonly used because of its simplicity, we predict that many duplicate failures will be occurring over the coming years. The consequence of repeated duplicate failures will be a loss of credibility, not only for the radon testers and laboratories involved, but also for the entire radon measurement industry.

A better method for interpreting duplicate measurements is to use a mean range control chart. This method is based on knowledge of the inherent uncertainties of individual radon measurements due to random counting (precision) errors. Therefore, each measurement is not treated as an absolute value and allows evaluations of duplicates with full recognition of differences that are due to random chance.

Better yet, we believe that more useful information for real estate testing could be obtained by measurements in two or more locations in a home. We believe that variations in radon measurements will occur more often due to factors involving device placement, occupancy, or weather rather than differences in precision of devices. The use of duplicates continues to invite people to interpret the results as a measure of accuracy. Better estimates of the true (accurate) radon concentration in a house can be obtained by measurements in several locations in the house.

REFERENCES

Goldin, A. Evaluation of Internal Quality Control Measurements and Radioassay, Health Physics, 47:361-374; 1994.

U.S. Environmental Protection Agency, Protocols for Radon and Radon Decay Product Measurements in Homes, Washington, DC; U.S. EPA Office of Radiation Programs; EPA 402-R-92-003; 1993

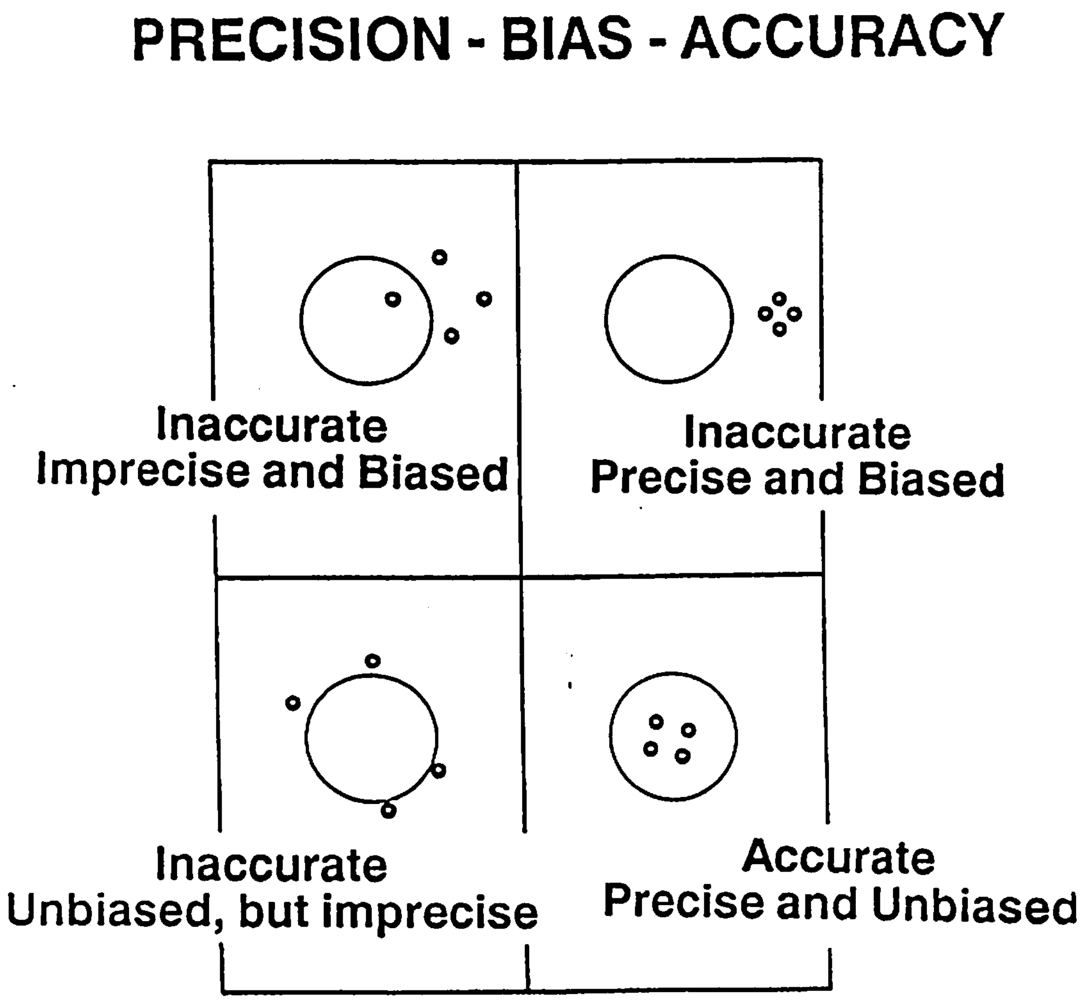

Figure 1. Illustration for defining the QA terms, Precision, Bias, and Accuracy

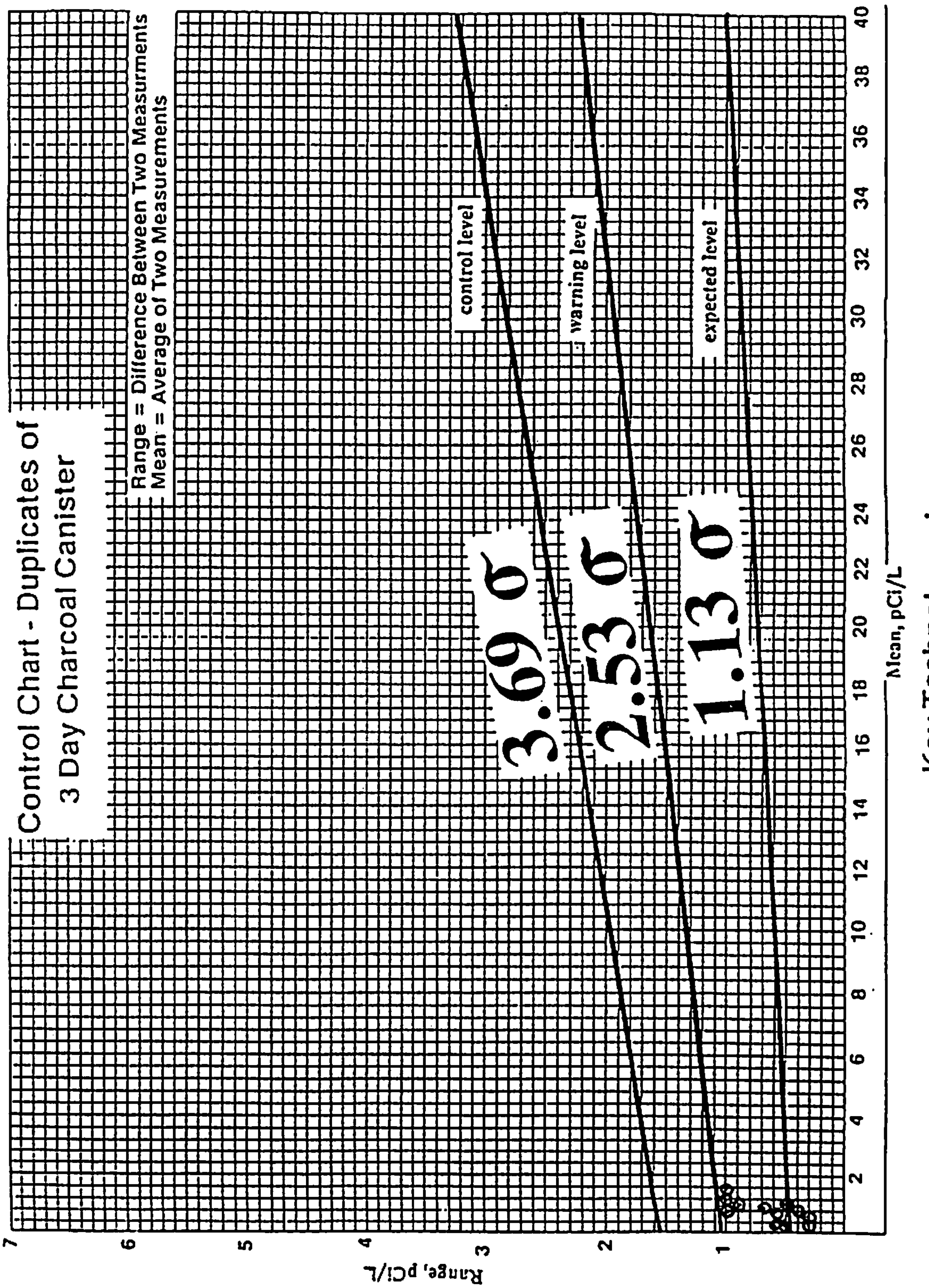

Figure 2. Mean range control chart for 3-day open face charcoal canisters, Key Technology, Inc.

Table 1. Evaluations of Selected Duplicate Radon Tests by the Relative Percent Difference Method.*

Radon Measurements in pCi/L				
No. 1	No.2	Difference	Average	RPD - %
0.2	1.2	1.0	0.7	143
1.0	2.0	1.0	1.5	67
0.7	0.4	0.3	0.6	50
1.0	0.6	0.4	0.8	50
0.2	0.5	0.3	0.4	75
0.4	0.2	0.2	0.3	67
1.2	0.6	0.6	0.9	67
1.3	0.8	0.5	1.0	50
0.8	0.2	0.6	0.5	120
1.0	1.9	0.9	1.4	64
1.5	0.9	0.6	1.2	50
1.7	0.7	1.0	1.2	83
0.3	0.9	0.6	0.6	100
0.2	0.3	0.1	0.2	50
0.7	0.2	0.5	0.4	125

* Data selected from about 100 sets of duplicate measurements

Table 2. Criteria for taking action for measurements outside the warning level (Goldin 1984).

Number Outside Warning Level	40% Level Alert	25% Level Investigate	5% Level Stop Measurements
1	1 - 9	X	X
2	20 - 27	8 - 19	2 - 7
3	35 - 45	17 - 34	8 -16
4	52-64	29 - 51	17 - 28
5	68 - 82	41 - 67	29 - 40
6	85 - 100	54 - 84	41 - 53
7	*	67 - 100*	54 - 66
8	*	81 - 100*	67 - 80
9	*	95 - 100*	81 - 95
10	*	*	96 - 100*

X - No action required when only one measurement is outside of the warning level

* - Calculation truncated at 100 measurements on the control chart.