Pathology (December 2019) 51(7), pp. 711–717

CHEMICAL PATHOLOGY

Detecting reagent lot shifts using proficiency testing data

Rui Zhen Tan[1], Wilson Punyalack[2], Peter Graham[2], Tony Badrick[2], Tze Ping Loh[3]

[1]*Engineering Cluster, Singapore Institute of Technology, Singapore;* [2]*Royal College of Pathologists of Australasia Quality Assurance Programs, Australia;* [3]*Department of Laboratory Medicine, National University Hospital, Singapore*

Summary

Clinically significant systematic analytical shifts can evade detection despite between-lot reagent verification, quality control and proficiency testing systems practiced by most laboratories.

Through numerical simulations, we present two methods to determine whether there has been a shift in the proficiency testing peer group of interest, peer group *i*, using the measurements from peer group *i* and *J* other peer groups. In method 1 ('group mean'), the distance of peer group *i* from the mean of the other *J* peer groups is used to determine whether a shift occurs. In method 2 ('inter-peer group' method), the distances of peer group *i* from each of the means of the other *J* peer groups are used to determine whether a shift has occurred.

The power of detection for both methods increases with the magnitude of systematic shift, the number of peer groups, the number of laboratories within the peer groups and the proportion of laboratories within the affected peer group, and a smaller analytical imprecision. When the number of peer groups is low, the power of detection for the group mean method is comparable to the inter-peer group method, using the m = 1 criterion (a single inter-peer group comparison that exceeds the control limit is considered a flag). At larger peer groups, the inter-peer group method using the same (m = 1) criterion outperforms the group mean method.

The proposed methods can elevate the professional role of the proficiency testing program to that of monitoring the peer group method on top of the performance of individual laboratories.

Key words: External quality assessment; proficiency testing; reagent lot; shift; drift; bias.

Received 17 April, revised 1 August, accepted 12 August 2019
Available online 10 October 2019

INTRODUCTION

Recently, there have been several commercial assay reagent recalls after systematic shifts were uncovered that adversely affected patient results and consequent clinical interpretation.[1–3] The clinically significant systematic shift evaded detection despite between-lot reagent verification, internal quality control (IQC) and proficiency testing systems practiced by most laboratories. In part, this is because individual laboratories use a relatively small number of samples for routine between-lot reagent verification, which significantly limits the statistical power of the process.[1,4] Systematic shifts can evade IQC detection when the target values are reassigned during reagent lot changes, or when the concentration of the lowest IQC material is not low enough to detect a minor shift.[3]

At the same time, it can be challenging for an individual laboratory to notice systematic shifts in the proficiency testing program if the entire peer group moves in the same direction as they adopt the same reagent (or calibrator) lot.[1] This view can also be confounded if staff believe the change is due to a matrix effect or represents non-commutability of the proficiency testing material.

A peer group can be defined as a grouping of laboratory method that share the same underlying analytical principle and system such that the matrix effects influence can be assumed to be same. Peer group comparison is often undertaken for analytes where no commutable materials with reference value assignment are available under the assumption that matrix effects influence the comparisons between results of different methods.

Proficiency testing programs are in the unique position of having data from many peer groups who are measuring the same material,[5] which can be analysed to detect such shifts. We hypothesised that the statistical distance (i.e., the between-method difference) between the different peer group results in a proficiency testing program is relatively stable and can be used as an anchor for one another. When one peer group drifts away, its relationship with the other peer groups can be detected statistically with appropriate control limits. We explored two different approaches to this concept through the numerical simulations described below.

METHODS

Assumptions

We present two methods to determine whether there has been a shift in the peer group of interest, peer group *i*, using the measurements from peer group *i*, and *J* other peer groups. In method 1, the distance of peer group *i* from the mean of the other *J* peer groups is used to determine whether a shift occurs. In method 2, the distances of peer group *i* from each of the means of the other *J* peer groups are used to determine whether a shift has occurred (Fig. 1).

