Reference 3

Tinch, Latroy D

From: Hare, Michael [Michael.Hare@solvay.com]
Sent: Friday, June 09, 2006 3:00 PM
To: Tinch, Latroy D
Subject: Docket No. 2006D-0083 and Docket No. 2006D-0088

Latroy,

The documents provided by Solvay Pharmaceuticals in response to Docket No. 2006D-0083 and Docket No. 2006D-0088 can be released to the public. The information contained in these documents are ***not*** confidential in nature including reference #3. This reference was inadvertently submitted to the Agency with the word "Confidential" printed on the first page.

Best regards,

Michael F. Hare
Manager, Regulatory Affairs
Solvay Pharmaceuticals, Inc.
Marietta, GA 30062
T: 770-578-5620
F: 770-578-5864
E: michael.hare@solvay.com

This e-mail is confidential.
If you are not the addressee or an authorized recipient of this message, any distribution, copying, publication or use of this information for any purpose is prohibited.
Please notify the sender immediately by e-mail and then delete this message.

Ce message est confidentiel.
Si vous n'etes pas le destinataire designe de ce message ou une personne autorisee a l'utiliser, toute distribution, copie, publication ou usage a quelques fins que ce soit des informations contenues dans ce message sont interdits.
Merci d'informer immediatement l'expediteur par messagerie electronique et

d'ensuite detruire ce message.

Title:

Statistical analysis of influenza vaccine lot consistency studies

CONFIDENTIAL

Author:

Jozef JP Nauta, M.Sc.

Affiliation:

Clinical Development Influenza Vaccines

Solvay Pharmaceuticals

Weesp, The Netherlands

e-mail: jos.nauta@solvay.com

Running title:

Influenza vaccine lot consistency studies

ABSTRACT

This paper presents an integrated statistical approach to the analysis of influenza vaccine lot consistency studies in which three lots are compared. The approach ensures that the overall type I error rate (i.e. the probability of wrongly concluding that the lots are similar) is controlled. It is argued that the optimum efficacy measure is the geometric mean titer. The approach is demonstrated using data from a randomized, double-blind lot consistency study in which three consecutive production lots of Solvay Pharmaceuticals' new, virosomal subunit influenza vaccine Invivac® were compared.

Key words: influenza vaccines, lot consistency, immunogenicity, statistical analysis, equivalence margin.

INTRODUCTION

The objective of influenza vaccine lot consistency studies is to show that the lots are similar with respect to the immunogenicity of the hemagglutinin (HA) of the three viral strains contained in the vaccine. Lot consistency studies are therefore an example of equivalence trials. For studies with two lots (i.e. two treatments) and a single outcome, the statistical theory of equivalence testing is well established. Briefly, let Δ be the true treatment difference and δ the equivalence margin. The two treatments are considered equivalent if $|\Delta| < \delta$. To demonstrate equivalence, the null hypothesis H_0: $|\Delta| = \delta$ is tested against the alternative hypothesis H_1: $|\Delta| < \delta$. If the null hypothesis is tested by the two one-sided tests procedure at the significance level α, this approach corresponds to checking that the two-sided $100(1-2\alpha)\%$ confidence interval for Δ lies within the equivalence range $-\delta$ to δ (Schuirmann, 1987).

Influenza vaccine lot consistency studies are, however, non-trivial. Firstly, there are two types of multiplicity in statistical testing which have to be dealt with. For licensure in the United States of America not two but three lots need to be compared (FDA Guidance, 1997). In addition there are three outcomes instead of one - one for each of the three strains in the vaccine. Ignoring these multiplicities may increase the overall type I error rate, i.e. the probability of wrongly concluding that the lots are equivalent. Secondly, there is the difficult aspect of the choice of the equivalence margin δ. Thirdly, there is the problem of the choice of the optimum efficacy measure.

