Reference-Scaled Margin for Lot Consistency Study

Jin Xu[1], David Li[2], G. Frank Liu[1]

[1]Merck & Co. Inc., 351 N. Sumneytown Pike, North Wales, PA 19454
[2]Pfizer, 500 Arcola Road, Collegeville, PA 19426

Abstract

Before a test vaccine can be approved by regulatory agencies for public use, a lot consistency study is usually required for the manufacturer to demonstrate its ability to produce the vaccine consistently. Three (3) vaccine lots are typically required for the study of consistency in terms of clinical endpoints, e.g. geometric mean titers (GMTs). And lot consistency is demonstrated only if 95% confidence intervals for the 3 pairwise lot-to-lot differences are all contained within a pre-specified consistency margin. Thus, the success of a lot consistency study is heavily hinged upon the consistency margin. Currently, a fixed consistency margin of 1.5, i.e., an interval of [1/1.5, 1.5], is commonly required by regulatory agencies, although a larger margin of 2.0 has also been allowed on a case by case basis. One concern with fixed margin is that it does not take consideration of within lot variance for the vaccine. Because the width of confidence interval is a function of both lot-to-lot difference and within lot group variance, a larger lot-to-lot difference may pass the consistency test if variances within lot groups are small but a smaller lot-to-lot difference may fail if variances within lot groups are large. For vaccines with high within lot variability a fixed margin of 1.5 becomes too stringent and requires a substantially larger sample size for it to pass the consistency test. This paper proposes a reference-scaled lot consistency margin based on variability of an active reference included in the lot consistency study. This active reference is a vaccine with antigens similar to those in the test vaccine, ideally one that has been approved by regulatory agencies for public use. Thus, the reference, being a vaccine of similar antigens, can provide an estimate of variance within lot group for the test vaccine antigens. The proposed consistency margin gradually widens based on intrinsic property of the vaccine, namely, the variability of the antibody titers it elicits in its target population. This approach is scientifically more appealing than a subjective decision to allow the use of a consistency margin of 2.0. Additionally, the proposed margin controls Type I error rate because the margin based on reference variance is independent of the lot consistent hypothesis testing.

Key Words: lot consistency, equivalence, consistency margin, reference-scaled margin, variability, vaccine

1. Introduction

Vaccines are large-molecule biologics with considerably more complex structures compared to small-molecule drugs. They are often produced by living organisms or systems from complex processes which are sensitive to minor changes such as air flow, temperature, or lighting in their environment. Compared with that for small-molecule drugs, the manufacturing process for vaccines are more variable, thus it is necessary for the manufacturer to demonstrate the consistency of vaccine production and performance of the vaccine lots in a randomized clinical trial with clinical endpoints.

In a typical lot consistency study, subjects from the target population are recruited and randomized to receive one of the 3 vaccine lots. After a suitable time period, blood samples are collected from subjects and measured for their antibody titers. The means of antibody titers for the 3 lot groups are then compared with each other in terms of log-transformed antibody titers, resulting in a total of 3 pairwise comparisons (Lot 1 vs. Lot 2, Lot 1 vs. Lot 3, and Lot 2 vs. Lot 3). The log-transformation is applied to make the distributions more normal because the distributions of antibody titers in the original scale are often skewed. A ratio of geometric means (GMTs) in the original scale (μ_i / μ_j) is equivalent to a difference between the two GMTs in the log scale $(\log(\mu_i) - \log(\mu_j))$, i and j index vaccine lots 1, 2, or 3. Similarly, a consistency interval of [1/1.5, 1.5] with a margin of 1.5 in the original scale is equivalent to [log (1/1.5), log (1.5)] = [-0.4055, 0.4055] in a natural log scale. Lot consistency is commonly assessed by a Two One-sided Test (TOST) proposed by Schuirmann (1987). The pairwise differences and their 95% confidence intervals are calculated under the normality assumption. The lot consistency is demonstrated only if the confidence intervals for the 3 pairwise differences are all contained within a pre-specified consistency interval, which is typically set by regulatory agencies to an fixed value of [-0.4055, 0.4055] with a margin of 1.5 (FDA, 2007). Occasionally, a wider consistency interval of [log (1/2.0), log (2.0)] = [-0.6093, 0.6093] with a margin of 2.0 is allowed by regulatory agencies, as in the cases of the lot consistency studies for Prevnar™13 (FDA, 2009) and BERXSERO™ (FDA, 2015), although guidance is absent on under what circumstances this wider consistency margin may be applied.

