Design and analysis of bridging studies with prior probabilities on the null and alternative hypotheses

Donglin Zeng[1] | **Zhiying Pan[2]** | **D. Y. Lin[1]**

[1]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina

[2]Amgen Inc, 1 Amgen Center Dr, Thousand Oaks, California

Correspondence
Donglin Zeng, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420.
Email: dzeng@email.unc.edu

Funding information
National Institute of General Medical Sciences, Grant/Award Numbers: R01 GM124104, R01 GM047845

Abstract

The pharmaceutical industry and regulatory agencies are increasingly interested in conducting bridging studies in order to bring an approved drug product from the original region (eg, United States or European Union) to a new region (eg, Asian-Pacific countries). In this article, we provide a new methodology for the design and analysis of bridging studies by assuming prior knowledge on how the null and alternative hypotheses in the original, foreign study are related to the null and alternative hypotheses in the bridging study and setting the type I error for the bridging study according to the strength of the foreign-study evidence. The new methodology accounts for randomness in the foreign-study evidence and controls the average type I error of the bridging study over all possibilities of the foreign-study evidence. In addition, the new methodology increases statistical power, when compared to approaches that do not use foreign-study evidence, and it allows for the possibility of not conducting the bridging study when the foreign-study evidence is unfavorable. Finally, we conducted extensive simulation studies to demonstrate the usefulness of the proposed methodology.

KEYWORDS

adaptive significance level, bridging studies, ethnic factors, foreign studies, optimal designs, sample size calculations

1 | INTRODUCTION

In an effort to expedite the availability of drug products to different populations, regulatory authorities and pharmaceutical companies are increasingly interested in bringing an approved drug product from the original region (eg, United States or European Union) to a new region (eg, Asian-Pacific countries) by utilizing the original clinical trial data to meet the regulatory standards and clinical trial practices of the new region. In 1998, the International Conference on Harmonization (ICH) issued a guideline, entitled "Ethnic Factors in the Acceptability of Foreign Clinical Data" (known as ICH E5), to provide guidance on regulatory and development strategies that will permit adequate evaluation of the influence of ethnic factors on efficacy, safety, dosage, and dose regimen, while minimizing duplication of clinical studies and expeditiously supplying medicines to patients. The guideline describes the use of "bridging studies" (when necessary) to extrapolate efficacy and/or safety data from the original, foreign region to a new region. A bridging study is defined as a supplementary study conducted in the new region to provide pharmacokinetic, pharmacodynamic, and/or clinical data on efficacy, safety, dosage, and dose regimen to enable extrapolation of clinical trial data from the foreign region to the new region. Although ethnic differences among populations may cause variability in a medicine's safety, efficacy, dosage, or dose regimen, many medicines have comparable pharmacokinetic and pharmacodynamic characteristics and clinical effects across regions (International Conference on Harmonization, 1998). Therefore, one can borrow evidence from clinical data of the original region and incorporate it into the design and analysis of the bridging study.

Several methods have been developed for the design and analysis of bridging studies; see Chow *et al.* (2012) for a review. Specifically, Shao and Chow (2002) proposed a sensitivity index to assess the reproducibility probability, and Shih (2001) suggested a sequential procedure to test the consistency between the foreign and bridging studies. To design a bridging study, Liu *et al.* (2002) and Hsiao *et al.* (2007) proposed Bayesian methods by assuming normal or mixture-normal priors for the drug effects based on the foreign study, such that the posterior distribution for the drug effect can be derived after combining data from the two studies. These Bayesian methods require strong distributional assumptions on both the data and priors. Alternatively, Lan *et al.* (2005), Lan and Pinheiro (2012), and Huang *et al.* (2012) proposed a weighted statistic that is a weighted sum of the Z-statistics from the foreign and bridging studies. The type I error or power for the bridging study is obtained by conditioning on the evidence from the foreign study and the prespecified weights. Although the statistics from the two studies are usually normalized, a weighted sum of the test statistics is not biologically interpretable. When the two statistics have opposite directions, the weighted sum may lead to lower power than using evidence from the bridging study only.

In this article, we propose a new framework for incorporating evidence from the foreign study into the design and analysis of the bridging study. In order to leverage the evidence from the foreign study, we assume certain prior probabilities on the relationship between the null and alternative hypotheses in the two studies. We use an adaptive significance level to design the bridging study based on the strength of the evidence in the foreign study. By accounting for the randomness of the foreign-study evidence and the uncertainty of the foreign-study truth, the proposed adaptive significance level for the bridging study guarantees that the overall type I error of the bridging study (ie, averaged over all possible scenarios of the foreign-study evidence) will be controlled at a prespecified level. In addition, we devise a grid-search algorithm to determine the optimal choice of the adaptive significance level that ensures higher power than the bridging study without using evidence from the foreign study.

2 | METHODS

We use subscript 1 to denote the foreign study and subscript 2 to denote the bridging study. For $k = 1$ and 2, let Δ_k denote the parameter of interest that quantifies the difference between the test and control groups (eg, the difference or the log-ratio of the response rates) in the kth study. We wish to test the null hypothesis

$$H_{k0} : \Delta_k \notin (L_k, U_k)$$

against the alternative hypothesis

$$H_{ka} : \Delta_k \in (L_k, U_k),$$

where L_k and U_k are specific margins. The constants L_k and U_k are chosen to yield different types of tests: the equivalence test corresponds to finite L_k and U_k; superiority test corresponds to $L_k = 0$ and $U_k = \infty$; and the noninferiority test corresponds to $L_k < 0$ and $U_k = \infty$.

2.1 | Priors for hypotheses

The main reason that one can incorporate evidence from the foreign study into the design or analysis of the bridging study for certain drugs is because the drug's safety, efficacy, dosage, and dose regimen are similar across regions. Intuitively, if the null (or alternative) hypothesis holds for the foreign study, then there is a high likelihood that the null (or correspondingly, alternative) hypothesis holds for the bridging study and vice versa. Thus, we wish to weigh the strength of the foreign-study evidence when deciding whether or not to reject H_{20}. Specifically, we consider H_{10} and H_{1a} as random variables whose distributions depend on whether H_{20} or H_{2a} holds, and we impose the following assumptions a priori:

Assumption 1. There exists some constant $p \in [0.5, 1]$ such that

$$Pr(H_{10} \mid H_{20}) = p.$$

Assumption 2. There exists some constant $q \in [0, 1]$ such that

$$Pr(H_{1a} \mid H_{2a}) = q.$$

In the above assumptions, p and q characterize the relationship between the two studies. Consider, for example, that $p = q = 90\%$. In this case, Assumptions 1 and 2 imply that if the null hypothesis holds for the bridging study, then there is a 90% chance that the null hypothesis also holds for the foreign study and that if the alternative hypothesis holds for the bridging study, then there is a 90% chance that the alternative hypothesis also holds for the foreign study. Thus, p and q reflect how confident we are in borrowing the evidence from the foreign study that favors the null or alternative hypothesis. Additionally, p and q can be understood as implicit weights for the foreign-study evidence that support the truth in the bridging study. Note that both p and q must be prespecified without looking at the evidence from the foreign study. The choice of p and q, although subjective, can be obtained from knowledge of the product's pharmacokinetic and pharmacodynamic properties and translation of those properties to clinical effectiveness and safety, or clinical experience with other members of the drug class in the new region. It is implicitly assumed that the data from the bridging study are independent of H_{10} and H_{1a}

conditional on H_{20} or H_{2a}. This is a reasonable assumption in practice.

