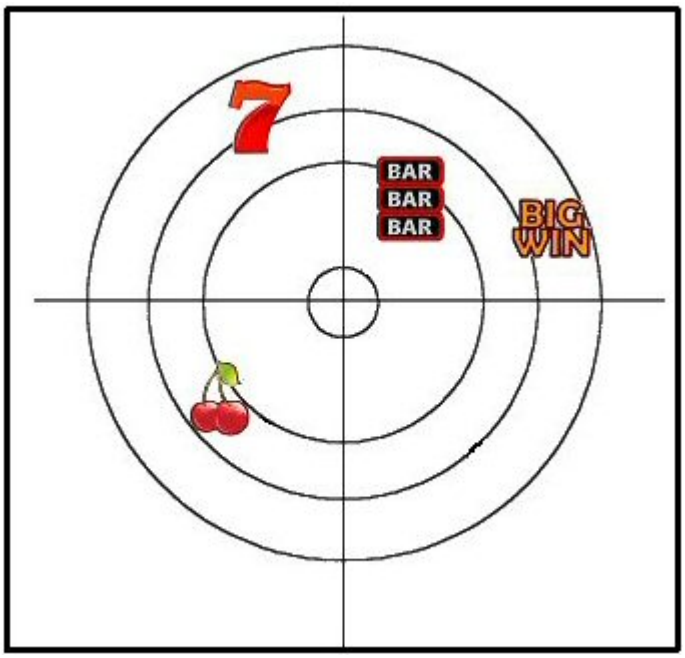
Rationalizing Representativeness



In this installment on the discussion of *Data Quality Objectives*, I'm going to address the "R" in PARCC – Representativeness.

Over the weekend, an individual emailed me and asked why PARCC parameters only seemed to apply to sampling for moulds. The concepts discussed in this series apply to virtually all kinds of sampling – sampling theory ultimately has its roots in characterizing uncertainty and bringing appropriate weight to bear on "data." I only use the mould industry as an example since, next to the assessment of contamination in drug houses, that is where I see the biggest abuses in sampling and analysis. So in this next discussion on "Representativeness" I'm going to change gears a bit, and hopefully the parallels will become more clear.

Change of Plan:

Posit - You have been asked to determine the average pocket change of the inhabitants of a small village on the outskirts of Chita, Russia. Your budget is small, and the train to the extremely remote village comes through once each month and remains in the village station for only one hour.

To have complete confidence in your conclusion you would have to interview everyone in the entire village. Obviously, with limited time and resources, that's not an option, so you are going to have to *sample* some of the inhabitants and use that data to determine the average pocket change of this population… but who do you select? How do you choose your sample with confidence?

You decide the best approach is to leave it to chance - brilliant! You will select three chance encounters and determine how much pocket change they carry – it's completely random - brilliant! On the train you develop a brilliant algorithm using dice and playing cards to generate three random numbers representing the people you will interview in order of meeting them. According to the algorithm, you are going to interview the 54th person you meet, the 2nd person you meet and the 12th person you meet. Completely random…Brilliant!

Here are the "results" of your sample:

The second person you met was a local fur trader from Chuluunkhoroot, just across the border to the south and he was carrying 50,000 Mongolian tögrög; about $25 US. (The village of Chita sits on the border of three countries – China, Russia and Mongolia).

The 12th person you met was a Chinese silk vendor who lives a couple miles to the east in Hulun Buir, China and he had 100 Yuan (about $15 US).

The 54th person you encounter looks a lot like Yul Brynner and he pulled up in a stunning Maserati, and you determine that Mr. Brynner was making a movie in the area and he was carrying 2 Million Russian rubles (about $30,300 US).

There's your raw data; (50,000 + 100 + 2,000,000)/3 = 683,367

Certainly, no one can argue the math. But what does the "average" *represent*? In fact, what are the units of expression? You have three different currencies each with different buying powers, and surely it's not every day filthy rich Yul Brynner is wandering the village. In fact, although your sampling plan *seemed* like a good idea, your samples don't *represent* anything; as a result, your "results" don't represent anything. In fact, you don't have results, and your average isn't. For all your efforts, you have no data.

In short, you have just performed the same kind of "sampling" that is performed by "certified" mould inspectors when they collect air samples. Mismatched entities, with differing characteristics, fluctuating amounts, uncertainty in accuracy (if you are a silk trader and a stranger approached you and asked you how much money you were carrying, are you going to tell them the truth?)

