Bayesian Hierarchical Modeling and the Integration of Heterogeneous Information on the Effectiveness of Cardiovascular Therapies

Heemun Kwok, MD, MS; Roger J. Lewis, MD, PhD

Abstract—When making therapeutic decisions for an individual patient or formulating treatment guidelines on a population level, it is often necessary to utilize information arising from different study designs, settings, or treatments. In clinical practice, heterogeneous information is frequently synthesized qualitatively, whereas in comparative effectiveness research and guideline development, it is imperative that heterogeneous data are integrated quantitatively and in a manner that accurately captures the true uncertainty in the results. Bayesian hierarchical modeling is a technique that utilizes all available information from multiple sources and naturally yields a revised estimate of the treatment effect associated with each source. A hierarchical model consists of multiple levels (ie, a hierarchy) of probability distributions that represent relationships between information arising within single populations or trials, as well as relationships between information arising from different populations or trials. We describe the structure of Bayesian hierarchical models and discuss their advantages over simpler models when multiple information sources are relevant. Two examples are presented that illustrate this technique: a meta-analysis of immunosuppressive therapy in idiopathic dilated cardiomyopathy and a subgroup analysis of the National Institute of Neurological Disorders and Stroke Intravenous Tissue Plasminogen Activator Stroke Trial. (Circ Cardiovasc Qual Outcomes. 2011;4:657-666.)

Key Words: multilevel analysis ■ comparative effectiveness research ■ practice guidelines

Health care providers must often decide whether to recommend a therapy for a patient for whom directly relevant treatment outcome data are limited or absent. For example, there may be limited experience with the therapy, the disease state may be rare, or the patient may be a member of a population that is underrepresented in available clinical studies. Similarly, those setting health care policy, authoring treatment guidelines, or making coverage determinations must also often make decisions regarding the effectiveness of treatments when directly relevant information is limited. An intuitive approach, frequently used by practicing clinicians, is to consider evidence from related populations and therapies, although the similarity to the patient or population of interest may vary significantly. For example, efficacy data for other drugs in the same class may be available, or the therapy may have been evaluated in populations with different comorbidities, demographic characteristics, or disease severities.

Whether at the individual patient or population level, this type of decision-making requires the synthesis of information from heterogeneous sources. Many common statistical methods assume either homogeneity (eg, fixed-effects meta-analysis) or complete independence across populations and are thus ill-suited for the quantitative synthesis of heterogeneous data. Hierarchical, or multilevel, modeling is a statistical method that can be used to quantitatively and coherently combine heterogeneous information. A hierarchical model consists of multiple levels (ie, a hierarchy) of probability distributions that represent relationships between information arising within single populations or trials, as well as relationships between information arising from different populations or trials.

Hierarchical models may be implemented within either a Bayesian or frequentist framework. This review focuses on Bayesian hierarchical modeling and is divided into 6 sections. The first section is a clinical example that illustrates how information from multiple, related populations may be synthesized qualitatively when clinicians make decisions about individual patients. The next section argues that a similar but quantitative approach has an important role in multiple settings, including comparative effectiveness research, the development of treatment guidelines, and insurance coverage decisions. The third section describes the structure of a Bayesian hierarchical model and discusses how it utilizes all available information from multiple sources and naturally...
yields a revised estimate of the treatment effect associated with each source. In the fourth section, as an example, we apply this technique to estimate the efficacy of immunosuppressive therapy for the treatment of virus-negative, idiopathic dilated cardiomyopathy, using data from related trials. In the fifth section, hierarchical modeling is applied to a subgroup analysis of the National Institute of Neurological Disorders and Stroke (NINDS) intravenous tissue plasminogen activator (t-PA) stroke trial. In the sixth and final section, we discuss how hierarchical modeling could be used to address the statistical issues raised in the clinical example from the first section.

Clinical Example: Drug-Eluting Stents
The COMPARE trial randomly assigned 1800 patients undergoing percutaneous coronary intervention (PCI) to treatment with an everolimus-eluting stent (EES) or a paclitaxel-eluting stent (PES). The primary end point, a composite of cardiac death, target vessel myocardial infarction, and target vessel revascularization within 12 months, occurred in 6% of the EES group and 9% of the PES group. The overall relative risk (RR) was 0.62 (95% confidence interval [CI], 0.46–0.82) in favor of EES (Figure 1). An exploratory subgroup analysis demonstrated that the observed treatment effect was qualitatively similar in all but 2 subgroups: those with diabetes (RR = 1.00; 95% CI, 0.53–1.89) and those with lesion length ≥20 mm (RR = 1.02; 95% CI, 0.40–2.61). There were 325 subjects with diabetes, and among the 1475 patients without diabetes, the RR was 0.61 (95% CI, 0.42–0.90).

Shortly after the results of the COMPARE study were published, a clinician was referred a 56-year-old man with hypertension, non–insulin-dependent diabetes, and stable angina, whose nuclear study demonstrated a moderate-sized area of reversible ischemia in the lateral wall with preserved systolic function. The patient did not have contraindications to dual antiplatelet therapy, and, should the patient require PCI, the clinician intends to use a drug-eluting stent, given the history of diabetes. The COMPARE trial enrolled a real-life practice population, and the clinician believes that its results apply to this patient—but, which result: the overall estimate or the diabetes subgroup estimate? Because the test for interaction between diabetic status and treatment was not statistically significant (P = 0.22), the clinician might decide that the overall estimate is more applicable. However, the clinician ought not to ignore the subgroup results and should consider that the degree of benefit may be attenuated among diabetic patients.

Four months later, the results of the multicenter SPIRIT IV trial, which also compared EES and PES, were published. There were 3687 patients, and the primary end point was similar to that of the COMPARE trial. The overall RR was 0.62 (95% CI, 0.46–0.82) in favor of EES (4%) compared with PES (7%). Among patients with diabetes (n = 1114), the RR was 0.94 (0.59–1.49), and, among those without diabetes (n = 2573), it was 0.47 (95% CI, 0.32–0.68); the test for interaction was significant (P = 0.02). The clinician was struck by the consistency in the results of the 2 trials and concluded that EES provide little benefit over PES among diabetic patients. However, considering the overall results of the 2 trials, which favor EES, the clinician found it unlikely that outcomes with EES are worse than those with PES, despite the confidence intervals for the diabetic subgroup estimates being relatively symmetrical, around an RR of 1.

The above clinical vignette is not intended to espouse a single, “correct” clinical reasoning process for a diabetic patient in need of PCI. Rather, the purpose is to illustrate the complexities involved in synthesizing evidence when formulating decisions for individual patients. In applying the evidence from each trial, the clinician considered both the overall and the diabetic subgroup estimates because both provided information regarding the efficacy of EES and PES in diabetics to some degree. In contrast, common statistical practice is to use a test of interaction to determine whether to apply the overall estimate (and ignore the subgroup estimate) or to apply the subgroup estimate (and ignore the overall estimate). However, neither estimate alone incorporates all of the available data nor provides a complete picture. The overall estimate ignores the observed differences between the diabetic and nondiabetic populations and considers data from both populations to be equally relevant. The diabetic subgroup estimate considers data from the nondiabetic population to be irrelevant. In clinical practice, clinicians are able to compensate for the limitations of standard statistical practice, to some degree, by synthesizing the overall and subgroup estimates qualitatively. However, in many settings, it is preferable to use statistical methods that are capable of integrating data quantitatively.

Comparative Effectiveness Research
According to the Federal Coordinating Council for Comparative Effectiveness Research, the primary purpose of com-
Comparative effectiveness research (CER) is "to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances." CER thus strives to accomplish for groups of patients what clinicians try to do for their individual patients: to determine the most effective intervention. Moreover, the phrase "for which patients under specific circumstances" emphasizes the goal of preserving and even focusing on the divergent treatment effects that may be seen in different populations or settings. Both CER and evidence-based clinical decision-making require the integration of heterogeneous evidence and the subsequent application to a new population. The available data may originate from different study designs, settings, and treatment comparisons. A fundamental difference is that whereas clinical decision-making is a qualitative, subjective, and often hidden process, CER should be transparent and quantitative in order for its results to be credible to decision-makers. Ideally, the quantitative methods would reflect the flexibility and complexity of clinical reasoning, particularly with respect to the handling of complex relationships among heterogeneous data sources.

Many common statistical methods are restricted by their tendency to treat information from different populations as either completely distinct or exactly identical. For example, in a stratified analysis, if heterogeneity of effect is judged to be present, stratum-specific estimates of treatment effects are generated independently; if not, the underlying parameters are assumed to be identical, and the data are pooled to estimate a common parameter. Similarly, in many regression models, the relationship between subpopulations is limited to a choice of homogeneity or independence. This restriction is a significant limitation within the context of CER, whose goal is to address the heterogeneity of treatment effects, while simultaneously using all available information in selecting the most effective treatment for each patient population. Outcomes from various study designs, settings, and interventions would rarely be expected to be identical; however, treating them independently precludes any meaningful synthesis. One notable consequence is the inability to make informative inferences about populations in which there are absent or limited data. In contrast, hierarchical modeling involves a more flexible assumption termed exchangeability, which may be regarded as a compromise between assuming independence and assuming identicality of the treatment effects arising from different sources. As a result, it is a natural method for aggregating data originating from different study designs, populations, outcomes, and/or treatments.