Remark 1. Since we are using the foreign study evidence to design and analyze the bridging study, it would seem more natural to consider $Pr(H_{20} \mid H_{10})$ and $Pr(H_{2a} \mid H_{1a})$ rather than $Pr(H_{10} \mid H_{20})$ and $Pr(H_{1a} \mid H_{2a})$. By Bayes' theorem; however, we can convert $Pr(H_{10} \mid H_{20})$ and $Pr(H_{1a} \mid H_{2a})$ to $Pr(H_{20} \mid H_{10})$ and $Pr(H_{2a} \mid H_{1a})$ and vice versa. Because we wish to control the type I error and improve the power for the bridging study, we are primarily interested in the rejection probabilities conditional on the null or alternative hypothesis in the bridging study. Specifically, we need to evaluate $Pr(H_{10} \mid H_{20})$ and $Pr(H_{1a} \mid H_{2a})$ rather than $Pr(H_{20} \mid H_{10})$ and $Pr(H_{2a} \mid H_{1a})$. Thus, we state Assumptions 1 and 2 in terms of $Pr(H_{10} \mid H_{20})$ and $Pr(H_{1a} \mid H_{2a})$. If we instead formulate Assumptions 1 and 2 in terms of $Pr(H_{20} \mid H_{10})$ and $Pr(H_{2a} \mid H_{1a})$, then we will need to specify the prior probabilities for the null and alternative hypotheses of the bridging study in order to obtain $Pr(H_{10} \mid H_{20})$ and $Pr(H_{1a} \mid H_{2a})$.

Remark 2. There is a close connection between (p, q) and the joint prior distribution for (Δ_1, Δ_2). In Appendix A, we show this mathematical relationship by assuming a bivariate-normal prior distribution. The proposed framework is based on p and q instead of the joint prior distribution of (Δ_1, Δ_2) for two major reasons. First, it is desirable not to specify the prior distribution function that involves a number of hyper-parameters. Second, the assumptions on p and q pertain only to the hypotheses and thus allow the parameters in the two studies to be different quantities, for which the joint prior distribution would be difficult to specify.

2.2 | Rejection region incorporating the foreign-study evidence

Let α denote the overall type I error to be controlled in the bridging study. Our approach is to adjust the type I error for the bridging study according to the empirical evidence from the foreign study, such that the overall type I error for the bridging study remains under α after accounting for the randomness of the foreign-study data. Specifically, Assumption 1 implies that when the null hypothesis is likely to hold for the bridging study, the null hypothesis likely holds for the foreign study. Thus, the evidence favoring the alternative hypothesis in the foreign study should also be considered as the evidence to reject the null hypothesis in the bridging study. By contrast, Assumption 2 implies that when the alternative hypothesis is likely to hold in the bridging study, the alternative hypothesis is also likely to be true for the foreign study. Thus, the evidence against the alternative hypothesis in the foreign study can also be treated as the evidence against the alternative hypothesis in the bridging study. In either way, we can use

the evidence from the foreign study to improve the decision making in the bridging study.

To formalize the above strategy, we introduce additional notation. For any constant $\eta \in [0, 0.5)$, let $\text{CI}_k(\eta) \equiv (\widehat{L}_k(\eta), \widehat{U}_k(\eta))$ denote the $(1 - 2\eta)$-confidence interval for Δ_k based on the data in the kth study when testing for equivalence or the one-sided confidence interval with level $(1 - \eta)$ when testing for superiority or inferiority. Thus, at significance level η, we would reject the null hypothesis in the kth study if this confidence interval lies within (L_k, U_k).

To characterize the strength of the foreign-study evidence, we prespecify a constant $\omega \in (0, 1]$ and construct an ω-level rejection region for the foreign study, such that the foreign study favors the alternative hypothesis H_{1a} if $\text{CI}_1(\omega) \subset (L_1, U_1)$ and favors the null hypothesis H_{10} otherwise. This ω is the significance level for borrowing the foreign-study evidence and has nothing to do with the significance level actually used in the foreign study for rejecting H_{10}. In some sense, ω represents our level of confidence in whether the foreign study favors the null or alternative hypothesis. If the evidence from the foreign study favors the alternative hypothesis, then we conduct the bridging study and set the adaptive type I error for the bridging study to

$$\widehat{\alpha}_2 = \alpha\gamma,$$

where $\gamma \geq 1$ is an adjustment factor. If the foreign-study evidence favors the null hypothesis, then, with a prespecified probability τ, we abandon the bridging study and, with the remaining $(1 - \tau)$ probability, we conduct the bridging study but set the adaptive type I error to

$$\widehat{\alpha}_2 = \alpha/\gamma.$$

In the above procedure, the null hypothesis in the bridging study (H_{20}) is rejected if and only if the bridging study is conducted and its null hypothesis is rejected at significance level $\widehat{\alpha}_2$. Thus, the proposed rejection region for the null hypothesis in the bridging study is

$$\mathcal{R}(\omega, \gamma) \equiv \{\text{the bridging study is conducted and } \text{CI}_2(\widehat{\alpha}_2)$$

$$\subset (L_2, U_2)\}.$$

2.3 | Choice of γ to preserve the type I error for a given ω

The constant ω describes the strength of the evidence in the foreign study, and the constant γ is the adjustment factor based on the strength of the foreign-study evidence. These two constants cannot be chosen independently; otherwise, the overall type I error over all possibilities of the foreign-study evidence (ie, the chance of rejecting the null hypothesis in the bridging

study when the null hypothesis is the truth) may not be appropriately controlled. In the following theorem, we provide an explicit form of γ as a function of ω, which may be subject to a prespecified upper bound, such that the type I error, that is, $\Pr(\mathcal{R}(\omega, \gamma) \mid H_{20})$, is controlled at α.

