the paper are to investigate the (a) effect of sampling methods and sizes on the PWL estimation of a lot, and (b) influence of sampling methods and sizes on the risks to agency and contractor.

BACKGROUND

Generally, statistical quality control in highway industry uses a variable sampling plan since the AQCs used are mostly measured on a numerical scale. One of the main advantages of using variable sampling plans is that the same OC curve can be obtained with a smaller sample size than would be required by an attributes-sampling plan. Thus, a variable acceptance sampling plan that has the same protection as an attribute acceptance sampling plan would require less sampling (Hughes et al. 2011; Montgomery 2013). In particular, when destructive testing is employed, variables sampling will help reduce the costs of inspection. In addition, measurement data usually provide more information about the construction process or the lot than do attributes data. Generally, numerical measurements of quality characteristics are more useful than simple classification of the item as defective or non-defective. Furthermore, when acceptable quality levels are very small, the sample sizes required by attributes sampling plans are very large. Under these circumstances, there may be significant advantages in switching to variables measurement. The most important advantage of adopting a variable acceptance plan for highway construction is its application in calculating a quality measure like percent within limits (PWL). The use of PWL as a quality measure further facilitates the implementation of pay factors. Based on the current literature (Fugro Consultants 2011; Hughes et al. 2011; J. L. Burati 2003; J.L. Burati 2004; Scott et al. 2014), PWL or percent deflective (PD) is the most widely used quality measure. PWL is briefly described below; however, its details can be found elsewhere (AASHTO 2009; Corrigan 2006; Hand and Epps 2006; Hughes 2006; Willenbrock and Kopac 1978).

PWL value of a lot is calculated using the quality index (Q-value) of the specification limits. The Q-statistics are calculated using Equations (1) and (2).

$$Q_L = \frac{\bar{x} - LSL}{s} \tag{1}$$

$$Q_U = \frac{USL - \bar{x}}{s} \tag{2}$$

where;

Q_L = quality index for the lower specification limit

Q_U = quality index for the upper specification limit

LSL = lower specification limit

USL = upper specification limit

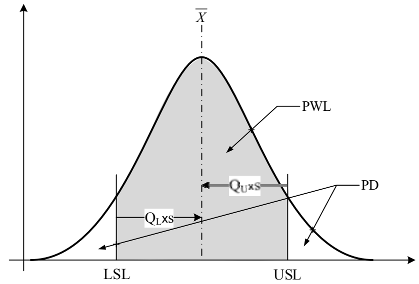
$\bar{x}$ = the sample mean for the lot

s = sample standard deviation for the lot

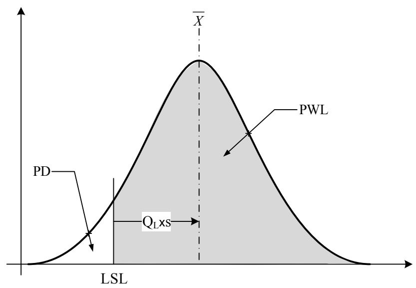
The lower and upper Q-values are used to find the PWL from tables (J. L. Burati 2003), or by using a beta distribution (Willenbrock and Kopac 1978) to estimate the PWL for two-sided specification limits by using the following equation:

$$PWL_T = PWL_U + PWL_L - 100 \tag{3}$$

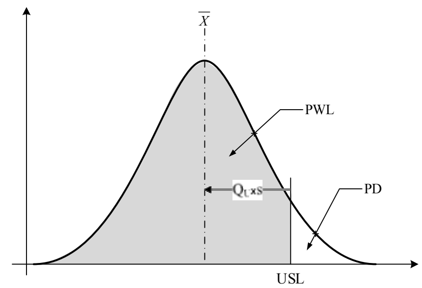
The PWL for double- and single-sided (upper or lower specification limits) are illustrated in Figure 1.



(a) PWL based on double-sided specification limits



(b) PWL based on lower specification limits



(c) PWL based on upper specification limits

Figure 1 Illustration of PWL and PD for different specification limits

TYPES OF SAMPLING METHODS

The units selected for inspection from a lot should be chosen at random, and they should be representative of all the items in the lot. The random-sampling concept is extremely important in acceptance sampling. Unless random samples are used, bias will be introduced and the effectiveness of the inspection process is destroyed. Sometimes the inspector may stratify the lot. This consists of dividing the lot into sublots and then taking random samples from each sublot. Although this stratification of the lot is usually a subjective activity performed by the inspector, it ensures that units are selected from all locations in the lot.