The assumptions of the measurements are as follows: there is a proportional bias (i.e., between-method difference) in the measurements of each peer group (denoted as α_i for the peer group of interest, peer group *i*, and as α_j for the *J* peer groups; the proportional bias, α_j, of the *J* peer groups being used to determine whether there is a shift in peer group *i* are constant (i.e., there is no

Print ISSN 0031-3025/Online ISSN 1465-3931
DOI: https://doi.org/10.1016/j.pathol.2019.08.002

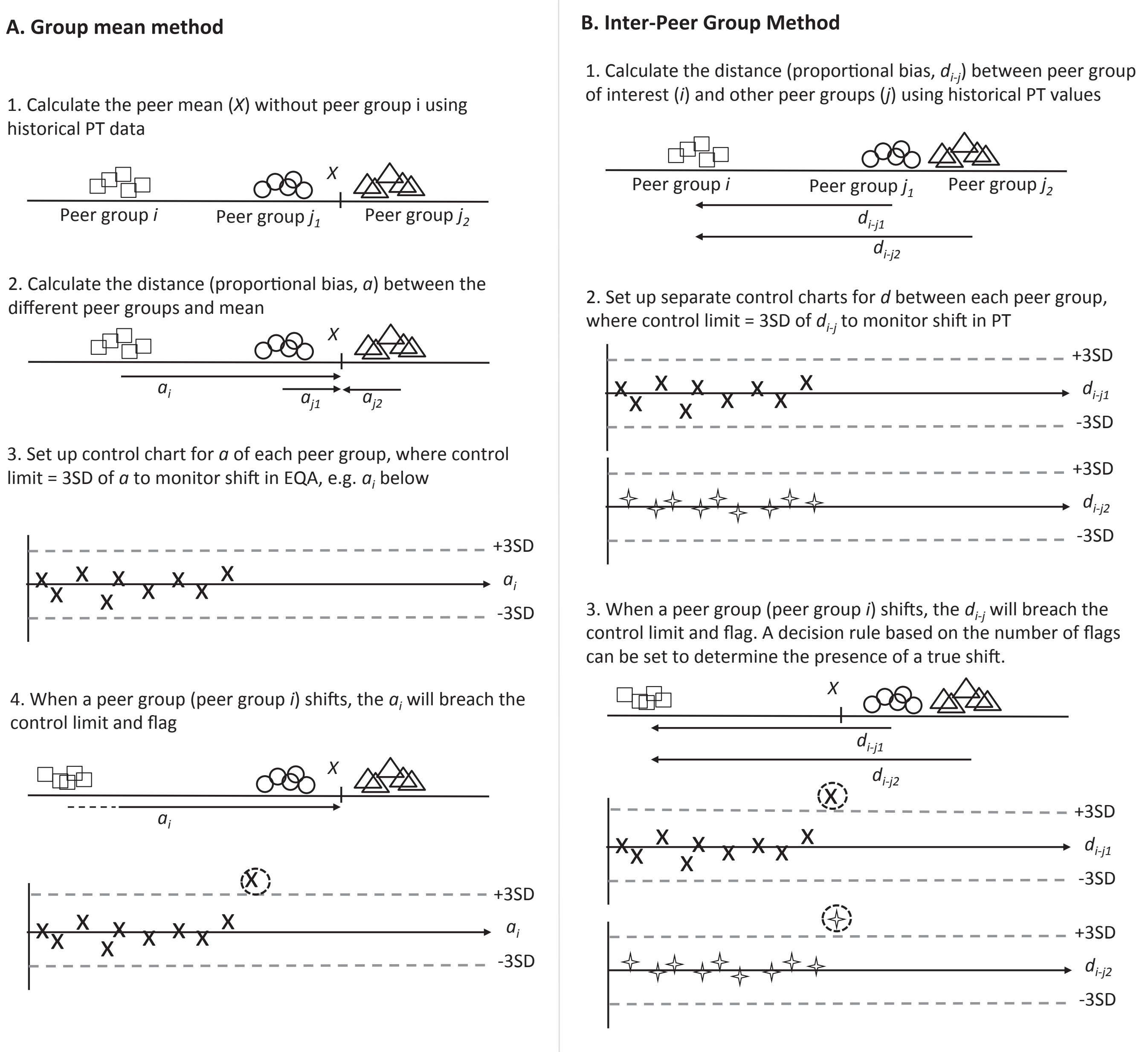

Fig. 1 Set up of (A) the group mean method and (B) the inter-peer group method. PT, proficiency testing program; SD, standard deviation.

shift in these other peer groups); and the analytical coefficient of variation of the peer group (CVa_J, i.e., the within-method CV, which includes within-laboratory CV as well as between-laboratory CV) in each peer group is the same and constant within the analytical measurement range of the assay. The number of laboratories within each peer group is assumed to be constant to simplify the simulation.