Theorem 1. *For any $\omega \in (0, 1]$, let*

$$\gamma = \begin{cases} \min\left(\gamma_{\max}, \dfrac{p(1-\omega)}{1 - p(1-\omega)}\right) & \text{if } 0 < \omega \leq 1 - \dfrac{1}{2p}, \\ 1 & \text{if } 1 - \dfrac{1}{2p} < \omega \leq 1, \end{cases} \tag{1}$$

where $\gamma_{\max}$ is a prespecified upper bound (larger than 1) for the type I error adjustment. Then under Assumption 1, $\Pr(\mathcal{R}(\omega, \gamma) \mid H_{20}) \leq \alpha$.

The proof of Theorem 1 is provided in Appendix B. According to Theorem 1, if ω is chosen to be close to 1, that is, our constructed evidence from the foreign study always favors the alternative hypothesis, then since the null hypothesis is true in the bridging study and thus also likely to hold in the foreign study, the constructed evidence tends to be of little use. Therefore, we should abandon the evidence from the foreign study by setting $\gamma = 1$. Note that the choice of γ given in (1) may not yield the tightest upper bound for the overall type I error. In fact, the optimal value for γ may depend on the discontinuation probability τ. However, a major advantage of formula (1) that it depends only on ω and thus is robust to the subjective choice of τ.

2.4 | Optimal choice of ω to maximize the power

Theorem 1 implies that the proposed test with a proper adjustment factor γ always preserves the overall type I error for any $\omega \in (0, 1]$. Therefore, a question naturally arises as to what value of ω is optimal in terms of yielding the highest power when the alternative hypothesis holds in the bridging study. To answer this question, it is necessary to calculate the power of the bridging study with the rejection region $\mathcal{R}(\omega, \gamma)$, where γ is given by expression (1).

Specifically, we let $Q_k(\alpha_0)$ denote the power function corresponding to the type I error α_0 in study k ($k = 1, 2$). Then the following theorem holds.

Theorem 2. *Under Assumption 2, $\Pr(\mathcal{R}(\omega, \gamma) \mid H_{2a}) = \widetilde{Q}_2(\omega)$, where*

$$\widetilde{Q}_2(\omega) = q[Q_2(\alpha\gamma)Q_1(\omega) + (1-\tau)Q_2(\alpha/\gamma)\{1 - Q_1(\omega)\}]$$

$$+ (1-q)\omega Q_2(\alpha\gamma) + (1-q)(1-\tau)(1-\omega)Q_2(\alpha/\gamma),$$

with γ given by expression (1). In addition, the optimal choice of ω that maximizes this power function, that is,

$$\omega^* = \operatorname{argmax}_{\omega \in [0,1]} \widetilde{Q}_2(\omega),$$

satisfies $\Pr(\mathcal{R}(\omega^*, \gamma^*) \mid H_{2a}) \geq Q_2(\alpha)$.

The proof of Theorem 2 is also given in Appendix B. This theorem implies that when the prior assumption on the alternative hypothesis holds, the proposed method based on the optimal choice of ω always yields higher power than any method that does not use evidence from the foreign study.

2.5 | Testing procedure for the bridging study

In light of Theorems 1 and 2, we propose the following steps to perform the hypothesis test for the bridging study with a given sample size:

1. Determine p, q, and τ, and calculate the power function $Q_k(\cdot)$ ($k = 1, 2$) based on the design parameters for the given sample size.
2. Calculate ω^* by grid search to maximize $\widetilde{Q}_2(\omega)$ and obtain the corresponding γ^*.
3. Upon completion of the foreign study, construct the confidence interval associated with ω^* using the data from the foreign study, that is, $\mathrm{CI}_1(\omega^*)$.
4. If $\mathrm{CI}_1(\omega^*)$ is contained in (L_1, U_1), conduct the bridging study and use $\widehat{\alpha}_2^* = \alpha\gamma^*$ as the significance level; otherwise, with probability τ, abandon the bridging study and, with probability $(1 - \tau)$, conduct the bridging study and set the significance level to $\widehat{\alpha}_2^* = \alpha/\gamma^*$.
5. If the bridging study is conducted and $\mathrm{CI}_2(\widehat{\alpha}_2^*)$ lies within (L_2, U_2), reject H_{20}.

The above testing procedure implies that the minimal sample size required to achieve power of $(1 - \beta)$ is the smallest sample size such that $\max_\omega \widetilde{Q}_2(\omega)$ is greater than $1 - \beta$.

3 | NUMERICAL STUDIES

To examine the performance of the proposed methods, we carried out simulation studies in the setting of establishing superiority for a test drug product over a reference drug, such that $L_k = 0$ and $U_k = \infty$ for $k = 1, 2$. In both the foreign and bridging studies, the null hypotheses pertained to 30% response rates in both products, and the alternative hypotheses pertained to response rates of 40% and 30% for the test and reference drugs, respectively. The foreign study had 400 patients on each arm, and the bridging study had 280 patients on each arm. The bridging study had 80% power without borrowing any evidence from the foreign study.
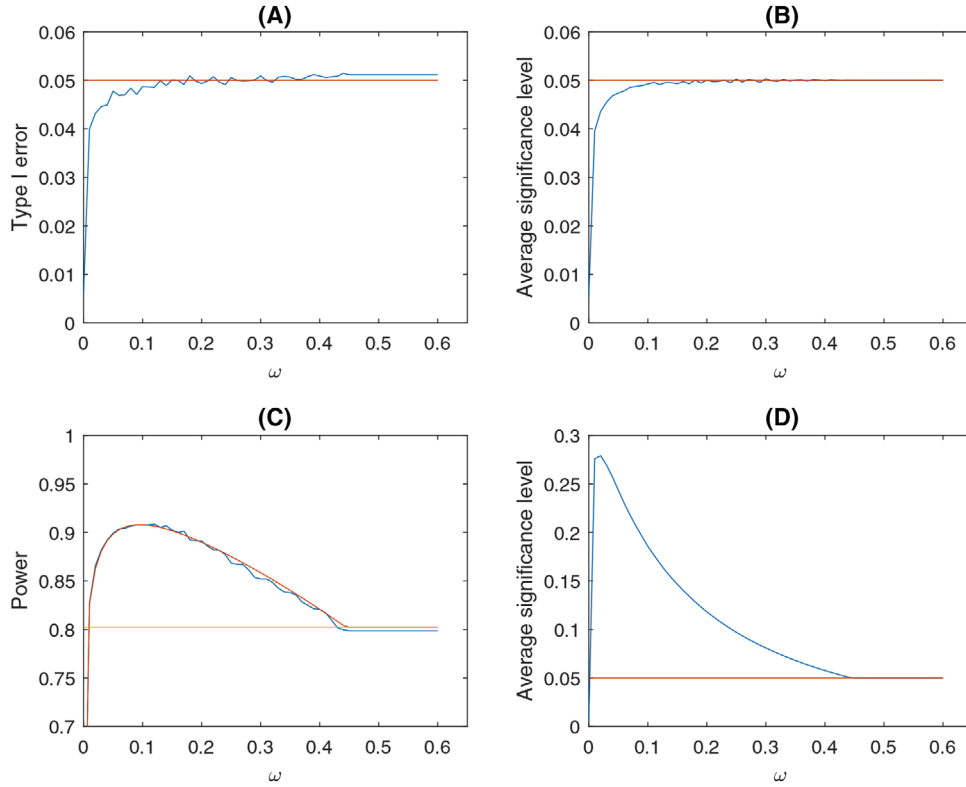
FIGURE 1 Type I error and power with varying ω: A, the overall type I error for different values of ω; B, the average significance level in the bridging study under the null hypothesis as ω varies; C, the average power for different values of ω, where the less smooth curve is the empirical power and the smooth curve is the theoretical power given in Theorem 2; and D, the average significance level in the bridging study under the alternative hypothesis as ω varies. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