Since, sample size is determined based on the mean AQC by using the statistical methods described above, there is a need to determine the impact of sample size on PWL for developing guidelines for PRS. In addition, the sampling method effects on PWL have not been quantified in the past. Therefore, the impacts of three different sampling methods and four sample sizes on PWL were evaluated in this study. The three sampling methods include: (a) random sampling with replacement, (b) stratified sampling, and (c) random sampling without replacement. For each sampling method, sample sizes of 3, 5, 10, and 20 were evaluated. A two-mile-long lot was assumed for this evaluation. The lot was divided into 20 sublots each 0.1-mile long.

Random Sampling with Replacement

In this sampling method, samples of different sizes are randomly selected from any location on the lot. So for a given sample size, all the samples could be collected from one sublot or different sublots, i.e., with replacement. Figure 2a shows examples of random sampling with replacement for sample sizes of 3, 5, 10, and 20, respectively. The disadvantage of this sampling method is that the samples obtained may not represent the entire lot. The shortcoming might not be a problem if the construction variability is low, i.e., there is uniform construction quality across the entire lot.

Stratified Sampling

In this sampling method, the sublots are grouped together spatially such that the number of samples required is equal to the number of grouped sublots. Note that for a given sample size, each sample has to come from a different group of sublots. Figure 2b shows examples of stratified sampling for sample sizes of 3, 5, 10, and 20, respectively. The aim of stratified sampling is to reduce the bias in the selection of samples. As a result, the samples collected are highly representative of the lot. Stratified sampling also prevents over representation of a particular part of the lot. For a given sample size, stratified sampling provides greater precision than random sampling, especially when construction variability is high. Therefore, it may be possible to use a smaller sample size in this method.

Random Sampling without Replacement

This method is very similar to method 1 except that for a given sample size, each sample has to be collected from a different sublot, i.e., without replacement. It is possible that this method could be more precise than method 1 in the sense that samples selected from the lot are spread more evenly along the lot. Figure 2c shows example random sampling without replacement for sample sizes of 3, 5, 10, and 20, respectively.

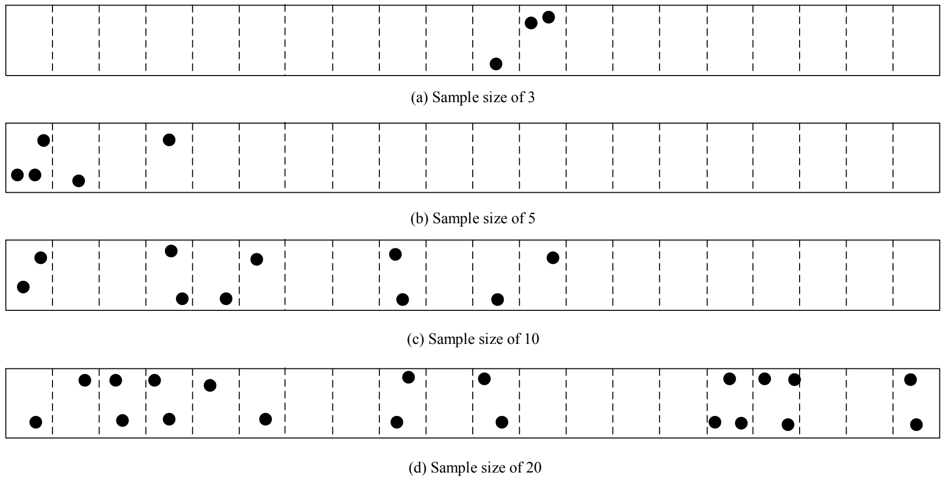
IMPACT OF SAMPLE SIZES AND SAMPLING METHODS ON PWL

The effects of the three sampling methods described above and sample sizes of 3, 5, 10, and 20 on the estimation of PWL of a lot were evaluated. As an example, air void content was chosen as an AQC. Four after-construction qualities (i.e. four different lots) were simulated. For each sampling method, sample of sizes 3, 5, 10, and 20 were used to evaluate the PWL estimations. The estimated PWL values were compared with the true PWL of the lot. The true PWL of a lot was determined based on the simulated values of the air void contents and specifications limits (i.e., lower and upper limits of 2% and 8%, respectively). The four simulated after-construction qualities can be seen in Figure 3.

