Résumé

Les programmes de prévention de chutes chez les aînés : une méta-analyse secondaire bayesienne

Joseph F. Lucke

Une méta-analyse secondaire bayesienne des programmes pour réduire les chutes chez les aînés est effectuée pour démontrer le processus d'analyse bayesienne. La tradition statistique bayesienne fait l'objet d'une importante distinction relativement à la tradition statistique standard de Neyman-Pearson-Wald (NPW). Dans le cadre des 12 études, l'ampleur de l'effet logit est utilisée pour comparer des groupes thérapeutiques adhérant à un programme de prévention à des groupes témoins qui ne sont pas soumis à un programme. Afin de mettre en contraste l'analyse bayesienne, des méta-analyses d'effets indépendants et d'effets fixes sont d'abord réalisées selon la tradition de NPW. Cette procédure est suivie de méta-analyses d'effets indépendants et d'effets fixes qui reproduisent numériquement les résultats NPW mais qui comportent des interprétations différentes sur le plan conceptuel. Les dernières analyses comprennent des méta-analyses prédictives et des méta-analyses d'effets aléatoires bayesiennes. Ces résultats diffèrent sur le plan numérique de toutes les autres méta-analyses antérieures et sur le plan conceptuel des analyses de NPW. Les analyses des effets aléatoires permettent une hétérogénéité en ce qui a trait à l'ampleur de l'effet. L'analyse prédictive génère une distribution d'effet hors-échantillon nouveau, qui favorise non seulement l'hétérogénéité des effets mais aussi l'imprécision dans les estimations de paramètres. Cette dernière analyse démontre que l'efficacité des nouveaux programmes de prévention des chutes est moins définitive que celle relevée dans l'échantillon. Les méthodes statistiques bayesiennes se prêtent particulièrement bien au traitement des complexités contenues dans les recherches en sciences infirmières.

Mots clés : tradition statistique bayesienne, ampleur de l'effet logit, tradition statistique de Neyman-Pearson-Wald, analyse prédictive, méta-analyse secondaire

CJNR 2004, Vol. 36 N° 3, 48–64

Fall-Prevention Programs for the Elderly: A Bayesian Secondary Meta-analysis

Joseph F. Lucke

A secondary meta-analysis of programs to reduce falls in the elderly is undertaken to demonstrate a Bayesian analysis. The Bayesian statistical tradition is carefully distinguished from the standard Neyman-Pearson-Wald (NPW) statistical tradition. In the 12 studies, the logit effect size is used to compare treatment groups using a prevention program to control groups without a program. To contrast the Bayesian analysis, independent-effects and fixed-effect meta-analyses are first conducted in the NPW tradition. This is followed by Bayesian independent-effects and fixed-effect meta-analyses that numerically replicate the NPW results but have conceptually different interpretations. The final analyses comprise Bayesian random-effects and predictive meta-analyses. These results differ numerically from all the previous meta-analyses and conceptually from the NPW meta-analyses. The random-effects analysis allows for heterogeneity in the effect sizes. The predictive analysis yields the distribution of a new, out-of-sample effect size, which accommodates not only the heterogeneity of the effects but also the imprecision in the parameter estimates. This last analysis shows that the effectiveness of new fall-prevention programs is less definitive than that found in the sample. Bayesian statistical methods are particularly well-suited for the complexities of nursing science studies.

Keywords: Bayesian statistical tradition, fixed-effect model, health-care outcomes, hierarchical model, independent-effects model, logistic regression, logit effect size, Neyman-Pearson-Wald statistical tradition, predictive analysis, random-effects model, secondary meta-analysis

Hill-Westmoreland, Soeken, and Spellbring (2002) conducted a fixed-effect meta-analysis of the success of programs to prevent falls in the elderly. Their statistical analysis was based on the familiar Neyman-Pearson-Wald (NPW) statistical tradition. The purpose of this article is to illustrate the alternative Bayesian statistical tradition by conducting a secondary analysis of the same data. Bayesian statistical methods are rarely used in nursing science. A *fin de millénaire* review of research using Bayesian inference ranging from "archeology" to "social sciences" found no reference to nursing science (Berger, 2000). I will first present a brief sketch of the principles of the Bayesian statistical tradition.