3.1 | Type I error and power with varying ω

In the first simulation study, we varied the significance level used to borrow information from the foreign study (ie, ω) and set $\tau = 0$, such that the bridging study is always conducted. In addition, we set $\gamma_{max} = \infty$, so there was no restriction on the type I error adjustment in the bridging study. This simulation was designed to investigate whether the adaptive test preserves the overall type I error for any ω and whether the optimal choice of ω guarantees higher than 80% power.

We set $p = q = 0.9$ in Assumptions 1 and 2 and considered two scenarios. In the first scenario, we generated data for the bridging study under the null hypothesis. To conform with Assumption 1, we generated data for the foreign study from the null and alternative hypotheses with probabilities 0.9 and 0.1, respectively. In the second scenario, we generated data for the bridging study under the alternative hypothesis. To conform with Assumption 2, we generated data for the foreign study from the null and alternative hypotheses with probabilities 0.1 and 0.9, respectively. In each scenario, we first simulated data for the foreign study and then followed the steps in Section 2.5 to generate data for the bridging study.

We constructed the confidence intervals and power functions using the difference between the estimated response rates of the two drug groups. Given the data from the bridging study, we rejected the null hypothesis if the lower bound of the one-sided confidence interval with level $(1 - \hat{\alpha}_2)$, where $\hat{\alpha}_2$ is given in Section 2.2, was above 0. We simulated 50 000 replicates to estimate the overall type I error and 10 000 replicates to estimate the power.

The simulation results are displayed in Figure 1. Clearly, the type I error is well preserved, and the power is greater than 80% for most values of ω. As shown in panel (C), the optimal ω is close to 0.1 and yields power as large as 90%. In addition, panels (B) and (D) show that the adaptive significance level used for the bridging study is below 5% if the null hypothesis holds and can be as high as 30% if the alternative hypothesis holds. We also observe that when ω is larger than 0.45, there is no power gain because the evidence from the foreign study is most likely not incorporated ($\gamma = 1$). This is consistent with the comment made at the end of Section 2.3. Thus, this simulation study showed that, by leveraging evidence from the foreign study, the proposed method with any value of ω maintains the overall type I error and can substantially improve the power when ω is chosen to be close to 0.1.
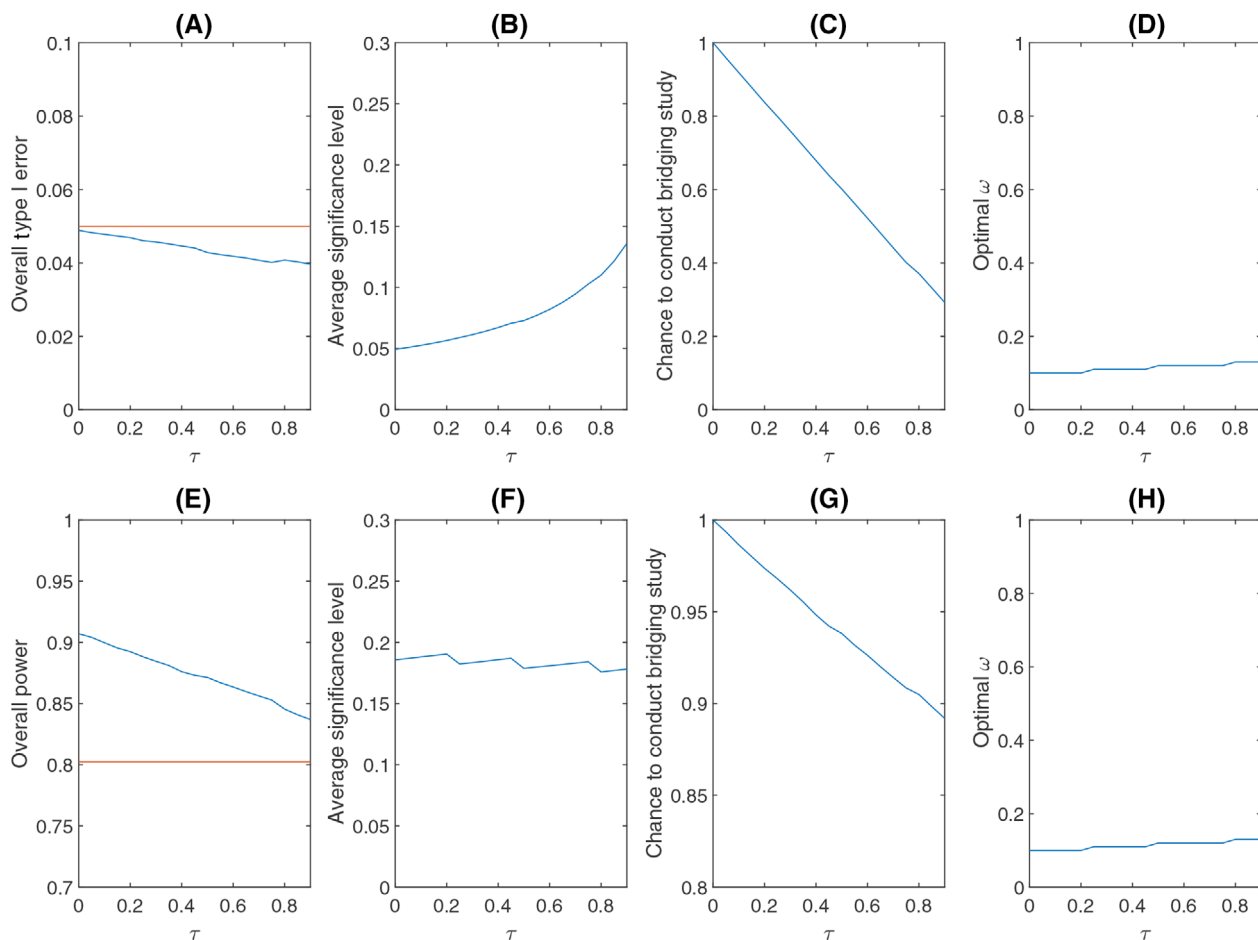
FIGURE 2 Type I error and power with optimal ω^* under $(p, q) = (0.9, 0.9)$: A, the overall type I error for different values of τ; b, the average significance level in the bridging study under the null hypothesis as τ varies; C, the chance of conducting the bridging study under the null hypothesis; D, the optimal value for ω for different values of τ under the null; E, the average power for different values of τ; F, the average significance level in the bridging study under the alternative hypothesis as τ varies; G, the chance of conducting the bridging study under the alternative hypothesis; H, the optimal value for ω for different values of τ under the alternative. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