Hierarchical Modeling

Structure
We will describe the structure of a hierarchical model in the context of an analysis whose objective is to estimate the rate of major adverse cardiac events (MACE) after treatment with a hypothetical drug-eluting stent. This analysis uses data from three registries whose underlying rates of MACE are unlikely to be identical due to differences in populations and practice settings. Furthermore, their underlying rates are unlikely to be identical to those of external populations to which we would like to generalize our results. Although not identical, the rates of MACE in the populations of interest (the specific registry populations as well as the external populations) are likely to be related, because the same intervention is being applied to all of the patients. A judgment that a group of parameters is related but not identical is termed exchangeability. More precisely, a set of parameters or observations is exchangeable if, a priori, it is not possible to predict which ones are larger or smaller. In other words, if one could reliably predict that a particular registry population had a higher or lower incidence of MACE than the others, then an assumption of exchangeability would be invalid. Of note, in conducting the analysis, the rates of MACE in the registries may be found to be quite different, but such a finding would not violate an assumption of exchangeability. Exchangeability does not require that the rates of MACE are similar to a certain degree, but only that a priori their relative order was not predictable. Exchangeability is a crucial assumption of hierarchical models and should be justifiable on biological and/or clinical grounds.

A hierarchical model includes multiple units of analysis. In this example, the first unit of analysis is the patient, and the second unit of analysis is the registry, with patients grouped or nested within registries. For each registry, the observed patient outcomes are modeled using a probability distribution which is a function of that registry’s true underlying rate of MACE. These probability distributions together comprise the first level of the hierarchical model (Figure 2). A defining characteristic of hierarchical models is that the parameters of the first level—the respective true rates of MACE within each registry—are themselves modeled using a probability distribution, which is termed the hyperdistribution and constitutes
the second level of the model. The parameters of the first level may be thought of as being drawn randomly from the hyperdistribution, which is commonly assumed to be normal with 2 hyperparameters: the overall rate of MACE, averaged across all possible registries, and the between-registry variance of the rate of MACE.

Various estimation methods may be used to determine the hyperparameters. In a fully Bayesian approach, the hyperparameters are treated as unknown random variables and assigned probability distributions. The prior distributions of the hyperparameters quantify knowledge based on information external to the present study and comprise the third level of the hierarchical model. In the absence of relevant external information, or to permit inferences to depend exclusively on the present data, vague, noninformative prior distributions may be used. A fully Bayesian approach naturally accounts for the prior and final uncertainty in the estimates of the unknown hyperparameters. In contrast, in an empirical Bayes approach, the hyperparameters are estimated from the data and subsequently treated as known or fixed parameters. Although the term “empirical Bayes” suggests that it is strictly a Bayesian method, this approach may also be implemented using frequentist techniques. It ignores the final uncertainty in the values of the hyperparameters and often results in estimates of the first-level parameters that are overly precise.

Estimation

Improved accuracy of estimation is a major advantage of hierarchical modeling relative to the use of conventional, independent estimators. In general, aggregating data decreases the variance associated with the estimator of a given parameter (e.g., an odds ratio), because variance is inversely related to the quantity of information. However, it can potentially introduce bias, which is the difference between the estimator’s expected value and the parameter’s true value. Because the accuracy of an estimator is related to both bias and variance, expected accuracy may increase or decrease when information is combined. For example, consider the Mantel-Haenszel method, which combines data from different groups to produce an estimate of a common odds ratio that applies to all of the groups. If the true odds ratios associated with the different groups are in fact identical, there will be no increase in bias, and the overall accuracy will increase by use of the Mantel-Haenszel method. If, however, at the other extreme, the parameters are in actuality substantially different, the common odds ratio, when applied to each of the groups, will have a large increase in bias and accuracy will decrease despite a reduction in variance. When the underlying parameters are neither identical nor substantially different, overall accuracy could increase or decrease with the Mantel-Haenszel method, relative to the independent estimates for each group, depending on the degree of similarity.

With a hierarchical model, information from all of the exchangeable groups is shared to some extent; alternatively, we can imagine the estimate for each group incorporating or borrowing information from the other groups. Unlike the Mantel-Haenszel or other methods that assume homogeneity, the amount of borrowing is flexible and results in the partial pooling of data. The effect of borrowing is that the independent, or naive, estimates are pulled toward one another with a narrowing of their intervals; this effect is known as shrinkage. Shrinkage does introduce bias, except in the unrealistic situation in which all of the underlying parameters are identical. However, on average, the increase in bias is more than offset by a reduction in variance, and total accuracy increases.

The hypervariance and the method used to estimate it play a critical role in determining the amount of shrinkage. A smaller value of the hypervariance implies that the parameters are more similar and results in greater shrinkage. A larger value implies that the parameters are less similar and results in less shrinkage. In an extreme case in which the hypervariance equals zero, the parameters are identical, and the result is a common, pooled estimate. At the other extreme, an infinite hypervariance corresponds to the parameters being unrelated, and there is no shrinkage. In both fully Bayesian and empirical Bayes methods, the observed consistency in the data are used to estimate the hypervariance. If a fully Bayesian approach is used, the prior distribution for the hypervariance also influences the estimate of the hypervariance and the amount of shrinkage, although shrinkage will occur even if a noninformative prior distribution is used.

In addition to the hypervariance, the degree of shrinkage for any particular naive estimate is also related to its distance from the hypermean. Under an assumption of exchangeability, the parameters are expected to be similar, so an observed estimate that is far from the other estimates (and the hypermean) is more likely to be further from its true underlying parameter due to random variation and will experience more shrinkage. In addition, an independent estimate that is less precise, that is, one that is based on fewer observations, will be shrunk to a greater degree. This is consistent with our intuition, because an imprecise estimate is also less reliable.

Example: Application to Meta-Analysis

Typically, the primary objective of a meta-analysis is to estimate the overall effect of an intervention by quantitatively integrating evidence from a set of related studies. With a fixed-effects model, the included studies are assumed to share a common treatment effect. A random-effects model recognizes that under most circumstances, the true treatment effects are likely to vary between studies due to differences in populations and how the intervention is applied. In contrast, in this example, our primary interest is in the study-specific (i.e., primary) treatment effects, rather than the overall mean effect. A hierarchical modeling approach to random-effects meta-analysis provides estimates of both the study-specific and overall effects and their associated uncertainty. We illustrate fully Bayesian and empirical Bayes methods and compare them with a fixed-effects meta-analysis and to the approach of analyzing each trial separately. Our focus is on the quantitative process of combining information, and we therefore defer discussion of other aspects of meta-analysis to reviews and texts dedicated to this topic.

Background

Although there is evidence that inflammation and autoimmunity may contribute to the development of cardiomyopathy in
patients with myocarditis, immunosuppression has not been shown to have a clinically significant benefit in multiple randomized, controlled trials (RCTs).17–19 As a result, immunosuppression is not recommended as routine therapy for patients with myocarditis and/or idiopathic dilated cardiomyopathy (IDC).20,21 However, the sample sizes of these RCTs have been small, and their power may have been diminished by heterogeneity of the study populations.22 Thus, it has been hypothesized that there are subsets of patients who could benefit from immunosuppression.

Frustraci et al recently conducted a trial that enrolled patients in whom endometrial biopsy (EMB) showed evidence of myocarditis, but no viral genomes.23 Eighty-five patients with left ventricular ejection fraction (LVEF) ≤45% and symptoms of congestive heart failure (CHF) for more than 6 months were randomly assigned to a regimen of prednisone and azathioprine versus conventional therapy alone.17 Randomization was stratified into “reactive” and “nonreactive” subgroups according to the presence of inflammation on EMB. Although there was a marginal improvement in LVEF at 3 months in both subgroups, they found no difference in LVEF at 9 months. Below, we consider the “reactive” and “nonreactive” patients as 2 separate populations.

In the Myocarditis Treatment Trial,18 110 patients with evidence of myocarditis on EMB and LVEF ≤45% were randomly assigned to either prednisone plus azathioprine or cyclosporine (considered as one treatment group) versus conventional therapy alone. All but 1 patient had symptoms for less than 1 month. There was no significant difference between the groups regarding the change in LVEF at 28 weeks, the primary end point.