Vaccines protect subjects by eliciting antibodies against the disease-causing agents. Many vaccines can elicit a wide range of efficacious antibody levels in its target population. As long as a subject achieved a protective antibody level, the subject is being protected; a higher antibody level does not necessarily confer additional clinical efficacy. PENTACEL™ is an example; the anti-polio antibody titers it elicits have a wide range in its target population. As reported in the FDA review package, the GMTs and 95% confidence intervals for the 3 lot groups were 731 (640, 834), 394 (338, 459), and 478 (417, 549) for Polio 1; 1628 (1445, 1834), 1212 (1062, 1384), and 1364 (1211, 1536) for Polio 2; and 1314 (1153, 1498), 1127 (990, 1282), 977 (853, 1119) for Polio 3 (FDA, 2008). Although the ranges of antibody titers were wide, they were well above the protective level of 8. Thus, the wide range of high antibody levels did not necessarily have clinically meaningful impact in terms of vaccine efficacy. However, the wide range resulted in higher variances within each lot group, which consequently led to a wider confidence interval for the pairwise lot-to-lot difference. If a fixed margin, e.g., 1.5, is required, a lot consistency study for vaccines with high variability has to enroll a substantially large number of subjects to secure a confidence interval narrow enough to

pass the stringent consistency criterion. Such an enrollment increase is arguably an inefficient use of resource and may raise concerns on ethnical use of subjects.

In this paper, we propose a reference-scaled consistency margin based on the variability of a concurrent reference vaccine. We then compare the performance of the reference-scaled margin with a fixed margin of 1.5, and provide discuss and conclusions in the last section.

2. Hypothesis Testing and Variability in clinical endpoints

The main objective of a lot consistency study is to rule out a difference larger than a pre-specified margin δ_0. Let μ_i and μ_j, $i, j = 1,2,3$ be the true means of log-transformed antibody titers for lot groups *i* and *j*, then the hypotheses for the Two One-Sided Test consists of the following two sets of null and alternative hypotheses.

$$H_{01}: \mu_i - \mu_j \geq \delta_0 \ vs. \ H_{a1}: \mu_i - \mu_j < \delta_0, \ i, j = 1,2,3 and \ i \neq j$$

and

$$H_{02}: \mu_i - \mu_j \leq -\delta_0 \ vs. \ H_{a2}: \mu_i - \mu_j \geq -\delta_0, \ i, j = 1,2,3 and \ i \neq j$$

where δ_0 is a pre-specified consistency margin value, e.g. log(1.5)=0.4055, or log (2.0)=0.6931.

Under the normality assumption, the 95% CI can be calculated as

$$(\hat{\mu}_i - \hat{\mu}_j) \pm z_{1-\alpha/2}\sqrt{(\hat{\sigma}_i^2 + \hat{\sigma}_j^2)/n} \qquad (1)$$

Where α=0.025 is the Type I error rate, n is the sample size in a lot group (assume lot groups have the same sample size which is a common practice but it is not required), and $\hat{\sigma}_i^2 = \sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.}) \Big/ (n-1)$ is the variance within *i*th lot group (y_{ij} is the log-transformed antibody titer for the *j*th subject in the *i*th lot group). In order to claim lot consistency, all 3 confidence intervals need to be contained within the consistency interval, e.g. [-0.4055, 0.4055] for δ_0=log(1.5).