In short, your data lacks "Representativeness." "Representativeness" is the ability to obtain an articulable reflection of actual conditions under investigation. *Articulable* means that the investigator will be capable of identifying the limits of confidence of representativeness before the samples are collected; and be able to make a statement regarding the decision threshold (whatever that threshold may be).

If one cannot put "representation" into context of the data, then one's samples are meaningless. Sometimes we can put representation into context by describing the "precision" and the "accuracy," thereby ensuring the limitations of the samples are understood. To illustrate, let's go back and look at some of the data discussed in "**The Problem With Precision**"

Time	Spores/m3
09:00	97
10:45	2016
12:15	129
13:45	807
15:00	353
17:00	129

These are the results of very real spore trap samples collected from a normal, ordinary, everyday home in Denver, Colorado. The sampling parameters are exactly the same for each datum except the sampling time. Just like our different currencies in the above example, the genera of the spore profiles was very different from sample to sample (exactly what we would expect).

Here is the question: What did the sample at 09:00 represent? Answer: Nothing.

How about the sample at 10:45; what did it represent? Answer: Nothing.

In fact, it isn't until we look at the data set as an whole that we can see that it represents *something*, but what? If we said the data represented the spore count in the property, clearly we are utterly incorrect.

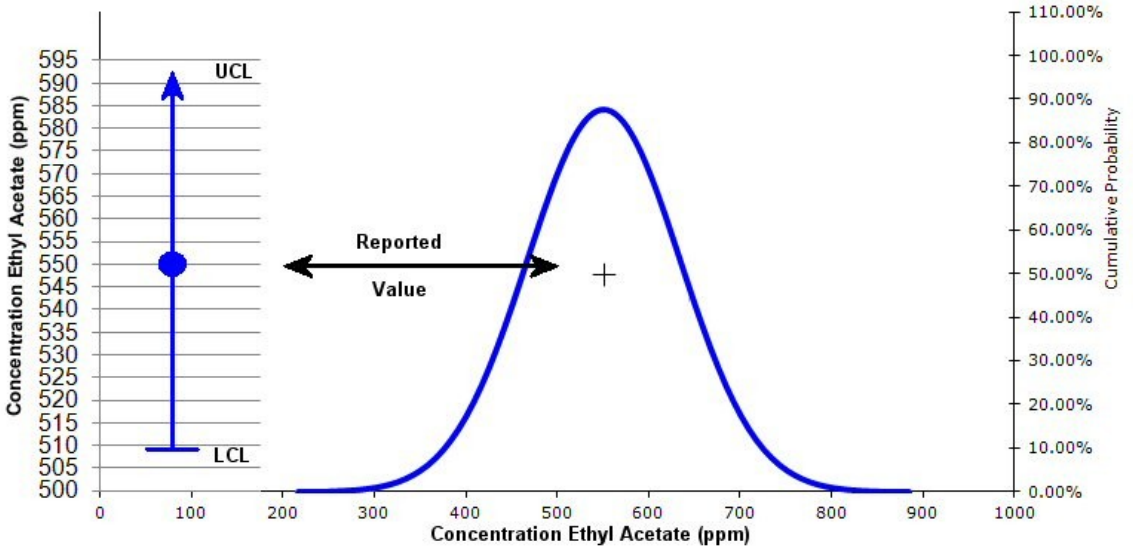
So what *does* the sampling represent? Answer – the data represents uncertainty; the variance of spore counts about a central, but unknown, amount. If a consultant told the homeowner that their samples "represent" the spore count in a study area, they would be wrong, since it does no such thing (and as previously discussed, can never represent such a thing without some very complex mathematical corrections).