In this paper an integrated approach to the statistical analysis of influenza vaccine lot consistency studies is presented, combining available methodology with new developments. It is explained how both types of multiplicity can be dealt with. The first type – multiple comparisons between lots – can be handled by applying a statistical method developed by Wiens and Iglewicz (1999, 2000) to test the equivalence of three populations. It will be argued that the second type – multiple comparisons because of more than one strain – can in fact be ignored: in other words, no adjustment is needed. Next, an equivalence margin is proposed. It is argued that the optimum efficacy measure is the geometric mean titer and not the mean fold increase or the seroprotection rate, for example. The approach is demonstrated using data from a randomized, double-blind lot consistency study in which three consecutive production lots of Solvay Pharmaceuticals' new, virosomal subunit influenza vaccine Invivac® were compared.

MATERIALS AND METHODS

Clinical Data

In 2001 Solvay Pharmaceuticals conducted a randomized, double-blind lot consistency study to compare three consecutive production batches of its new, virosomal subunit influenza vaccine Invivac®. The tested formulation contained 15 micrograms HA of each of the three strains recommended by the World Health Organization for the season 2001/2002. After informed consent had been obtained and the inclusion and exclusion criteria had been checked, single blood samples were taken for baseline anti-HA antibody titration. Next, the subjects were randomized to be immunized with a single dose from lot 1, lot 2 or lot 3. Questionnaires were handed out so that subjects could record local and systemic reactions. After three weeks the subjects returned for a second blood sample for anti-HA antibody titration. The questionnaires were collected and any other adverse events were recorded in the case report forms. In total, 373 subjects were randomized: 124 to lot 1, 125 to lot 2 and 124 to lot 3. Following ICH guideline E9 (1998), the data of this equivalence study were analyzed according to the per-protocol principle. Prior to unblinding the database, the Blind Review Committee excluded 10 subjects (2.7%) from the per-protocol sample because of an intercurrent infection or a major protocol violation: one subject randomized to lot 1, two subjects randomized to lot 2 and seven subjects randomized to lot 3.

Statistical methods

HI titers

The immunogenicity of influenza vaccines is determined by the hemagglutination inhibition (HI) test for anti-HA antibody titration (Dowdle et al., 1997). The HI titer is defined as the dilution factor of the highest dilution that still completely inhibits hemagglutination. The starting dilution is usually 1:10. From this dilution, further twofold dilutions are prepared: 1:20, 1:40, ..., 1:2560. Thus, HI titers can take the values 10, 20, 40, ..., 2560. If the starting dilution does not inhibit hemagglutination a titer value of 5 is assigned. It is standard procedure to titrate every blood sample twice, and the HI titer assigned to the blood sample is the geometric mean of the two titrations.

The standard statistic for summarizing HI titers is the geometric mean titer (GMT). Similarity of the lots can be tested by comparing the GMTs between lots for each strain. Establishing equivalence is easiest if the data are normally distributed. HI titers tend to be skewed to the right, but a log transformation usually makes the observations approximately normal. A convenient transformation for HI titers is

$$\text{log HI titer} = \log_2(\text{HI titer}/5)$$

Note that the log HI titers are now identical to the dilution steps: 0 (HI titer = 5), 1 (HI titer = 10), 2 (HI titer = 20), ..., 9 (HI titer = 2560). There is a one-to-one relationship between the arithmetic mean of the log-transformed HI titers (AMLT) and the GMT of the untransformed titers

$$GMT = 2^{AMLT} \times 5$$

Furthermore, let GMT_i and GMT_j denote the geometric mean titer of the ith and the jth lot respectively, and $AMLT_i$ and $AMLT_j$ the arithmetic means of the log-transformed HI titers. Then

$$2^{AMLTi - AMLTj} = GMT_i / GMT_j$$

Thus, if equivalence of the log-transformed HI titers is demonstrated by showing that $AMLT_i - AMLT_j \cong 0$, equivalence of the HI titers is demonstrated at the same time by showing that the geometric mean ratio (GMR) is close to one, i.e. $GMT_i / GMT_j \cong 1$.

