

QUALITY ASSURANCE SAMPLING PLANS IN US STOCKPILES FOR PERSONAL PROTECTIVE EQUIPMENT

Patrick L. Yorio, Dana R. Rottach, and Mitchell Dubaniewicz

Personal protective equipment (PPE) stockpiles in the United States were established to facilitate rapid deployment of medical assets to sites affected by public health emergencies. Large quantities of PPE were introduced into US stockpiles because of the need to protect healthcare and other professionals during these events. Because most stockpiled PPE was acquired during, or immediately following, large-scale public health events, such as pandemic influenza planning (2005-2008), SARS (2003), H1N1 (2009-10), and Ebola (2014-15), aging PPE poses a significant problem. PPE such as N95 filtering face piece respirators were not designed to be stored for long periods, and much of the currently stored PPE has exceeded its manufacturer-assigned shelf life. Given the significant investment in the procurement and storage of PPE, along with projections of consumption during public health emergencies, discarding large quantities of potentially viable PPE is not an attractive option. Although shelf-life extension programs exist for other stockpiled medical assets, no such option is currently available for stockpiled PPE. This article posits stockpile quality assurance sampling plans as a mechanism through which shelf-life extension programs for stockpiled PPE may be achieved. We discuss some of the nuances that should be considered when developing a plan tailored to stockpiles and provide basic decision tools that may be used in the context of a quality assurance program tailored to stockpiled PPE. We also explore basic information by comparing and contrasting different sample size options.

STOCKPILES REPRESENT an important component of the United States' public health infrastructure and are a strategic component of many federal, state, and local public health agencies.^{1,2} Throughout the United States, stockpiles were established to store large quantities of inventory to quickly supply healthcare and emergency response professionals responding to public health events. The placement of stockpiles in strategic locations facilitates the rapid deployment of medical assets needed in countermeasure efforts. US stockpiles were established through federal funding and initiatives associated with the Public Health

Security and Bioterrorism Preparedness and Response Act of 2002 (P.L. 107-188), which followed the terror attacks of 2001.¹ This initiative enhanced appropriations for public health throughout the United States and facilitated the creation of stockpiles at all levels of government and in nongovernment healthcare settings.¹

Since their widespread establishment in the early 2000s, the role of US stockpiles in the nation's public health strategy and infrastructure has become institutionalized. Over time, in many instances, stored inventory has been expanded to include large quantities of personal protective

Patrick L. Yorio, PhD, is a Health Statistician, and Dana R. Rottach, PhD, is a Physical Scientist; both at the National Personal Protective Technology Laboratory, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Pittsburgh, PA. Mitchell Dubaniewicz is a student researcher, Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, and was on assignment with CDC/NIOSH/NPPTL.

equipment (PPE). PPE—including respirators, gloves, and surgical gowns—was introduced into US stockpiles primarily to protect healthcare professionals and first responders on the frontlines of treatment and containment.³ Large volumes of PPE were introduced due to utilization rates and shortages experienced by hospitals and healthcare providers during actual and simulated public health emergencies and pandemics.⁴⁻¹³

Much of the currently stockpiled PPE was acquired in response to high-consequence public health events such as severe acute respiratory syndrome (SARS) in 2003, 2005-2008 pandemic influenza planning, the 2009-10 H1N1 influenza pandemic, and the 2014-15 Ebola outbreak.^{2,4} Thus, due to the significant financial investments required for procurement and storage, which are often viewed as prohibitive, stockpiled PPE have, in some cases, been stored beyond their manufacturer-recommended shelf life.

Although shelf-life extension programs exist for other stockpiled medical assets, no such manufacturer-approved or government-sanctioned program is currently available for stockpiled PPE. Formal programs are in place to extend the manufacturer-assigned shelf life for stockpiled pharmaceuticals, since it has been empirically demonstrated that their actual shelf life can be much longer if they are stored properly.^{14,15} The motivation to develop and implement the pharmaceutical Shelf-Life Extension Program (SLEP) emerged during the mid-1980s when the military realized that extending the shelf life of medicines could help solve the supply and demand challenges they experienced.¹⁴ Since its adoption, SLEP has proved to be effective at reducing the logistical burden and replacement costs of valuable medical assets held in US stockpiles.^{14,15} An important contributor to the success of this program, however, was verification of product quality prior to shelf-life extensions being granted. In the context of stockpiled PPE, the paucity of research and initiatives related to quality assurance may be a hindrance in the development of a PPE-tailored SLEP.

In this article, we posit stockpile quality assurance sampling plans as a mechanism through which shelf-life extension programs for stockpiled PPE may be realized. In doing so, we discuss some of the nuances that a plan tailored to stockpiles should consider and provide basic decision tools that may be used in the context of a quality assurance program tailored to stockpiled PPE. We also compare and contrast different sample size options.

QUALITY ASSURANCE OF STOCKPILED PPE

An effective effort to develop a quality assurance initiative suitable to stockpiled PPE should consider the context. Stockpiles typically store numerous manufacturers' models of similar PPE. For each model, a manufacturer assigns a single batch or lot number under the same set of manufacturing parameters during the same period. There-

fore, PPE units with the same lot number share a consistent expiry date and may be expected to have relatively uniform quality. Stockpiles are likely to store numerous lots of various sizes for each model.

An adequate evaluation of PPE performance typically requires destructive testing in order to assess its protective properties. While testing samples composed of a large number of units will result in higher confidence in population estimates, depleting the stock of critical assets through excessive testing is not ideal. Further, given the associated expense of testing stockpiled PPE, a performance evaluation program should balance costs of testing and respirator attrition with a level of accuracy that is sufficient to ensure stockpiled PPE continue to meet regulatory guidelines.