Equation 1 shows that the width of confidence interval is an increasing function of both lot-to-lot difference $\hat{\mu}_i - \hat{\mu}_j$ and variances within individual lot groups; a larger difference between lot groups or larger variances within lot groups can both lead to a wider confidence interval, thus a lower probability of claiming lot consistency. For a fixed δ_0 it is possible that a larger lot-to-lot difference with smaller lot variances can pass a lot consistency test but a smaller lot-to-lot difference with larger lot variance may fail.

Define between-lot variance as

$$\hat{\sigma}_B^2 = \sum_{i=1}^{3}(\bar{y}_{i.} - \bar{y}_{..})/2$$

and within-lot variance as the average of variances within lot groups,

$$\hat{\sigma}_W^2 = \sum_{i=1}^{3}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.}) \Big/ \sum_{i=1}^{3}(n-1),$$

where $\boldsymbol{y_{ij}}$is the log-transformed antibody titer for the *j*th subject in the *i*th lot group. The between-lot variance $\hat{\sigma}_B^2$ is a direct consequence of the differences among the 3 vaccine lots; larger differences among the 3 lot groups lead to larger $\hat{\sigma}_B^2$. By the time a vaccine developing program reaches the stage of conducting a lot consistency study, its vaccine manufacturing process is usually stabilized enough to produce vaccine lots consistently. Thus, the differences among lot groups are expected to be small and the purpose of a lot consistency study is to formally demonstrate it in a randomized clinical study. The within-lot variance $\hat{\sigma}_W^2$, on the other hand, is not related to lot-to-lot differences; it comes from factors such as variability in subjects' antibody titers and variability in assay measurements. Biologically, each vaccine or vaccine antigen acts uniquely on its target population, resulting in a unique characteristic variability in subjects' antibody titers. This subject variability is an intrinsic property of the vaccine antigen for a target population which cannot be reduced by any amount of improvement in assay precision, although it is not feasible to separate subject variability from assay variability in a lot consistency study.

We surveyed within-lot variance and between-lot variances in some recently approved vaccines, namely, Prevnar™ 13 (approved in 2010), PENTACEL™ (approved in 2008), and RotaTeq™ (approved in 2006). A summary of this survey is provided in Table 1 with highlighted rows indicating the antigens that would have failed lot consistency test with a fixed consistency margin of 1.5. Table 1 shows that the between-lot variances $\hat{\sigma}_B^2$ were small, as expected; and the within-lot variances $\hat{\sigma}_W^2$ were much larger, $\hat{\sigma}_W^2 >> \hat{\sigma}_B^2$ **.** Consequently, $\hat{\sigma}_W^2$ will have a large influence on the width of confidence intervals. Additionally, it is clear from Table 1 that $\hat{\sigma}_W^2$ varies in magnitude from one antigen to the next. For example, $\hat{\sigma}_W^2$ for PRP was 2.5015, more than five times that for FHA, which was 0.4715. Furthermore, an interesting observation made on Table 1 highlighted the influence of $\hat{\sigma}_W^2$ in the outcome of the lot consistency test. Two antigens in Prevnar™ 13, PCV4 and PCV6B had opposite outcomes: PCV4 which had a larger $\hat{\sigma}_B^2$ of 0.0244 passed the consistency test but PCV6B which had a smaller $\hat{\sigma}_B^2$ of 0.022 failed the test, partly because $\hat{\sigma}_W^2$ for PCV6B was twice as large as that for PCV4. The significant influence of $\hat{\sigma}_W^2$ and its varied magnitudes for different antigens underscore the needs for lot consistency margins to be based on intrinsic property of vaccine antigen's variability.