Wojnicz et al enrolled 84 patients with increased expression of human leukocyte antigen, a sensitive marker of inflammation, on EMB, symptoms of CHF for more than 6 months, and LVEF ≤40%.19 Patients were randomly assigned to either prednisone and azathioprine or placebo. There was no difference in the primary outcome, a composite of cardiac death, transplant and hospital readmission at 2 years, but the proportion of patients whose LVEF increased more than 5% was significantly higher in the immunosup-
pression group than the control group at 3 months (71% versus 21%), 6 months (81% versus 24%), and 1 year (69% versus 28%). We use the 6-month outcome for our analysis.

Methods
The metric of interest for this example is the odds ratio for improvement in LVEF in the immunosuppressive treatment arm compared with the control arm. We applied 4 different analytic approaches to illustrate the relationship between modeling assumptions and inferences. In the first approach, which we will refer to as an independent approach, no assumption was made regarding the relationship among the treatment effects existing in the 5 trial populations (recall that the Parrillo trial has 2 populations). An odds ratio for each trial population was calculated using the Fisher exact method, or, in the case of the Frustaci trial, a small sample correction method as described above. The software program R (v. 12.2.1), in conjunction with the package Epitools (v. 0.5–6), was used to conduct this portion of the analysis.

The second approach used a fully Bayesian hierarchical model, which assumed that the treatment effects were exchangeable, that is, that there was no reason to presume that immunosuppressive therapy was more or less efficacious in any particular population. In some analyses, the magnitude of the treatment effect may be thought to be related to study-level characteristics such as sample size or publication year. If covariates that explain part of the heterogeneity of treatment effects are included in the model, then the assumption becomes partial exchangeability, that is, that the treatment effects are exchangeable after controlling for these variables. We considered the inclusion of publication year in the model as a surrogate for secular trends in medical care. However, doing so did not appreciably change the results, so this covariate was therefore not included in the final model.

Specifying a fully Bayesian hierarchical model required making an assumption regarding the form of the hyperdistribution (the second level) and selecting prior distributions for the hypermean and hypervariance (the third level). The following choices are fairly conventional but are based on mathematical convenience more so than any strong theoretical justification; reasonable alternatives were therefore explored with a sensitivity analysis (see below). The logarithm of the odds ratios were assumed to be drawn from a normal hyperdistribution. The prior distribution for the hypermean (on a logarithmic scale) was Normal \((0, 1000)\), that is, a normal distribution with a mean of 0 and variance of 1000, which states mathematically that over a large range of values centered around the null, no particular value was believed to be more likely to equal the true mean effect than any other. The prior distribution for the hypervariance requires a form that is restricted to nonnegative values, because a variance is always greater than or equal to zero. Specifying a prior distribution for the standard deviation of the hyperdistribution (ie, the square root of the hypervariance) is an alternative to specify a prior distribution for the hypervariance. We used a half-Normal \((0, 10000)\) distribution, which is a Normal \((0, 10000)\) distribution that is restricted to nonnegative values, for the standard deviation. This distribution is noninformative and allows the data to primarily determine the degree of similarity among the different treatment effects.

To make inferences in Bayesian analysis, the prior distribution is combined with the likelihood function, which indicates how likely the data are given values of the model parameters, to produce the posterior distribution. In many analyses, the posterior distribution cannot be calculated exactly and is therefore approximated using a numeric procedure called Markov Chain Monte Carlo (MCMC). A discussion of MCMC techniques is beyond the scope of this review, but 3 chains containing 5000 samples from the posterior distribution were obtained after a burn-in of 1000, and convergence was assessed using the Gelman-Rubin statistic (convergence ensures that the posterior distribution of interest is sampled properly, and the Gelman-Rubin statistic is one method to diagnose convergence). The fully Bayesian analysis was performed using R (v. 12.2.1), the package R2WinBUGS, and the software program OpenBUGS (v. 3.1.2); the software code is provided in the online-only Data Supplement.

The third approach involved the application of empirical Bayes methods to hierarchical modeling. Like the fully Bayesian approach described above, the treatment effects were assumed to be exchangeable and to follow a normal hyperdistribution. An empirical Bayes approach involved the additional assumption that the hyperparameters were known with certainty, despite being estimated from the data. Many software programs are capable of implementing hierarchical models and use a variety of frequentist, likelihood-based estimation methods. The performance of any particular software procedure may vary according to conditions such as the number of studies and number of observations within studies, so it is crucial to understand the strengths and limitations of each. We used the glmer function from the R package lme4 (v. 0.999375–39), which can fit hierarchical logistic regression models, to derive empirical Bayes estimates of the study-specific effects.

The final approach entailed a fixed-effects meta-analysis, which made the strongest possible assumption regarding the treatment effects—identity. This assumption is equivalent to assuming that the hypervariance of a hierarchical model equals zero. We used the Mantel-Haenszel method to compute a single common odds ratio.

Results
The independent, fully Bayesian, empirical Bayes and fixed-effects estimates for the treatment effect in each trial population are provided in Figure 3. The independent estimates for the odds ratios varied from 1.5 to 600, and their intervals were the widest of any method, since the results from each trial population were considered in isolation. In comparison to the independent estimates, the fully Bayesian hierarchical estimates were shrunken toward the overall mean, and their intervals were more narrow, but only to a small extent. The small degree of shrinkage reflected the large degree of observed heterogeneity in the trial results and the use of a hypervariance prior that permitted consideration of large variability in the treatment effects. Nevertheless, the independent estimate from the Frustaci trial population was shrunken to
the greatest degree, due to its poor precision and its extreme value relative to the other estimates (note that without the small sample correction, the independent Frustaci estimate would have been infinitely large, because there were zero successes in the control arm). An empirical Bayes approach produced narrower 95% intervals than a fully Bayesian approach, because it ignored the uncertainty in the values of the hyperparameters. The common estimate (5.7; 95% CI, 3.7–9.0) excluded 4 of the 5 independent point estimates, strongly suggesting that it did not accurately represent the true treatment effects in the individual trial populations.

**Sensitivity Analysis**

As discussed above, specifying a fully Bayesian hierarchical model involved a number of distributional assumptions. When there are multiple, reasonable alternatives, the sensitivity of the results to the assumptions should be assessed by comparing the results from different models. It is particularly important to examine the form of the hyperdistribution and the prior distribution for the hypervariance, because both can influence the degree of shrinkage to a substantial degree. With respect to the hyperdistribution, in addition to assuming normality, we investigated a Student $t$ distribution with 4 degrees of freedom, which has heavier tails than the normal distribution and thus allows greater heterogeneity in the treatment effects. Using a Student $t$ distribution did not substantially affect the results (Figure 4).

The form of the prior distribution for the hypervariance (or the square root of the hypervariance) can also potentially affect inferences, even if one uses distributions that are intended to be noninformative. We repeated our analysis using 2 other common noninformative prior distributions: (1) a Uniform (0, 100) distribution, which places equal probability on all values between 0 and 100, as the prior distribution for the square root of the hypervariance; and (2) an Inverse Gamma (0.001, 0.001) distribution as the prior distribution for the hypervariance. Using these prior distributions resulted in estimates that were similar to those obtained with a half-Normal (0, 10000) distribution.

Sensitivity analysis may also be used to evaluate the influence of prior distributions that are informative, for example, that incorporate knowledge derived from data that were not collected as part of the present study and/or from expert opinion. As an example, we considered a Uniform (0, 2.3) distribution, which is a mathematical statement that plausible values of the standard deviation of the hyperdistribution range from 0 (ie, the treatment effects are identical in
all studies) to 2.3. Because a difference of 2.3 between natural log odds ratios is equivalent to a 10-fold difference between their corresponding odds ratios, a Uniform (0, 2.3) is compatible with a large degree of heterogeneity and, in fact, confers only a small degree of information. For example, if a normal hyperdistribution has a mean of 0 and a standard deviation of 2.3, the 95% interval of odds ratios includes values from 0.01 to 100. Using this prior distribution resulted in slightly more shrinkage than the noninformative prior distributions discussed above, but the inferences were not materially different.

Example: Application to Subgroup Analysis
The purpose of subgroup analysis is to explore how a treatment effect varies among subsets of patients within a larger population. Subgroup analyses may be difficult to interpret and are subject to both false-positive and false-negative results; however, when conducted, reported, and interpreted with caution, they have the potential to inform clinical practice and future research. For example, consider the NINDS intravenous t-PA stroke trial, which demonstrated a benefit from t-PA in acute ischemic stroke patients who were treated within 3 hours of symptom onset. Since the publication of this trial’s primary results, a number of post hoc subgroup analyses have supported the general efficacy of t-PA. A novel analysis explored how the effect of t-PA varied according to the baseline NIH Stroke Scale (NIHSS) value, which is a measure of stroke severity. It stratified patients into approximate quintiles and, following convention, estimated the effect of t-PA within each quintile independently from the other effects. In this section, we replicate this analysis and then apply a fully Bayesian hierarchical model in order to demonstrate how shrinkage can potentially improve the accuracy of these subgroup estimates.