Fold increases

Instead of using HI titers to demonstrate lot consistency, a preference has been expressed for using the baseline-corrected fold increases

$$\text{Fold Increase} = HIT_P / HIT_B$$

with HIT_B and HIT_P being the baseline and the post-vaccination HI titer respectively (Zangwill et al., 2001). Note that

$$\log \text{Fold Increase} = \log HIT_P - \log HIT_B$$

Let MFI_i denote the geometric mean of the fold increases (= mean fold increase) and GMT_{Bi} and GMT_{Pi} the geometric mean of the baseline and the post-vaccination HI titers of the ith lot respectively. It is easy to see that

$$MFI_i / MFI_j = [GMT_{Pi} / GMT_{Pj}] / [GMT_{Bi} / GMT_{Bj}]$$

If there is no baseline imbalance in HI titers between the ith and the jth lot (i.e. $GMT_{Bi} / GMT_{Bj} \sim 1$), then the MFI ratio is approximately equal to the GMR, i.e.

$$MFI_i / MFI_j \sim GMT_{Pi} / GMT_{Pj}$$

Thus only in the case of baseline imbalance will there be a difference between the analysis of the fold increases and that of the post-vaccination HI titers. Contrary to common opinion, the use of the fold increases does not eliminate bias in the GMT due to baseline imbalance; it actually introduces bias. Fold increases are negatively correlated with baseline HI titers: the higher the

baseline HI titer the smaller the fold increase (Beyer et al., 2004). This means that if the baseline HI titers are imbalanced the MFI will be highest for the lot with the lowest baseline values. Thus, if there is no baseline imbalance between the lots the analysis of the fold increases will yield the same results as the analysis of the post-vaccination HI titers. However, the analysis of the HI titers would be expected to be more powerful because of the lower variability between the observations. The explanation for this is that log-transformed fold increase is a difference, namely between a log-transformed post-vaccination and a log-transformed baseline HI titer. If the variances σ^2 at baseline and post-vaccination are the same, then the variance of the log difference is $2\sigma^2(1-\rho)$, which implies that if the correlation ρ between the post-vaccination and the baseline HI titers is less than 0.5, the variance of the difference is larger than that of the post-vaccination HI titers.

Equivalence of three influenza vaccine lots

The first type of multiplicity mentioned in the introduction – multiple comparisons between the three lots – can be handled by applying a statistical method developed by Wiens and Iglewicz (1999, 2000). To demonstrate equivalence, the differences between the lot means (of the log-transformed HI titers or the log-transformed fold increases) must be shown to be small. This is done by testing the null hypothesis H_0: max $|\Delta_{ij}| = \delta$ against the alternative hypothesis H_1: max $|\Delta_{ij}|$

$< \delta$, where Δ_{ij} is the true but unknown difference between the ith and the jth lot. The null hypothesis is tested using the test statistic

$$Z_{min} = \min \{ (\delta - |D_{ij}|) / se_{ij} \}$$

where D_{ij} is the observed difference between the ith and the jth lot and se_{ij} the standard error of this difference, i.e.

$$se_{ij} = v \, [sd_i^2/n_i + sd_j^2/n_j]$$

with sd_i and sd_j the standard deviations of the lot means and n_i and n_j the lot sizes. As critical values for Z_{min} the upper centiles z_a of the standard normal distribution can be used, in which case the test will be conservative. Less conservative critical values c_α are given in Table 1 of Wiens and Iglewicz (1999). Both types of critical values account for the multiple comparisons between the three lots, and no further adjustment is needed.

It has been explained above how the equivalence of three lots can be demonstrated in the case of a single strain, i.e. in the case of a single outcome. Current influenza vaccines, however, contain HA of three strains (an A-H_1N_1 strain, an A-H_3N_2 strain and a B strain), and consequently there are three outcomes. The question then arises of whether and how to adjust for this second

type of multiplicity. The answer to this question is as simple as it is straightforward, if the point of view is taken that equivalence of the lots can only be claimed if equivalence is demonstrated for all three strains. In that case no adjustment of the test-wise significance levels is needed. If H_0 is tested per strain at the significance level a and equivalence of the lots is concluded only if the null hypothesis is rejected for all three strains, then the probability of wrongly concluding that the three lots are equivalent is α at most. This was first pointed out by Berger (1982) in a quality control context. Berger and Hsu (1996) generalized this to the concept of intersection-union tests (IUTs). An IUT is a test in which each component of an overall null hypothesis must be rejected separately; if each component is tested at significance level α, the size of the overall test is also α.