The term "statistical tradition" is used here to emphasize the conceptual discontinuity between the two approaches. A *research tradition,* similar

to a paradigm (Kuhn, 1962), is a global cluster of beliefs about the entities and processes that make up the domain of inquiry and the methodological norms by which the domain is to be investigated (Laudan, 1981). A *statistical tradition* is by analogy a global set of ideas about the nature of probability and principles for statistical applications. Bayesian statistics is not another branch of NPW statistics but a comprehensive, alternative stance on probability and statistics. Use of the word "tradition" also emphasizes that concepts in the NPW tradition (e.g., significance, power) may have no meaning in the Bayesian tradition, that concepts in the Bayesian tradition (e.g., Bayes's factor, prior distribution) may have no meaning in the NPW tradition, and that identical numerical results may have radically different interpretations (e.g., confidence intervals, hypothesis tests). Table 1 provides a brief overview of the principal differences, for which only a brief comparison can be provided here.

The NPW statistical tradition interprets probability as relative frequency. Statistical inference is based on the fundamental concept of *inductive behaviour,* behaviour that can be repeatedly evaluated as correct or incorrect. The purpose of statistics is to develop rules for inductive behaviour and to assess their performance according to their relative frequency of success. With this approach, observations are considered random variables that are controlled by fixed but unknown parameters. For example, a statistical test will be acceptable if the rejection (behaviour) of a truly null (fixed and unknown) hypothesis will be incorrect only 5% of the time (performance) in the long run (probability). Confidence intervals, now favoured over significance tests (Altman, Machin, Bryant, & Gardner, 2000), provide the set of points that would not have been rejected by the corresponding significance test.

Table 1 ***Comparison of the Neyman-Pearson-Wald and Bayesian Traditions***

Topic	**NPW**	**Bayesian**
Probability	Relative frequency	Logic of judgement
Inference	Inductive behaviour	Degrees of belief
Statistic	Performance	Evidence
Observations	Random	Fixed
Parameters	Fixed	Random
Confidence interval	Random interval covers fixed parameter	Fixed interval contains random parameter

The Bayesian tradition interprets probability as a *logic of judgement* for those opinions that can be represented as *degrees of belief* (Howson & Urbach, 1993). This logic is regulated by *coherence,* a counterpart to consistency in deductive logic. Coherence ensures that a person cannot hold degrees of belief that are uniformly disadvantageous. The remarkable Ramsey-de Finetti Theorem shows that for judgements to be coherent the corresponding degrees of belief must satisfy the axioms of probability (Howson & Urbach). Thus, judgements can be represented by *subjective probabilities.* This approach considers observations to be fixed and the parameters of models that account for the data to be random. Bayesian confidence intervals or *credible intervals* give the fixed interval in which the random parameter can be found with the specified probability. It is often the case that NPW confidence intervals have the exact same numerical values as Bayesian credible intervals, but with different interpretations.

Because coherent judgements are subjective probabilities, the full use of the probability calculus is available for statistical inference. The Bayesian principle of inference is *Bayes's Theorem,* first posthumously published by Bayes and Price (1763) but a trivial theorem in modern probability theory. The application of Bayes's Theorem is a four-step process. First, the investigator develops the *likelihood,* which provides the probability of the observations y given the values of a parameter θ, written $\Pr(y \mid \theta)$. The likelihood provides the evidential link between the parameter and the observations (Royall, 1997). Second, because it is uncertain, the investigator represents the parameter θ as a random variable with a *prior* distribution $\Pr(\theta)$ that reflects uncertainty regarding its possible values. The prior distribution represents the information the investigator possesses before any observations are made in the current study. Third, the study is undertaken and the observations are made. Fourth, Bayes's Theorem provides the "updated" *posterior* distribution regarding the parameter conditioned on the observations. Bayes's Theorem states

$$\Pr(\theta \mid y) = \frac{\Pr(y \mid \theta)\Pr(\theta)}{\Pr(y)} . \qquad (1)$$

Roughly, Equation 1 states that

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{data}} ,$$

or, sardonically, that "Bayesians can be recognized by their posteriors."

Bayesian statistical inference requires a prior distribution representing the investigator's information regarding the parameter of interest. The primary advantage of a prior is that it encapsulates previous knowledge regarding the parameter under investigation. The primary disadvantage is that an investigator's prior could also be quite idiosyncratic. However, several considerations moderate the choice of a prior. First, the chosen prior is exposed to the scrutiny of the scientific community. Determining the prior distribution forces the investigator to confront and make explicit his or her beliefs, both justifiable and speculative, regarding the phenomenon under investigation (Kadane, 1995). Second, the prior distribution is a *representation,* not a measurement, of one's degrees of belief (Hacking, 2001). Thus the investigator may *temper* the prior to more reasonably represent degrees of belief of the scientific community rather than actual beliefs (Shimony, 1970/1993). And third, a prior distribution can be made *diffuse* or *vague* so that it has little influence on the data (Edwards, Lindman, & Savage, 1963). The requirement of a publicly presented prior makes the subjective evaluation of prior knowledge more objective.