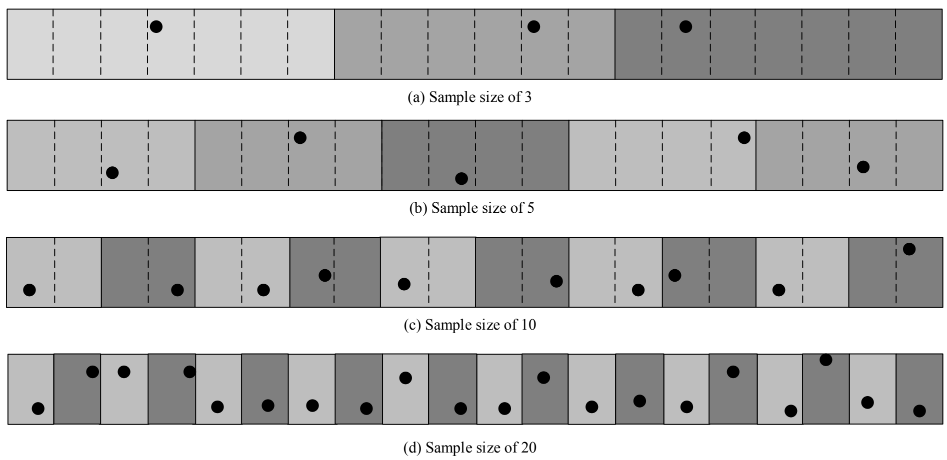
For this evaluation, an acceptable quality level (AQL) of 90 PWL and a rejectable quality level (RQL) of 60 PWL were used. For each construction quality, samples of sizes 3, 5, 10, and 20 were selected from the lot using the different sampling methods. Each time the samples were randomly picked, the PWL value of the lot was estimated. This procedure was simulated 5000 times. The percentage of PWL values below RQL, between AQL and RQL, and above AQL for each sampling method and size were calculated as shown in Table 1. The effect of the sampling methods and sample sizes on the PWL values are discussed below relative to the true construction quality.

Constriction Quality Below RQL

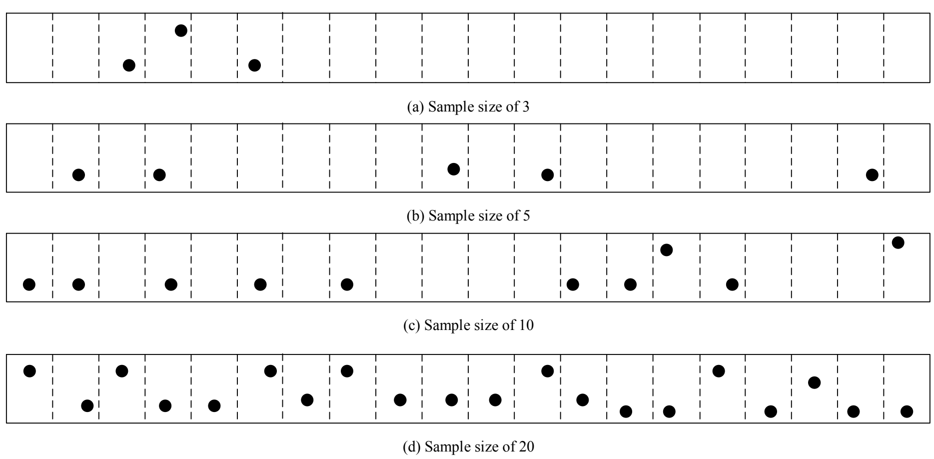
The true PWL value of construction quality 3 is 55% as seen in Table 1. Recall that the established RQL value is 60%. Out of 5000 simulations, stratified sampling (Method 2) has a higher percentage of PWL values less than RQL, thus better representing the true quality.