Choice of the equivalence margin

A difficult aspect of equivalence studies is the choice of the equivalence margin. Many strategies have been suggested as to how to choose equivalence limits, either on clinical or statistical grounds (Wiens, 2002). However, no golden rule exists. If too small a margin is chosen, the sample size required to secure sufficient statistical power will be prohibitively large. If too large a margin is chosen, it will be meaningless.

The GMR is a ratio, thus the value for exact equality of two GMTs is 1.0. Because proof of exact equality is impossible, in an equivalence study a range of values must be defined (a priori) indicating how different treatments can be before the difference is of concern. This is the equivalence range. An equivalence range for the GMR and the MFI ratio found in the literature is 0.25 to 4.0. Zangwill et al. (2001) report on a lot consistency study of a cold-adapted, intranasal influenza vaccine. Their analysis involves comparing, per strain, the mean fold increases between the lots. The lots are considered equivalent if all 95% confidence intervals for the mean fold increase ratios fall within the range 0.25 to 4.0. Nolan et al. (2003) report the results of a study comparing the immunogenicity of an intranasal influenza vaccine blended and filled at two different manufacturing facilities. Equivalent immunogenicity was pre-specified as the 90% confidence interval for the GMR falling within the range 0.25 to 4.0. Neither publication justifies this range, but many readers will consider it to be very broad. Here, a somewhat smaller range will be proposed: 0.35 to 2.83.

The equivalence range 0.25 to 4.0 for the GMR corresponds to an equivalence range of $\log_2(0.25) = -2.0$ to $\log_2(4.0) = +2.0$ for the log-transformed HI titers.

As explained above, it is standard procedure to titrate blood samples twice. In the laboratory Solvay Pharmaceuticals uses, the sample is re-analyzed if two intra-individual HI titers differ by two or more dilution steps (e.g. 320 and 1280). However, if the two HI titers differ by only one step (e.g. 40 and 80) they are considered to be identical. The justification for this protocol is that a difference of one dilution can be regarded as being due to random errors and an inherent variability in biological responses. Examples of possible causes of variations include, amongst others, the exact amount of antigen used in the test, the accuracy of dilutions, the temperature and the time allowed for the reactions to occur.

It is justifiable to allow at least the same difference between inter-individual HI titers and thus between lot means. A GMT of 640 would therefore be equivalent to a GMT of 1280 but not to a GMT of 2560. Returning to the log-transformed HI titers this would mean that it is allowed that $1 < |\Delta_{ij}| < 2$. There are of course an infinite number of choices which satisfy this criterion, but given the discrete nature of the observations $d = 1.5$ may be the preferred choice. This margin corresponds to the equivalence range of $2^{-1.5} = 0.35$ to $2^{+1.5} = 2.83$ for the GMR or the MFI ratio.

RESULTS

Table 1 gives the GMTs of the post-vaccination HI titers, by lot and per strain. GMTs were highest for the B strain and, with one exception, lowest for the A-H_1N_1 strain. Note that the differences between the lots per strain were small.

In Table 2, the arithmetic means and standard deviations of the log-transformed HI titers are shown. For the A-H_1N_1 strain, the observed differences are: D_{12} = +0.11, D_{13} = -0.01 and D_{23} = -0.12. The standard deviations of the lot means are: sd_1 = 1.69 and sd_2 = sd_3 = 1.65. Hence, the standard errors of the differences are

$$s_{12} = \sqrt{[1.69^2/123 + 1.65^2/123]} = 0.21$$

$$s_{13} = \sqrt{[1.69^2/123 + 1.65^2/117]} = 0.22$$

$$s_{23} = \sqrt{[1.65^2/123 + 1.65^2/117]} = 0.21$$

This gives (where $\delta = 1.5$)

$$Z_{min} = \min \{(1.5 - 0.11)/0.21, (1.5 - 0.01)/0.22, (1.5 - 0.12)/0.21\} = 6.57$$

Even if the conservative critical value based on normal distribution is used, $z_{0.025} = 1.96$, the null hypothesis that the three lots are not equivalent can be rejected for the A-H_1N_1 strain. The reader is invited to check that Z_{min} is 5.90 and 8.88 for the A-H_3N_2 and the B strains respectively.

