

# DETERMINING THE OPTIMUM SAMPLE SIZE FOR QUALITY ASSURANCE (QA) OF ASPHALT MIXTURES: A CASE STUDY

---

HAYDAR AL-KHAYAT<sup>1\*</sup>, CHARLES GURGANUS<sup>1</sup>,  
DAVID E. NEWCOMB<sup>1</sup>, MARYAM S. SAKHAEIFAR<sup>2</sup>

<sup>1</sup>*Materials and Pavements Division, Texas A&M Transportation Institute,  
Bryan TX, United States*

<sup>2</sup>*Turner-Fairbank Highway Research Center, Federal Highway Administration,  
SE Washington DC, United States*

Received 27 July 2021; accepted 1 April 2022

**Abstract.** Acceptance plans for asphalt mixtures use a certain sample size that is often established based on the purpose of sampling, population size, risk, and allowable error for evaluation. The rate of quality control (QC) sample size is often higher than the quality assurance (QA) sample size. The test results obtained from the QA samples are commonly used to validate the QC test results and to assist the state department of transportation (DOT) with payment decisions. However, if the QA sample size is insufficient to make accurate judgments, the probability of making incorrect decisions regarding acceptance increases. On the other hand, oversampling needlessly consumes both time and cost. To identify the appropriate sample size for QA testing, a balance must be struck between a number of variables. In this case study, two models were developed using the Oregon Department of Transportation (ODOT) data to determine the appropriate QA sample size. The need for this work was realized when a review of ODOT paving projects revealed a large variability in lot size.

---

\* Corresponding author. E-mail: haydar@tamu.edu

Haydar AL-KHAYAT (ORCID ID 0000-0003-3438-2978)

Copyright © 2022 The Author(s). Published by RTU Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

These ranged from 3000 to more than 100 000 tons with commensurate QA sample size rates. The typical standard deviation (*STDEV*) values of asphalt content (AC) and in-place density were determined. The developed models show that using the *STDEV* values that represented more than 90 percent of the projects, ODOT needed to increase QA sample size for both AC and density in lots of less than 22 000 tons. The results also show that sample can be decreased for AC and remain as is for density in projects of more than 22 000 tons of asphalt mixtures. The proposed models can be used to determine the optimum sample size for different lots sizes.