3.2 | Type I error and power with optimal ω^*

In the second simulation study, we varied the discontinuation probability τ while fixing the significance level for the foreign study at the optimal value ω^*. We set (p, q) to $(0.9, 0.9)$, $(0.8, 0.9)$, or $(0.9, 0.8)$ and set $\gamma_{\max} = \infty$. This study was designed to investigate how the values of τ, p, and q affect the performance of the proposed methods. For each (p, q), we simulated the data in the same manner as before except that we varied τ from 0 to 0.9. Given the data from the bridging study, we rejected the null hypothesis if the study was conducted and the lower bound of the one-side confidence interval with level $(1 - \hat{\alpha}_2)$, where $\hat{\alpha}_2$ is given in Section 2.2 with $\omega = \omega^*$, was above 0. We simulated 50 000 replicates for estimating the overall type I error and 10 000 replicates for estimating the power.

The simulation results are displayed in Figures 2-4. Evidently, the type I error is always kept below 0.05 regardless

of the value of τ, and the power is at least 80%. When the alternative hypothesis holds in the bridging study, under which the alternative hypothesis is likely to hold in the foreign study, the probability of conducting the bridging study is at least 0.85, and the average significance level for the bridging study is approximately 0.15 to 0.20. As q becomes smaller, the power gain is reduced, and the chance of discontinuation is increased. We also observe that the optimal value for ω, the confidence measure of using the foreign study evidence, is mostly between 0.1 and 0.2 except for the scenarios of very high chance of discontinuation. The sharp changes when τ is larger than 0.6 in Figure 4 ($q = 0.8$) are due to the thresholds for the type I error (below 0.05) and power (above 0.8). In other words, when anticipating that there is a high chance of not conducting the bridging study and without strong confidence in borrowing the foreign study to gain power, we would rather not use the foreign study evidence at all.
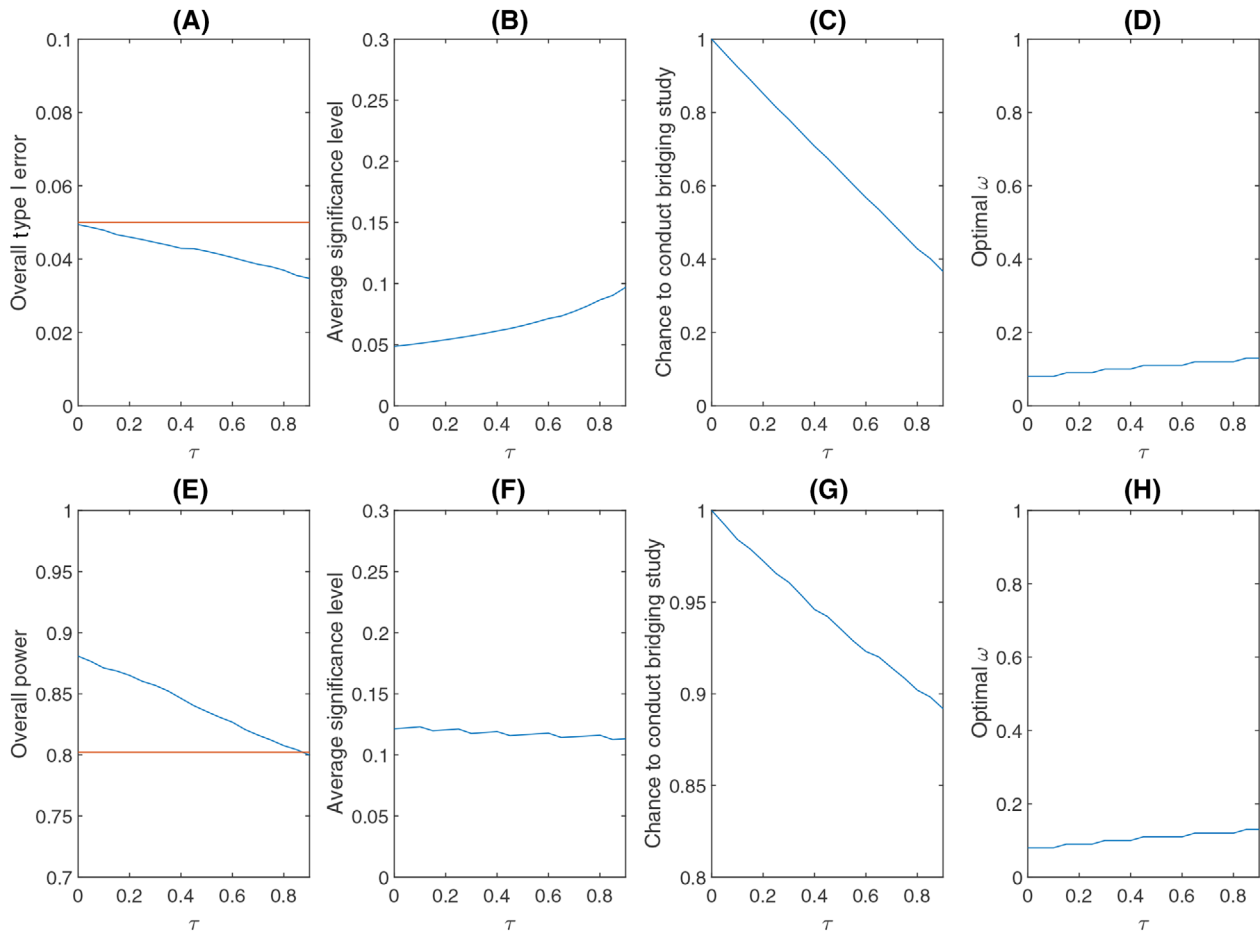
FIGURE 3 Type I error and power with optimal ω^* under $(p, q) = (0.8, 0.9)$: A, the overall type I error for different values of τ; B, the average significance level in the bridging study under the null hypothesis as τ varies; C, the chance of conducting the bridging study under the null hypothesis; D, the optimal value for ω for different values of τ under the null; E, the average power for different values of τ; F, the average significance level in the bridging study under the alternative hypothesis as τ varies; G, the chance of conducting the bridging study under the alternative hypothesis; H, the optimal value for ω for different values of τ under the alternative. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

We repeated the above simulation study but restricted the adjustment factor to be less than 3 (ie, $\gamma_{max} = 3$ in Theorem 2) when computing the optimal type I error for the bridging study. The results are shown in Figure S2 of the Supporting Information. Clearly, the proposed method continues to preserve the type I error. With the restriction, the type I error for the bridging study is always below 0.05 and the power is still larger than 0.8. There is a power loss of about 2% compared to the case of no restriction.