Let X be the target value of the measurement, the relationship between $\overline{x_j}$, the mean of the measurement in each peer group j, and X, is given by

$$\overline{x_j} = (1+\alpha_j)X$$

where α_j is the proportional bias for peer group j.

The measurement obtained by each laboratory n, in peer group j, is given by

$$x_{n,j} = \overline{x_j} + \xi_{n,j}$$

where $\xi_{n,j} \sim N(0, \overline{x_j} xCVa)$ denotes the random measurement error, when each laboratory in the peer group performs the measurement.

Similarly for the peer group of interest, peer group i, the mean of the measurement of the peer group, $\overline{x_i}$, and the measurement obtained by each laboratory, n, in the peer group are given by

$$\overline{x_i} = (1+\alpha_i)X$$

and

$$x_{n,i} = \overline{x_i} + \xi_{n,i}$$

respectively, where α_i is the proportional bias for peer group i and $\xi_{n,i} \sim N(0, \overline{x_i} xCVa)$.

Using the measurements from the J peer groups, an estimate (mean value) for X will be obtained as follows:

$$\overline{X} = \frac{\sum_j \sum_n x_{n,j}}{\sum_j N_j} = \frac{\sum_j \sum_n x_{n,j}}{NJ}$$

where J and N_j are the number of peer groups and the number of laboratories in each peer group, respectively. Since the number of laboratories in each peer group is constant, given by N, the total number of laboratories is simply the product of N and J. Measurements from peer group i are not being used in the calculation of $\overline{X}$ as the problem involves detection of drift in α_i.

From $\overline{X}$, the value of α_j can be deduced as follows

$$\widehat{\alpha_j} = \frac{1}{N}\left[\sum_{n=1}^{N}\frac{x_{n,j}-\overline{X}}{\overline{X}}\right]$$

In general, each laboratory will measure more than one concentration per proficiency testing cycle. For K measurements, the best estimate for $\widehat{\alpha_j}$ is generalised to

$$\widehat{\alpha_j} = \frac{1}{KN}\left[\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{x_{k,n,j}-\overline{X_k}}{\overline{X_k}}\right]$$

where the average of the estimate at each measurement is taken.

Similarly, for the peer group of interest, peer group i, the best estimate for $\widehat{\alpha_i}$ is given by

$$\widehat{\alpha_i} = \frac{1}{KN}\left[\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{x_{k,n,i}-\overline{X_k}}{\overline{X_k}}\right]$$

Method 1 (the 'group mean' method)

In method 1, we obtained $\widehat{\alpha_i}$ for peer group i. Using the results from many proficiency testing cycles, the mean $\overline{\alpha_i}$, and standard deviation s_{α_i}, of $\widehat{\alpha_i}$ are obtained. To determine where a shift in the value of α_i has occurred, we set the upper control limit UCL, and lower control limit LCL, for α_j as follows:

$$UCL_\alpha = \overline{\alpha_i}+3s_{\alpha_i} \text{ and } LCL_\alpha = \overline{\alpha_i} - 3s_{\alpha_i}$$

If the value of $\widehat{\alpha_i}$ obtained falls outside the upper or lower control limit, a drift in α_i is said to have occurred (Fig. 1).

Method 2 (the 'inter-peer group' method)

In method 2, the distance d_{ij}, between the peer group of interest i, and each of the other peer groups j, are obtained by

$$d_{ij} = \widehat{\alpha_i} - \widehat{\alpha_j}$$

Using the results from historical proficiency testing cycles, the mean $\overline{d_{ij}}$ and standard deviation $s_{d_{ij}}$ of d_{ij} are obtained (Fig. 1).

To determine where a drift in the value of α_j has occurred, we set the upper control limit (UCL) and lower control limit (LCL) for each of the distances (d_{ij}) as follows:

$$UCL_{d_{ij}} = \overline{d_{ij}}+3s_{d_{ij}} \text{ and } LCL_{d_{ij}} = \overline{d_{ij}} - 3s_{d_{ij}}$$

If the value of d_{ij} obtained falls outside the upper or lower control limit, the test is positive for peer group j. A parameter m, the minimum number of positive tests required to conclude that a drift in α_i has occurred is introduced. The higher the value of m the more stringent (specific) the requirement is.

Numerical simulations

To evaluate the two decision rules, we estimated their power in detecting a peer group shift by numerical simulations. In the simulations, we assume that only a fraction of the laboratories within peer group i has switched to using the affected reagent. The parameters in each simulation are as follows: J, the number of peer groups in an proficiency testing program which will be used to deduce the shift in the affected peer group; N, the number of laboratories in each peer group; η, CVa of each peer group; β, the shift of the affected peer group and; f, the fraction of laboratories affected within peer group i.