(a) Random sampling with replacement



(b) Stratified sampling



(c) Random sampling without replacement

Figure 2 Sampling methods for various sample sizes

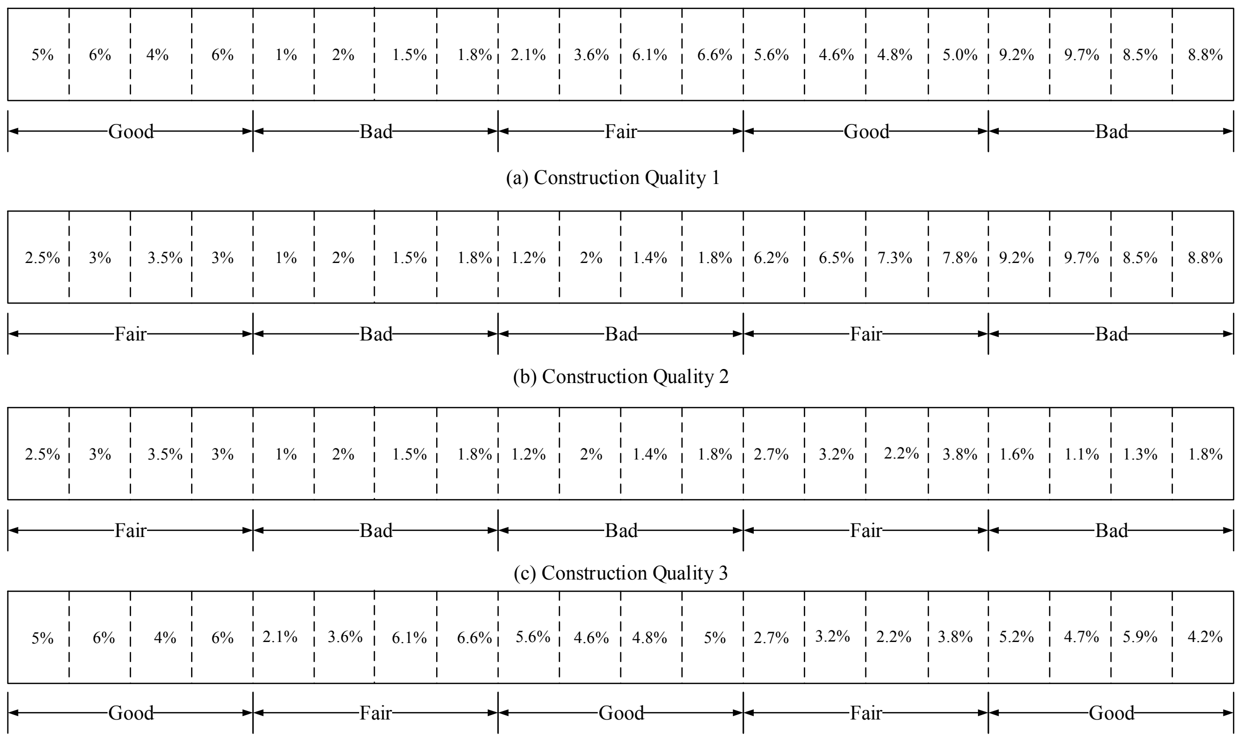
As the sample size increases, the percentage of PWL values less than RQL increases for each of the methods. However, the rate of increase is higher for Method 2. The percentage values between AQL and RQL, and above AQL for Method 2 are lower, which represents the true quality better than the other methods. For a sample size of 3, there is not much difference between the methods, but at a sample size of 5 or more, Method 2 represents the true quality better than the other two methods. This means that, for a sample size of 3 for a lot with poor construction quality, none of the methods will be appropriate.

Construction Quality Between RQL and AQL

The true PWL values of construction qualities 1 and 2 are 75% and 65%, respectively (see Table 1). The construction qualities are between the RQL of 60% and the AQL of 90%. The results of simulations show that stratified sampling (Method 2) has a higher percentage of PWL values between the category RQL and AQL. As the sample size increases, the percentage of PWL values between the category RQL and AQL increases for each method. Once again, this percentage increases at a higher rate for Method 2 as compared to the other methods, as the sample size increases. As expected, the stratified sampling has higher precision than the other methods, especially for a sample size of 5 or more. For Method 2, the percentage PWL values below RQL decreases more rapidly than for other methods as the sample size increases. The percentage of PWL values less than RQL is higher for construction quality 2 than quality 1 because the true PWL value is closer to the RQL value for construction quality 2. For example, for construction quality 2, there is a 65% chance of rejecting the lot when the sample size is 3 for Method 2. When the sample size increases to 5, there is a 28% chance of rejecting the lot even though the actual lot quality if 65% PWL. In addition, for a sample size of 10, there is no chance of rejecting the lot. Therefore, if the true quality is closer to RQL, a larger sample size is needed to represent the actual quality.