**Keywords:** optimum sample size, lot size, quality control, quality assurance, standard deviation.

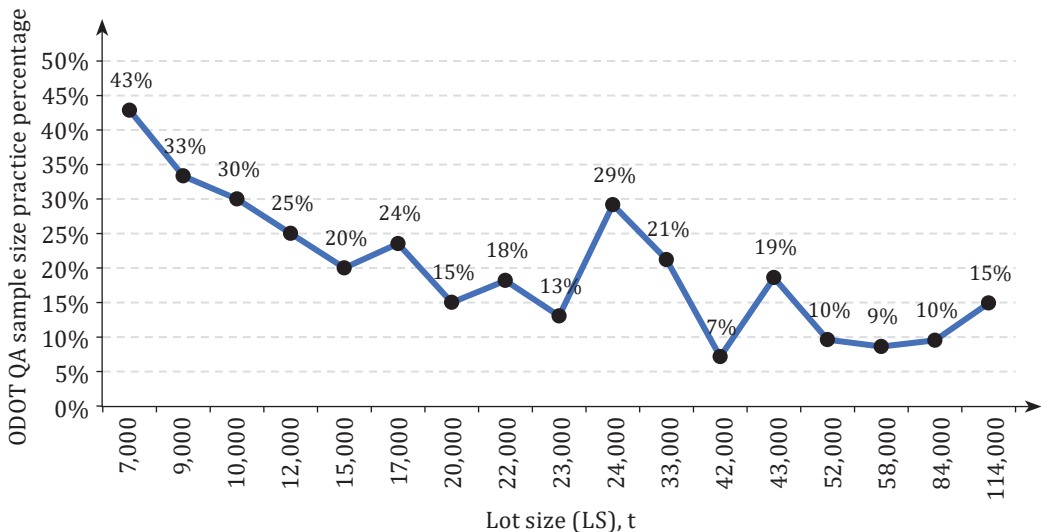
## Introduction

Sample size ( $n$ ) refers to as a number of samples/tests taken randomly from the asphalt mixture lot (also called population) and used to assess the asphalt mixture quality. Typically, quality control (QC) (i.e., testing performed by the contractor) and quality assurance (QA) (i.e., testing performed by the state department of transportation (DOT)) are sampled and tested at different rates. The rate of QC  $n$  is often higher than the QA  $n$ . While a larger  $n$  yields more reliable results, it is impractical for QC or QA testing to sample a large population of materials such as the entire asphalt mixture lot (Winter, 2013). Therefore, a smaller  $n$  is used to save time, cost, and speed up the paving process. In most cases, QC collects one sample per subplot. A subplot is defined as a portion of the lot, while QA collects one sample per 10 000 tons or sampling at a specific rate range between 5–10% of QC  $n$  (Elseifi, 2007). In general, the QA  $n$  ranges from three to seven units per lot. Typically, results from QA  $n$  are used to verify QC results (Gharaibeh et al., 2010). After verification, the Federal Highway Administration (FHWA) allows DOTs to use the QC data for asphalt mixture acceptance and to determine the percent of the lot within limits and the pay factors. Verification of QC results by comparing them with the QA is usually done by running t-test and F-test. DOTs can use QC data for payment when the t-test and F-test are passed. A study with 42 state DOTs responding showed that 27 DOTs used QC data for acceptance of the asphalt mixture after verification (Schmitt et al., 1998; Al-Khayat, 2018).

It is very important to utilize the optimum QA  $n$  for evaluation and verification processes. Too large QA  $n$  (more than needed) is costly and may not be necessary. On the contrary, too small QA  $n$  (less than optimum) may increase the probability of making incorrect decisions such as using the QC data for acceptance and payment when it is not valid. In most cases, QC and QA are testing the asphalt mixture

aggregate gradation, asphalt content (AC), air voids (AV), and density. These variables are often tested because studies have shown that these properties are strongly related to pavement performance. Therefore, DOTs monitor and test them to ensure the asphalt mixture quality, and then decide to reject or accept all or part of the asphalt mixture lot, and adjust the payment as necessary (Winter, 2013; Willenbrock, 1976; Newcomb et al., 2016).

The Oregon Department of Transportation (ODOT) uses QC tests after verifying with QA tests to accept or reject the asphalt mixture. The ODOT standard specification defines the lot size as a total quantity of asphalt mixture per project with the same job mix formula. The subplot size is 1000 tons on asphalt mixture (Oregon Department of Transportation, 2018). Therefore, the lot size and the QA  $n$  vary from project to other. In Oregon, the pay elements of the asphalt mixture are aggregate gradation, AC, and density. To test them, QC collects one sample per subplot (i.e., one sample for every 1000 tons of asphalt mixture). According to the ODOT specification, QA samples at a minimum rate of 10% of QC  $n$ , or a minimum of three samples, whichever is larger (Oregon Department of Transportation, 2018; Oregon Department of Transportation, 2009). For instance, the QC must collect 10 samples from a project consisting of 10 000 tons (10 sublots) of asphalt mixture, while QA is required to collect at least three. Statistically, 10% of the QC  $n$  or three samples may not be enough to represent the lot quality or to make acceptance decisions. The gradation, AC, and density which are



**Figure 1.** ODOT QA  $n$  performed practice percentage for AC and density

considered as pay elements may need to be sampled at different rates corresponding to their variability (Al-Khayat, 2018; Willenbrock, 1976).