Methods
The NINDS t-PA trial randomly assigned patients to t-PA or placebo. The primary outcome was neurological status at 90 days, with a favorable outcome defined as minimal or no disability according to 4 measures of neurological status: the modified Rankin scale, the Barthel Index, the NIHSS, and the Glasgow Outcome Scale. The overall result, which was calculated using a global test statistic that combined information from all 4 measures, was an odds ratio of 1.7 (95% CI, 1.2–2.6; P = 0.008) in favor of t-PA. In a post hoc subgroup analysis conducted by outside investigators, patients were stratified by baseline NIHSS (higher values indicate greater stroke severity): 0–5 (Quintile 1), 6–10 (Quintile 2), 11–15 (Quintile 3), 16–20 (Quintile 4), and >20 (Quintile 5). These investigators reported the subgroup effects for all 4 measures of neurological status separately, but we will present only the results using a modified Rankin scale < 2 as the definition of a favorable outcome.

We estimated the odds ratio after treatment with t-PA in comparison to placebo, using independent and fully Bayesian hierarchical modeling approaches. The independent approach assumed no relationship among the effects in the various subgroups and used the Fisher exact method. A fully Bayesian hierarchical modeling approach assumed that the effects were exchangeable. Although therapies are generally more efficacious when disease severity is greater, the risk of intracranial hemorrhage following treatment with t-PA increases with stroke severity, so a priori it would have been difficult to predict how the effect of t-PA varied with baseline stroke severity. Thus the exchangeability assumption is reasonable. Similar to the above immunosuppressive therapy meta-analysis example, the logarithm of the odds ratios were assumed to be drawn from a normal hyperdistribution, the prior distribution for the hypermean was a noninformative Normal (0, 10000) distribution, and the prior distribution for the square root of the hypervariance was a noninformative half-Normal (0, 10000) distribution. The posterior distribution was approximated using MCMC simulation with 3 chains containing 5000 samples and a burn-in of 1000. Convergence was assessed using the Gelman-Rubin statistic.

Results
A total of 622 patients were included in the analysis, and the sample sizes within the baseline NIHSS quintiles ranged from 58 to 150. In both the t-PA and placebo arms, the proportion of patients with a favorable outcome was inversely related to baseline stroke severity (Table 2). The independent and fully Bayesian hierarchical estimates for the odds ratio within each NIHSS quintile are shown in Figure 5. The independent odds ratio estimates were all greater than 1 and favor t-PA, except for that of Quintile 1, which equaled 0.85 (95% CI, 0.13–4.2). The independent estimates generally lacked precision because, like many trials, the NINDS t-PA was not powered to precisely estimate subgroup effects.

The fully Bayesian hierarchical estimates were shrunk toward an estimated hypermean value of 1.9, which may be interpreted as the t-PA effect averaged across all baseline NIHSS quintiles. The independent odds ratio estimates that were larger than the hypermean (Quintiles 2, 3, and 5) were shrunk downward, whereas those smaller than the hypermean (Quintiles 1 and 4) were pulled upward. In addition, the hierarchical estimates were more precise than the indepen-

<table>
<thead>
<tr>
<th>Baseline NIHSS Quintile</th>
<th>Treatment</th>
<th>Favorable Outcome*/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–5 (Quintile 1)</td>
<td>t-PA</td>
<td>33/42</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>13/16</td>
</tr>
<tr>
<td>6–10 (Quintile 2)</td>
<td>t-PA</td>
<td>46/67</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>38/83</td>
</tr>
<tr>
<td>11–15 (Quintile 3)</td>
<td>t-PA</td>
<td>27/65</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>15/66</td>
</tr>
<tr>
<td>16–20 (Quintile 4)</td>
<td>t-PA</td>
<td>21/73</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>14/70</td>
</tr>
<tr>
<td>&gt;20 (Quintile 5)</td>
<td>t-PA</td>
<td>6/63</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>3/77</td>
</tr>
</tbody>
</table>

NIHSS indicates National Institute of Health Stroke Scale; t-PA, tissue plasminogen activator.
*A favorable outcome was defined as a score of 0 or 1 on the modified Rankin Scale at 90 days.

From Ingall et al.36
in the first section, we presented a clinical example in which data from 2 trials (COMPARE and SPIRIT IV) and 2 additional assumptions, and preferred in some circumstances. For example, the odds ratio of a favorable outcome comparing t-PA to placebo was estimated using 2 approaches (1): an independent approach that used the Fisher exact method and (2) a fully Bayesian hierarchical modeling approach. Relative to the independent estimates, the hierarchical estimates are shrunken toward the hypermean (dotted line) because information is borrowed across the subgroups.

The hierarchical estimates (and the independent estimates) did not vary monotonically, that is, they did not increase or decrease uniformly as baseline NIHSS quintile increased. Exchangeability was a relatively weak assumption that ignored the inherent ordering of the quintiles and thus permitted a large degree of flexibility. For example, the odds ratio in Quintile 2 was not restricted to values between those in Quintiles 1 and 3. Approaches that consider the ordinal nature of the quintiles may be more efficient, albeit at the expense of additional assumptions, and preferred in some circumstances. For example, a logistic regression model that includes baseline NIHSS as a numeric variable, assigned treatment, and their interaction term assumes that the t-PA effect varies monotonically and is additive on a logarithmic scale. Additional discussions regarding the application of hierarchical modeling to subgroup analysis may be found in Lipsky and Jones.

Clinical Example: Revisiting Drug-Eluting Stents
In the first section, we presented a clinical example in which subgroups (diabetes and no diabetes) were presented. Because of differences in populations and practice settings, it is unlikely that the true overall treatment effects in the 2 trials are identical. On the other hand, if it would not have been possible to predict how the true overall effects differed before observing the data, an assumption of exchangeability of the overall effects is reasonable. Similarly, before the trials were conducted, if there was uncertainty regarding the relative magnitudes of the treatment effects in the diabetic and nondiabetic populations (ie, one could not have predicted whether the treatment effect was larger in diabetics or nondiabetics), then an assumption of exchangeability across subgroups is appropriate. Note that exchangeability pertains to the relative risk in the 2 subgroups and not to the baseline risk, which is known to be greater in diabetics and is thus not exchangeable across the subgroups. It is possible to structure a hierarchical model that incorporates the above assumptions of exchangeability. However, with only 2 studies, it is not possible to reliably estimate the between-study hypervariance from the data, and it would be necessary to use a strongly informative prior for the between-study hypervariance. For the same reason, a strongly informative prior distribution for the between-subgroup hypervariance would also be required.

Conclusion
Bayesian hierarchical modeling may be used to combine information from different sources in a coherent manner that is analogous to the clinical synthesis of information. It has the potential to serve an important role in multiple settings, including comparative effectiveness research, the development of treatment guidelines, and insurance coverage decisions. Although this review focused on assessing the effectiveness of interventions, hierarchical modeling is similarly useful for investigating other topics in outcomes research, such as safety, equity, and efficiency. Fortunately, improvements in computing power and the availability of software have increased the ease and accessibility of this technique. As a result, use of this methodology is likely to grow in health sciences research.

Sources of Funding
This publication was supported by a fellowship grant from the Emergency Medicine Foundation to Dr Kwok. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the Emergency Medicine Foundation.

Disclosures
Dr Lewis is the Senior Medical Scientist for Berry Consultants, LLC, a statistical consulting group that specializes in the application of Bayesian methods, including Bayesian hierarchical modeling, to the design and analysis of clinical trials.

References
Bayesian Hierarchical Modeling and the Integration of Heterogeneous Information on the Effectiveness of Cardiovascular Therapies
Heemun Kwok and Roger J. Lewis

doi: 10.1161/CIRCOUTCOMES.111.960724
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/4/6/657

Data Supplement (unedited) at:
http://circoutcomes.ahajournals.org/content/suppl/2011/11/15/4.6.657.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/
BUGS Code for Immunosuppressive Therapy Example

The following code may be used to conduct the fully Bayesian hierarchical modeling analysis of immunosuppressive therapy in idiopathic dilated cardiomyopathy using the software programs OpenBUGS and WinBUGS.