Lot Quality Assurance Sampling

Lot quality assurance sampling (LQAS), a quality assurance process that may be most suitable for stockpiled PPE, originated in the manufacturing industry for quality control purposes and has been adopted in numerous quality assurance contexts since its inception.^{16,17} LQAS uses a sample of items to estimate the quality of the lot instead of testing each of the items. LQAS is designed to facilitate decisions to accept or reject the entire lot based on the results of single or multiple samples, an approach that keeps sampling costs to a minimum.¹⁸ This approach is consistent with a stratified random sampling technique, but the samples are purposefully smaller than would be necessary to provide narrow confidence intervals for each stratum or lot.^{18,19} In contrast, LQAS enables lot quality decisions to be made based on probabilities.^{18,19} Given that stockpiles store a population of PPE items that can be delineated by numerous lots of distinct manufacturer models, LQAS's approach of randomly sampling a small numbers of units in each stratum fits the context nicely. Figure 1 depicts a hypothetical stockpile stratified random sampling design for N95 filtering face piece respirators using a quality assurance program grounded in the LQAS approach. This technique divides the entire population of N95 respirators into nonoverlapping subpopulations of lots that are expected to share quality commonalities. The manufacturer-assigned lot number represents the homogenous level quality needed to characterize each stratum in a stockpile's LQAS plan.

Sampling done at the lot level implies that the test results will generalize to the lot; however, LQAS is also designed to allow information at each level to be combined and inferences to be made at higher levels. For example, the estimates of quality for all units that share a common manufacturer model can be made by combining the results of samples corresponding to particular lots. Accurate estimates at the higher level can be obtained by weighting the results obtained at lower levels according to their relative percentage in the population.

In a stratified random sampling approach context, the random selection of units from subpopulations ensures that the estimates of quality derived are unbiased. This means that, although the estimate from a single sample can be very different from the true quality level, the average of the sample proportions from repeated sampling should equal the true quality level that exists in the subpopulation. When units are not randomly selected from a population, bias in the estimated population parameter is likely. A potential challenge in using a stratified random sampling approach is the need to fully account for the inventory of PPE units in each stockpile. A true random process would require that each PPE item has an equal probability of being selected for sampling.

A potentially useful adaptation of the LQAS method is the adoption of “double sampling.” This approach divides the chosen sample size into 2 equal, smaller samples: n_1 and n_2 . A random sample of size n_1 is selected during the first stage. If the results of this sample are not conclusive, a second random sample is obtained (n_2), and conclusions are based on the results of the combined sample ($n_1 + n_2$).^{18,19}

Quality Evidence for Stockpiled PPE

In the context of a PPE-specific LQAS, focused information should be provided as evidence relating to its quality. Numerous consensus standards designed to evaluate the initial quality of the various types of PPE may be referenced as a starting point. Table 1 shows the tests that may be considered when assessing N95 filtering face piece respirators and gowns for their protective qualities.

For N95 respirators, the National Institute for Occupational Safety and Health’s (NIOSH) standard testing procedures (STPs) may be used to assess and compare the quality of stockpiled respirators to approved levels of initial quality. These standard testing procedures have an assigned level of quality for breathing resistance and filtration efficiency that can be used to assign a dichotomous quality attribute to each item. In its attribute form, each unit is categorized as either being defective or not in relation to the criteria in each standard. Any of the tests shown in Table 1 can result in this assigned attribute. The LQAS method requires that the test results corresponding to each attribute be tabulated in the form of the proportion of defective units in each sample. This sample proportion is then used to make decisions regarding the overall quality of the lot.

Decision Rationales in LQAS

The decision rationales center on the number of defective items that need to be found in a sample before the lot is deemed unacceptable; the rationales are grounded in the risks that stockpile managers are willing to accept.¹⁸ Implied in this framework is the premise that stockpile managers should first have some sense of the minimum

proportion of defective units in a lot they are not willing to exceed. It can be assumed that the desired quality level in each stockpile lot is 0% failing units. Given that there is some probability of defect in new units, however, more realistic expectations that incorporate some level of defective units should be considered.

Once the desired level of quality is chosen, the acceptable level of risk of rejecting a lot that meets the desired quality level must be selected. Because rules governing attribute quality data are based on the binomial distribution, the probability of obtaining an exact number of defective units in any sample size can be calculated, assuming that the true number of defective units in a lot was at a desired level.* The rejection risk, often identified as the type I error rate (α), is the probability of rejecting a lot that actually meets the quality level deemed appropriate by stockpile management. This level of risk is commonly set at 5%.²⁰ In other words, a satisfactory lot would be incorrectly rejected less than 5% of the time. Figure 2 depicts the expected probabilities and the cumulative probability for obtaining an exact number of defective units in a sample of 32 taken from a lot in which 15% of the PPE units were defective.