We conducted additional simulation studies with other choices of p and q, and the results are displayed in Figures S3 and S4 of the Supporting Information. When the prior probability for p or q is low, there is very small power gain when $\tau = 0$; when there is a chance of not performing the bridging study, we hardly use any evidence from the foreign study.

To assess the robustness of the proposed method to misspecification of p and q, we set the true values of p and q to 0.8 and 0.9, respectively, but assumed p to be 0.9 or 0.7 or q to

be 0.8 when implementing the proposed method. The results are displayed in Figures S5-S8 of the Supporting Information. When p is misspecified to be too large, the type I error can be as high as 0.065, but the inflation decreases as τ increases. When p is misspecified to be too small, the type I error is deflated, and the power is reduced as compared to using the correct p. When q is misspecified, the type I error is maintained; however, the power gain is not as substantial as using the correct q, and the power gain decreases when p is also misspecified.

3.3 | Sample size calculations

To quantify the efficiency gain of the proposed methodology, we performed a numerical study to examine the sample sizes for different choices of (p, q), as compared to the sample size without using any information from the foreign
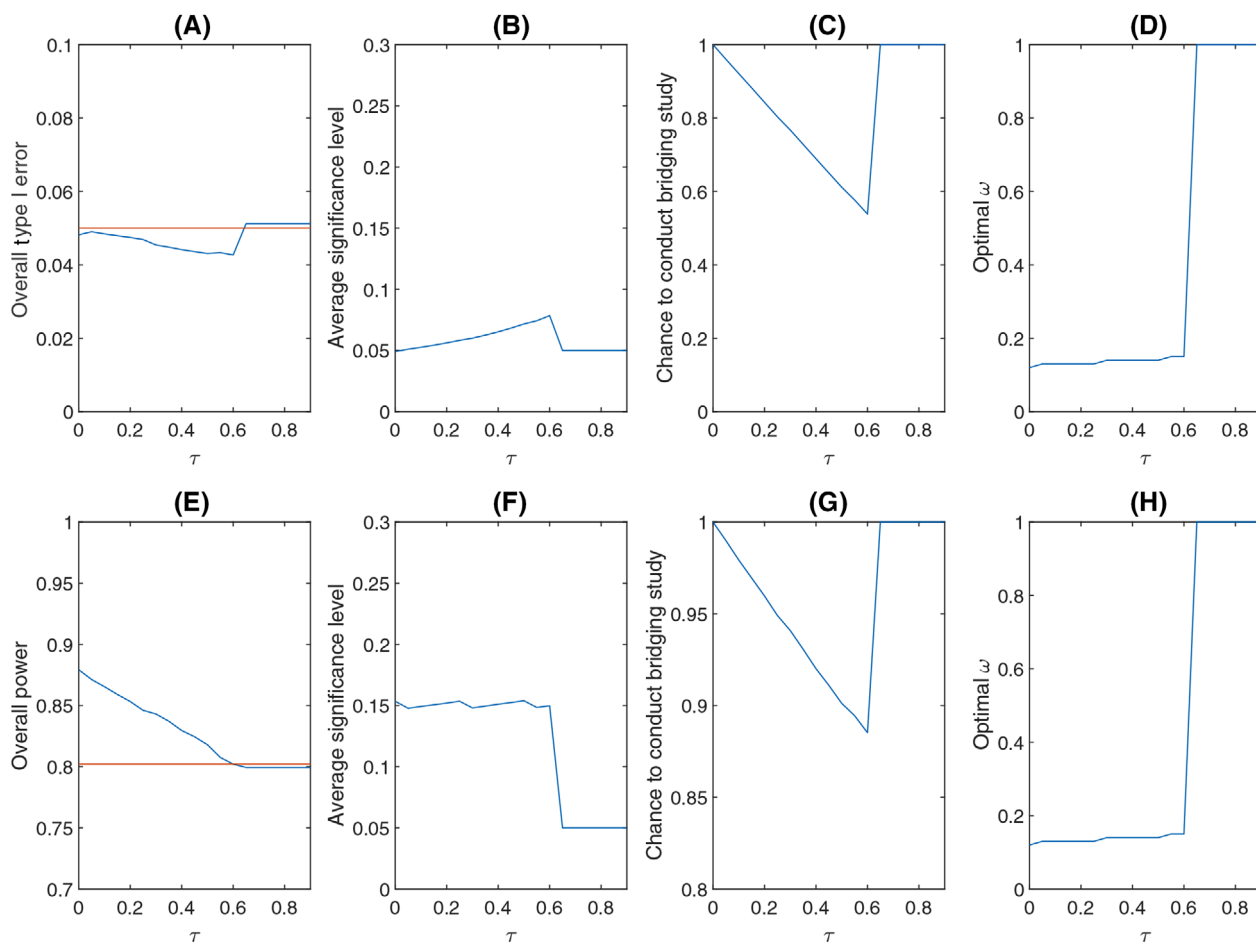
FIGURE 4 Type I error and power with optimal ω^* under $(p, q) = (0.9, 0.8)$: A, the overall type I error for different values of τ; B, the average significance level in the bridging study under the null hypothesis as τ varies; C, the chance of conducting the bridging study under the null hypothesis; D, the optimal value for ω for different values of τ under the null; E, the average power for different values of τ; F, the average significance level in the bridging study under the alternative hypothesis as τ varies; G, the chance of conducting the bridging study under the alternative hypothesis; H, the optimal value for ω for different values of τ under the alternative. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

study. We used the same parameter values under the null and alternative hypotheses as before, and let $\tau = 0$. We set the number of patients per arm to 400 for the foreign study and varied p and q from 0.6 to 0.9. For each choice of (p, q), we used the procedure described in Section 2.5 to determine the minimal sample size for achieving 80% power. We then conducted simulation with the resulting sample size to evaluate the power for the proposed testing procedure and compared it to the power without using the foreign-study evidence. Note that without using the foreign-study evidence, the required sample size to achieve the same power is at least 280 patients per arm.

The results are summarized in Table 1. The proposed method maintains the empirical power of 80%, and the efficiency gain, in terms of sample sizes, varies substantially with (p, q). The gain is small when p or q is less than 0.8. We can save the sample size by almost one half when both p and q are equal to 0.9.