The simulations were performed in two parts. The first part represents a stable period where peer group i does not exhibit any systematic shift. This part of the simulation generates data for the calculation of the key statistics of the decision rules (e.g., the control limits). In the second part, a systematic shift is introduced into peer group i, and methods 1 and 2 are used to determine whether the shift is detected.

In total, 10,000 simulations were performed for each set of parameters. In each simulation, α of each peer group was drawn randomly from a uniform distribution –from −0.2 to 0.2. Next, we simulated data where the affected peer group did not exhibit drift. These data are simulated for 20 cycles. In each cycle, target values for 2 samples, x_k for $k = 1, 2$, are drawn randomly from a uniform distribution from 1 to 100. The measured values for each laboratory in each peer group j are then drawn from a normal distribution with the mean centred at $(1+\alpha_j)x_k$ and CVa equal to η (assumed as 2% or 10%).

Similarly for each laboratory in the peer group of interest, the measured values are being drawn from a normal distribution with the mean centred at $(1+\alpha_i)x_k$ for $k = 1,2$ and CV equal to η. $\overline{x_k}$, the best estimate for x_k, is then calculated using the measurements from the laboratories in the J peer groups.

Best estimates for α_j and α_i are determined as follows for $k = 2$:

$$\widehat{\alpha_j} = \frac{1}{2n}\left(\frac{\sum_n\left(x_{1,j}-\overline{x_1}\right)}{\overline{x_1}}+\frac{\sum_n\left(x_{2,j}-\overline{x_2}\right)}{\overline{x_2}}\right)$$

$$\widehat{\alpha_i} = \frac{1}{2n}\left(\frac{\sum_n\left(x_{1,i}-\overline{x_1}\right)}{\overline{x_1}}+\frac{\sum_n\left(x_{2,i}-\overline{x_2}\right)}{\overline{x_2}}\right)$$

The distance of our peer group of interest i to each of the other peer groups is also determined.

After 20 cycles there will be 20 sets of values. Using these values, the mean and standard derivation of $\widehat{\alpha_i}$ and d_{ij} are calculated. In the final cycle a systematic shift β will be included into peer group i. As before, two target values are drawn and the measured values for each laboratory in each peer group j, are then drawn from a normal distribution with the mean centred at $(1+\alpha_j)x_k$ for $k = 1,2$ and CVa equal to η.

Next, we will determine n_{aff}, the number of laboratories affected by the drift as $n_{aff} = fn$. The number of unaffected laboratories will then be given by

$$n_{unaff} = n - n_{aff}$$

For the unaffected laboratories, the measured values are being drawn from a normal distribution with mean centred at $(1+\alpha_i)x_k$ for $k = 1, 2$ and CVa equal to η. For the affected laboratories the measured values are being drawn from a normal distribution with the mean centred at $(1+\alpha_i+\beta)x_k$ for $k = 1,2$ and CVa equal to η.

Calculations of $\overline{x_k}$, $\widehat{\alpha_j}$, $\widehat{\alpha_i}$ and d_{ij} are determined as before. The decision rules for methods 1 and 2 will then be used to determine if a shift has occurred.

Simulations with no systematic shift ($\beta = 0$) in the final cycle are also performed. The decision rules for methods 1 and 2 are applied to obtain the false detection rates associated with the analysis.

RESULTS

The results of the simulations with an assumed CVa of 2% for the laboratory methods are summarised in Table 1 and Supplementary Tables 1–3 (Appendix A). The results of the simulations where a CVa of 10% is assumed are summarised in Supplementary Tables 4–7 (Appendix A). In general, the power of detection for both methods increased with the magnitude of the systematic shift, the number of peer groups, the number of laboratories within the peer groups and the proportion of laboratories within the affected peer group (Fig. 2). A smaller CVa was also associated with higher power of detection.