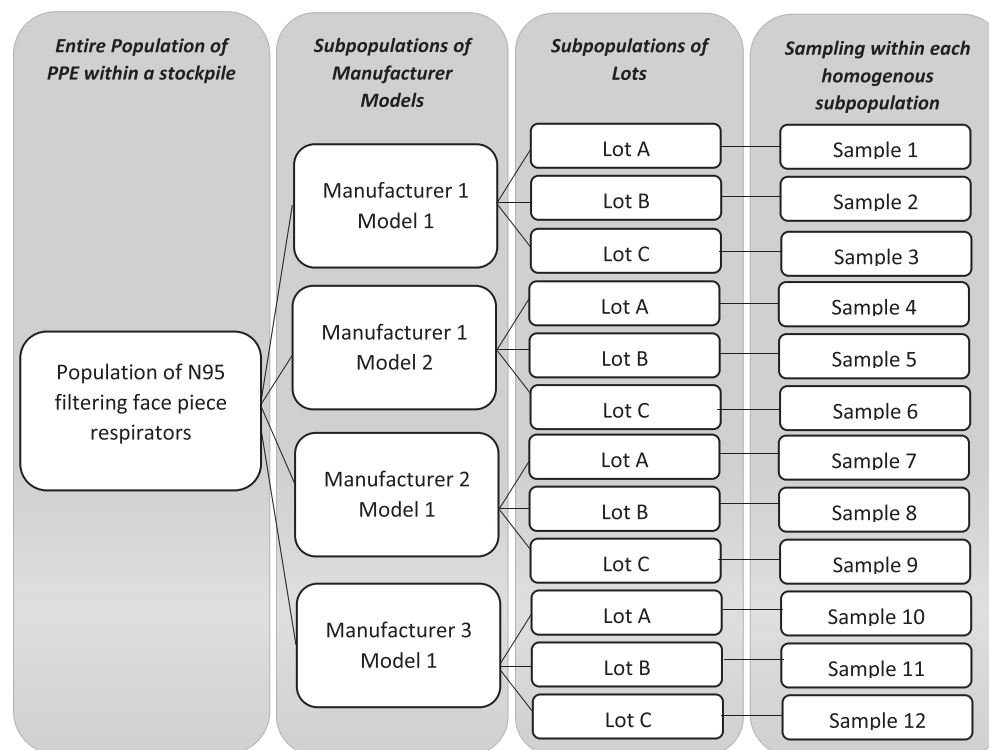
After some consideration is given to the desired level of quality and the rejection risk, units from the lot can be sampled and tested. The number of units that failed in the sample can then be used to determine the likelihood that the sample was taken from a lot that exceeded the proportion of defective units deemed acceptable. As depicted in Figure 2, if a rejection risk of 5% is selected, it may be concluded that a sample of 32 that resulted in 9 or more defective units was likely sampled from a lot that exceeded 15% defective units. If stockpile management determined 15% defective units in any given lot to be the maximum number they are willing to accept and set the rejection risk (α) to 5%, a finding of 9 or more defects in the sample indicates that the lot may be rejected or that a second round of testing may be helpful to obtain additional evidence.

In the LQAS method, the probabilities for finding a different number of defective units can be computed for any sample size and maximum number of defects that stockpile managers deem acceptable. Figure 2 shows the expected probabilities for only a single scenario in which a sample of 32 PPE items was taken from a lot with 15% defective units. In Table 2, probabilities are provided for obtaining a specific number of failures in sample sizes of 10, 12, 18, 24, and 32, assuming the true number of defects in the lot were 15% and 5%.

Although not exhaustive of all possibilities, the table is meant to represent a simple tool that can be used to determine when the proportion of defects in a lot is likely to have exceeded the maximum number they are

* $P(x|\rho) = \left(\frac{n!}{x!(n-x)!} \right) (1-\rho)^{(n-x)} \rho^x$, where $P(x|\rho)$ = the probability of observing a given number of PPE items fail (x) in a sample of size (n) given the true lot fail rate (ρ).

Figure 1. PPE sampling depiction for stockpiles based on the lot quality assurance sampling (LQAS) technique



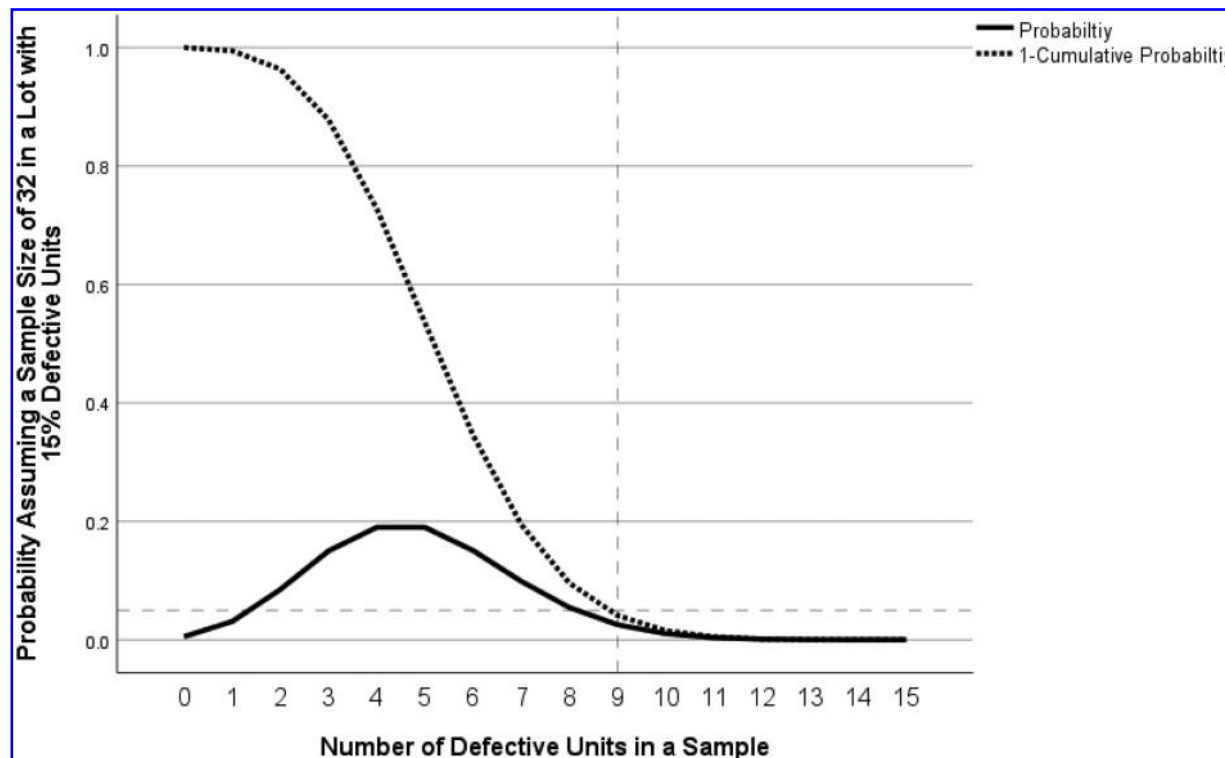
willing to accept. To illustrate, assume a stockpile management team deemed that a sample of 18 PPE items could be used to determine if the number of defects in a lot was likely greater or less than 5%. If 18 PPE items were randomly sampled from the lot and tested, and 3 of the 18 failed the performance test, the best guess estimate of the true lot fail rate is 3/18 or ~16.7%. However, we know that, based on sampling variability, 3 defects out of a sample of 18 PPE items can occur for a wide range of true lot quality. By consulting Table 2, it can be seen that samples of 18 PPE items from a lot in which 5% of the units are defective will result in 0, 1, or 2 failures a little over 94% of the time (ie, the sum of the probabilities of