4 | DISCUSSION

In this article, we present a new framework for incorporating the strength of evidence from a foreign study into the design and analysis of a bridging study. This framework relies on some Bayesian priors on the relationship between the hypotheses of the two studies and focuses on the adaptive type I error or power of the bridging study. The proposed methods account for randomness in the evidence of the foreign study so as to preserve the overall type I error. We develop a simple procedure to determine the optimal parameter that maximizes the conditional power for the bridging study.

The choice of p and q plays an important role in the proposed methodology. A premise for borrowing the foreign study evidence when conducting a bridging study is that p and q are high. In practice, the sponsor should work closely with stakeholders to examine the operating characteristics of the proposed method over a reasonable range of p and q and then

T A B L E 1 Sample sizes for different choices of (p, q) when $\tau = 0$

| | | | Empirical power | |
| | | Number of patients per arm | Without foreign study | Proposed method |
p	q			
0.6	0.6	276	0.787	0.797
	0.7	270	0.779	0.804
	0.8	266	0.774	0.802
	0.9	260	0.766	0.802
0.7	0.6	268	0.776	0.801
	0.7	258	0.763	0.801
	0.8	246	0.745	0.797
	0.9	234	0.727	0.801
0.8	0.6	260	0.766	0.799
	0.7	246	0.745	0.796
	0.8	224	0.710	0.798
	0.9	202	0.671	0.805
0.9	0.6	252	0.754	0.799
	0.7	230	0.720	0.801
	0.8	196	0.660	0.796
	0.9	156	0.576	0.796

decide on a clinically meaningful value of p and q, such that there is a modest gain in power.

The proposed methodology can be extended to incorporate multiple sources of evidence. When there are several foreign studies, the prior distributions for p and q are multivariate. The choice of the type I error on the prior distribution is also multivariate and becomes more complicated because there may be some evidence favoring the null hypothesis and other evidence favoring the alternative hypothesis. The search for the optimal choice will be a grid search on a multidimension set. Alternatively, one may combine the multiple sources of evidence into a single measure (eg, some prespecified weighted combination), such that the methodology presented in this article can be applied directly.

Although this work is focused on incorporating foreign-study evidence into the design and analysis of a bridging study, the basic ideas are applicable to other contexts, where some prior study evidence is used to improve the design and analysis of a new study, for example, when historical data are incorporated into the design and analysis of a new clinical trial. By treating the prior study evidence as the "foreign study" evidence and the new study as the "bridging study," the proposed methods can be applied directly.

Note that the proposed methodology allows the endpoints of the foreign and bridging studies to be different. This is a very useful feature because the indications for the drug product may be different between the new region and the original region or between the new study and the prior study within the same region. By contrast, Bayesian methods typically require the same endpoint for the two studies so as to combine their likelihood functions.

ORCID
Donglin Zeng https://orcid.org/0000-0003-0843-9280
D. Y. Lin https://orcid.org/0000-0002-0150-3115

REFERENCES
Chow, S.C., Chiang, C., Liu, J.-P. and Hsiao, C-F. (2012). Statistical methods for bridging studies. *Journal of Biopharmaceutical Statistics*, 22, 903–915.

Hsiao, C.-F., Hsu, Y.-Y., Tsou, H.-H. and Liu, J.-P. (2007). Use of prior information for Bayesian evaluation of bridging studies. *Journal of Biopharmaceutical Statistics*, 17, 109–121.

Huang, Q., Chen, G., Yuan, Z. and Lan, K.K.G. (2012). Design and sample size considerations for simultaneous global drug development program. *Journal of Biopharmaceutical Statistics*, 22, 1060–1073.

International Conference on Harmonization. (1998). Tripartite GUIDANCE E5: Ethnic factors in the acceptability of foreign data. *Federal Register*, 83, 31790–31796.

Lan, K.K.G. and Pinheiro, J. (2012). Combined estimation of treatment effects under a discrete random effects model. *Statistics and Biosciences*, 4, 235–244.

Lan, K.K.G., Soo, Y., Siu, C. and Wang, M. (2005). The use of weighted Z-tests in medical research. *Journal of Biopharmaceutical Statistics*, 15, 625–639.

Liu, J.-P., Hsueh, H. and Chen, J. J. (2002). Sample size requirements for evaluation of bridging evidence. *Biometrical Journal*, 44, 969–981.

Shih, W. J. (2001). Clinical trials for drug registration in Asian-Pacific countries: proposal for a new paradigm from a statistical perspective. *Controlled Clinical Trials*, 22, 357–366.

Shao, J. and Chow, S.-C. (2002). Reproducibility probability in clinical trials. *Statistics in Medicine*, 21, 1727–1742.

SUPPORTING INFORMATION
Figure S1 referenced in Appendix A and Figures S2-S8 referenced in Section 3 are available with this article at the Biometrics website on Wiley Online Library. The simulation codes in this article are also available online with this article.

APPENDIX A: RELATIONSHIP BETWEEN (p,q) AND JOINT PRIOR DISTRIBUTION OF (Δ_1, Δ_2)

Assume that the prior distribution for (Δ_1, Δ_2) is the bivariate normal distribution with mean (δ_1, δ_2) and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Then

$$p = \Pr(H_{10} \mid H_{20}) = \Pr\{\Delta_1 \notin (L_1, U_1) \mid \Delta_2 \notin (L_2, U_2)\}$$

$$= \frac{\int_{(-\infty, L_1] \cup [U_1, \infty)} \int_{(-\infty, L_2] \cup [U_2, \infty)} (2\pi)^{-1} |\Sigma|^{-1/2} \exp\{-(x - \delta_1, y - \delta_2)\Sigma^{-1}(x - \delta_1, y - \delta_2)^{\mathrm{T}}\} dy\, dx}{\Phi((L_2 - \delta_2)/\sigma_2) + \Phi(-(U_2 - \delta_2)/\sigma_2)},$$

and

$$q = \Pr(H_{1a} \mid H_{2a}) = \Pr\{\Delta_1 \in (L_1, U_1) \mid \Delta_2 \in (L_2, U_2)\}$$

$$= \frac{\int_{L_1}^{U_1} \int_{L_2}^{U_2} (2\pi)^{-1} |\Sigma|^{-1/2} \exp\{-(x - \delta_1, y - \delta_2)\Sigma^{-1}(x - \delta_1, y - \delta_2)^{T}\} dy\, dx}{-\Phi((L_2 - \delta_2)/\sigma_2) + \Phi((U_2 - \delta_2)/\sigma_2)}.$$

We provide an example to illustrate the above relationship. Let $L_1 = L_2 = -0.75$ and $U_1 = U_2 = 0.75$. Also, let $\delta_1 = \delta_2 = 0$ and $\sigma_1 = \sigma_2 = 1$. Figure S1 in the Supporting Information shows how p and q vary with the correlation coefficient ρ. The higher the correlation in the joint prior distribution is, the larger the values of p and q are.