When the number of peer groups was low, the power of detection for the group mean method was comparable to the inter-peer group method using the m = 1 criterion, where a single inter-peer group comparison that exceeds the control limit is considered a flag (Fig. 2). At larger numbers of comparator peer groups, the inter-peer group method using the same (m = 1) criterion outperforms the group mean method. If the flagging criterion is tightened for the inter-peer

Table 1 Power of shift detection using the group mean method under different scenarios; the analytical coefficient of variation is set at 2%

No. of peer group in proficiency testing program	No. of laboratory in each peer group	Fraction of laboratory affected	Magnitude of shift				
			0.02	0.05	0.1	0.2	0.4
2	5	0.2	0.0197	0.0857	0.4051	0.963	1
		0.5	0.054	0.4156	0.9616	1	1
		1	0.4228	0.9967	1	1	1
	10	0.1	0.0133	0.042	0.1847	0.7402	0.9995
		0.2	0.0267	0.1832	0.7462	0.9999	1
		0.5	0.1919	0.9134	1	1	1
		1	0.7434	1	1	1	1
	20	0.1	0.0196	0.0782	0.3998	0.9662	1
		0.2	0.0527	0.4038	0.9667	1	1
		0.5	0.4056	0.9966	1	1	1
		1	0.964	1	1	1	1
	50	0.1	0.0351	0.2385	0.8479	0.9998	1
		0.2	0.1473	0.845	1	1	1
		0.5	0.8371	1	1	1	1
		1	0.9999	1	1	1	1
5	5	0.2	0.0197	0.1125	0.5141	0.9847	1
		0.5	0.0638	0.5116	0.986	1	1
		1	0.5132	0.9996	1	1	1
	10	0.1	0.0112	0.05	0.2319	0.843	1
		0.2	0.0338	0.2427	0.8374	0.9999	1
		0.5	0.2465	0.9607	1	1	1
		1	0.8505	1	1	1	1
	20	0.1	0.0198	0.1063	0.5104	0.9858	1
		0.2	0.0638	0.5084	0.9868	1	1
		0.5	0.5102	0.9992	1	1	1
		1	0.9871	1	1	1	1
	50	0.1	0.0473	0.3103	0.9146	1	1
		0.2	0.1861	0.9197	1	1	1
		0.5	0.9208	1	1	1	1
		1	1	1	1	1	1
10	5	0.2	0.0232	0.1152	0.5568	0.9919	1
		0.5	0.0738	0.5636	0.9921	1	1
		1	0.5464	0.9998	1	1	1
	10	0.1	0.0138	0.057	0.2659	0.8696	1
		0.2	0.0361	0.2607	0.8715	0.9999	1
		0.5	0.26	0.9744	1	1	1
		1	0.8783	1	1	1	1
	20	0.1	0.022	0.1107	0.5481	0.9929	1
		0.2	0.0692	0.5543	0.9927	1	1
		0.5	0.5584	0.9997	1	1	1
		1	0.9933	1	1	1	1
	50	0.1	0.0447	0.3441	0.9379	1	1
		0.2	0.2041	0.9394	1	1	1
		0.5	0.9392	1	1	1	1
		1	1	1	1	1	1

group method to m > 1 (i.e., two or more inter-peer group comparisons must exceed the control limit), the power of detection becomes lower than the group mean method.

On the other hand, the false detection rate is lower for the group mean method compared to the inter-peer group method using the m = 1 criterion (Table 2; Supplementary Table 8, Appendix A). The false rejection rate reduces when the number m criterion, the number of failed inter-peer group comparisons, is raised. In other words, a lower m criterion favours sensitivity (higher power of detection or true positive rate), whereas a higher m criterion favours specificity (higher true negative rate). The false rejection rate for the group mean method is relatively independent of the number of peer groups and the number of laboratories within the peer group. The inter-peer group method is significantly affected by the number of peer groups and relatively independent of the number of laboratories within each peer group.

For the group mean method, the number of laboratories within the peer group has a greater influence on the power of detection as compared to the number of peer groups (Fig. 2). By contrast, this effect is somewhat attenuated in the inter-peer group method. It is possible to select the m criterion for the inter-peer group method depending on the number of peer groups and the number of laboratories within the peer group by optimising the trade-off between sensitivity and specificity (Fig. 3).

DISCUSSION

As clinical laboratories become more automated and reliant on commercial reagents, they face increasing challenges in managing systematic shifts between reagent lots. In part, it has been suggested that this is due to under-appreciation of the clinical laboratory quality requirements by the reagent manufacturers.[6] Moreover, the manufacturers lack access to clinical materials to perform a robust reagent lot quality control check with sufficient statistical power for error detection before introducing a new lot into the market.