The need to develop models to assist ODOT in determining the optimum QA arose from the variability in QA  $n$  and lot size. In projects constructed in 2014–2018, the QA  $n$  ranged from three (i.e., minimum required  $n$ ) to 17 samples (i.e., 9% to 43% of QC  $n$ ) for lot sizes ranging from 7000 to 114 000 tons. In this case study, two models were developed to determine the optimum  $n$  for the AC and density. The typical standard deviation ( $STDEV$ ) values, lot size, allowable error, and a 95% confidence level were considered. The AC data were obtained from 20 lots, while density data were obtained from 17 lots constructed by ODOT contractors during the 2014–2018 paving seasons. The data were representing most contractors in Oregon state. The formula used to determine the optimum QA  $n$  and develop the models is standardized in ASTM D 6433 (ASTM, 2020). The formula in ASTM D 6433 is widely used to determine the optimum sample size for pavement condition index surveys. Figure 1 shows the ODOT QA  $n$  as a percentage of QC  $n$  for the AC and density of 17 lots paved during the 2014–2018 paving seasons.

## 1. Objective

The objectives of this study are to:

- Propose two models that can assist ODOT in determining the optimum QA  $n$  to test the AC and the density of asphalt mixtures for different lot sizes to ensure the quality.
- Verify the current ODOT QA sampling plan.

## 2. Research method

Determining a statistically relevant  $n$  ensures the ability to properly compare the QC and the QA results when t-test and F-test are used. The optimum QA  $n$  can be determined for various levels of confidence and lot sizes from Eq. (1).

$$n = \frac{STDEV^2 N}{STDEV^2 + \frac{(N-1)e^2}{4}}, \quad (1)$$

where

$n$  – the required number of QA samples;

$N$  – the population size represented by the number of sublots or number of QC tests;

$STDEV$  – the assumed or calculated (Lot) standard deviation;

$e$  – the acceptable level of precision.

The above equation assumes a normal distribution associated with the parameters of interest (i.e., AC and density). In fact, and based on historical construction materials test results, material testing data follow the normal distribution (Al-Khayat, 2018).

The *STDEV* value represents the distribution of the data. A larger *STDEV* value means higher variability (less homogeneous), requiring a larger QA *n*. Selecting the *STDEV* value to determine the optimum QA *n* is important and can be critical if it is not close to the true *STDEV*. Therefore, AC QC results from 20 lots and density QC results from 17 lots were used to determine the typical *STDEV* values. For AC, two levels of *STDEV* were selected. The first *STDEV* value was 0.2 and represented 90% of the studied lots. Thus, 90% of the studied lots had a *STDEV* less than or equal to 0.2. The second *STDEV* value was 0.3, which represented 100% of the studied lots. Thus, in the worst case, the largest *STDEV* value was 0.29. Table 1 shows the *STDEV* values of AC QC results obtained

Table 1. *STDEV* values of AC from 20 lots

Lot number	AC measurements / number of sublots	<i>STDEV</i> of AC
1	114	0.17
2	33	0.14
3	43	0.11
4	7	0.13
5	9	0.21
6	10	0.14
7	17	0.24
8	20	0.15
9	22	0.29
10	24	0.12
11	42	0.13
12	30	0.17
13	30	0.13
14	39	0.16
15	69	0.11
16	56	0.17
17	78	0.19
18	15	0.12
19	65	0.16
20	32	0.13

from 20 lots built by ODOT contractors during the 2014–2018 paving seasons.

A similar calculation was performed on the average QC density results. This calculation has an intricacy as the QC value reported is the average of five individual density measurements within each subplot and reported as a single value for subplot density. The *STDEV* values were determined from 17 asphalt mixture lots. The maximum *STDEV* value was 0.83, and the minimum was 0.11. Two *STDEV* values of 0.76 and 0.84 were selected to develop the density model. The first *STDEV* value of 0.76 was greater than 94% of *STDEV* values found within the studied lots (Table 2), while the second *STDEV* value of 0.84 was used to represent the worst-case scenario (greater than the maximum *STDEV* of 0.83 found within the studied lots). Table 2 shows the *STDEV* values of density results obtained from 17 asphalt mixture lots built by ODOT contractors during the 2014–2018 paving seasons.

*Population Size (N)* was obtained by dividing the lot size (asphalt mixture quantity in the entire project in tons) by subplot size (1000 tons).