Table 1. Within-Lot Variance and Between-Lot Variance Observed in Lot Consistency Studies for Recently Approved Vaccines

Vaccine/Margin/ Year of Approval	Antigen (Lot Sample Size)	$\hat{\sigma}_W^2$	$\hat{\sigma}_B^2$	$\hat{\sigma}_B^2/(\hat{\sigma}_B^2+\hat{\sigma}_W^2)$
PENTACEL™ (Margin 1.5) (in 2008)	PRP (n=382, 378, 367)	2.5015	0.0148	0.59%
	FHA (n=336, 334, 325)	0.4715	0.0017	0.35%
	FIM (n=335, 332, 325)	0.6715	0.0118	1.73%
	PRN (n=336, 334, 325)	1.0128	0.0061	0.60%
	D (n=381, 378, 366)	1.3692	0.0125	0.90%
	T (n=380, 379, 366)	0.6691	0.0108	1.59%
	IPV1 (n=377, 369, 358)	1.8977	0.0998	5.00%
	IPV2 (n=376, 368, 358)	1.4609	0.0221	1.49%
	IPV3 (n=374, 367, 359)	1.6624	0.0220	1.31%
Prevnar™ 13 (Margin 2.0) (in 2010)	PCV4 (n=411, 404, 398)	0.5188	0.0244	4.50%
	PCV6B (n=409, 401, 396)	1.3786	0.0220	1.57%
	PCV9V (n=411, 403, 396)	0.4644	0.0010	0.22%
	PCV14 (n=398, 387, 387)	0.7652	0.0005	0.06%
	PCV18C (n=413, 401, 398)	0.5220	0.0046	0.87%
	PCV19F (n=408, 399, 398)	0.6159	0.0293	4.55%
	PCV23F (n=411, 402, 399)	0.9409	0.0131	1.38%
	PCV1 (n=411, 403, 395)	0.7018	0.0071	0.99%
	PCV3 (n=406, 391, 393)	0.5701	0.0064	1.11%
	PCV5 (n=412, 402, 393)	0.7416	0.0211	2.76%
	PCV6A (n=413, 402, 398)	0.7368	0.0056	0.75%
	PCV7F (n=412, 401, 397)	0.4907	0.0010	0.20%
	PCV19A (n=411, 403, 397)	0.6399	0.0017	0.26%
RotaTeq™ (Margin=2.0) (in 2006)	G1 (n=185, 195, 171)	2.0075	0.0058	0.29%
	G2 (n=185, 195, 171)	1.4030	0.0146	1.03%
	G3 (n=185, 195, 171)	1.3145	0.0168	1.26%
	G4 (n=185, 195, 171)	1.0997	0.0049	0.44%
	P1 (n=185, 195, 171)	1.3747	0.0027	0.20%
	IgA (n=186, 196, 172)	1.6933	0.0024	0.14%

n=number of subjects contributed to the analysis from the 3 lot groups.

Highlight indicates that the antigen would have failed the lot consistency test if a fixed margin of 1.5 were used in conjunction with 95% confidence interval.

3. Direct Widening of Consistency Margin

A fixed consistency margin of 1.5 is often required by regulatory agencies, as seen in the lot consistency studies for Menatra™ or PENTACEL™. But we also see a fixed margin of 2 used in lot consistency studies for recently approved vaccines, e.g., BERXSERO™, GARDSIL™ 9 and Prevnar™ 13 (Table 2). Since the rationale of this direct widening is not clear to us, we will not speculate except noting that such a widening from 1.5 to 2.0 is possible and a direct communication with regulatory agencies is highly recommended since the wider margin can reduce the sample size and save valuable resources.

Table 2. Summary of Consistency Margin and Inclusion of an Active Reference Group in Lot Consistency Studies

Vaccine	Date of Approval	Consistency Margin	Purpose of Including Active Reference Group[1]
BERXSERO	1/23/2015	2	NA
GARDSIL9	12/10/2014	2	NA
Prevnar 13	2/24/2010	2	Concomitant use with Pediarix
Pentacel	6/28/2008	1.5	Non-inferiority of Pentacel vs. component controls
RotaTeq	2/3/2006	2	NA
Menactra	1/14/2005	1.5	Safety comparison with active reference
DAPTACEL	5/14/2002	1.5	Non-inferiority of Pentacel vs. DAPTACEL, and safety comparison
[1] Refer to control group using reference vaccine which had antigens similar to those in the test vaccine.			