Ideally, this problem should be managed at source, where the reagent manufacturer, clinical laboratories, laboratory

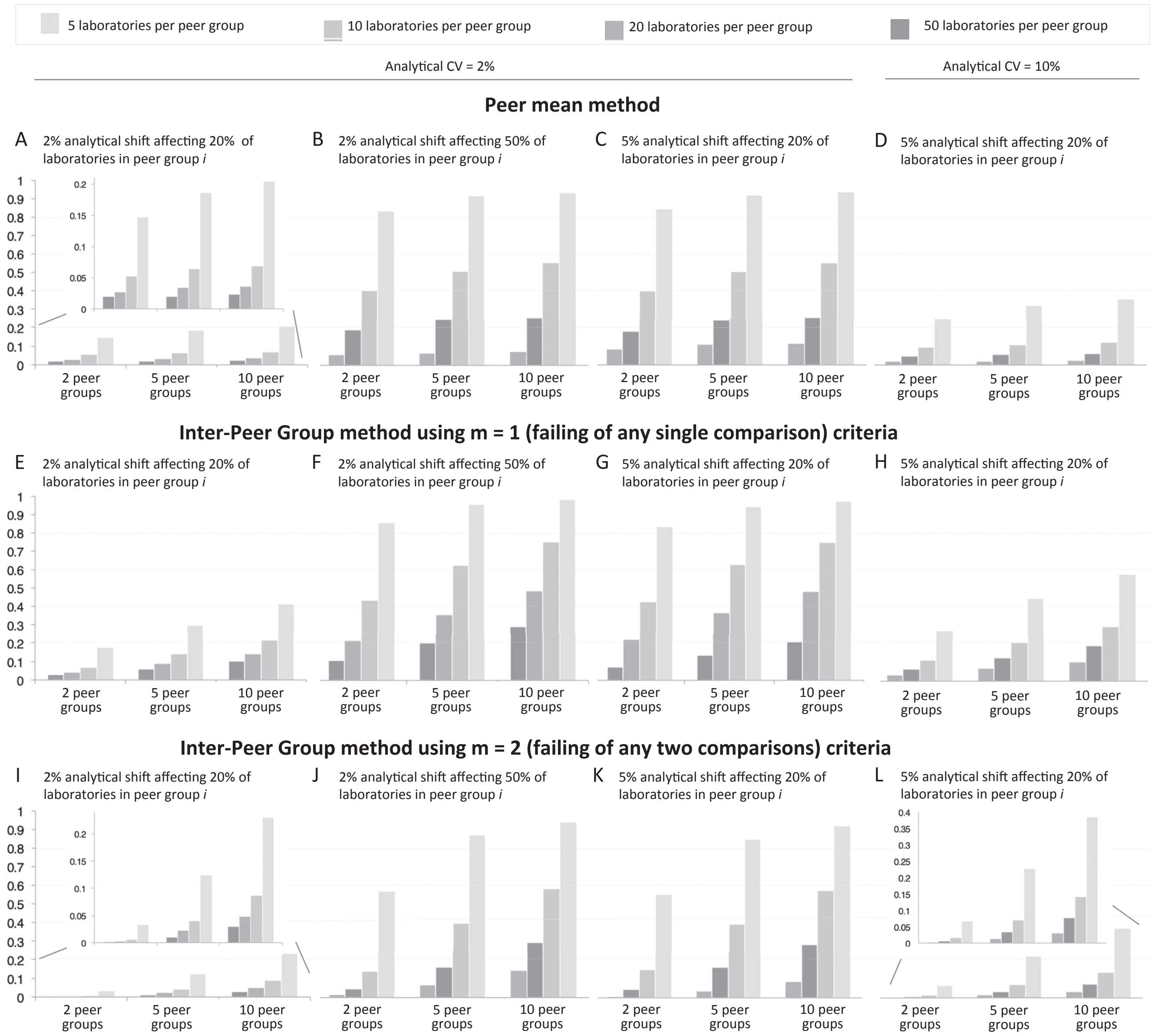

Fig. 2 Comparison of power of systematic shift detection under different simulation parameters [analytical coefficient of variation (CV), number of peer group, number of laboratories per peer group, magnitude of analytical shift, and proportion of laboratories affected by the shift within peer group of interest (peer group *i*)] by the group mean and inter-peer group methods. All panels are drawn to the same scale. Insets in A, I and L are enlarged figures.

professional bodies and proficiency testing providers work together to set the quality criteria and provide clinically relevant materials before a reagent lot is released for use. For example, the laboratory professional bodies and proficiency testing providers can set the quality goals, based on the Milan Consensus with the manufacturers.[6] The clinical laboratories can work through the professional bodies or proficiency testing providers (to avoid direct conflict of interest) to provide sufficient numbers of anonymised leftover clinical samples to evaluated a new lot of reagents before they leave the factory. This will ensure quality is assured at source. In reality, this process will take significant coordination and effort to realise.