Meta-analysis is particularly well-suited for an introduction to Bayesian statistical analysis. The data sets tend to be small but fraught with inferential problems not found in usual analyses. Bayesian meta-analysis has a growing and increasingly sophisticated literature (Beard, Curry, Edwards, & Adams, 1997; Berlowitz et al., 2002; Brophy, Belisle, & Joseph, 2003; Brophy & Joseph, 2000; Brophy, Joseph, & Rouleau, 2001; Burr, Doss, Cooke, & Goldschmidt-Clermont, 2003; DuMouchel, 1990; Higgins & Spiegelhalter, 2002; Louis & Zelterman, 1994; Nam, Mengersen, & Garthwaite, 2003; Spiegelhalter & Best, 2003; Stangl & Berry, 2000; Warn, Thompson, & Spiegelhalter, 2002). The presentation here is intended to serve as an entry to Bayesian meta-analysis and to other Bayesian methods.

The remainder of this paper is organized as follows. First, the observed fall-prevention data and a metric for assessing treatment effects, namely the logit effect size, are presented. Second, two brief NPW analyses of the data, an independent-effects analysis and a fixed-effect analysis, are undertaken to provide a comparison for the Bayesian analyses. Third, two Bayesian analyses, also an independent-effects analysis and a fixed-effect analysis, are presented to compare to the NPW analyses. These two sets of analyses yield similar numerical results. The purpose is to demonstrate the conceptual differences between the two traditions even when the numbers are the same. And fourth, a random-effects Bayesian analysis coupled with a predictive analysis is presented. This last analysis shows how the Bayesian approach can utilize more realistic assumptions.

The Data and the Logit Effect Size

Table 2 presents the data on 12 select studies of falls in the elderly (Hill-Westmoreland et al., 2002). The first column contains the study number as enumerated by those authors. Each study i, $i = 1, \ldots, 12$, comprises two groups, a treatment group $j = 1$ which had an intervention for the prevention of falls, and a control group $j = 0$ which had no such intervention. Let y_{ij} denote the number of subjects experiencing one or more falls from a sample of n_{ij} subjects in group j of study i. Let $p_{ij} = y_{ij}/n_{ij}$ denote the proportion of falls. The *Control* columns give the observed y_{i0}, n_{i0}, and p_{i0}; the *Treatment* columns give the observed y_{i1}, n_{i1}, and p_{i1}.

Table 2 ***Meta-analytic Data on 12 Studies of Falls in the Elderly***

	Control		**Treatment**		**Comparison**	
Study	Falls/Sample	Prop.	Falls/Sample	Prop.	Logit	SE
1	62/117	0.53	53/116	0.46	-0.29	0.26
2	111/213	0.52	59/184	0.32	-0.84	0.21
3	45/94	0.48	34/75	0.45	-0.10	0.31
4	40/92	0.43	42/88	0.48	+0.17	0.30
5	129/261	0.49	91/221	0.41	-0.33	0.18
6	17/50	0.34	72/180	0.26	+0.40	0.34
7	61/81	0.75	56/79	0.71	-0.23	0.36
8	3/15	0.20	3/30	0.10	-0.81	0.89
9	6/13	0.46	2/14	0.14	-1.64	0.94
10	68/144	0.47	52/147	0.35	-0.49	0.24
11	8/47	0.17	68/332	0.20	+0.23	0.41
12	223/607	0.37	268/952	0.28	-0.39	0.11

Hill-Westmoreland et al. (2002) used the *difference effect*, $d_i = p_{i1} - p_{i0}$, to estimate the treatment effect (Rosenthal, 1994). Although the difference effect has a simple and straightforward interpretation, d_i has a number of defects, not least of which is that it is confounded with the baseline proportion found in the control group (Fleiss, 1994). A better measure of treatment effect, and the one used here, is the *logit effect*

$$w_i = \log\left(\frac{p_{i1}}{1 - p_{i1}}\right) - \log\left(\frac{p_{i0}}{1 - p_{i0}}\right) \qquad (2)$$

with standard error

$$SE(w_i) = \sqrt{\frac{1}{p_{i0}(1 - p_{i0})n_{i0}} + \frac{1}{p_{i1}(1 - p_{i1})n_{i1}}}$$

(Rosenthal). Although the logit is less intuitive than the difference, it is preferable on several grounds, two of which are given here. First, the logit effect is not confounded with the baseline proportion and may assume any value regardless of the proportions being compared (Fleiss). And second, the logit effect is a natural parameter of log-linear and logistic regression models (Fleiss). The logit effects and their standard errors for the 12 studies are also given in the last two columns of Table 2. Negative logits favour treatment over control.