Table 2. *STDEV* values of density measurements from 17 lots

Lot number	Density measurements / number of sublots	<i>STDEV</i> of average density
1	114	0.56
2	33	0.52
3	43	0.66
4	7	0.76
5	10	0.54
6	17	0.50
7	23	0.11
8	11	0.45
9	13	0.66
10	20	0.75
11	18	0.52
12	27	0.83
13	84	0.68
14	23	0.72
15	58	0.44
16	15	0.42
17	52	0.57

The model was developed based on different lot sizes for applicability to most scenarios of lot sizes in Oregon state (Fig. 1).

The sampling error,  $e$ , is defined as the maximum allowable difference between the true average and the sample mean value. A large allowable error reduces the  $n$ . The allowable difference was assumed at an acceptable level of precision,  $\Delta$ , of 0.15 for AC and 0.5 for density.

### 3. Results and discussion

The sample size formula presented in the research method section was used to develop the models (lot size vs.  $n$ ). Figure 2 shows the models that can be used to determine the optimum  $n$  for AC. The red curve was created by using the greater  $STDEV$  value of 0.3, while the blue curve was created by using the  $STDEV$  value of 0.2. The 0.3  $STDEV$  value leads to a larger  $n$ . The black dots in Fig. 2 represent the ODOT QA  $n$  that took place in the actual construction practice of the studied lots built during the 2014–2018 paving seasons. The green curve represents the minimum ODOT  $n$  requirements (i.e., 10% of QC  $n$ , or a minimum of three samples, whichever is larger). The sampling model developed using a  $STDEV$  of 0.2 (blue curve) indicates that the ODOT QA  $n$  on the studied lots is below the proposed  $n$  in lots consisting of 23 000 tons or less. On the other hand, based on actual practices, ODOT oversampled

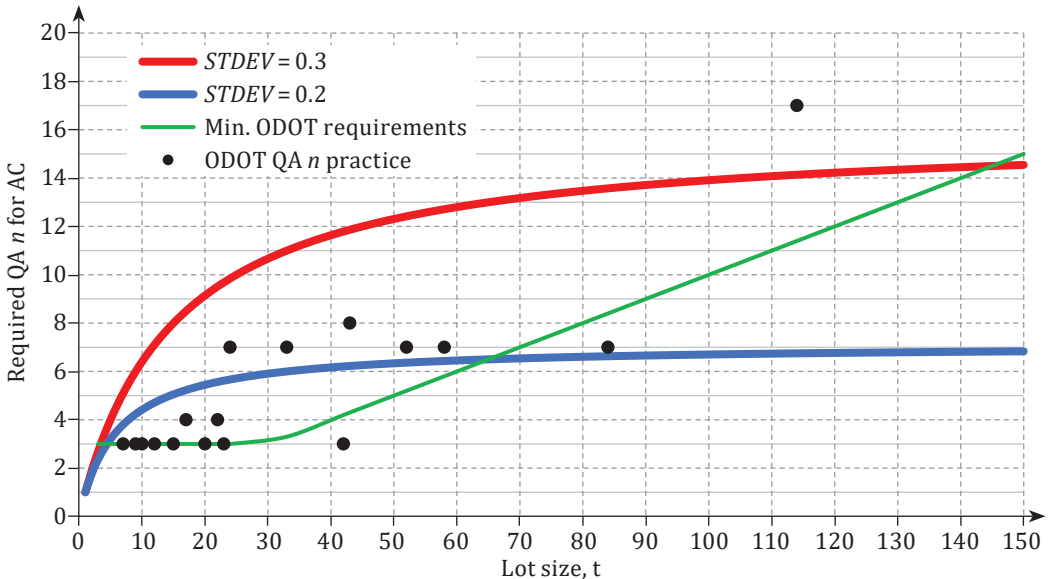


Figure 2. Required QA  $n$  for AC

on lots consisting of 24 000 tons and more. Additional samples are required when a *STDEV* value of 0.3 is used. All studied lots fall below the red curve except the project that consisted of 114 000 tons. Table 3 compared ODOT QA *n* performed practice on studied lots versus the ODOT minimum requirements versus the proposed QA *n* based on the two values of *STDEV* (i.e., 0.2 and 0.3).