In the meantime, the existing proficiency testing data can be leveraged to help monitor and provide early flags of systematic shifts in the reagent, as well as calibrator shift/drift that manifest in the proficiency testing. These methods are relatively simple and conceptually familiar to most laboratory practitioners. If adopted they have the potential to help minimise delay in triggering a product recall and reduce the number of patients affected or magnitude of harm. This will help maintain the trust of patients in clinical laboratory services. It will also help elevate the professional role of the proficiency testing program to that of monitoring the peer group method on top of the performance of individual laboratories.

From the results of the simulation, the ability to detect for a systematic shift in a peer group of interest is mainly dependent on the number of laboratories within each peer group, the proportion of laboratories affected by the shift and the analytical imprecision of the measurand. As such, the methods described in this study are more optimal for general chemistry assays which better fulfil the above conditions, compared to immunoassays. On the other hand, many immunoassays have a larger tolerance for analytical shift commensurate with the wider biological variation of the measurands. Consequently, the magnitude of systematic shift tolerable is also higher in

Table 2 False positive rate for the group mean method and the inter-peer group method with m = 1 and 2 criteria, where any one or two inter-peer group comparison that exceeds the control limit, respectively, is considered a flag, under different scenarios; the analytical coefficient of variation is set at 2%

Simulation parameters		False detection rate		
No. of peer group in proficiency testing program	No. of laboratory in each peer group	Group mean method	Inter-peer group method (m = 1)	Inter-peer group method (m = 2)
2	5	0.0087	0.0181	0.0007
	10	0.0086	0.0175	0.0004
	20	0.0095	0.0203	0.0004
	50	0.0105	0.0186	0.0006
5	5	0.0114	0.0432	0.0038
	10	0.0113	0.0437	0.003
	20	0.0101	0.0454	0.0037
	50	0.0115	0.0417	0.0035
10	5	0.0099	0.0731	0.0076
	10	0.0106	0.0832	0.0088
	20	0.0118	0.076	0.0099
	50	0.0104	0.0781	0.0089

these assays. In practice, the number of laboratories within each peer group may change with time and may affect the general performance of this approach, particularly with reduced number of laboratories for a peer group.

To increase the number of laboratories within each peer group and reduce the lag time for reagent lot adoption (or increase the proportion of laboratories affected by the analytical shift), proficiency testing programs may consider pooling their data for analysis. However, for this approach to work optimally, the proficiency testing programs that are sharing the data should use materials with similar characteristics to avoid issues with discrepant observations due to material non-commutability. This can only be realised by ongoing efforts to harmonise the practice among proficiency testing providers.[5]

Measurands using laboratory methods with larger CVa (e.g., immunoassays) can significantly limit the power of detection (Fig. 2). One possible approach to reduce the CVa of the peer group and improve the power of detection is to remove the poorest performing laboratories for analysis. This should be done with care to avoid discarding excessive

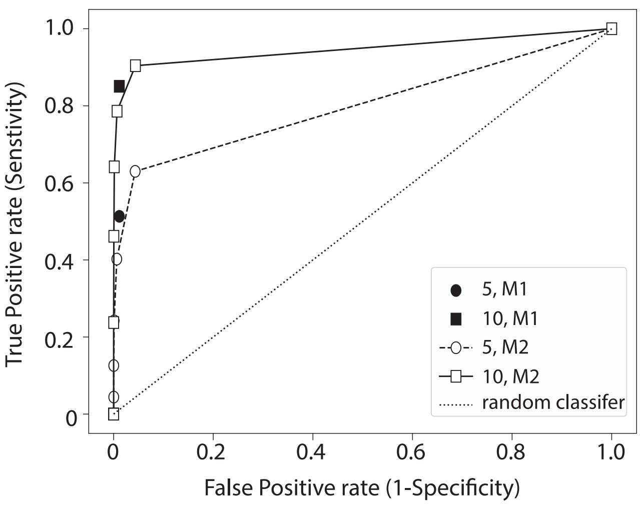

Fig. 3 Plot of false positive rate (1−specificity) versus true positive rate (sensitivity) for 5 peer group, CV = 2%, magnitude of analytical drift = 2%, proportional of lab = 100% for 5 (circle) and 10 (square) laboratory per peer group, respectively. M1 and M2 corresponds to method 1 (‘group mean’) and method 2 (‘inter-peer group’), respectively. In method 2, by varying the m criteria, a range of false positive rate and true positive rate is obtained. The dotted line shows the performance for a random classifier.

amounts of data while ensuring sufficient representation of the reagent lots in use.