4. Reference-Scaled Consistency Margin

Rationales and Proposed Margin

The reference-scaled consistency margin is applicable to a lot consistency study that includes an active reference group in addition to the 3 lot groups for the test vaccine. The active reference vaccine used in the study should have already been approved for public use and should have antigens similar to those in the test vaccine. Subjects recruited from the target population are randomized to the 4 study groups, not necessarily in equal allocation. An unequal allocation to the lot groups and reference group, e.g., in the ratio of 2:2:2:1, can be used depending on the study objectives. The inclusion of the reference vaccine of similar antigens provides us a reference point. As the reference vaccine has already been approved for public use, it is reasonable to assume its manufacturing process is consistent. From that view point, we formulate a putative lot consistency study for the reference vaccine with a reasonable sample size, e.g. 300 per lot group for a total of 900 subjects in the study. We envision that this consistency study for the reference should conclude with lot consistency correctly with a high probability, e.g. a power of 90%, and incorrectly with a low probability, e.g. a Type I error rate of 2.5%, with the variance for the reference group observed in the current study. With these parameters set,

we then reverse engineered corresponding consistency margin. Under the normality assumption of the log-transformed antibody titers, this margin is

$$\delta_0^{'} = \sqrt{2/n_{ref}}\hat{\sigma}_{ref}\left(z_{1-\alpha/2} + z_{\beta}\right),$$

where n_{ref} is a pre-specified sample size for the reference group, e.g. 300;

$\hat{\sigma}_{ref}$ is the square root of the variance for the reference group observed in the current lot consistency study; and

$Z_{1-\alpha/2}$ and Z_{β} are the corresponding percentiles of standard normal distribution, α and β are 2-sided Type I error and power, respectively.

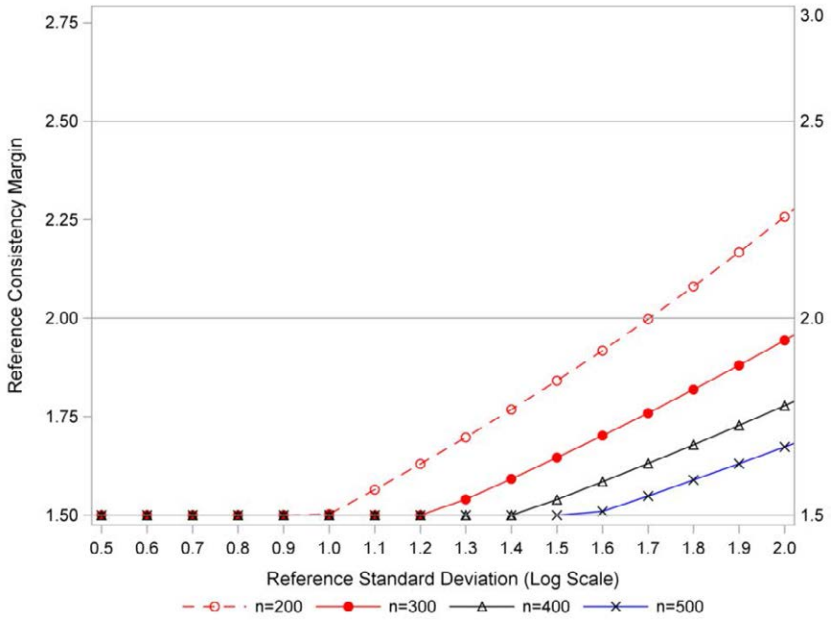
Furthermore, because a fixed margin of 1.5 is commonly used, we use it as the minimum margin, that is, we propose a reference-scaled consistency margin as

$$\delta_0 = \max\left(\log(1.5),\quad \sqrt{2/n_{ref}}\hat{\sigma}_{ref}\left(z_{1-\alpha/2} + z_{\beta}\right)\right) \qquad (2).$$