Thus, the null hypothesis that the three lots are not equivalent can also be rejected for these two strains. Because the null hypotheses have been rejected for all three strains, equivalence of the three production lots of Invivac® has been demonstrated.

A critical value based on normal distribution was used in the above. As outlined, less conservative critical values c_α are available. The c_α depend on two parameters: δ /se and ρ. The first parameter is defined as

$$\delta/se = \delta / v\,[2 \min sd_i^2/n_i]$$

The second parameter is defined as

$$\rho = \Delta_{12}/\Delta_{13}.$$

Because this ratio is unknown, Wiens and Iglewicz (1999) suggest setting ρ to ½. For the A-H_1N_1 strain, $\delta/se = 7.13$. For $\delta/se > 5$, $c_{0.025} \cong 1.96$, so there is no advantage here. However, if δ/se had been equal to, say, 2.75, then $c_{0.025} = 1.71$, which would have meant a considerable gain in statistical power.

As a comparison, the arithmetic means and standard deviations of the log-transformed fold increases are shown in Table 3. For the A-H_1N_1 strain, the observed differences are: D_{12} = -

0.12, D_{13} = -0.03 and D_{23} = +0.09. The standard deviations of the lot means are: sd_1 = 1.82, sd_2 = 1.69 and sd_3 = 1.89. The observed differences are of the same magnitude as those found for the log-transformed HI titers. However, as expected, the standard deviations are larger. This confirms that for the A-H_1N_1 strain an analysis based on fold increases would have been less powerful (i.e. smaller, less significant value for Z_{min}) than that based on HI titers. The same applies to the two other strains.

DISCUSSION AND CONCLUSION

This paper was motivated by a concern about the overall type I error rate of influenza vaccine lot consistency studies, i.e. the probability of wrongly concluding that the lots are similar. One of the first principles of statistics is that this error rate must be controlled. In clinical studies with two treatments and a single outcome, the error rate is controlled because it is equal to the significance level α at which the null hypothesis is tested. However, if more than one statistical test is carried out the probability of making at least one test-wise type I error is greater than α. The method developed by Wiens and Iglewicz solves this problem by having one statistical test for three comparisons. However, because influenza vaccines contain HA of three viral strains, the test statistic Z_{min} has to be evaluated three times. If equivalence of the lots were claimed per strain, this would imply an overall type I error rate greater than the test-wise significance level α, and an adjustment would be needed - the Bonferroni-Holm correction for multiple comparisons, for example. However, since it is proposed here that lot consistency can only be claimed if Z_{min} exceeds the critical value for all three strains, the overall type I error rate is at most α, and thus controlled.

A popular method to demonstrate the equivalence of influenza vaccine lots is to show that two-sided intervals for the geometric mean ratios GMT_1/GMT_2, GMT_1/GMT_3 and GMT_2/GMT_3 fall within a predefined range (Zangwill et al., 2001; Nolan et al., 2003). It is easy to see that if the critical values z_α are used to test the significance of Z_{min}, the Wiens and Iglewicz method

corresponds to checking that all three two-sided 100(1-2α)% confidence intervals for the GMRs lie within the equivalence range. The reader might conclude from this that the two methods are identical. This is only partly true. If, instead of $z_{\alpha,}$ the critical values c_α are used, the two methods will yield different results. In that case the Wiens and Iglewicz method is the more powerful of the two.

Lachenbruch et al. (2004) have proposed an alternative to the Wiens and Iglewicz method, which they criticize for two reasons. Firstly, the normality assumption may not hold. Secondly, equivalence is defined in terms of group means, but equivalence with respect to group variances may be important as well. The following reply may be made to these criticisms. Firstly, if log-transformed HI titers are not normally distributed, the most likely reason is that the observations in the right tail of the distribution are truncated because HI titers > 2560 are not determined. In that case, unbiased estimates of the mean and the standard deviation of the underlying normal distribution can be obtained by the method of maximum likelihood estimation for censored observations, and the Wiens and Iglewicz method can still be applied. Secondly, one hypothesis put forward here is that if log-transformed HI titers are normally distributed and group means are similar, the standard deviations will also be similar, as suggested by Table 2. The explanation would be that the range of possible values is limited. Lachenbruch et al. propose a nonparametric alternative to the test statistic Z_{min}. The drawback of their test statistic G is that the critical values have to be simulated. However, if the normality assumption does not hold it is not clear how this should be done, especially if sample sizes are unequal.