# Definition of variables
# N: number of trial populations
# control_s: number of clinical responders in control arm
# control_n: total number of subjects in control arm
# treat_s: number of clinical responders in active treatment arm
# treat_n: total number of subjects in active treatment arms
# control_p and treat_p: the probabilities of clinical response in the control and treatment arms, respectively
# theta: logarithm of the odds ratio
# mu: hypermean
# tau: precision of the hyperdistribution (reciprocal of the hypervariance)
# sigma: standard deviation of the hyperdistribution (square root of the hypervariance)

# Model 1: Normal hyperdistribution and Half-normal (0, 10000) prior distribution for sigma, the square root of the hypervariance

model{
  for(i in 1:N) {
    control_s[i] ~ dbin(control_p[i],control_n[i]);
    treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
    control_p[i] ~ dbeta(1,1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] ~ dnorm(mu, tau);
  }
  mu ~ dnorm(0, 0.001)
  tau <- pow(sigma,-2)
  sigma ~ dnorm(0, 0.0001)I(0,)
  # Half-normal prior distribution with variance of 10000 and precision of 0.0001
}

# Enumerate the data set
list(N = 5, control_s = c(0,9,5,14,10), control_n = c(42,32,20,35,41), treat_s = c(38,18,8,27,30), treat_n = c(43,27,22,54,37) )

# Initial values for all random variables
list(control_p=c(0.2,0.2,0.2,0.2,0.2), theta = c(0,0,0,0,0), mu = 0, sigma=1)
Bayesian hierarchical modeling of heterogeneous information
Kwok and Lewis

#Model 2: Student-t hyperdistribution and Half-normal (0, 10000) prior distribution for sigma

model{
  for(i in 1:N) {
    control_s[i] ~ dbin(control_p[i],control_n[i]);
    treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
    control_p[i] ~ dbeta(1,1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] ~ dt(mu, tau, 4);  #Student-t hyperdistribution with #4 degrees of freedom
  }
  mu ~ dnorm(0, 0.001)
  tau <- pow(sigma,-2)
  sigma ~ dnorm(0, 0.0001)I(0,1)  #Half-normal prior distribution with variance #of 10000 and precision of 0.0001
}

#The list of data and list of initial values are the same as for Model 1.
list(N = 5, control_s = c(0,9,5,14,10), control_n = c(42,32,20,35,41), treat_s = c(38,18,8,27,30),
     treat_n = c(43,27,22,54,37) )

#Initial values for all random variables
list(control_p=c(0.2,0.2,0.2,0.2,0.2), theta = c(0,0,0,0,0), mu = 0, sigma=1)
Bayesian hierarchical modeling of heterogeneous information
Kwok and Lewis

#Model 3: Normal hyperdistribution and Uniform (0, 100) prior distribution for sigma

model{
for(i in 1:N) {
  control_s[i] ~ dbin(control_p[i],control_n[i]);
  treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
  control_p[i] ~ dbeta(1,1);
  logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
  theta[i] ~ dnorm(mu, tau);
}
mu ~ dnorm(0, 0.001)
tau <- pow(sigma,-2)
sigma ~ dunif(0, 100)   #Uniform (0,100) distribution
}

#Lists of data and initial values do not change from above:
list(N = 5, control_s = c(0,9,5,14,10), control_n = c(42,32,20,35,41), treat_s = c(38,18,8,27,30),
treat_n = c(43,27,22,54,37) )
list(control_p=c(0.2,0.2,0.2,0.2,0.2), theta = c(0,0,0,0,0), mu = 0, sigma=1)
Bayesian hierarchical modeling of heterogeneous information
Kwok and Lewis

#Model 4: Normal hyperdistribution and Inverse Gamma (0.001, 0.001) prior distribution

model{
for(i in 1:N) {
  control_s[i] ~ dbin(control_p[i],control_n[i]); #binomial likelihood, control arm
  treat_s[i] ~ dbin(treat_p[i],treat_n[i]);  #binomial likelihood, treatment arm
  control_p[i] ~ dbeta(1,1);    #non-informative beta prior
    #distribution for the control rate
  logit(treat_p[i]) <- logit(control_p[i]) + theta[i]; #define log-odds  treatment rate
  theta[i] ~ dnorm(mu, tau);    #normal hyperdistribution
}
mu ~ dnorm(0, 0.001)  #prior for hypermean (in BUGS, the parameters of the normal #distribution are the mean and #precision)

  tau ~ dgamma(0.001, 0.001) #Gamma prior for the precision, tau, which is equivalent to an inverse gamma prior for the #hypervariance

  sigma <- pow(tau,-0.5) #sigma is 1 / (square root of tau)

}

#Enumerate the data set

list(N = 5, control_s = c(0,9,5,14,10), control_n = c(42,32,20,35,41), treat_s = c(38,18,8,27,30),
    treat_n = c(43,27,22,54,37) )

#Initial values for all random variables

list(control_p=c(0.2,0.2,0.2,0.2,0.2), theta = c(0,0,0,0,0), mu = 0, tau=1)
Bayesian hierarchical modeling of heterogeneous information
Kwok and Lewis

Model 5: Normal hyperdistribution and Uniform (0, 2.3) prior distribution for sigma

model{
for(i in 1:N) {
  control_s[i] ~ dbin(control_p[i], control_n[i]);
  treat_s[i] ~ dbin(treat_p[i], treat_n[i]);
  control_p[i] ~ dbeta(1, 1);
  logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
  theta[i] ~ dnorm(mu, tau);
}
mu ~ dnorm(0, 0.001)
tau <- pow(sigma, -2)
sigma ~ dunif(0, 2.3) \quad \# \text{Uniform}(0, 2.3) \text{ distribution}
}

Model 5 uses the same data and initial values as Model 3:

list(N = 5, control_s = c(0, 9, 5, 14, 10), control_n = c(42, 32, 20, 35, 41),
treat_s = c(38, 18, 8, 27, 30), treat_n = c(43, 27, 22, 54, 37))

list(control_p = c(0.2, 0.2, 0.2, 0.2, 0.2), theta = c(0, 0, 0, 0), mu = 0, sigma = 1)
BUGS Code for NINDS t-PA Therapy Example

The WinBUGS code to perform the subgroup analysis of the NINDS t-PA stroke trial is provided below. Similar to the variables definitions above, “control_s” and “control_n” represent the number of subjects with a favorable outcome and the total number of subjects in the placebo (control) arm, respectively. “treat_s” and “treat_n” are the corresponding values in the t-PA (active treatment) arm. The code is identical to Model 1 above.

model{
for(i in 1:N) {
control_s[i] ~ dbin(control_p[i],control_n[i]);
treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
control_p[i] ~ dbeta(1,1);
logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
theta[i] ~ dnorm(mu, tau);
}
mu ~ dnorm(0, 0.001)
tau <- pow(sigma,-2)
sigma ~ dnorm(0, 0.0001)I(0,)
}
#Enumerate the data set (in order of quintiles)
list(N = 5, control_s = c(13,38,15,14,3), control_n = c(16,83,66,70,77), treat_s = c(33,46,27,21,6), treat_n = c(42,67,65,73,63) )
#Initial values for all random variables
list(control_p=c(0.5,0.5,0.5,0.5,0.5), theta = c(0,0,0,0,0), mu = 0, sigma=1)
#Contents

#Example meta-analysis of immunosuppressive therapy in idiopathic cardiomyopathy

I. Define Dataset
II. Create dataframes for results
III. Independent Model
IV. Fixed-effects analysis (Mantel-Haenszel)
V. Fully Bayesian analysis, Model 1
VI. Add covariate publication year to Model 1
VII. Empirical Bayes analysis
VIII. Sensitivity analysis (Models 2-5)
IX. Figure 3
X. Figure 4

#Example subgroup analysis of NINDS t-PA stroke trial

I. Define Dataset
II. Analysis (Independent and fully Bayesian)
III. Figure 5

# I. Define Dataset

N_dz=5; #number of trial populations

#Summary data from trials (1.Frustaci; 2.Parrillo: reactive; 3. Parrillo: nonreactive; 4.Mason; 5 Wojnicz)
control_s = c(0,9,5,14,10);
treat_s = c(38,18,8,27,30);
control_n = c(42,32,20,35,41);
treat_n = c(43,27,22,54,37);

covariate: publication year
year.log = log(year - 1989 + 1) #log transformation
year.st = year.log - mean(year.log) #mean-centered

# II. Create dataframes for results

# results = formatted results of primary analysis
n.r=4 #number of rows
results <- data.frame(character(n.r), character(n.r), character(n.r), character(n.r), character(n.r));
rownames(results) <- NA
rownames(results) <- c('Independent', 'Fully Bayesian', 'Empirical Bayes', 'Mantel-Haenszel')
colnames(results) <- studynames <- c("Frustaci", "Reactive", "Non-reactive", "Mason", "Wojnicz")
#results.sens = formatted results of sensitivity analysis
#IG, t dist, uniform [0,100], half-Normal, uniform[0,2.3]
n.r.2=5   #number of rows
results.sens <- data.frame(character(n.r.2), character(n.r.2),character(n.r.2), character(n.r.2), character(n.r.2), character(n.r.2));
results.sens[,] <- NA
colnames(results.sens) <- c('Form', 'Prior', "Frustaci", "Reactive", "Non-reactive", "Mason", "Wojnicz")

#Vectors used to plot Figure 3
top.f <- vector(length=N_dz+1)
bottom.f <- vector(length=N_dz+1)
center.f <- vector(length=N_dz)
top.b <- bottom.b <- center.b <- top.eb <- bottom.eb <- center.eb <-
vector(length=N_dz)

# III Independent Model

library(epitools)  #load package epitools

#loop to calculate independent OR for each disease
for (i in 1:N_dz) {
  #first setup contingency table
  table.1 <- matrix(c(control_n[i]-control_s[i], control_s[i], treat_n[i]-treat_s[i], treat_s[i]), nrow=2, ncol=2, byrow=T)