Spread Betting

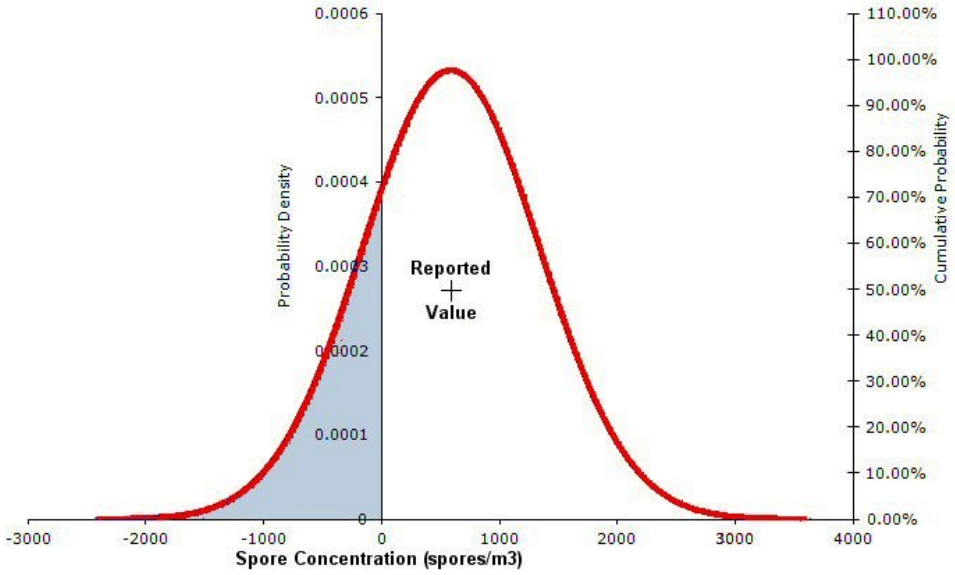
When we look at the confidence intervals of spore trap data, we see that the variation about the mean is so large that the *probable* concentrations are so divergent as to render the samples virtually meaningless. For example, let's compare the above data set with legitimate Industrial Hygiene exposure monitoring.

In our discussion, "**The Problem with Precision**," we used a comparison between ethyl acetate and spore counts; let's use that again here. Imagine we have collected a representative sample for

ethyl acetate according to good sampling protocols. For that one full shift sample, we see the following probability curve (representing random error) and the confidence intervals (error bar on the left) for that day's
sample:

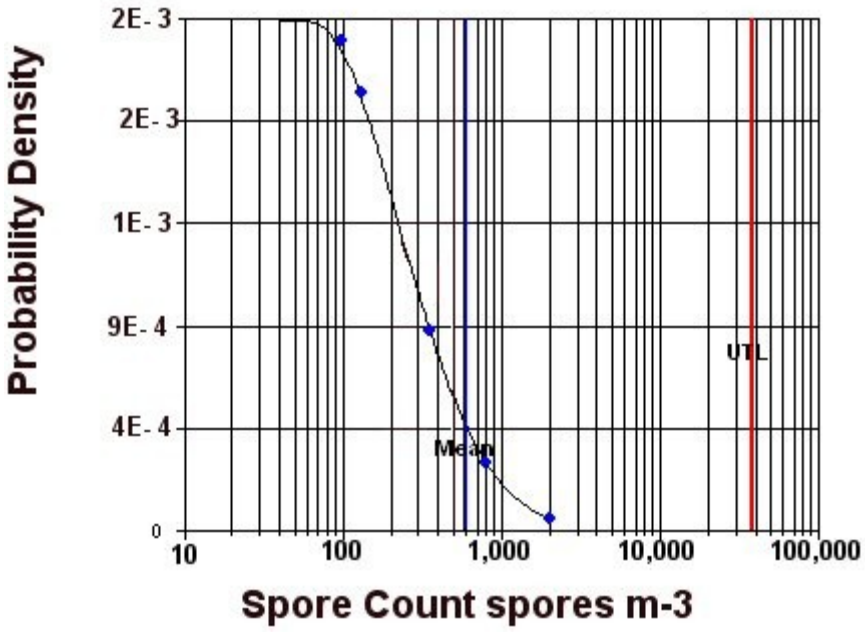


We see the lab reported a value of 550 ppm and therefore, we know with confidence the most probable concentration was somewhere between 509 ppm and 592 ppm. Now, let's do the same thing for our spore trap data:

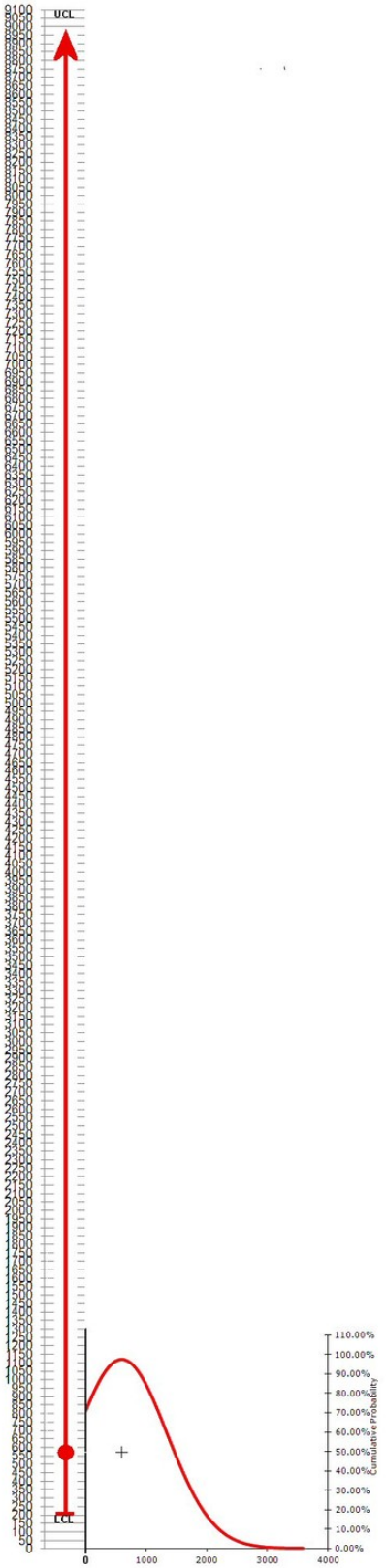


We immediately encounter the problem that fully 21% of the area under the curve is negative. How could that be? How can we have a negative spore count? The problem lies in the presumption that the data exhibit a Gaussian distribution. In fact, in virtually all cases, the spore counts are lognormally distributed and so it is with this data set.

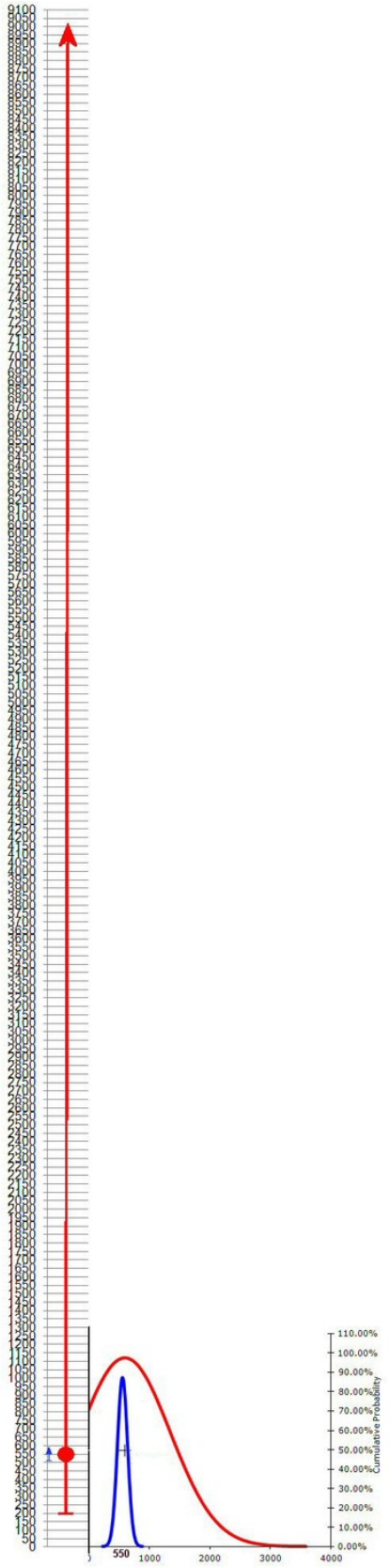
A quick and dirty test for normality is to plot the log of the data; if the log of the data are Gaussian, then the data may be lognormal. But here, we are going to log-plot the data:



We now realize that the LCL and UCL cannot be based on a normal distribution, so we are going to select Land's confidence intervals, and when we do that we see that the error associated with the samples is HUMUNGOUS! We see that the LCL is 270 spores/m3 and the UCL is 9,030 spores/m3. That is, we only have confidence in any single result within those two limits, and we must honestly report those limits to the client, and use those limits in our decision making processes. (Which is just one of the reasons why such samples cannot be used for "clearance" purposes after remediation jobs – the consultant has no idea what the actual spore concentrations are. We will address that issue in the next discussion on "Comparability").