observing exactly 0, 1, and 2 failures). The likelihood of observing 3 failures out of 18 PPE items is relatively small—only 4.7% of the time, and a little over 5% of the total samples will result in 3 or more failures. Thus, while there is a possibility that this sample of 18 PPE units may have been taken from a lot with 5% defective units, it is reasonable to conclude that the number of defects in the lot likely exceeds 5%. The rows for samples of 18 that resulted in 3 or more failures are shaded in Table 2, given that the sum of the expected probabilities for observing 3 or more failures out of a sample of 18 items from a lot with 5% defective units is around 5%. In Table 2, 5% cutoffs are highlighted for reference.

Table 1. Examples of quality evidence that can be provided for PPE commonly stored in US stockpiles

<i>N95 Filtering Face Piece Respirators</i>	<i>Protective Gowns</i>
<ul style="list-style-type: none"> • Visually inspect: damage, degradation, molding, etc. • NIOSH Standard Testing Procedures (STP) 3 & 7: Inhalation and Exhalation Resistance • NIOSH STP 59: Particulate Filter Efficiency for N95 • NIOSH STP 53: Liquid Particulate Filter Efficiency for P95 • ASTM D412: Rubber/Elastomer Tensile Strength • OSHA 1910.134 Fit Testing 	<ul style="list-style-type: none"> • Visually inspect: damage, degradation, molding, etc. • AATCC 42: Water Resistance: Impact Penetration—Level 3 and front of Level 4 gowns • AATCC 127: Water Resistance: Hydrostatic Pressure Test—Level 3 gowns • ASTM F1671: Penetration by Bloodborne Pathogens Using Phi-X174 Bacteriophage Penetration as a Test System—Level 4 gowns • ASTM F1929: Standard Test Method for Detecting Seal Leaks in Porous Medical Packaging by Dye Penetration

Figure 2. Probabilities and the cumulative probability of obtaining an exact number of defective units in a sample of 32 from a lot with 15% defective units



Note: The solid line in Figure 2 represents the probabilities (on the Y axis) that a sample of 32 will result in a specific number of defective units (on the X axis) when taken from a lot in which 15% of the units are defective. The probability of obtaining 4 and 5 defective units in the sample is greatest—both being approximately 19%. The probability of finding 6 or more defective units in the sample decreases steadily. The dotted line is the sum of the probabilities from right to left in the figure. This line shows that there is nearly a 0% chance of obtaining a sample of 32 with 11 or more defective units if the true number of defects in the lot is 15%. The intersection between the horizontal ($\alpha = 0.05$) and vertical dotted line (9 defective units in the sample) indicates that obtaining 9 or more defective units in a sample of 32 from a lot with 15% will happen less than 5% of the time.

Sample Size Considerations

There are tradeoffs that need to be managed between the resources (time and money) available to administer the quality assurance program and the precision desired to draw quality conclusions. The goal of the LQAS method is to determine which lots have acceptable and unacceptable levels of quality. This allows for less emphasis to be placed on establishing population parameters within a desired level of precision—which requires large samples to derive narrow confidence intervals around the estimated population proportion of defective units.¹⁸ This nuance allows for relatively smaller sample sizes to be used and, in-turn, reduces the waste experienced through destructive testing and the cost associated with testing stockpiled PPE.

Since a random sampling approach provides unbiased estimates of lot quality, eventually the average quality level revealed through repeated samples will equal the true quality level of the lot. However, the results of in-

dividual samples may be quite different from the true quality level in the lot. In the context of stockpile quality assurance sampling programs—with a large number of lots, destructive testing, and limited resources—decisions regarding lot quality will likely be made based on a small number of samples. One area of interest, therefore, concerns the accuracy of different sample sizes in estimating the true quality level when the number of repeated samples is restricted. Although LQAS is a well-developed quality assurance technique, little empirical attention has been devoted to comparing and contrasting different sample sizes.

In order to compare and contrast different sample sizes considering an LQAS specific to stockpiled PPE, we ran a series of simple computer-based experiments and examined the sampling distribution of the proportion of defective units for different sample sizes taken from lots with known quality levels. Three different lots consisting of 100,000 PPE items were created. Each lot was simulated to have a different number of defective units—1%, 5%, and 15%.

Table 2. Probability that a sample of n PPE units will result in a specific number of defective units when the true number of defective units in the lot is 15% and 5%

Sample Size	Number of Failures	Probability (in %) for a Lot with 15% Defective Units	Probability (in %) for a Lot with 5% Defective Units
10	0	19.7	59.9
	1	34.7	31.5
	2	27.6	7.5
	3	13.0	1.0
	4	4.0	0.1
	5	0.8	—
	6	0.1	—
	7	—	—
12	0	14.2	54.0
	1	30.1	34.1
	2	29.2	9.9
	3	17.2	1.7
	4	6.8	0.2
	5	1.9	—
	6	0.4	—
	7	0.1	—
18	0	5.4	39.7
	1	17.0	37.6
	2	25.6	16.8
	3	24.1	4.7
	4	15.9	0.9
	5	7.9	0.1
	6	3.0	—
	7	0.9	—
	8	0.2	—
	9	0.0	—
24	0	2.0	29.2
	1	8.6	36.9
	2	17.4	22.3
	3	22.5	8.6
	4	20.9	2.4
	5	14.7	0.5
	6	8.2	0.1
	7	3.7	—
	8	1.4	—
	9	0.4	—
	10	0.1	—
32	0	0.6	19.4
	1	3.1	32.6