  #calculate OR using Fisher method
  x <- oddsratio(table.1, method="fisher")
  #for Frustaci trial (i=1), add 0.5 to each cell and use Wald method to estimate OR
  if (i==1) {x <- oddsratio(table.1 + 0.5, method="wald")}

  #round to 2 significant digits and format
  est <- signif(x$measure[2,], 2)  #round to 2 significant digits
  results[i,i] <- paste(est[1], " (", est[2], ",-", est[3], ",")", sep=""")

  #add results to vectors for figure 3
  center.f[i] <- log(x$measure[2,1]);  #logOR point estimate
  top.f[i] <- log(x$measure[2,3]);  #logOR upper limit
  bottom.f[i] <- log(x$measure[2,2]);  #logOR lower limit
}

# IV. Fixed-effects analysis (Mantel-Haenszel)

#format data into array to be used for MH test
table.2 <- array(NA, dim=c(2,2,5))
table.2[1,1,] <- control_n - control_s
table.2[1,2,] <- control_s
table.2[2,1,] <- treat_n - treat_s
table.2[2,2,] <- treat_s

#Mantel-Haenszel test
CMH <-mantelhaen.test(table.2)
# Format results and add to dataframe results

est <- signif( c(CMH$estimate, CMH$conf.int),2)
for (i in 1:N_dz) {
  results[4,i] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
}

# add results to vectors for Figure 3
center.f[N_dz+1] <- log(CMH$estimate)
bottom.f[N_dz+1] <- log(CMH$conf.int[1])
top.f[N_dz+1] <- log(CMH$conf.int[2])

# V. Fully Bayes Model: Normal hyperdistribution, Half-normal distribution for sigma
# BUGS model

my.model.1 <- function(){
  for(i in 1:N_dz) {
    control_s[i] ~ dbin(control_p[i],control_n[i]);
    treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
    control_p[i] ~ dbeta(1,1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] ~ dnorm(mu, tau);
  }
  mu ~ dnorm(m, prec)
  tau <- pow(sigma,-2)
  sigma ~ dnorm(0,prior.prec) %_%I(0,)
}
filename.half <- file.path(tempdir(), "modell.bug")
write.model(my.model.1, filename.half)

# Model parameters
m=0; prec=0.001;
prior.sigma=100
mydata = list("N_dz", "control_s", "control_n", "treat_s", "treat_n", "m","prec", "prior.prec")
myparameters=c("theta", "mu", "sigma")

# Initial Values
myinits = list(
  list(theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
    control_p=runif(N_dz ) ),
  list(theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
    control_p=runif(N_dz ) ),
  list(theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
    control_p=runif(N_dz ) )
)

# Call OpenBUGS
bugs.out1 <- openbugs(data=mydata, inits=myinits, parameters.to.save=myparameters, model.file = filename.half, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1, DIC = TRUE, bugs.directory = "c:/Program Files/OpenBUGS/", working.directory = NULL, digits = 3, over.relax = FALSE, seed=NULL)

# add to results dataframe, Fully Bayesian
for (i in 1:N_dz) {
  est <- signif(exp(bugs.out1$summary[i,c(1,3,7)]), 2)
results[2,i] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")

# Extract results and add to results.sens
results.sens[1,1:2] = c('Normal', 'Half-normal(0,10000)')
for (i in 1:N_dz) {
  est <- signif(exp(bugs.out1$summary[i,c(1,3,7)]), 2)
  results.sens[1,i+2] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
}

#add results to plotting vectors
bottom.b <- bugs.out1$summary[1:N_dz,3]; #2.5%
center.b <- bugs.out1$summary[1:N_dz,1]; #mean
top.b <- bugs.out1$summary[1:N_dz,7]; #97.5%

#VI. Add covariate publication year to model

#df.cov = dataframe for results
df.cov <- data.frame(character(2), character(2), character(2), character(2), character(2));
df.cov[,] <- NA
rownames(df.cov) <- c('No covariates', 'Year')
colnames(df.cov) <- studynames <- c("Frustaci", "Reactive", "Non-reactive", "Mason", "Wojnicz")

#add results from Model 1 (no covariates) to results
for (i in 1:N_dz) {
  est <- signif(exp(bugs.out1$summary[i,c(1,3,7)]), 2)
  df.cov[i,1] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
}

#BUGS model with log year as covariate
my.model.cov <- function(){
  for(i in 1:N_dz) {
    control_s[i] ~ dbin(control_p[i],control_n[i]);
    treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
    control_p[i] ~ dbeta(1,1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] <- alpha[i] + beta*year.st[i]  #add year
    alpha[i] ~ dnorm(mu, tau);
  }
  mu ~ dnorm(m,prec)
  tau <- pow(sigma,-2)
  sigma ~ dnorm(0,prior.prec)% %I(0,)
  beta ~ dnorm(0,0.001)   #noninformative prior for regression coefficient
}
filename.cov <- file.path(tempdir(), "cov.bug")
write.model(my.model.cov, filename.cov)

#parameters and data list
m=0; prec=0.001;
prior.sigma=100
prior.prec = (1/prior.sigma)^2
mydata = list("N_dz", "control_s", "control_n", "treat_s", "treat_n", "m","prec", "prior.prec", "year.st")
myparameters=c("theta", "mu", "sigma")

#initial values
myinits = list(
    list(alpha = rnorm(N_dz,0,2), mu=runif(1,0,2), sigma=runif(1,0,prior.sigma),
        control_p=runif(N_dz ) ),
    list(alpha = rnorm(N_dz,0,2), mu=runif(1,0,2), sigma=runif(1,0,prior.sigma),
        control_p=runif(N_dz ) ),
    list(alpha = rnorm(N_dz,0,2), mu=runif(1,0,2), sigma=runif(1,0,prior.sigma),
        control_p=runif(N_dz ) )
)

#call OpenBUGS
bugs.out.cov <- openbugs(data=mydata, inits=myinits, parameters.to.save=myparameters, model.file = filename.cov, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1, DIC = TRUE, bugs.directory = Files/OpenBUGS/", working.directory = NULL, digits = 3, over.relax = FALSE, seed=NULL)

#Extract results
for (i in 1:N_dz) {  
est <- signif(exp(bugs.out.cov$summary[i,c(1,3,7)]), 2)  
df.cov[2,i] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
}

# VII. Empirical Bayes analysis
library(lme4)

#first reshape data into "long" format
n.tot <- sum(control_n, treat_n)
long.df <- data.frame(study = factor(NA, levels=studynames), treat = logical(n.tot), response=logical(n.tot));
x = 1;    #indexing variable
for (i in 1:N_dz) {
    n = control_n[i] + treat_n[i]
    long.df$response[x:(x+n-1)] <- c(rep(1, control_s[i]), rep(0, (control_n[i]-control_s[i])), rep(1, treat_s[i]), rep(0, (treat_n[i]-treat_s[i]) )
    long.df$treat[x:(x+n-1)] <- c(rep(0, control_n[i]), rep(1, treat_n[i]))
    long.df$study[x:(x+n-1)] <- rep(studynames[i], n)
    x <- x + n
}

#call glmer
#treatment is random
#set number of points for adaptive Gauss-Hermite approximation to 5
model.eb = glmer(response ~ -1 + study + (treat + 0|study), nAGQ=5, family="binomial", data=long.df)

#extract empirical Bayes estimates and CI
thetas <- ranef(model.eb,postVar=TRUE)    #study-specific estimates
thetas.sd <- sqrt(attr(thetas$study, "postVar")[1,1])

#add estimate to results in third row
for (i in 1:N_dz) {
    point <- thetas$study$treat[i]
limits <- point + c(-1.96, 1.96) * thetas.sd[i]
est <- signif(exp(c(point, limits)), 2)
results[3, i] <- paste(est[1], " (", est[2], ",-", est[3], ")", sep="")
center.eb[i] <- point
bottom.eb[i] <- limits[1]
top.eb[i] <- limits[2]