UCL
LCL
0
1000
2000
3000
4000
110.00%
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Cumulative Probability

Now let’s superimpose the two data sets in the same scale to better appreciate the error (believe it or not, the error bars for the ethyl acetate are in the following figure and is the blue smudge to the left of the red error bars);


110.00%
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Cumulative Probability
550
1000
2000
3000
4000

This tells us that even with the collection of six data spore traps collected throughout the day from one room, we are still left with huge uncertainty and seemingly unmanageable “results” for that study area.

This concept is valid regardless of the type of sampling, be it sampling for pocket change in Chita, sampling widgets off a production line or even something as seemingly simple as measuring the airflow through a duct.

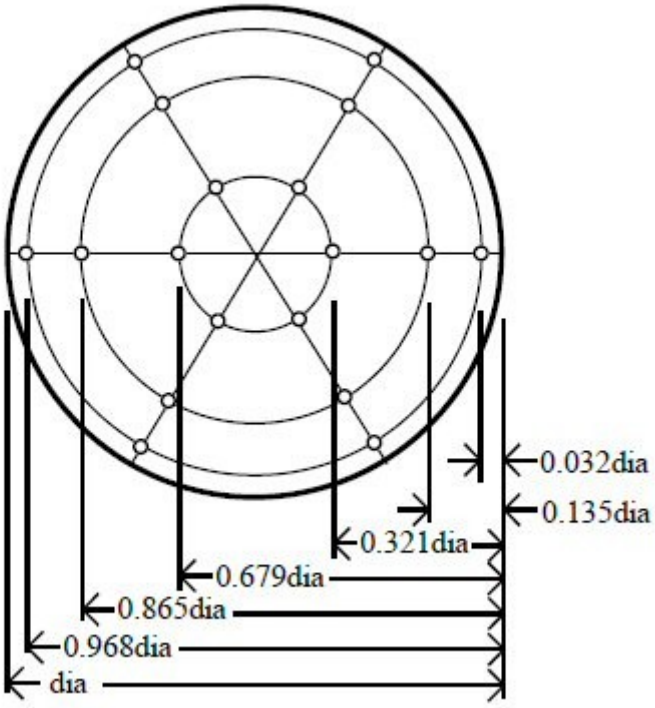
Monkey Around with Samples

Like spore trap sampling, or wipe sampling in a meth-lab, a chimpanzee can be taught the *mechanisms* of performing the motions of sampling, but what will the data mean? Imagine we want to teach a Chimp how to measure the air flow in a duct, and give the Chimp instructions on how to use an hot-wire anemometer for the task. Within two minutes, just as in collecting surface wipes in a meth-lab or air sampling for spores, the Chimp has mastered the mechanics behind collecting a sample of air flow from a duct.

But, we want the Chimp’s sample data to *represent* something - the actual air flow in an air duct. For this, we must now instruct the chimpanzee of the *form* (the concept) of the sample collection and tell the Chimp to meet the sample DQOs and collect his samples at least 7.5 duct diameters downstream from any turns or obstructions and at least 3 duct diameters upstream from any turns or obstructions and then run a traverse … unless of course, the duct is rectangular, then the Chimp needs to calculate the “equivalent diameter” using

$$\sqrt{\frac{4HV}{\pi}}$$

Then, we tell our Chimp, the samples need to be collected according to the following log-Tchebycheff schematic: (1)



Eventually, we come to the realization that mechanistically “collecting a sample” and “collecting a legitimate sample within the DQOs” are not the same thing and our Chimp may be capable of collecting a sample, but he lacks the skill set needed to collect a meaningful sample. It is for this reason we see some “certified meth-lab” inspectors and “certified” mould inspectors trying to argue they collected a “real” “scientific” sample– when, in fact, all they have done is fulfill the role of a Chimp with an anemometer and they are left with a very fancy “laboratory report,” but no data because their sample didn’t actually represent anything at all.

General Industrial Hygiene Considerations

Generally, the geometric standard deviation (GSD) of interday and intraday airborne contaminants lies between 1.2 and 2.5 geometric standard deviations. (2) These large variations have been known to legitimate Industrial Hygienists for decades, (3)(4)(5)(6)(7) and are similar to those seen by other authors, specific to airborne mould concentrations (8)(9)(10) some of whom have reported even higher fungal variations in indoor air. (11)

Due to these kinds of variations, classic Industrial Hygiene sampling strategy indicates that *reasonable* confidence in estimating an average ambient airborne concentration in an environment is achieved when at least 70% of the exposure time is measured. (12) Therefore, random grab samples are the least desirable technique for estimating the average exposure. (13) Spore traps, as used by virtually all “certified mould consultants”, are in fact grab samples whose normal total sampling time is usually less than one half of 1% of the anticipated exposure time in the structure.