(continued)

Table 2. (Continued)

Sample Size	Number of Failures	Probability (in %) for a Lot with 15% Defective Units	Probability (in %) for a Lot with 5% Defective Units
	2	8.5	26.6
	3	15.0	14.0
	4	19.0	5.3
	5	19.0	1.6
	6	15.1	0.4
	7	9.9	0.1
	8	5.5	—
	9	2.6	—
	10	1.0	—
	11	0.4	—
	12	0.1	—
	13	—	—

For example, in the 5% defective lot, 5% or 5,000 of the numbers were set to be ones—representing defective PPE items—and the remaining 95,000 (95%) were set to be zeros—nondefective PPE items. Once the 3 lots were generated, random samples of 5, 10, 12, 18, 24, 32, 40, 50, and 100 units were selected from them; 75 distinct samples of each sample size with replacement were taken from each simulated lot. Although 75 repeated samples from the same lot far exceeds a realistic number of repeated samples, they were used to generate estimates of sample-to-sample variability that could be statistically compared and contrasted among the different sample sizes.

We initially examined quality estimate for each sample size, noting the number of repeated samples required to derive an accurate estimate of the true defective rate in the lot. The estimated number of defective units in the lot derived from samples 1 through 10 is presented in Table 3. The table shows that, after a single sample, the estimated proportion of defects in each lot can be quite different from the actual true number for most of the sample sizes. In Table 3, the column corresponding to 2 repeated samples shows the average of the first and second fail rates. Further, the column corresponding to 3 repeated samples shows the average of the first, second, and third samples, and so on. These columns reveal that even after numerous repeated samples, the estimated number of defective units in the lot can be much different from the true proportion of defective units in the lot. This suggests that relying on a small number of repeated samples to estimate lot quality levels can be problematic with relatively smaller sample sizes.

These observations support the use of probabilities to determine when lot quality is likely to exceed acceptable levels included in the LQAS approach. However, the differences in the variability of quality estimates for each sample size across

Table 3. Estimated lot defective proportion—the average proportion defective—after repeated samples when the true defective proportion is 1%, 5%, and 15%

		<i>Number of Repeated Samples</i>									
	<i>Sample Size</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
1% True Lot Defective Proportion	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
	12	0%	4%	3%	2%	2%	3%	1%	1%	1%	1%
	18	0%	0%	0%	1%	2%	3%	2%	1%	1%	1%
	24	4%	4%	3%	2%	3%	2%	2%	2%	1%	2%
	32	3%	3%	2%	2%	1%	1%	1%	1%	1%	1%
	40	0%	0%	1%	1%	1%	1%	1%	1%	1%	1%
	50	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
	100	1%	3%	3%	2%	2%	2%	2%	2%	1%	1%
5% True Lot Defective Proportion	5	0%	10%	13%	10%	8%	7%	6%	5%	4%	4%
	10	0%	0%	0%	3%	6%	7%	7%	6%	6%	6%
	12	8%	8%	11%	10%	8%	7%	6%	6%	6%	7%
	18	0%	6%	4%	4%	4%	4%	6%	6%	6%	6%
	24	0%	2%	1%	2%	4%	6%	7%	6%	6%	6%
	32	6%	5%	5%	5%	4%	5%	4%	4%	3%	3%
	40	3%	4%	2%	3%	3%	3%	3%	3%	4%	3%
	50	0%	5%	3%	4%	5%	5%	5%	5%	5%	5%
	100	3%	4%	4%	5%	5%	5%	5%	5%	5%	5%
15% True Lot Defective Proportion	5	20%	20%	27%	20%	16%	13%	11%	10%	9%	10%
	10	50%	35%	27%	25%	22%	18%	17%	18%	17%	17%
	12	25%	21%	19%	17%	13%	11%	16%	10%	12%	13%
	18	11%	14%	17%	15%	17%	15%	12%	13%	12%	13%
	24	38%	17%	12%	19%	18%	18%	13%	18%	18%	18%
	32	6%	13%	13%	15%	16%	17%	16%	15%	14%	15%
	40	10%	12%	12%	11%	13%	12%	12%	12%	14%	14%
	50	14%	13%	13%	12%	12%	14%	13%	13%	12%	12%
	100	15%	16%	15%	14%	14%	15%	15%	15%	14%	14%

the number of repeated samples shown in Table 3 suggests that some sample sizes may be better suited for use in an LQAS approach specific to stockpiled PPE.

Corresponding to each sample size, Table 4 highlights the sample-to-sample variability that was observed across the entire set of 75 repeated samples. The range depicts how extreme the observed fail rates can get depending on sample size. For each of the simulated lots, a sample size of 5 resulted in the greatest range in the percent of defective units observed. In general, the range decreased as the sample size increased. Similarly, the sample-to-sample dispersion was greatest where the sample size was equal to 5 and generally decreased as the size of the random sample increased. The average percent of defective units across the 75 samples and their standard deviations are shown in Figure 3. Consistent with the results presented in Table 4, Figure 3 shows that for each sample size, the mean percentage of defective items observed across the 75 samples was fairly accurate, while the standard deviation generally decreased as the sample size increased.

Next, the sample-to-sample variances were statistically compared and contrasted. This process allows us to group sample sizes that resulted in similar and different variances (ie, those not statistically different and those statistically

different, respectively). The results of these pairwise comparisons are also reported in Table 4. Considering the results in each of the 3 simulated lots collectively, 3 distinct groups of sample sizes could be delineated based on similar characteristics in their variability: group 1 consisted of sample sizes of 10 and 12 PPE units; group 2 consisted of 18, 24, and 32 units; and group 3 consisted of 40 and 50 units. The magnitude of the benefits of group 3 were dependent on the true number of defective units in the lot and were maximized as this number increased. The sample sizes of 5 and 100 displayed outcome characteristics that differed from the overall groupings and resulted in the greatest and least variability, respectively.