Figure 3 shows the sampling model that can be used to determine the optimum *n* for density. The first model (red curve in Fig. 3) was created using a *STDEV* of 0.84, while a *STDEV* of 0.76 was used to develop the second model (blue curve in Fig. 3). The block dots represent the ODOT QA *n* that took place in the actual construction practice of the studied lots built during the 2014–2018 paving seasons. The density sampling model using a *STDEV* of 0.76 (blue curve) indicates that the ODOT QA *n* in practice was below the proposed *n* in most cases. However, lots with 24 000, 33 000, and 43 000 tons fall within the proposed sampling guidelines presented in the model. The very large project, consisting of 114 000 tons, was sampled more than needed according to the proposed

Table 3. The proposed *n* based on 0.2 and 0.3 *STDEV* values

Lot size (LS)- thousands ton	ODOT QA <i>n</i> performed practice on studied lots, %	ODOT QA <i>n</i> minimum requirements	Proposed <i>n</i> , when using 0.2 <i>STDEV</i> , %	Proposed <i>n</i> , when using 0.3 <i>STDEV</i> , %
LS – 7 000	3 (43%)	3 (43%)	4 (57%)	6 (86%)
LS – 9 000	3 (33%)	3 (33%)	5 (56%)	6 (67%)
LS – 10 000	3 (30%)	3 (30%)	5 (50%)	7 (70%)
LS – 12 000	3 (25%)	3 (25%)	5 (42%)	7 (58%)
LS – 15 000	3 (20%)	3 (20%)	5 (33%)	8 (53%)
LS – 17 000	4 (24%)	3 (18%)	6 (35%)	9 (53%)
LS – 20 000	3 (15%)	3 (15%)	6 (30%)	10 (50%)
LS – 22 000	4 (18%)	3 (14%)	6 (27%)	10 (45%)
LS – 23 000	3 (13%)	3 (13%)	6 (26%)	10 (43%)
LS – 24 000	7 (29%)	3 (13%)	6 (25%)	11 (46%)
LS – 33 000	7 (21%)	4 (12%)	6 (18%)	11 (33%)
LS – 42 000	3 (7%)	5 (12%)	7 (17%)	12 (29%)
LS – 43 000	8 (19%)	5 (12%)	7 (14%)	12 (33%)
LS – 52 000	7 (10%)	6 (12%)	7 (13%)	13 (23%)
LS – 58 000	7 (9%)	6 (10%)	7 (12%)	13 (22%)
LS – 84 000	7 (10%)	9 (11%)	7 (8%)	14 (17%)
LS – 114 000	17 (15%)	12 (11%)	7 (6%)	15 (13%)