In Equation 2, the reference group is entirely independent of the 3 lot groups, so is its variance. Consequently, the reference-scaled margin is totally independent of the lot consistency test. This independency distinguished this reference-scaled consistency margin from similar reference-scaled margins in bioequivalence test (Boddy, et al. 1995) or in biosimilar test (Zhang, et al., 2012) where the reference variance used to determine the margin is also part of the test statistics. The concept of an objective margin using a reference had also been proposed by Li and Xu (2014) for biosimilar studies.

Figure 1 presents reference-scaled margins for standard deviation, σ_{ref} from 0.5 to 2 (i.e. σ_{ref}^2 from 0.025 to 4) covering the range of $\hat{\sigma}_W^2$ observed in lot consistency studies for recently approved vaccines (Table 1), and reference sample sizes from 200 to 500, which covers sample sizes typically used. Figure 1 displays one desirable feature of the proposed reference-scaled margin—it yields consistency margins within the boundaries set by the 2 currently used margins of 1.5 and 2.0 in nearly all cases, except for low reference sample size of n_{ref}=200 and standard deviation >1.7 ($\hat{\sigma}_{ref}^2 > 2.89$) which is larger than all $\hat{\sigma}_W^2$ observed in Table 1. So although the reference-scale margin increases as the reference variance increases, the proposed margin is still conservative enough and will not be larger than a margin of 2, a more liberal margin used in some approved vaccines. Another desirable feature of the proposed margin is its continuity; there is no jumping point at any variance values. This continuity provides certain protection against estimation error of the reference variance that may be near the jumping point. A small overestimate or underestimate of σ_{ref} does not change consistent margin dramatically as it could near a jumping point.

Figure 1. Reference-Scaled Margin in Relation to Reference Standard Deviation and Sample Size



Performance in Simulation

We compared the performance of the reference-scaled consistency margin with a fixed margin of 1.5 in simulated studies of 300 subjects for the reference group and the each of the 3 lot groups. The true study parameters used in the simulations are as follows:

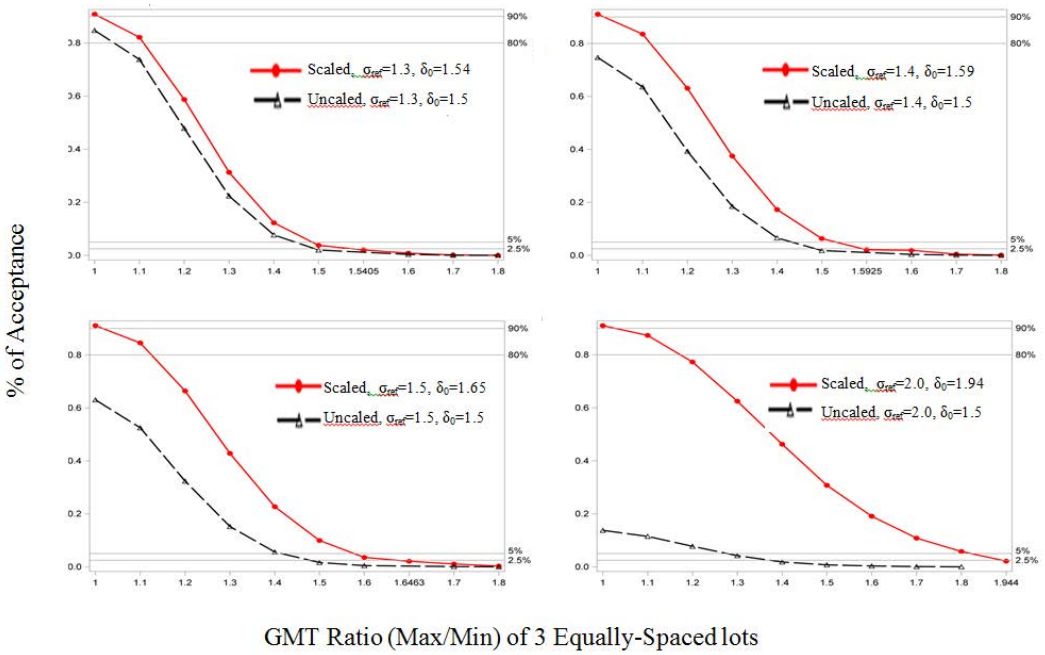
a. GMT ratios (largest/smallest) from 1 to 1.5 (or to the scaled margin). An equal-spacing configuration for the 3 GMTs, i.e., largest/middle = middle/smallest, was used.