# VIII. Sensitivity Analysis

# Model 2: Model t dist, half-Normal prior on sigma
deg.free = 4;

model.t <- function(){
  for(i in 1:N_dz) {
    control_s[i] ~ dbin(control_p[i], control_n[i]);
    treat_s[i] ~ dbin(treat_p[i], treat_n[i]);
    control_p[i] ~ dbeta(1, 1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] ~ dt(mu, tau, deg.free);
  }
  mu ~ dnorm(m, prec)
tau <- pow(sigma, -2)
sigma ~ dnorm(0, prior.sigma) %_%I(0,)
}

filenamet <- file.path(tempdir(), "modelt.bug")
write.model(model.t, filenamet)
#file.show(filenamet)

m=0; prec=0.001;
prior.sigma=100
prior.prec = (1/prior.sigma)^2

mydata = list("N_dz", "control_s", "control_n", "treat_s", "treat_n", "m", "prec", 'deg.free', 'prior.prec')
myparameters = c("theta", "mu", "sigma")
myinits = list(
  list(theta = rnorm(N_dz, 0, 2), mu=rnorm(1,0,2), sigma=rnorm(1,0,prior.sigma),
       control_p=rnorm(N_dz)),
  list(theta = rnorm(N_dz, 0, 2), mu=rnorm(1,0,2), sigma=rnorm(1,0,prior.sigma),
       control_p=rnorm(N_dz)),
  list(theta = rnorm(N_dz, 0, 2), mu=rnorm(1,0,2), sigma=rnorm(1,0,prior.sigma),
       control_p=rnorm(N_dz)))

bugs.out.t <- openbugs(data=mydata, inits=myinits, parameters.to.save=myparameters,
                       model.file = filenamet, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1,
                       DIC = TRUE, bugs.directory = "c:/Program Files/OpenBUGS/", working.directory = NULL,
                       digits = 3, over.relax = FALSE, seed=NULL)

# Extract results
results.sens[2,1:2] = c('Student-t', 'half-N(0,10000)')
for (i in 1:N_dz) {
est <- signif(exp(bugs.out.t$summary[i,c(1,3,7)]), 2)
results.sens[2,i+2] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
"

# VIII B. Model 3: Uniform distribution for sigma U[0,100]

#Write BUGS model
my.model.u <- function(){
  for(i in 1:N_dz) {
    control_s[i] ~ dbin(control_p[i],control_n[i]);
    treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
    control_p[i] ~ dbeta(1,1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] ~ dnorm(mu, tau);
  }
  mu ~ dnorm(m,prec)
  tau <- pow(sigma,-2)
  sigma ~ dunif(0,prior.sigma)
}
model.u <- file.path(tempdir(), "model.bug")
write.model(my.model.u, model.u)

#parameters
m=0; prec=0.001; prior.sigma = 100
mydata = list("N_dz", "control_s", "control_n", "treat_s", "treat_n", "m", "prec", "prior.sigma")
myparameters=c("theta", "mu", "sigma")

#Initial values generated randomly
myinits = list(
  theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
  control_p=runif(N_dz ),
  theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
  control_p=runif(N_dz ),
  theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
  control_p=runif(N_dz ))

#Call OpenBUGS
bugs.3 <- openbugs(data=mydata, inits=myinits, parameters.to.save=myparameters, model.file = model.u, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1, DIC = TRUE, bugs.directory = "c:/Program Files/OpenBUGS/", working.directory = NULL, digits = 3, over.relax = FALSE, seed=NULL)

#Extract results and add to results.sens in third row
results.sens[3,1:2] = c('Normal', 'Uniform(0,100)')
for (i in 1:N_dz) {
  est <- signif(exp(bugs.3$summary[i,c(1,3,7)])), 2)
  results.sens[3,i+2] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
  }

# VIIC. Fully Bayes Model 4: Normal hyperdistribution, prior hypervariance
IG(0.001,0.001)

#write OpenBUGS model as an R function
model.IG <- function()
{
  for(i in 1:N_dz) {
    control_s[i] ~ dbin(control_p[i],control_n[i]);
    treat_s[i] ~ dbin(treat_p[i],treat_n[i]);
    control_p[i] ~ dbeta(1,1);
    logit(treat_p[i]) <- logit(control_p[i]) + theta[i];
    theta[i] ~ dnorm(mu, tau);
  }
  mu ~ dnorm(m,prec)
  tau ~ dgamma(0.001,0.001)
  sigma <- pow(tau,-0.5)
}
filename3 <- file.path(tempdir(), "model3.bug")
write.model(model.IG, filename3)

#model parameters and data to send to BUGS
m=0; prec=0.001
mydata = list("N_dz", "control_s", "control_n", "treat_s", "treat_n", "m","prec")
myparameters=c("theta", "mu", "sigma")

#initial values for 3 chains generated randomly
myinit = list(
  list(theta = rnorm(N_dz,0,2), mu=runif(1,0,100),
    control_p=runif(N_dz ),),
  list(theta = rnorm(N_dz,0,2), mu=runif(1,0,100),
    control_p=runif(N_dz ),),
  list(theta = rnorm(N_dz,0,2), mu=runif(1,0,100),
    control_p=runif(N_dz ))
)

#Call openbugs
bugs.out4 <- openbugs(data=mydata, inits=myinit, parameters.to.save=myparameters, model.file = filename3, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1,
DIC = TRUE, bugs.directory = "c:/Program Files/OpenBUGS/", working.directory = NULL, digits = 3, over.relax = FALSE, seed=NULL)

results.sens[4,1:2] = c('Normal','IG (0.001,0.001)')
for (i in 1:N_dz) {
  est <- signif(exp(bugs.out4$summary[i,c(1,3,7)]), 2)
  results.sens[4,i+2] <- paste(est[1], " (", est[2], ",", est[3], ")", sep="")
}

# VIID. Model 5: uniform distribution for SD of hyperdistribution = U[0,2.3]

#parameters
m=0; prec=0.001; prior.sigma = 2.3
mydata = list("N_dz", "control_s", "control_n", "treat_s", "treat_n", "m","prec", "prior.sigma")
myparameters=c("theta", "mu", "sigma")

#generate random initial values
myinit = list(
  list(theta = rnorm(N_dz,0,2), mu=runif(1,0,100), sigma=runif(1,0,prior.sigma),
    control_p=runif(N_dz ),),
  list(theta = rnorm(N_dz,0,2), mu=runif(1,0,100), sigma=runif(1,0,prior.sigma),
    control_p=runif(N_dz )),
list(theta = rnorm(N_dz,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
control_p=runif(N_dz ))}

#call BUGS
bugs.5 <- openbugs(data=mydata, inits=myinits, parameters.to.save=myparameters,
model.file = model.u, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1,
DIC = TRUE, bugs.directory = "c:/Program Files/OpenBUGS/", working.directory =
NULL, digits = 3, over.relax = FALSE, seed=NULL)

#add estimate to results.sens in third row
results.sens[5,1:2] = c('Normal', 'Uniform(0,2.3')
for (i in 1:N_dz) {
    est <- signif(exp(bugs.5$summary[i,c(1,3,7)]), 2)
    results.sens[5,i+2] <- paste(est[1], " (", est[2], ",-", est[3], ")", sep=")
}

#IX. Figure 3: Plot comparing analytic approaches -------------------------
library(plotrix)
size = 1;
x.axis.lim = log(10000)
op <- par(mar=c(4,10,2,2)+0.1, font=2, font.axis=2, xpd=NA, cex=1)

#Frequentist Error Bars: Fisher + MH
N = seq(N_dz+1:1, by=-1);
plotCI(y=N+0.2, x=center.f, li=bottom.f, ui=top.f, pch=c(16,16,16,16,16,8),
err="x", xlab="", ylab="", gap=0, axes=F, xlim=c(-1.4,x.axis.lim), ylim=c(1,7),
sfrac=0.005, cex=size*1.2, lwd=2)
mtext(at = -2.9, adj = 0, side=1,"Odds Ratio for Improvement in Left
Ventricular Ejection Fraction",line=2.5,cex=size)

#add fully Bayesian error bars
M = seq(N_dz+1,2,by=-1)  #there are only 9 error bars to plot
plotCI(y=M, x=center.b, li=bottom.b, ui=top.b, pch=17, err="x", xlab="", ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)

#add empirical Bayesian error bars
M = seq(N_dz+1,2,by=-1)  #there are only 9 error bars to plot
plotCI(y=M-0.2, x=bottom.eb, li=bottom.eb, ui=top.eb, pch=15, err="x",
lab="", ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)

#legend and axes
legend(5.2,4.3, c("Independent", "Fully Bayesian", "Empirical Bayes "), cex=size,
pch=c(16,17,15), pt.lwd=2, pt.cex=1.2,lty=c(1,1,1), horiz=F, bty="o")
ticks = log(c(0.1,1,10,100, 1000,10000))
axis(1,at=ticks,label=exp(ticks), cex.axis=size*0.9, lwd=2)
full.names = c("Frustaci", "Parrillo: Reactive", "Parrillo: Non-Reactive", "Mason",
"Wojnicz")
    axis(2, at=N[1:5], labels=c(full.names), cex.axis=size, las=2, tick=F, pos=-6, hadj=0)  #las orients labels
axis(2, at=1.2, labels="Common Estimate", cex.axis=size, las=2, tick=F, pos=-6, adj=0)

lines(x=c(0,0), y=c(0.8,6.4), lty=2, lwd=2)

par(op)  # resets par

# X. Figure 4: Sensitivity analysis

# dataframe for sensitivity analysis plot (Figure 4)
sens = array(NA, dim=c(N_dz,3,n.r.2))
rownames(sens) = c(studyNames)
colnames(sens) = c('top','bottom','center')