CONCLUSIONS

Throughout this article, we positioned quality assurance as a potential mechanism to manage aging PPE currently stored in US stockpiles. We demonstrated that relying on a small number of repeated samples to estimate lot quality levels can be problematic with relatively smaller sample sizes. In response, we posited LQAS as a quality assurance framework that has potential applicability in this context.

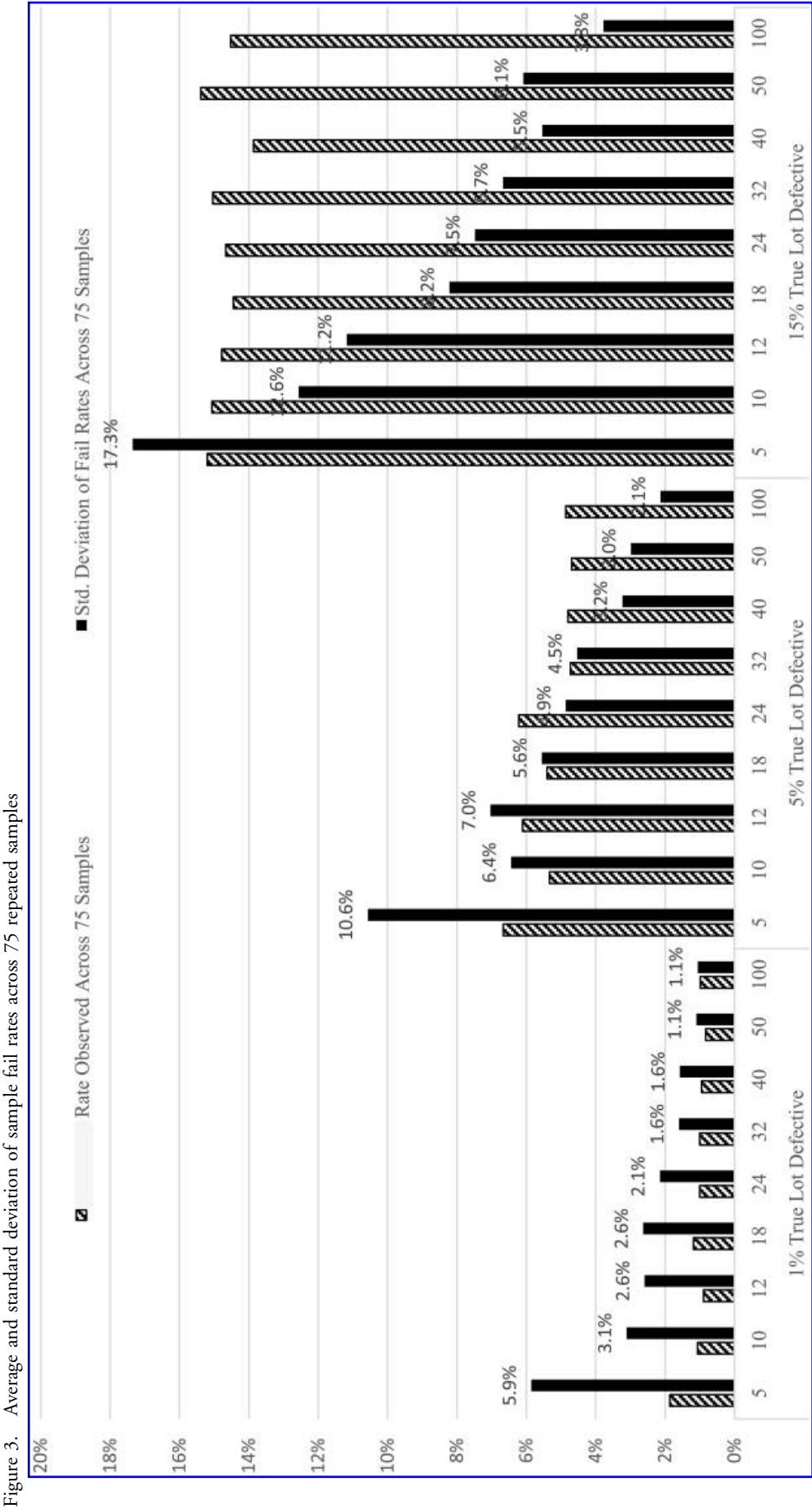
Table 4. Descriptive statistics of sample fail rates and mean/median based pairwise comparisons of the variances across 75 trials of 5, 10, 12, 18, 24, 32, 40, 50, and 100 samples sizes