When grab samples are used, we know the error associated with them is *extremely* elevated, and is reduced somewhat when multiple grab samples are collected (this is "reducing the sampling error").

So, how many samples would one have to take to overcome the sampling error of grab samples like this? Although a counsel of perfection, one consideration called "Shannon's Sampling Theorem" (14) estimates that the number of measurements needed to achieve "perfect information" about airborne concentrations is roughly 250,000 measurements per cubic meter per hour. While working on an early ASTM sampling draft, a colleague, Dr. Joe Spurgeon, did some high-faluttin' mental gyrations and was able to whittle that number down to a more manageable 12,000 samples per house.

As also discussed in "**The Problem with Precision**" not only is the enormous deviations in air concentrations seen from minute to minute, but, there are enormous differences in many air contaminants (e.g. spores) over the distance of just a few inches. That is, we must be able to characterize both spatial representation and temporal representation.

Generally, by characterizing the error, and the uncertainty, and placing the uncertainty into perspective, we are allowed to collected much fewer samples that permit comparability (the topic of the next installment). Accepted classic industrial hygiene references(15)(16) have estimated that for each daily study period (when expressed as any 8 hour period), between eight and eleven random grab samples are needed for every study area, (such as every room in a house) to obtain adequate confidence in knowing the distribution of the concentration of airborne contaminants (notice here the minimum number is for knowing the distribution, not necessarily knowing the actual concentration).

Therefore, for an anticipated 16 hour occupancy for an home, if we wanted to characterize the variation in spore concentrations, we would shoot for a minimum of 16 to 22 grab samples per area. So for a three bedroom, two bathroom house with living room, kitchen, etc, we would probably need no fewer than 112 to 154 air samples. How many "certified" mould inspectors are doing anything even remotely close to this? Answer: None.

Some authors (17) state that as many samples as necessary to determine the distribution should be taken, and therefore, the actual number of samples will be incumbent on the size and topography of the study area. I find that for grab samples when looking for "averages" no fewer than seven samples are needed to characterize the distribution for any one compartment in the study area.

The spatial dimensions, as well as the size of the study population, impact the decision on the number of samples needed. According to NIOSH, when looking at equal at-risk populations, in order to ensure, with 90% confidence, that at least one individual in the top 10% of the exposure range will be included in the study, the following number of samples is required; each sample representing 70% of the estimated exposure time.

Size of Group	Number of Samples Required
n<8	n
8	7
9	8
10	9
11-12	10
13-14	11
15-17	12
18-20	13
21-24	14
25-29	15
30-37	16
38-49	17
50	18

Then, having determined the minimum number of samples required, one is still faced with the problem of which individuals within the study group are to be selected. NIOSH provides a guidance table to help with the selection process since humans are notoriously poor with true random selections.

Representation

The point here is that simply collecting a “sample” using “scientific equipment” does not inherently mean the sample has intrinsic meaning or even speaks to the tacit reason for collecting the sample in the first place. Just because an investigator claims the sample *represents* an aspect under investigation, does not mean the sample actually represents any such thing.

We see this sort of silliness even in regulatory sampling. For example, Colorado has new “meth-lab” cleanup regulations wherein the State of Colorado made history when it intentionally removed all legitimate data quality objectives. In Colorado, it is now lawful for an untrained consultant to enter an heavily contaminated meth-lab chocker-block full of hazardous chemicals, explosives, and illegal equipment, and like the Chimp, collect nonsensical “samples” that represent nothing at all, and declare the property “compliant” with State regulations. The reason this can happen (and does happen) is because the State *intentionally* disallowed all scientific sampling theory when it promulgated the regulations. As a result, Colorado Citizens and their families are now moving into heavily contaminated properties falsely believing the State regulations have been fulfilled and falsely believing that “scientific sampling” has confirmed the properties are safe. When problems are encountered, staff in the Colorado Department of Public Health and Environment merely falsify the public record, lie to the Citizens and grant “variations” based on documents that otherwise don’t exist. (18) Apparently the CDPHE has never heard of “Flint, Michigan.”