In general, the inter-peer group method with m = 1 criterion is more sensitive than the group mean method. However, the increased sensitivity comes at a cost of increased false positive signals. When early signs of systematic drift are suspected for a particular peer group, the apparently affected laboratories can be notified to assess their IQC data and where available, archived patient samples. This verification exercise may be performed by several laboratories with pooled data analysed by a sensitive statistical method to reduce the burden of resources.[4]

To determine the effects of variation in the number of laboratories in each peer group, we repeated the simulations for the same average number of laboratories but with the laboratories now being randomly assigned to each peer group. From Supplementary Tables 9–12 (Appendix A), we found that the power of detection is in general lower with the variation in the number of laboratories.

Additionally, patient-based QC techniques, such as moving average and other patient based real time QC techniques should be used to assess the analytical performance during the period under suspicion.[2,3,7,8] These techniques can identify issues that may not be apparent on IQC or proficiency testing due to material commutability or statistical practice (e.g., changing the target values). In particular, patient-based real time QC techniques are important for measurands where the clinical interpretation may be affected by a small shift at the lowest concentration (i.e., physiological nadir concentration) that may evade detection as the IQC and proficiency testing materials may not adequately cover those concentrations.[2]

The proposed methods in this study do not overcome the issue of commutability traditionally associated with the use of manipulated proficiency testing materials. More recently, there have been attempts to use pooled patient data to perform moving statistics analysis to monitor analytical performance of thyroid assays. The results thus far are promising,[9,10] and may address the issue of commutability.

In summary, we have described two different statistical methods to detect analytical shifts using proficiency testing data, which uniquely pool data derived from the same sample material and laboratory method. While analytical shifts related to changes in reagent or calibrator lot should ideally be arrested at source (manufacturer) before distribution to the end user, it is possible to develop a monitoring system through close cooperation between proficiency testing programs and participating laboratories.

Conflicts of interest and sources of funding: The authors state that there are no conflicts of interest to disclose.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pathol.2019.08.002.

Address for correspondence: Dr Tze Ping Loh, National University Hospital, 5 Lower Kent Ridge Road, 119074, Singapore. E-mail: tploh@hotmail.com

References

1. Algeciras-Schimnich A, Bruns DE, Boyd JC, *et al*. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. *Clin Chem* 2013; 59: 1187–94.
2. Liu J, Tan CH, Badrick T, *et al*. Moving sum of number of positive patient result as a quality control tool. *Clin Chem Lab Med* 2017; 55: 1709–14.
3. Koerbin G, Liu J, Eigenstetter A, *et al*. Missed detection of significant positive and negative shifts in gentamicin assay: implications for routine laboratory quality practices. *Biochem Med (Zagreb)* 2018; 28: 010705.
4. Liu J, Tan CH, Loh TP, *et al*. Detecting long-term drift in reagent lots. *Clin Chem* 2015; 61: 1292–8.
5. Jones GR. The role of EQA in harmonization in laboratory medicine - a global effort. *Biochem Med (Zagreb)* 2017; 27: 23–9.
6. Thompson S, Chesher D. Lot-to-lot variation. *Clin Biochem Rev* 2018; 39: 51–60.
7. Liu J, Tan CH, Badrick T, *et al*. Moving standard deviation and moving sum of outliers as quality tools for monitoring analytical precision. *Clin Biochem* 2018; 52: 112–6.
8. Badrick T, Cervinski M, Loh TP. A primer on patient-based quality control techniques. *Clin Biochem* 2019; 64: 1–5.
9. Goossens K, Van Uytfanghe K, Twomey PJ, *et al*. Monitoring laboratory data across manufacturers and laboratories – a prerequisite to make "Big Data" work. *Clin Chim Acta* 2015; 445: 12–8.
10. De Grande LAC, Goossens K, Van Uytfanghe K, *et al*. Monitoring the stability of the standardization status of FT4 and TSH assays by use of daily outpatient medians and flagging frequencies. *Clin Chim Acta* 2017; 467: 8–14.