APPENDIX B: PROOFS OF THEOREMS

Proof of Theorem 1. Note that

$\Pr(\mathcal{R}(\omega, \gamma) \mid H_{20})$

    $= \Pr(\text{the bridging study is conducted and } \mathrm{CI}_2(\hat{\alpha}_2)$

        $\subset (L_2, U_2) \mid H_{20})$

    $= \Pr(\mathrm{CI}_1(\omega) \subset (L_1, U_1), \mathrm{CI}_2(\alpha\gamma) \subset (L_2, U_2) \mid H_{20})$

        $+ (1 - \tau)\Pr(\mathrm{CI}_1(\omega) \not\subset (L_1, U_1), \mathrm{CI}_2(\alpha/\gamma)$

        $\subset (L_2, U_2) \mid H_{20}).$

Under Assumption 1, the last expression is equal to

    $\Pr(H_{10} \mid H_{20})\Pr(\mathrm{CI}_1(\omega) \subset (L_1, U_1) \mid H_{10})$

        $\times \Pr(\mathrm{CI}_2(\alpha\gamma) \subset (L_2, U_2) \mid H_{20})$

    $+ \Pr(H_{1a} \mid H_{20})\Pr(\mathrm{CI}_1(\omega) \subset (L_1, U_1) \mid H_{1a})$

        $\times \Pr(\mathrm{CI}_2(\alpha\gamma) \subset (L_2, U_2) \mid H_{20})$

    $+ \Pr(H_{10} \mid H_{20})(1 - \tau)\Pr(\mathrm{CI}_1(\omega) \not\subset (L_1, U_1)|H_{10})$

        $\times \Pr(\mathrm{CI}_2(\alpha/\gamma) \subset (L_2, U_2) \mid H_{20})$

    $+ \Pr(H_{1a}|H_{20})(1 - \tau)\Pr(\mathrm{CI}_1(\omega) \not\subset (L_1, U_1)|H_{1a})$

        $\times \Pr(\mathrm{CI}_2(\alpha/\gamma) \subset (L_2, U_2) \mid H_{20}),$

which can be bounded by

    $p\omega\alpha\gamma + (1 - p)\alpha\gamma\Pr(\mathrm{CI}_1(\omega) \subset (L_1, U_1)|H_{1a})$

        $+ p(1 - \tau)(1 - \omega)\alpha/\gamma$

        $+ (1 - p)(1 - \tau)\alpha/\gamma\Pr\big(\mathrm{CI}_1(\omega) \not\subset (L_1, U_1)|H_{1a}\big).$

Thus,

$\Pr(\mathcal{R}(\omega, \gamma) \mid H_{20})$

    $\leq p\omega\alpha\gamma + p(1 - \tau)(1 - \omega)\alpha/\gamma + (1 - p)\alpha \max(\gamma, (1 - \tau)/\gamma)$

    $\leq p\omega\alpha\gamma + p(1 - \tau)(1 - \omega)\alpha/\gamma + (1 - p)\alpha\gamma$

    $\leq \alpha[\{1 - p + p\omega\}\gamma + p(1 - \omega)(1 - \tau)/\gamma]$

    $\leq \alpha[\{1 - p + p\omega\}\gamma + p(1 - \omega)/\gamma]. \hspace{2cm} \text{(A1)}$

The conclusion of Theorem 1 follows upon verifying that γ given in Theorem 1 ensures that (A1) is bounded by α. $\square$

Proof of Theorem 2. Clearly,

    $\Pr(\mathcal{R}(\omega, \gamma) \mid H_{2a})$

        $= \Pr(\text{the bridging study is conducted and } \mathrm{CI}_2(\hat{\alpha}_2)$

            $\subset (L_2, U_2) \mid H_{2a})$

$$= \Pr(\text{CI}_1(\omega) \subset (L_1, U_1), \text{CI}_2(\alpha\gamma) \subset (L_2, U_2) \mid H_{2a})$$
$$+ (1 - \tau)\Pr(\text{CI}_1(\omega) \not\subset (L_1, U_1), \text{CI}_2(\alpha/\gamma)$$
$$\subset (L_2, U_2) \mid H_{2a}).$$

Under Assumption 2,

$$\Pr(\text{the bridging study is conducted and } \text{CI}_2(\hat{\alpha}_2)$$
$$\subset (L_2, U_2) \mid H_{2a})$$
$$= \Pr(H_{1a} \mid H_{2a})\Pr(\text{CI}_1(\omega) \subset (L_1, U_1) \mid H_{1a})$$
$$\times \Pr(\text{CI}_2(\alpha\gamma) \subset (L_2, U_2) \mid H_{2a})$$
$$+ \Pr(H_{1a} \mid H_{2a})(1 - \tau)\Pr(\text{CI}_1(\omega) \not\subset (L_1, U_1) \mid H_{1a})$$
$$\times \Pr(\text{CI}_2(\alpha/\gamma) \subset (L_2, U_2) \mid H_{2a})$$
$$+ \Pr(H_{10} \mid H_{2a})\Pr(\text{CI}_1(\omega) \subset (L_1, U_1) \mid H_{10})$$
$$\times \Pr(\text{CI}_2(\alpha\gamma) \subset (L_2, U_2) \mid H_{2a})$$
$$+ \Pr(H_{10} \mid H_{2a})(1 - \tau)\Pr(\text{CI}_1(\omega) \not\subset (L_1, U_1) \mid H_{10})$$
$$\times \Pr(\text{CI}_2(\alpha/\gamma) \subset (L_2, U_2) \mid H_{2a}).$$

It follows from the definition of $Q_k(\alpha_0)$ that

$$\Pr(\mathcal{R}(\omega, \gamma) \mid H_{2a})$$
$$= q\{Q_2(\alpha\gamma)\Pr(\text{CI}_1(\omega) \subset (L_1, U_1) \mid H_{1a})$$
$$+ (1 - \tau)Q_2(\alpha/\gamma)[1 - \Pr(\text{CI}_1(\omega) \subset (L_1, U_1) \mid H_{1a})]\}$$
$$+ (1 - q)\omega Q_2(\alpha\gamma) + (1 - q)(1 - \tau)(1 - \omega)Q_2(\alpha/\gamma)$$
$$= q\{Q_2(\alpha\gamma)Q_1(\omega) + (1 - \tau)Q_2(\alpha/\gamma)[1 - Q_1(\omega)]\}$$
$$+ (1 - q)\omega Q_2(\alpha\gamma) + (1 - q)(1 - \tau)(1 - \omega)Q_2(\alpha/\gamma).$$

If $\omega = 1$, then γ in Theorem 1 is equal to 1, such that we do not use the evidence from the foreign study. It follows that $\widetilde{Q}_2(1) = Q_2(\alpha)$. Therefore, the second part of Theorem 2 holds. $\qquad\square$