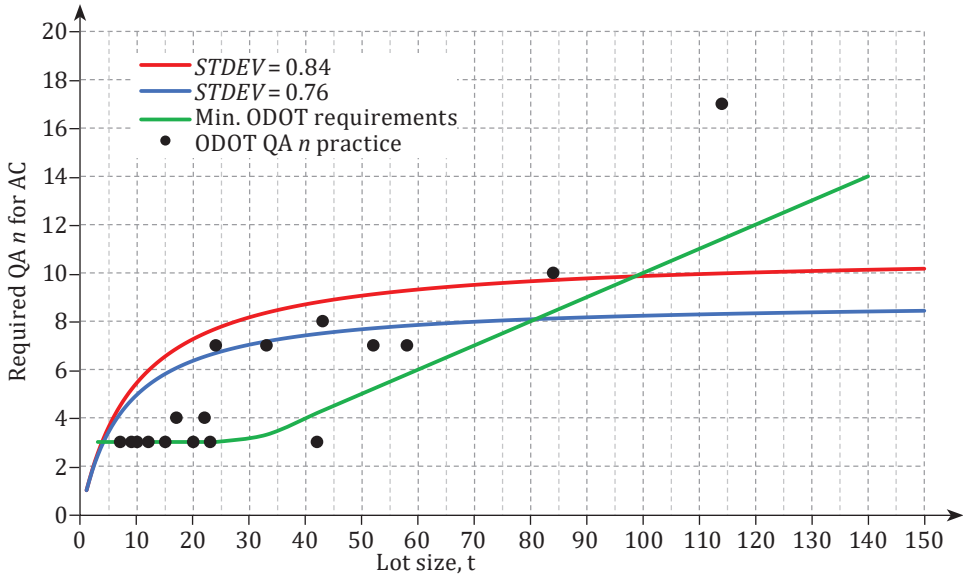


Figure 3. Required QA  $n$  for density

Table 4. The proposed  $n$  based on 0.76 and 0.84  $STDEV$  values

Lot size (LS)- thousands ton	ODOT QA $n$ performed practice on studied lots, %	ODOT QA $n$ minimum requirements	Proposed $n$ , when using 0.76 $STDEV$ %	Proposed $n$ , when using 0.84 $STDEV$ %
LS – 7	2 (43%)	3 (43%)	4 (57%)	4 (57%)
LS – 9	3 (34%)	3 (34%)	5 (56%)	5 (56%)
LS – 10	3 (30%)	3 (30%)	5 (50%)	5 (50%)
LS – 12	3 (25%)	3 (25%)	6 (50%)	6 (50%)
LS – 15	3 (20%)	3 (20%)	6 (40%)	7 (47%)
LS – 17	4 (24%)	3 (18%)	6 (35%)	7 (41%)
LS – 20	3 (15%)	3 (15%)	7 (35%)	8 (41%)
LS – 22	4 (18%)	3 (14%)	7 (32%)	8 (36%)
LS – 23	3 (13%)	3 (13%)	7 (30%)	8 (35%)
LS – 24	7 (29%)	3 (13%)	7 (29%)	8 (33%)
LS – 33	7 (21%)	4 (12%)	7 (21%)	9 (27%)
LS – 42	3 (7%)	5 (12%)	8 (19%)	9 (21%)
LS – 43	8 (19%)	5 (12%)	8 (19%)	9 (21%)
LS – 52	7 (13%)	6 (12%)	8 (15%)	9 (17%)
LS – 58	7 (12%)	6 (10%)	8 (14%)	10 (17%)
LS – 84	7 (8%)	9 (11%)	9 (11%)	10 (12%)
LS – 114	17 (15%)	12 (11%)	9 (8%)	10 (8%)

sampling models. Using either model (i.e., model that was created based on 0.76 or 0.84) leads to the same results for lots with less than or equal to 12 000 tons of asphalt mixture, and more by one or two samples for lots with more than 12 000 tons when the greater *STDEV* (0.84) was used. Table 4 shows the proposed *n* based on two values of *STDEV* (0.76 and 0.84) and ODOT QA *n* actual practice on studied lots.

In 2015 and 2016, the ODOT contractors paved 130 projects with approximately 74% having less than 20 000 tons (20 sublots) of asphalt mixture. The current ODOT specification requires a minimum QA rate of 10% of QC *n*, or a minimum of three samples, whichever is larger, for verification and to ensure asphalt mixture quality. The current requirements (green curve in Figs. 2 and 3) may introduce risk on a large percentage of paving projects. This risk comes when only three QA samples are taken, which has many limitations from a statistical perspective by reducing the power of the data and increasing the margin of error. The current requirements are applied to all pay elements, including aggregate gradation, AC, and density without considering the differences in their data variability.

The proposed sampling models could assist ODOT QA department to determine the optimum *n* needed to validate the QC data and perform t-test and F-test correctly, calculate the percent within limits and pay factors, and ensure the asphalt mixture quality.

## Conclusions and recommendations

This paper has presented a case study to develop a model that can be used to determine the optimum QA *n*. The ODOT historical data were used to develop two sampling models. The models considered the lot sizes, the purpose of samples, variability in AC and density data, confidence level, and allowable error.