Neyman-Pearson-Wald Analyses

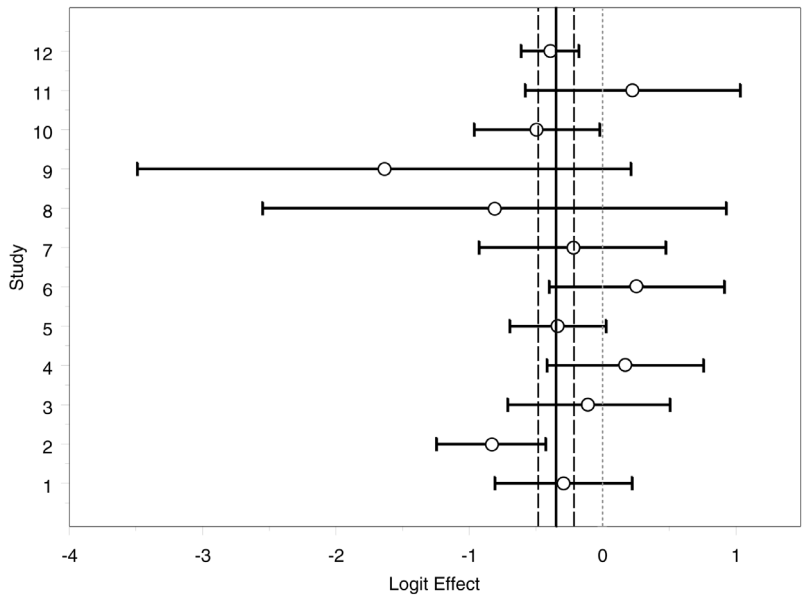
Independent-Effects Analysis

The independent-effects analysis addresses the treatment effect of each study separately from any other study. It is not a meta-analysis itself, but such analyses are invariably included in a meta-analysis to display the individual effects. The sample logit effects follow an asymptotic normal distribution with mean w_i and standard deviation $SE(w_i)$. Figure 1 displays the logit effect for each study together with the respective 95% confidence intervals. Only three studies, 2, 10, and 12, had 95% confidence intervals that excluded zero, or, equivalently, rejected the null hypothesis of $w_i = 0$ with a two-tailed significance at .05. Hill-Westmoreland et al. (2002), using difference effects, found Study 9 to be significant and not Study 10. Thus, the difference effect and logit effect need not produce identical results.

Fixed-Effect Analysis

A fixed-effect meta-analysis assumes the existence of a fixed but unknown population treatment effect. Each study provides a sample estimate of the population effect. The estimator of the population fixed effect is the weighted mean of the study sample effects and likewise follows an asymptotic normal distribution (Fleiss, 1994). The first row of Table 3 presents the weighted mean logit effect of $w = -0.35$, its standard error, and its 95% confidence interval. Because the interval for this statistic excludes 0, the null hypothesis of no overall treatment effect is rejected. This fixed-effect re-analysis using logit effects yields the same conclusion as Hill-Westmoreland et al.'s (2002) original analysis.

Figure 1 ***Forest Plot of Logit Effect Sizes for the NPW Analyses***



Note: Circles with horizontal bars represent independent logit effects and their respective 95% confidence intervals. The solid vertical line is the fixed logit effect and the dashed vertical lines are its 95% confidence interval.