b. σ_{ref} =1.3, 1.4, 1.5, and 2.0

c. $\sigma_W = \sigma_{ref}$.

A total of 100,000 simulations were run for each scenario. The percent of runs that resulted in acceptance of lot consistency was plotted in Figure 2. This percentage is study power when GMT ratio=1, and Type I error when GMT ratio=1.5 for lot consistency test with a fixed margin of 1.5, or reference-scaled margins for tests using these margins. As shown in Figure 2, the study power values were appreciably higher for consistency tests with reference-scale margin than with a fixed margin, which is expected given that the reference-scaled margin is wider. The power gain increases as the within-lot variance increases. Also as expected, the Type I error is controlled at the pre-specified 2.5% level. The guaranteed control of Type I error of consistency tests using the reference-scaled margin is by design because it is determined entirely independent of the lot consistency test.

Figure 2. Percent of Acceptance by GMT Ratio (300 subjects per group)



Case Study

We applied the reference-scaled margin to retest lot consistency for the 4 antigens that did not pass the consistency test with a fixed margin of 1.5 as occurred in the original study (Table 1). In the lot consistency study for PENTACEL™, an active reference group was included to show that PENTACEL™, a combination vaccine, was non-inferior to its antigen components. We calculated the reference variance $\hat{\sigma}^2_{ref}$ based on the sample size, mean and 95% confidence interval reported in the FDA review package, and calculated reference-scaled margin according to Equation 2 with n_{ref}=300, α=0.025 and β=0.1. As shown in Table 3, the scaled margin increased to 1.63 for PRP because of its large $\hat{\sigma}^2_{ref}$ but remained at 1.5 for IPVs because their $\hat{\sigma}^2_{ref}$ values were relative small. PRP would have passed lot consistency test if the reference-scale margin of 1.63 was used. Because the between-variance for PRP was <1% of the total variance, one could argue that the vaccine lots were similar enough in terms of PRP antibody titers, given the context of large variance within lot groups. A reference-scale consistency margin allowed the antigen with small lot-to-lot difference but large within-lot variance to pass the consistency test.

Table 3. Lot Consistency Test for the 4 Antigens that Failed Their Original Test

Antigen	Test Vaccine Variance			$\hat{\sigma}^2_{ref}$	Reference-Scaled Margin	Lot Consistent?
	$\hat{\sigma}^2_W$	$\hat{\sigma}^2_B$	$\frac{\hat{\sigma}^2_B}{\hat{\sigma}^2_B + \hat{\sigma}^2_W}$			
PRP	2.44	0.01	0.60%	2.13	1.63	Yes
IPV1	1.89	0.10	5.02%	1.11	1.50	No
IPV2	1.46	0.02	1.49%	0.99	1.50	No
IPV3	1.66	0.02	1.31%	1.32	1.50	No