True Population Fail Rate	Number of PPE Items Sampled	Mean (Median) Fail Rate Across 75 Trials	Range in Mean Fail Rate in Samples		Std. Deviation of Fail Rates Across 75 Trials	P value for difference in variance in observed fail rates across 75 trials. Results of both Brown Forsythe (median based) variance tests reported for the 1% and 5% true fail rate simulated lots. Results for both the Levene (mean based) and Brown Forsythe tests are reported for 15% true fail rate lot.										
			Min	Max		5 PPE items sampled	10 PPE items sampled	12 PPE items sampled	18 PPE items sampled	24 PPE items sampled	32 PPE items sampled	40 PPE items sampled	50 PPE items sampled			
1%	5	1.87%(0.0%)	0%	20.0%	5.86%	—	—	—	—	—	—	—	—	—	—	
	10	1.07%(0.0%)	0%	10.0%	3.11%	0.30	—	—	—	—	—	—	—	—	—	
	12	0.89%(0.0%)	0%	8.30%	2.59%	0.19	0.70	—	—	—	—	—	—	—	—	
	18	1.19%(0.0%)	0%	11.1%	2.63%	0.36	0.80	0.48	—	—	—	—	—	—	—	
	24	1.00%(0.0%)	0%	12.5%	2.15%	0.23	0.88	0.78	0.64	—	—	—	—	—	—	
	32	0.96%(0.0%)	0%	6.0%	1.57%	0.20	0.79	0.84	0.53	0.90	—	—	—	—	—	
	40	0.95%(0.0%)	0%	5.0%	1.58%	0.19	0.77	0.87	0.50	0.86	0.96	—	—	—	—	
	50	0.83%(0.0%)	0%	4.0%	1.10%	0.13	0.53	0.85	0.28	0.54	0.55	0.59	—	—	—	
	100	0.99%(1.0%)	0%	5.0%	1.06%	0.12	0.45	0.74	0.21	0.42	0.39	0.43	0.79	—	—	
	5	6.67%(0.0%)	0%	40.0%	10.57%	—	—	—	—	—	—	—	—	—	—	—
5%	10	5.33%(0.0%)	0%	20.0%	6.44%	0.35	—	—	—	—	—	—	—	—	—	—
	12	6.11%(8.3%)	0%	25.0%	7.03%	0.61	0.46	—	—	—	—	—	—	—	—	—
	18	5.41%(5.6%)	0%	22.0%	5.55%	0.05	0.17	0.01	—	—	—	—	—	—	—	—
	24	6.22%(4.2%)	0%	16.7%	4.86%	0.04	0.10	0.00	0.73	—	—	—	—	—	—	—
	32	4.73%(3.0%)	0%	19.0%	4.54%	0.01	0.01	0.00	0.07	0.12	—	—	—	—	—	—
	40	4.80%(5.0%)	0%	15.0%	3.23%	0.01	0.01	0.00	0.01	0.01	0.46	—	—	—	—	—
	50	4.69%(4.0%)	0%	12.0%	2.99%	0.01	0.00	0.00	0.00	0.01	0.21	0.41	—	—	—	—
	100	4.87%(5.0%)	1%	11.0%	2.13%	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01

(continued)

Table 4. (Continued)

True Population Fail Rate	Number of PPE Items Sampled	Mean (Median) Fail Rate Across 75 Trials	Range in Mean Fail Rate in Samples		Std. Deviation of Fail Rates Across 75 Trials	P value for difference in variance in observed fail rates across 75 trials. Results of both Brown Forsythe (median based) variance tests reported for the 1% and 5% true fail rate simulated lots. Results for both the Levene (mean based) and Brown Forsythe tests are reported for 15% true fail rate lot.								
			Min	Max		5 PPE items sampled	10 PPE items sampled	12 PPE items sampled	18 PPE items sampled	24 PPE items sampled	32 PPE items sampled	40 PPE items sampled	50 PPE items sampled	
15%	5	15.20%(20%)	0%	60.0%	17.35%	—	—	—	—	—	—	—	—	—
	10	15.07%(10%)	0%	50.0%	12.56%	0.00, 0.00	—	—	—	—	—	—	—	—
	12	14.78%(17%)	0%	41.7%	11.18%	0.00, 0.00	0.19, 0.42	—	—	—	—	—	—	—
	18	14.44%(11%)	0%	33.3%	8.22%	0.00, 0.00	0.00, 0.01	0.01, 0.02	—	—	—	—	—	—
	24	14.67%(13%)	0%	37.5%	7.48%	0.00, 0.00	0.00, 0.00	0.00, 0.01	0.45, 0.70	—	—	—	—	—
	32	15.04%(16%)	3%	31.0%	6.68%	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.14, 0.36	0.50, 0.56	—	—	—	—
	40	13.87%(15%)	3%	30.0%	5.54%	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.01, 0.01	0.01, 0.02	0.02, 0.03	—	—	—
	50	15.34%(14%)	2%	30.0%	6.09%	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.01, 0.05	0.06, 0.10	0.05, 0.12	0.38, 0.42	—	—
	100	14.52%(14%)	3%	24.0%	3.78%	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.01, 0.02	0.01, 0.01	



The LQAS approach advocates the use of probabilities to determine when lot quality is likely to have exceeded an acceptable, a priori, agreed upon level.

Grounded in the LQAS approach, Table 2 illustrates how the results of small samples can be assessed in relation to expected probabilities if the true lot quality was assumed. The tables represent a set of tools that can be used as a basis for decision making in an LQAS specific to stockpiled PPE. In Table 2 expectations for sample sizes of 10 to 32 and true lot fail rates of 5% and 15% were included. The approach used to develop Table 2, however, can be used to derive theoretical expectations for any sample size and any maximum number of defective units in a lot that is deemed acceptable.

Although LQAS is a well-developed quality assurance technique, little empirical attention has been devoted to comparing and contrasting different sample sizes. Thus, we also examined the question of sample size through a set of simple experiments. We found that sample sizes could be grouped together in terms of the variability with which they estimated the true number of defective units in a lot. Collectively, the findings related to sample size and the tools provided can be used by stockpile managers to begin the process of developing an LQAS specific to stockpiled PPE.

In order to maximize the prospect that the quality evidence accumulated by stockpiles can be used to inform a SLEP specific to stockpiled PPE, most of the parameters in the LQAS (eg, levels of risk assumed, sample size, schedules of repeated sampling, and the maximum number of defects in a lot deemed acceptable) should be subject to a formal consensus process overseen by representatives from PPE manufacturers, state and local departments of health, and privately held stockpiles. Manufacturer involvement in the design of a PPE-specific LQAS for stockpiles may maximize the prospect of a formal shelf-life extension program.