In Colorado, the State Department of Public Health and Environment is then astonished when a legitimate Industrial Hygienist performs legitimate environmental sampling, and discovers that the properties are in fact heavily contaminated. (19) In Colorado, unfortunately, the Colorado Department of Public Health and Environment overcomes this problem by simply lying to the

public, with the criminal intent to defraud, and falsifies the real estate record with impunity and to the chemical harm of Colorado citizens.

In conclusion, when performing sampling of *any* kind, even like that in Chita, Russia, if one does not have DQOs, one does not have data. Without ensuring that the samples are collected pursuant to *a priori* DQOs with specified precision, accuracy, and representativeness, one cannot conceivably move to the next stage of data acquisition, which is "Comparability." "Comparability is the first "C" in the PARCC parameters, and will be the topic of my next discussion.

References

1) *Traversing A Duct To Determine Average Air Velocity Or Volume*, APPLICATION NOTE AF-106 (A4) Airflow Instruments, TSI Instruments Ltd. AF-106-A4 (8/11/2014) ©2014 TSI Incorporated Copyright 2014, used here without permission in good faith according to the Fair Use Doctrine.

2) NIOSH Occupational Exposure Sampling Strategy Manual, HEW Publication Number 77-173 (1977)

3) Larsen R.I, *A Method for Determining Source Reduction Required to Meet Quality Standards* JAPCA, 11, 71, 1961

4) Larsen R.I, *A New Mathematical Model of Air Pollutant Concentration Averaging Time and Frequency*, JAPCA, 19, 24 (1969)

5) Breslin AJ, Ong, Glauberman H, *et al*, *The Accuracy of Dust Exposure Estimates Obtained from Conventional Air Sampling* J AIHA, Vol. 8, pp 56-61, (1967)

6) Sherwood RJ *On the Interpretation of Air Sampling for Radioactive Particles* Health Physics and Medical Division Atomic Energy Research Establishment, Presented at the AIHA Conference in Philadelphia, 1964 and appearing in its peer reviewed form in J of AIHA Vol. 27, pp 98-109 (1966)

7) Phinney DE, Newman JE, *The Precision Associated with the Sampling Frequencies of Total Particulate at Indianapolis, Indiana* JAPCA, 22, 9, (1972)

8) Spurgeon, J; Data submitted to the ASTM D22.08.02 Committee for review, October 2005

9) Connell, CP, *Sample results: What do they really tell us?* Presented at the IAQ in Schools Lecture Series, Corpus Christi, TX, 2003

10) Eudey L, Su HJ, Burge HA. *Biostatistics and bioaerosols*. In Bioaerosols, Burge HA, ed. Boca Raton: Lewis Publishers, pp. 269-307. 1995.

11) Reponen T, Nevalainen A, Jantunen M, et al, *Normal Range Criteria for Indoor Air Bacteria and Fungal Spores in a Subarctic Climate; Indoor Air,* 2:26-31 (1992). Referenced by Macher JM, Chatigny MA, Burge HA. *Sampling airborne microorganisms and aeroallergens*. In: Cohen BS, Hering SV, eds. *Air sampling instruments for evaluation of atmospheric contaminants*, 8th ed. Cincinnati, OH: American Conference of Governmental Industrial Hygienists, Inc., pp. 589-617, but not reviewed by this author (Connell).

12) NIOSH Occupational Exposure Sampling Strategy Manual, HEW Publication Number 77-173 (1977)

13) Ibid.

14) As referenced in Rock JC; *Occupational Air Sampling Strategies*, Chapter 2 of Air Sampling Instruments for Evaluation of Atmospheric Contaminants (ACGIH, 2001)

15) NIOSH Technical Information Exposure Measurement Action Level and Occupational Environmental Variability, HEW Publication 76-131, Cincinnati OH, 45226, (1975)

16) NIOSH Occupational Exposure Sampling Strategy Manual, HEW Publication Number 77-173 (1977)

17) Macher JM, Chatigny MA, Burge HA *Sampling airborne microorganisms and aeroallergens*. In: Cohen BS, Hering SV, eds. *Air sampling instruments for evaluation of atmospheric contaminants*, 8th ed. Cincinnati, OH: American Conference of Governmental Industrial Hygienists, Inc., pp. 589-617.

18) See discussion at http://www.forensic-applications.com/meth/coloregs.html

19) See for example: Regulatory Audit- 4893 S Johnson Street, Denver, http://www.forensic-applications.com/meth/Johnson_Critical_review.pdf