Another controversial issue in equivalence trials is the choice of the significance level α. In the case of a superiority trial in which an active treatment is compared with a placebo, most regulatory agencies will require a two-sided significance level of $\alpha = 0.05$. Given that in the case of a drug application the only outcome of interest is where the active treatment is significantly better than the placebo, the risk for the regulatory agency is 0.025. For this reason, it is sometimes argued that α should be set to 0.025 for equivalence trials for the sake of consistency, because of the one-sided nature of the null hypothesis being tested. Thus, the choice of significance level a should be carefully considered when the study protocol is written.

As argued above, it is not advisable to use fold increases to demonstrate lot consistency. The frequent assertion that fold increases eliminate the bias caused by baseline imbalance between lots is not true, and – with some exceptions – fold increases show a greater variability than HI titers. As a consequence, the analysis using the HI titers will be more powerful than that using fold increases if the same equivalence margin is used.

Other serological efficacy measures can be derived from HI titers. Standard efficacy measures are the mean fold increase discussed above, the seroprotection rate and the seroconversion rate (CPMP, Note for Guidance, 1997). The seroprotection rate is defined as the proportion of subjects with an HI titer = 40. The seroconversion rate is defined as the proportion of subjects with either a pre-vaccination HI titer < 10 and a post-vaccination HI titer = 40 or a pre-vaccination HI titer = 10 and post-vaccination increase in HI titer of at least four-fold. The question

is thus whether these might not be equally suitable or even better parameters for evaluating lot consistency. There are at least two arguments against using the seroprotection rate to evaluate the consistency of influenza vaccine lots. Firstly, the seroprotection rate is a surrogate measure for protection, because HI titers = 40 are considered to be protective (De Jong et al., 2003). The difference between the geometric mean titer and the seroprotection rate is that the geometric mean titer expresses the intrinsic immunogenicity of the vaccine, while the seroprotection rate is more a translation from the intrinsic immunogenicity into a marker of protection (Beyer et al., 2004). There could be considerable debate on which is more important: immunogenicity or protection. One could argue that in case of lot consistency it is the intrinsic immunogenicity that matters. Furthermore, the seroprotection rate is a proportion, and statistical analyses involving proportions are (usually) less powerful than those involving normal data such as log-transformed HI titers. In other words, much larger sample sizes would be required to demonstrate lot consistency. Secondly, it would probably be very difficult to justify any equivalence margin for the difference in seroprotection rates. For non-inferiority studies, the CPMP Note for Guidance on Clinical Evaluation of New Vaccines (1999) advises a margin of 10 percentage points, or even less in the case of very high seroprotection rates. Returning to the model data, the seroprotection rates for the A-H_1N_1 strain were: 90.2% (lot 1), 93.5% (lot 2) and 90.6% (lot 3). A two-sided 95% confidence interval for the seroprotection rate dfference (Armitage and Berry, 1994) of lot 1 minus lot 2 is (-10.1%, +3.6%). Hence, with d = 10%, at the a = 0.025 significance level equivalence could not be concluded, while the analysis based on the GMT gave overwhelming

evidence in favour of equivalence. Because the seroconversion rate is also a proportion, two of the three objections to the seroprotection rate also apply to this efficacy measure.

In conclusion, of the four serological efficacy measures, the GMT is the optimal one to use to demonstrate lot consistency.

ACKNOWLEDGMENT

The author wishes to thank Jeroen Tolboom of the Biometrics department of Solvay Pharmaceuticals, Weesp, the Netherlands, for his valuable comments and suggestions, as well as two anonymous referees.

REFERENCES

Armitage, P., Berry, G. (1994). *Statistical Methods in Medical Research, Third Edition*. Blackwell Science. 128-30.

Berger, R.L. (1982). Multiparameter hypothesis testing and acceptance sampling. T*echnometrics* 24:295-300.

Berger, R.L., Hsu J.C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Sciences* 11:283-319.