Construction Quality Above AQL

The true PWL value of construction quality 4 is 97% (see Table 1). The construction quality is above the AQL of 90%. The percentage of PWL values greater than AQL increases for each method with an increase in sample size. However, stratified sampling (Method 2) represents the true quality better, as seen in Table 1. There is not much difference in percentage of PWL values for all the methods in all the three categories at sample size 3. In case of stratified sampling, for a sample size of 3, the quality of the lot is above AQL only around 66% of the time even though the true quality is 97%. The estimate quality increases to 93% when the sample size is increased to 5. The stratified sampling method will capture the true quality better as compared to other methods, especially when the sample size is 5 or more.



(a) Construction Quality 1

(b) Construction Quality 2

(c) Construction Quality 3

(d) Construction Quality 4

Figure 3 Simulated construction qualities

Quantification of Risks

The PWL value obtained from a given sampling method and size can be used to determine whether the lot quality is below RQL, between RQL and AQL, or above AQL. However, an incorrect decision can sometimes be made regarding the lot quality. For example, if an agency tests three samples from a lot and if the PWL value obtained is below the RQL, the lot is rejected. It could have been the case that the three samples tested are from a sublot of a bad quality and the remaining sublots are of good quality, and therefore an incorrect decision would have been made. In any sampling process, an error can be made by falsely rejecting a good quality material or falsely accepting a bad quality material. The error of falsely rejecting a good quality material is a risk to the contractor and is called contractor's risk (α) and the error of falsely accepting a bad quality material is a risk to the agency and is called agency's risk (β). The complementary of the agency's risk ($1-\beta$) is called the power of the sampling method. The agency's risk can be calculated for each sampling method. As mention above, samples of different sizes were selected 5000 times for each construction quality within a method. The PWL value was calculated each time samples were selected. The mean PWL value and its standard error were calculated. The contractor's risk was set at 0.05, which means that out of one hundred times, the agency is going to reject a good quality lot 5 times, which is an acceptable level of risk. The null hypothesis would be, the average PWL of the lot is above RQL or above AQL. The agency's risk would be accepting the null hypothesis even though the lot quality is less than RQL or AQL respectively.

Table 1 Percentage of PWL values for various construction qualities

Quality Measure	Construction Quality 1					Construction Quality 2				
	Sample size	Percent of PWL values			True PWL	Sample size	Percent of PWL values			True PWL
		Method 1	Method 2	Method 3			Method 1	Method 2	Method 3	
< RQL	3	25%	27%	27%	75%	3	61%	65%	61%	65%
	5	11%	0%	10%		5	43%	28%	41%	
	10	1%	0%	0%		10	18%	0%	8%	
	20	0%	0%	0%		20	3%	0%	0%	
RQL - AQL	3	56%	56%	58%		3	38%	35%	38%	
	5	75%	100%	81%		5	57%	72%	59%	
	10	92%	100%	99%		10	82%	100%	92%	
	20	98%	100%	100%		20	97%	100%	100%	
> AQL	3	19%	17%	15%		3	2%	0%	1%	
	5	14%	0%	9%		5	0%	0%	0%	
	10	7%	0%	1%		10	0%	0%	0%	
	20	2%	0%	0%		20	0%	0%	0%	
Quality Measure	Construction Quality 3					Construction Quality 4				
	Sample size	Percent of PWL values			True PWL	Sample size	Percent of PWL values			True PWL
		Method 1	Method 2	Method 3			Method 1	Method 2	Method 3	
< RQL	3	61%	65%	62%	55%	3	0%	0%	0%	97%
	5	62%	74%	63%		5	0%	0%	0%	
	10	65%	79%	70%		10	0%	0%	0%	
	20	72%	100%	100%		20	0%	0%	0%	
RQL - AQL	3	33%	33%	34%		3	32%	34%	30%	
	5	35%	26%	36%		5	20%	7%	15%	
	10	35%	21%	30%		10	8%	0%	1%	
	20	28%	0%	0%		20	2%	0%	0%	
> AQL	3	6%	2%	4%		3	68%	66%	70%	
	5	2%	0%	1%		5	80%	93%	85%	
	10	0%	0%	0%		10	92%	100%	99%	
	20	0%	0%	0%		20	98%	100%	100%	