5. Discuss and Conclusion

The success of a lot consistency study depends on the consistency margin required for the test. Currently, fixed margins of 1.5 or 2.0 are commonly used; 1.5 is typically required by regulatory agencies, but 2.0 is occasionally allowed on a case by case basis. For TOST or other methods that require confidence intervals to be contained within a pre-specified margin, it is well known that a larger lot-to-lot difference may pass the consistency test if its associated variance is small but a smaller lot lot-to-lot difference may fail if its variance is large. When encounter the problem of demonstrating consistency for vaccine or vaccine antigen with large variability, some manufacturers with economic means enroll a large size of subjects to narrow the confidence intervals enough in order to pass the consistency test with a fixed margin of 1.5. There were cases when the sample size reached 800 subjects for each lot group. This required a total of 2400 subjects for a lot consistency study. A large sample size utilized this way may not be an efficient or even ethical use of valuable resources. But a regulatory decision to widen the consistency margin to 2.0 leaves too much for negotiation.

We propose a reference-scaled consistency margin based on an intrinsic property of the vaccine, namely the variability of the antibody titers elicited by a reference vaccine in the target population. The reference-scaled consistency margin widens as the reference variance increases. At the same time, it is conservative in the sense that it widens the margin gradually from 1.5 to 2.0, but not beyond 2.0 in most cases. A gradual widening based the variance from a concurrent reference is scientifically more appealing than a subjective decision. Additionally, the proposed margin is continuous with no jumping points at any variance values, which prevents dramatic change in the margin by slight overestimate or underestimate in the reference variance. Additionally, the proposed margin controls Type I error because the margin based on reference variance is independent of the lot consistent hypothesis testing.

To determine a reference-scale consistency margin, an active reference group will need to be added to the lot consistency study. Even though a control group is not required for a lot consistency study, it is quite common based on observations from FDA review packages of recently approved vaccines (Table 2). The active reference group in a lot consistency study can play a key role in expanding the study objectives beyond demonstrating lot consistency. As Table 2 shows, it can serve as a control for the non-inferiority of test vaccine, for safety assessment of the test vaccine, or for assessment of concomitant use of the test vaccine with standard vaccines required for the target population. The reference, being a vaccine of similar antigens as those in the test vaccine, can provide estimates of subject variability and assay variability for the test vaccine

antigens. Getting variance estimate from a reference group in the concurrent study is advantageous because that the same assay is used to measure the antibody titers for both the reference group and the lot groups, and the randomization provides comparability of subjects between the reference group and lot groups. Obviously, the accuracy of the reference variability in reflecting the test vaccine's variability depends on the similarity between the reference and test vaccines. The reference vaccine should be selected as similar to test vaccine as possible. Additionally, to ensure a reasonable accuracy, the sample size for the reference group should be sufficiently large. In general, we recommend a sample size of no less than 200 to provide reasonable accuracy in the estimation of variability.

In summary, reference-scaled consistency margin is a viable approach to set a reasonable consistency margin for vaccines with high within lot variance.

References

Boody, A.W., Snikeris, F.C., Kringle, R.B., et al. (1995) An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. *Pharmaceutical Research*, 12 (12) 1865-1868.

FDA (2006) Clinical review Part 2 – RotaTeq http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM142306.pdf (assessed Oct 2, 2015)

FDA (2007) Guidance for industry—Clinical data needed to support the licensure of seasonal inactivated influenza vaccines. http://www.fda.gov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/ucm091990.pdf (accessed Sep 27, 2015)

FDA (2008) Immunogenicity Review – Pentacel http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM125900.pdf (assessed Oct 2, 2015)

FDA (2009). Statistical review and evaluation for BLA 125324/0. http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM206200.pdf (accessed Sep 29, 2015)

FDA (2015) Clinical review of BERXOSER, STN 125546/0. http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM434714.pdf (accessed Sep 29, 2015)

Li, J. and Xu, J. (2014) Bridging a new biological product to its references. (To appear in book Clinical and Statistical Considerations in Biosimilar Study)

Schuirman D. J. (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, Vol 15, No. 6.

Zhang, N., Yang, J., Chow, S.C., Endrenyi. L,. & Chi, E. (2012) Impact of variability on the choice of biosimilarity limits in assessing follow-on biologics. *Statistics in Medicine*, 32, 424-433.