For the AC sampling model, ODOT, or any DOT developing such models, will need to decide on the appropriate *STDEV* value. For the studied lots, 0.2 and 0.3 *STDEV* were used for AC. The *STDEV* value of 0.2 was greater than 90% of the studied lot *STDEV* values. The *STDEV* value of 0.3 represents the worst-case scenario (i.e., the greater *STDEV* value found within the studied lots). The ODOT would need to increase the AC QA *n* for lots consisting of 65 000 tons or less and could decrease the *n* for lots with more than 65 000 tons of asphalt mixture when 0.2 *STDEV* value was used and compared to the current ODOT minimum requirements. On the other hand, using a 0.3 *STDEV* value means more QA *n* is required for lots consisting of less than 145 000 tons of asphalt mixture when compared to the current ODOT minimum requirements.

For density, ODOT previous project data indicated that *STDEV* values of 0.76 or 0.84 could be used. Using either value (0.76 or 0.84) leads to the same QA *n* for lots consisting of less than 12 000 tons, and one to two samples more (i.e., *n* when using 0.76 + 1 or 2 sample) when using 0.84 for projects consisting of more than 12 000 tons of asphalt mixture. According to the studied lots, more QA *n* was needed in lots consisting of less than 22 000 tons. Lots with 24 000 to 43 000 tons of asphalt mixture were within the acceptable rate of sampling based on the sampling model. The lot with 114 000 tons can be sampled at a lower rate with 10 samples instead of 17. Typically, ODOT tests aggregate gradation and AC from the same sample which is more practical and widely used among state DOTs. Further study needs to include the gradation and determine the *STDEV* values of each sieve size considered as a pay element in ODOT specification. This will help determine the optimum QA *n* for gradation and compare it to the AC *n*. A larger *n* can then be selected and used for both tests.

## Acknowledgments

The authors of this paper would like to thank the review panel of the Oregon Department of Transportation that generously provided feedback, funded the data collection and initial analysis, and offered technical assistance for this study.

### Funding

This work was supported by the Oregon Department of Transportation (ODOT).

### Disclosure statement

The authors do not have any competing financial, professional, or personal interests from other parties.

## REFERENCES

- Al-Khayat, H. (2018). A Case Study for Developing Specification Limits for Hot-Mix Asphalt and the Impact on Pay Factors. MSc Thesis, Texas A&M University. <https://oaktrust.library.tamu.edu/handle/1969.1/173949>
- ASTM. (2020). Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys (2020. D. 6433-20)
- De Winter, J. C. F. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation*, 18, Article 10.

- Elsefi, M. (2007). Evaluation of Hot-Mix Asphalt Sampling Techniques. Research Report FHWA-ICT-07-010, 2007. <http://hdl.handle.net/2142/45993>
- Gharaibeh, N. G., Garber, S. I., & Liu, L. (2010). Determining optimum sample size for percent-within-limits specifications. *Transportation Research Record*, 2151(1), 77–83. <https://doi.org/10.3141/2151-10>
- Newcomb, D., Gurganus, C., Al-Khayat, H., Sakhaeifar, M., & Epps, J. (2016). Review of Oregon Department of Transportation Asphalt Mix Specification, Phase II. <https://digital.osl.state.or.us/islandora/object/osl:81863>
- Oregon Department of Transportation. (2018). Oregon Standard Specifications for Construction. Oregon Department of Transportation. [https://www.oregon.gov/odot/Business/Specs/2018\\_STANDARD\\_SPECIFICATIONS.pdf](https://www.oregon.gov/odot/Business/Specs/2018_STANDARD_SPECIFICATIONS.pdf)
- Oregon Department of Transportation. (2015). Quality Assurance Program. Oregon Department of Transportation. [https://www.oregon.gov/ODOT/Construction/Doc\\_ManualofFieldTestProcedures/2015/06\\_QA\\_Program.pdf](https://www.oregon.gov/ODOT/Construction/Doc_ManualofFieldTestProcedures/2015/06_QA_Program.pdf)
- Schmitt, R. L., Russell, J. S., Hanna, A. S., Bahia, H. U., & Jung, G. A. (1998). Summary of Current Quality Control/Quality Assurance Practices for Hot-Mix Asphalt Construction. *Transportation Research Record*, 1632(1), 22–31. <https://doi.org/10.3141/1632-03>
- Willenbrock, J. H. (1976). A manual for statistical quality control of highway construction. Department of Transportation, Federal Highway Administration.