These risks are calculated using Equations (4) and (5).

$$\beta(PWL_{True}) = P\left[z < z_{\alpha} - \frac{|PWL_{RQL} - PWL_{True}|}{SE}\right] \tag{4}$$

$$\beta(PWL_{True}) = P\left[z < z_{\alpha} - \frac{|PWL_{AQL} - PWL_{True}|}{SE}\right] \tag{5}$$

where,

β = Agency's risk
PWL_{True} = Measured PWL value from samples
PWL_{RQL} = PWL value at RQL (60%)
PWL_{AQL} = PWL value at AQL (90%)
z_{α} = z-value at α =0.05
SE = Standard error of the average PWL value

The results of risk calculations are shown in Table 2. The power of a sampling method increases as the sample size increases. This is because as the sample size increases, the standard error of the mean PWL value decreases. The results also show that Method 2 (stratified sampling) has a higher power as compared to the other methods. As mentioned before, Method 2 requires collecting samples along the entire lot, and hence the variability in PWL values is less. Also, the closer the average PWL value to the RQL value, the lower the power. This means, that there is a high chance of wrongly accepting a lot if its PWL value obtained from the sampling is near the RQL.

SUMMARY OF FINDINGS

The statistical approaches for determining sample size for a lot generally estimate a higher sample size to represent a lot. None of the statistical approaches adequately addresses the optimum sample size. However, in practice, people responsible for QA specification development and use typically use a sample size of at least 5, although they recognize that a higher sample size is desirable, since they consider that a larger sample size may not be practical. If this sample size is too small, the probability of making erroneous acceptance or pay adjustment decisions would be too high for agencies. If this sample size is too large, the cost of sampling and testing would be unnecessarily high, especially where destructive testing is used.

Three sampling methods, (a) random sampling with replacement, (b) stratified sampling, and (c) random sampling without replacement, for sample sizes of 3, 5, 10, and 20 were evaluated. For a given sample size, stratified sampling provides greater precision than random sampling, especially when construction variability is high. Therefore, it may be possible to use a smaller sample size in this method. It is possible that random sampling without replacement could be more precise than random sampling with replacement since samples selected from the lot are spread more evenly along the lot.

Table 2 Power of various sampling methods and sizes at RQL and AQL for different construction qualities

Construction Quality	Sample Size	Method 1				Method 2				Method 3			
		Mean	SE	β	Power	Mean	SE	β	Power	Mean	SE	β	Power
1	3	71%	18%	0.85	15%	71%	16%	0.83	17%	71%	17%	0.84	16%
	5	73%	12%	0.71	29%	72%	4%	0.09	91%	73%	11%	0.68	32%
	10	75%	8%	0.41	59%	73%	2%	0.00	100%	74%	6%	0.25	75%
2	3	56%	17%	0.92	8%	58%	8%	0.92	8%	57%	14%	0.92	8%
	5	61%	9%	0.94	6%	62%	3%	0.84	16%	62%	7%	0.91	9%
	10	64%	5%	0.80	20%	64%	1%	0.01	99%	64%	3%	0.62	38%
3	3	50%	27%	0.90	10%	48%	25%	0.88	12%	51%	25%	0.90	10%
	5	53%	20%	0.90	10%	55%	8%	0.85	15%	53%	18%	0.90	10%
	10	54%	13%	0.88	12%	56%	6%	0.84	16%	55%	9%	0.86	14%
4	3	94%	8%	0.00	100%	94%	7%	0.00	100%	93%	7%	0.00	100%
	5	95%	5%	0.00	100%	96%	4%	0.00	100%	95%	5%	0.00	100%
	10	96%	4%	0.00	100%	96%	2%	0.00	100%	96%	2%	0.00	100%