Beyer, W.E.P., Palache, A.M., Lüchters, G., Nauta, J., Osterhaus, A.D.M.E. (2004). Seroprotection rate, mean fold increase, seroconversion rate: which parameter adequately expresses seroresponse to influenza vaccination? *Virus Res.* 103:125-132.

De Jong, J.C., Palache, A.M., Beyer, W.E.P., Rimmelzman, G.F., Boon, A.C.M., Osterhaus, A.D.M.E. (2003). Haemagglutination-inhibiting antibody to influenza virus. *Dev. Biol.* 115:63-73.

Dowdle, W.E., Kendal, A., Noble, W.R. Influenza Viruses. In *Diagnostic Procedures for Viral, Rickettsial and Chlamydial Infections 5th edition* Lennette, E.H., Schmidt, N.J., Eds., Washington DC: American Public Health Association, 1979.

Guidance for Industry for the Evaluation of Combination Vaccines for Preventable Diseases: Production, Testing, and Clinical Studies. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 1997.

Lachenbruch, P.A., Rida, W., Kou, J. (2004). Lot consistency as an equivalence problem. *J. Biopharm. Stat.* 14:275-290.

Nolan, T., Lee, M.S., Cordova, J.M., Cho, I., Walker, R.E., August, M.J., Larson, S., Coelingh, K.L., Mendelman, P.M. (2003). Safety and immunogenicity of a live-attenuated influenza vaccine blended and filled at two manufacturing facilities. *Vaccine* 21:1224-1231.

Note for Guidance on Harmonisation of Requirements for Influenza Vaccines. Committee for Proprietary Medicinal Products (CPMP). CPMP/BWP/214/96, March 1997.

Note for Guidance on Clinical Evaluation of New Vaccines. Committee for Proprietary Medicinal Products (CPMP). CPMP/EWP/463/97, May 1999.

Schuirmann, D. (1987). A comparison of the two one-sided tests procedure and the power for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* 15:657-80.

Statistical Principles for Clinical Trials. International Conference on Harmonization (ICH) E9 Expert Working Group, 1998.

Wiens, B.L., Iglewicz, B. (1999). On testing equivalence of three populations. *J. Biopharm. Stat.* 9:465-483.

Wiens, B.L., Iglewicz, B. (2000). Design and analysis of three treatment equivalence trials. *Control. Clin. Trials* 21:127-137.

Wiens, B.L. (2002). Choosing an equivalence limit for noninferiority or equivalence studies. *Control. Clin. Trials* 23:2-14.

Zangwill, K.M., Droge, J., Mendelman, P., Marcy, M., Partridge, S., Chiu, C.Y., Jing, J., Chang, S.J., Cho, I., Ward, J.I. (2001). Prospective, randomized, placebo-controlled evaluation of the safety and immunogenicity of three lots of intranasal trivalent influenza vaccine among young children. *J. Pediatr. Infect. Dis.* 20: 740-746.

Table 1. Post-vaccination Geometric Mean Titers

Viral strain type	Lot 1 (n = 123)	Lot 2 (n = 123)	Lot 3 (n = 117)
A-H_1N_1	151.4	163.4	150.3
A-H_3N_2	192.9	162.2	202.5
B	352.6	365.0	372.7

Table 2. Arithmetic means and standard deviations of the log-transformed post-vaccination HI titers

Viral strain type	Lot 1 (n = 123)	Lot 2 (n = 123)	Lot 3 (n = 117)
A-H_1N_1	4.92 (1.69)	5.03 (1.65)	4.91 (1.65)
A-H_3N_2	5.27 (1.57)	5.02 (1.60)	5.34 (1.57)
B	6.14 (1.20)	6.19 (1.21)	6.22 (1.28)

Table 3. Arithmetic means and standard deviations of the log-transformed fold increases

Viral strain type	Lot 1 (n = 123)	Lot 2 (n = 123)	Lot 3 (n = 117)
A-H_1N_1	4.51 (1.82)	4.39 (1.69)	4.48 (1.89)
A-H_3N_2	3.73 (2.04)	3.64 (1.98)	3.65 (2.21)
B	3.78 (2.27)	3.46 (2.38)	3.59 (2.38)