# first enter results of Model 1: Normal hyperdistribution with half-Normal prior
sens[,1,1] = top.b
sens[,2,1] = bottom.b
sens[,3,1] = center.b

# Results from Model 2 (t distribution)
sens[,1,2] <- bugs.out.t$sensitivity[1:N_dz,7]; #97.5%
sens[,2,2] <- bugs.out.t$sensitivity[1:N_dz,3]; #2.5%
sens[,3,2] <- bugs.out.t$sensitivity[1:N_dz,1]; # mean

# results from Model 3 (U[0,100])
sens[,1,3] <- bugs.3$sensitivity[1:N_dz,7]; #97.5%
sens[,2,3] <- bugs.3$sensitivity[1:N_dz,3]; #2.5%
sens[,3,3] <- bugs.3$sensitivity[1:N_dz,1]; # mean

# results from Model 4 (IG 0.001,0.001)
sens[,1,4] <- bugs.out4$sensitivity[1:N_dz,7]; #97.5%
sens[,2,4] <- bugs.out4$sensitivity[1:N_dz,3]; #2.5%
sens[,3,4] <- bugs.out4$sensitivity[1:N_dz,1]; # mean

# results from Model 5
sens[,1,5] <- bugs.5$sensitivity[1:N_dz,7]; #97.5%
sens[,2,5] <- bugs.5$sensitivity[1:N_dz,3]; #2.5%
sens[,3,5] <- bugs.5$sensitivity[1:N_dz,1]; # mean

size = 1;
x.axis.lim = log(10000)

op <- par(mar=c(4,10,2,2)+0.1, font=2, font.axis=2, xpd=NA, cex=1)

N = c(5,4,3,2,1);
plotCI(y=N+0.3, x=sens[,3,1], li=sens[,2,1], ui=sens[,1,1], pch=16, err="x",
      xlab="", ylab="", gap=0, axes=F, xlim=c(-1.4,x.axis.lim), ylim=c(0.5,5),
      sfrac=0.005, cex=size*1.2, lwd=2)

mtext(at = -2.9, adj = 0, side=1,"Odds Ratio for Improvement in Left Ventricular Ejection Fraction",line=2.5,cex=size)

plotCI(y=N+0.15, x=sens[,3,2], li=sens[,2,2], ui=sens[,1,2], pch=17, err="x",
      xlab="",ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)

plotCI(y=N, x=sens[,3,3], li=sens[,2,3], ui=sens[,1,3], pch=8, err="x",
      xlab="",ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)

plotCI(y=N-0.15, x=sens[,3,4], li=sens[,2,4], ui=sens[,1,4], pch=15, err="x",
      xlab="",ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)
plotCI(y=N-0.3, x=sens[,3,5], li=sens[,2,5], ui=sens[,1,5], pch=2, err="x", xlab="", ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)

#Legend
legend(6.5,3.45, c("Model 1", "Model 2", "Model 3", "Model 4", "Model 5"), cex=size, pch=c(16,17,8,15,2), pt.lwd=2, pt.cex=1.2,lty=1, horiz=F, bty="o")

#X-Axis
ticks = log(c(0.1,1,10,100, 1000,10000))
axis(1,at=ticks,label=exp(ticks), cex.axis=size*0.9, lwd=2)

#Y-Axis
full.names = c("Frustaci", "Parrillo: Reactive", "Parrillo: Non-Reactive", "Mason", "Wojnicz")
axis(2, at=N[1:5], labels=c(full.names), cex.axis=size, las=2, tick=F, pos=-6, hadj=0) #las orients labels
lines(x=c(0,0), y=c(0.3, 5.3),lty=2, lwd=2)
par(op)  #resets par

#I. Define dataset
#Summary outcomes from : quintiles 1-5
N_sub=5;  #number of subgroups
tpa_n = c(42,67,65,73,63);  #tpa arm total
tpa_s = c(33,46,27,21,6);  #tpa arm success
p_n = c(16,83,66,70,77);  #placebo arm total
p_s = c(13,38,15,14,3);  #placebo arm success
res.tpa <- matrix(nrow=6, ncol=5)
colnames(res.tpa) = c('q1', 'q2', 'q3', 'q4', 'q5')
rownames(res.tpa) = c('indep.ll', 'indep.pt', 'indep.ul','bayes.ll', 'bayes.pt', 'bayes.ul')

#Analysis

#independent analysis
for (i in 1:N_sub) {
  table.tpa <- matrix(c(p_n[i]-p_s[i], p_s[i], tpa_n[i]-tpa_s[i], tpa_s[i]), nrow=2, ncol=2, byrow=T)  #contingency table
  x <- oddsratio(table.tpa, method="fisher") #add results to res.tpa
  res.tpa[c(2,1,3),i] <- log(x$measure[2,])
}

#fully Bayesian with half-Normal prior on sigma
model.tpa <- function(){
for(i in 1:N_sub) {
  p_s[i] ~ dBin(p_p[i],p_n[i]);
}
#Frequentist Error Bars: Fisher + MH

```
tpa_s[i] ~ dbin(tpa_p[i],tpa_n[i]);
p_p[i] ~ dbeta(1,1);
logit(tpa_p[i]) <- logit(p_p[i]) + theta[i];
theta[i] ~ dnorm(mu, tau);
}
mu ~ dnorm(m, prec)
tau <- pow(sigma,-2)
sigma ~ dnorm(0,prior.prec)*%I(0)
```

```
filename.tpa <- file.path(tempdir(), "model.tpa")
write.model(model.tpa, filename.tpa)
```

```
m=0; prec=0.001;
prior.sigma=100
prior.prec = (1/prior.sigma)^2
myparameters = list("N_sub", "p_s", "p_n", "tpa_s", "tpa_n", "m","prec", "prior.prec")
myinits = list(
  list(theta = rnorm(N_sub,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
       p_p=runif(N_sub) ),
  list(theta = rnorm(N_sub,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
       p_p=runif(N_sub) ),
  list(theta = rnorm(N_sub,0,2), mu=rnorm(1,0,2), sigma=runif(1,0,prior.sigma),
       p_p=runif(N_sub) ))

bugs.out.tpa <- openbugs(data=mydata, inits=myinits,
parameters.to.save=myparameters, model.file = filename.tpa, n.chains = 3, n.iter = 6000, n.burnin = 1000, n.thin = 1, DIC = TRUE, bugs.directory = Files/OpenBUGS/", working.directory = NULL, digits = 3, over.relax = FALSE, seed=NULL)
```

```
assign results to res_tpa
res.tpa[5,] <- bugs.out.tpa$summary[1:5,1]  #pt est
res.tpa[4,] <- bugs.out.tpa$summary[1:5,3]  #lower limit
res.tpa[6,] <- bugs.out.tpa$summary[1:5,7]  #upper limit
hypermean <- bugs.out.tpa$summary[6,1]
```

#III. Plot: Figure 5

```
library(plotrix)
size = 1;

op <- par(mar=c(5,4,2,3)+0.1, font=2, font.axis=2, xpd=NA, cex=1)
#Frequentist Error Bars: Fisher + MH
N = c(5,4,3,2,1)
plotCI(y=N+0.1, x=res.tpa[2,], li=res.tpa[1,], ui=res.tpa[3,], pch=16, err="x",
xlab="",ylab="", gap=0, axes=F, xlim=c(log(0.1),log(10)), ylim=c(0.5,6),
sfrac=0.005, cex=size*1.2, lwd=2, minbar=-0.7, maxbar=2.3)
mtext(at = 0, adj=0.5, side=1, "Odds Ratio for Favorable Outcome at 3 Months",
line=2.5, cex=size)
mtext(at = 3, adj=0.5, side=2, "Baseline NIHSS Quintile", line=3, cex=size)
```

#add Bayesian error bars
```
plotCI(y=N-0.1, x=res.tpa[5,], li=res.tpa[4,], ui=res.tpa[6,], pch=17, err="x",
xlab="",ylab="", gap=0, axes=F, add=T, sfrac=0.005, cex=size*1.2, lwd=2)
```
# plot annotation

legend(x=-1.9, y=3.8, c("Independent", "Fully Bayesian "), cex=size,
pch=c(16,17), pt.lwd=2, pt.cex=1.2, lty= c(1,1), horiz=F, bty="o")
ticks = log(c(0.125,0.25, 0.5,1,2,4,8, 16))
axis(1, at=ticks, label=exp(ticks), cex.axis=size*0.9, lwd=2)

y.label = c('0-5 (Q1)', '6-10 (Q2)', '11-15 (Q3)', '16-20 (Q4)', '>20 (Q5)')

axis(2, pos=-2.6, at=N[1:5], labels=c(y.label), cex.axis=size, las=2, tick=F, hadj=0, font=2) # las orients labels

lines(x=c(hypermean,hypermean), y=c(0.25,5.5), lty=2, lwd=2)
text('Hypermean', x=hypermean, y=5.6)

par(op) # resets par