In this formal consensus setting process, contingencies also need to be established that define steps to be taken in lieu of the quality evidence found. If quality evidence is found that suggests that the lot tested is acceptable, decision rationales should be in place to inform possible extension for products beyond the manufacturer's labeled or recommend use date. PPE cannot provide expected protections indefinitely, thus decisions to hold product should reflect extensions based on quality evidence coupled with any expected degradation. As such, if quality evidence warrants a shelf-life extension for particular lots of PPE, effort should be made to use the older products and replenish them with new stock. Consistent with the currently implemented SLEP for pharmaceuticals, PPE products with granted extensions to the shelf life that are released to the public may require relabeling and/or communication strategies to ensure that the end users understand that the product may be used beyond its

labeled expiration date. Conversely, if quality evidence is found that suggests that the lot tested is or may be unacceptable, contingencies should include the option of discarding the lot or for a repeat testing to gather additional evidence.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dana Grau, Jeff Ballard, Joseph Reppucci, David Balbi, Ron Shaffer, Megan Casey, Lew Radonovich, and Mark Nicas for critical reviews of earlier versions of this manuscript. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention. In addition, citations to websites external to NIOSH do not constitute NIOSH endorsement of the sponsoring organizations or their programs or products. Furthermore, NIOSH is not responsible for the content of these websites. All web addresses referenced in this document were accessible as of the publication date.

REFERENCES

1. Lister SA. *An Overview of the US Public Health System in the Context of Emergency Preparedness*. Washington, DC: Congressional Research Service; 2005. <https://fas.org/sgp/crs/homsec/RL31719.pdf>. Accessed March 20, 2019.
2. Esbitt D. The Strategic National Stockpile: roles and responsibilities of health care professionals for receiving the stockpile assets. *Disaster Manag Response* 2003;1(3):68-70.
3. Institute of Medicine; Board on Health Sciences Policy; Committee on Personal Protective Equipment for Health-care Personnel to Prevent Transmission of Pandemic Influenza and Other Viral Respiratory Infections: Current Research Issues; Larson EL, Liverman CT, eds. *Preventing Transmission of Pandemic Influenza and Other Viral Respiratory Diseases: Personal Protective Equipment for Health-care Personnel: Update 2010*. Washington, DC: National Academies Press; 2011.
4. Pyrek KM. PPE utilization in a pandemic: more research needed to fuel preparedness. *Infection Control Today* March 2014;1-26.
5. Swaminathan A, Martin R, Gamon S, et al. Personal protective equipment and antiviral drug use during hospitalization for suspected avian or pandemic influenza. *Emerg Infect Dis* 2007;13(10):1541-1547.
6. Hashikura M, Kizu J. Stockpile of personal protective equipment in hospital settings: preparedness for influenza pandemics. *Am J Infect Control* 2009;37(9):703-707.
7. Hui Z, Jian-Shi H, Xiong H, Peng L, Da-Long Q. An analysis of the current status of hospital emergency preparedness for infectious disease outbreaks in Beijing, China. *Am J Infect Control* 2007;35(1):62-67.

8. Kaji AH, Lewis RJ. Hospital disaster preparedness in Los Angeles County. *Acad Emerg Med* 2006;13(11):1198-1203.
9. Carias C, Rainisch G, Shankar M, et al. Potential demand for respirators and surgical masks during a hypothetical influenza pandemic in the United States. *Clin Infect Dis* 2015; 60(Suppl 1):S42-S51.
10. Abramovich MN, Hershey JC, Callies B, Adalja AA, Tosh PK, Toner ES. Hospital influenza pandemic stockpiling needs: a computer simulation. *Am J Infect Control* 2017; 45(3):272-277.
11. Murray M, Grant J, Bryce E, Chilton P, Forrester L. Facial protective equipment, personnel, and pandemics: impact of the pandemic (H1N1) 2009 virus on personnel and use of facial protective equipment. *Infect Control Hosp Epidemiol* 2010;31(10):1011-1016.
12. Radonovich LJ, Magalian PD, Hollingsworth MK, Baracco G. Stockpiling supplies for the next influenza pandemic. *Emerg Infect Dis* 2009;15(6):e1.
13. World Health Organization. *Avian Influenza, Including Influenza A (H5N1), in Humans: WHO Interim Infection Control Guideline for Health Care Facilities*. Amended April 24, 2006. <http://pandemicflu.utah.gov/docs/HCIInfectionControlAlinHumansWHOInterimGuidelinesfor.pdf>. Accessed March 20, 2019.
14. Courtney B, Easton J, Inglesby TV, SooHoo C. Maximizing state and local medical countermeasure stockpile investments through the Shelf-Life Extension Program. *Biosecur Bioterror* 2009;7(1):101-107.
15. Khan SR, Kona R, Faustino PJ, et al. United States Food and Drug Administration and Department of Defense Shelf-Life Extension Program of Pharmaceutical Products: progress and promise. *J Pharm Sci* 2014;103(5):1331-1336.
16. Dodge HF, Roming HG. Sampling inspection tables (No. 311.21 D63 1959).
17. Robertson SE, Valadez JJ. Global review of health care surveys using lot quality assurance sampling (LQAS), 1984-2004. *Soc Sci Med* 2006;63(6):1648-1660.
18. Hoshaw-Woodard S. *Description and Comparison of the Methods of Cluster Sampling and Lot Quality Assurance Sampling to Assess Immunization Coverage*. Geneva: World Health Organization; 2001. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.409.3852&rep=rep1&type=pdf>. Accessed March 20, 2019.
19. Lemeshow S, Hosmer DW, Klar J, Lwanga SK; World Health Organization. *Adequacy of Sample Size in Health Studies*. New York: John Wiley & Sons; 1990. https://apps.who.int/iris/bitstream/handle/10665/41607/0471925179_eng.pdf?sequence=1&isAllowed=y. Accessed March 20, 2019.
20. Cowles M, Davis C. On the origins of the .05 level of statistical significance. *Am Psychol* 1982;37(5):553.

*Manuscript received December 14, 2018;
revision returned January 15, 2019;
accepted for publication January 22, 2019.*

Address correspondence to:
Patrick L. Yorio, PhD
National Institute for Occupational Safety and Health
Bruceton Research Center
626 Cochran's Mill Road
Pittsburgh, PA, 15236
E-mail: pyorio@cdc.gov