(a) RQL

Construction Quality	Sample Size	Method 1				Method 2				Method 3			
		Mean	SE	β	Power	Mean	SE	β	Power	Mean	SE	β	Power
1	3	71%	18%	0.72	28%	71%	16%	0.68	32%	71%	17%	0.70	30%
	5	73%	12%	0.59	41%	72%	4%	0.00	100%	73%	11%	0.54	46%
	10	75%	8%	0.41	59%	73%	2%	0.00	100%	74%	6%	0.15	85%
2	3	56%	17%	0.36	64%	58%	8%	0.01	99%	57%	14%	0.24	76%
	5	61%	9%	0.06	94%	62%	3%	0.00	100%	62%	7%	0.01	99%
	10	64%	5%	0.00	100%	64%	1%	0.00	100%	64%	3%	0.00	100%
3	3	50%	27%	0.56	44%	48%	25%	0.49	51%	51%	25%	0.53	47%
	5	53%	20%	0.42	58%	55%	8%	0.00	100%	53%	18%	0.34	66%
	10	54%	13%	0.13	87%	56%	6%	0.00	100%	55%	9%	0.01	99%
4	3	94%	8%	0.87	13%	94%	7%	0.86	14%	93%	7%	0.89	11%
	5	95%	5%	0.74	26%	96%	4%	0.56	44%	95%	5%	0.74	26%
	10	96%	4%	0.56	44%	96%	2%	0.09	91%	96%	2%	0.09	91%

(b) AQL

For a sample size of 3, there is not much difference between the sampling methods, but at a sample size of 5 or more, the stratified sampling method represents the true construction quality better than the other two methods. If the true quality of a lot is closer to RQL, a larger sample size (i.e., 5 or more) is needed to represent the actual quality. The power of a sampling method increases as the sample size increases. Stratified sampling method has a higher power as compared to other sampling methods. Stratified sampling requires collecting samples along the entire lot, and hence the variability in PWL values is less. Also, the closer the average PWL value to the RQL value, the lower the power. This means, that there is a high chance of wrongly accepting a lot if its PWL value obtained from the sampling is near the RQL. Therefore, agencies can start testing with five samples using stratified sampling by dividing a lot into five sublots (one sample from each sublot) if destructive testing is needed. If the variability within the samples is high or the PWL estimates are close to RQL, additional samples could be collected. This decision of testing additional samples can be made by evaluating the costs of falsely accepting bad quality material and the costs of additional testing.

REFERENCES

AASHTO (2009). "Acceptance Sampling Plans for Highway Construction."

Corrigan, M. R. (2006). "Percent Within Limits as Part of Federal Highway Administration's Quality Initiatives." *Journal of the Association of Asphalt Paving Technologists*, 75.

Fugro Consultants, I., Arizona State University (2011). "NCHRP Report 704 - A Performance-Related Specification for Hot-Mixed Asphalt." *Transportation Research Board*(NCHRP Report 704).

Hand, A. J., and Epps, J. A. (2006). "Fundamentals of Percent within Limits and Quality Control–Quality Assurance Compaction Specifications." *Transportation Research E-Circular No. EC105*, 140-162.

Hughes, C. S. (2006). "What in the World is PWL?" *Journal of the Association of Asphalt Paving Technologists*, 75.

Hughes, C. S., Moulthrop, J. S., Tayabji, S., Weed, R., and Burati, J. (2011). "NCHRP 10-79 - Guidelines for Quality-related Pay Adjustment Factors for Pavements."

J. L. Burati, R. M. W., C. S. Hughes, H. S. Hill (2003). "Optimal Procedures for Quality Assurance Specifications."

J.L. Burati, R. M. W., C.S. Hughes, and H.S. Hill (2004). "Evaluation of Procedures for Quality Assurance Specifications."

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*, John Wiley & Sons.

Scott, S., Ferragut, T., Anderson, S., Damjanovic, I., Huber, G., Katafanas, J., McGhee, K., Sprinkel, M., Ozyilirim, C., Diefenderfer, B., Merrit, D., Dawood, D., Molenar, K., Loulakis, M., White, D., and Schaeffer, V. (2014). "Performance Specifications for Rapid Highway Renewal."

Willenbrock, J. H., and Kopac, P. A. (1978). "Development of a Highway Construction Acceptance Plan." *Transportation Research Record*(691), 16-22.