Table 3 ***Meta-analytic Results from NPW and Bayesian Analyses***

Analysis		Logit Effect	Standard Error	Confidence Interval	Hypothesis: Effect < 0
NPW	Fixed	-0.35	0.07	-0.48, -0.21	$z = -5.07$
Bayesian	Fixed	-0.35	0.07	-0.48, -0.21	$\Pr(\beta < 0) > 0.999$
Bayesian	Random	-0.32	0.10	-0.51, -0.11	$\Pr(\mu_\beta < 0) = 0.996$
Bayesian	Predictive	-0.32	0.23	-0.79, +0.20	$\Pr(B < 0) = 0.93$

Bayesian Analyses I

Independent-Effects Analysis

A Bayesian independent-effects analysis is presented here for comparison with the NPW version. Again, this analysis is, strictly speaking, not a meta-analysis, and is not even needed as an intermediary step in the

Bayesian analysis. Nevertheless, it is frequently presented to display the individual effects. The likelihood is based on the binomial-logistic regression model (McCaullagh & Nelder, 1989). The propensity of a subject to experience at least one fall in the control group of study i is the logistic function of α_i. The propensity to fall in the treatment group is the logistic function of $\alpha_i + \beta_i$, where β_i is the treatment effect. If the treatment reduces the propensity to fall, β_i will be negative. The Bayesian analysis requires prior distributions for the parameters α_i and β_i. Because there is no prior information on these parameters and because the desire here is for the observations to have maximal influence on the posterior distributions, diffuse priors were chosen for both sets of parameters. In particular, the priors were chosen to be normal distributions with mean of zero and standard deviation of 1,000. These priors stipulate that this investigator was 95% sure that the parameter values lay between -2,000 and 2,000.

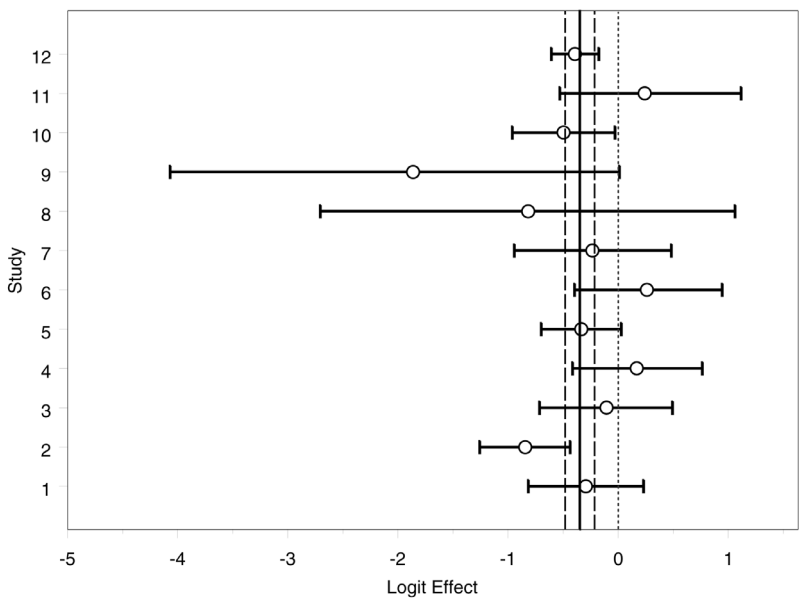
Equation 1 in principle supplies the posterior distributions of the 12 α_i's and the 12 β_i's. However, substituting the appropriate formulae from the distributional assumptions for the likelihood of γ_{ij} and the priors for the α_i and the β_i yields a formula that is analytically intractable. Numerical methods must be employed to obtain the posterior distributions. One popular approach is to numerically simulate the posterior distributions of each parameter to a high degree of accuracy. Given the simulated posteriors, summary information regarding the parameters can be readily obtained. The current widely used procedure for obtaining numerical solutions is *Markov Chain Monte Carlo* (MCMC) simulation using Gibbs sampling (Gelman, Carlin, Stern, & Rubin, 2004). MCMC is an algorithm in which samples are taken from each parameter's distribution in a round-robin fashion and used to update the other distributions in the model. This sampling procedure is iterated several thousand times, and under suitable mathematical conditions the distributions of the parameters will converge to a unique set of limit distributions. These limit distributions can then be used to extract statistical information regarding the parameters. This independent-effects model was programmed in WinBUGS 1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2002). To stabilize the starting distributions, an initial (burn-in) run of 2,000 iterations was taken without sampling the parameters. The sampling simulation was then run for another 20,000 iterations to obtain the limit distributions. The simulation takes about 12 seconds on a reasonably fast (1 GHz) desktop computer. S-PLUS Professional 6.2 (Insightful Corporation, 2001) and Microsoft Excel were used for additional computations and graphics.

Figure 2 displays the independent effects together with their 95% Bayesian credible intervals. These are very similar in numerical value to

the 95% confidence intervals of the NPW analysis in Figure 1. The single exception appears to be study 9, which has a smaller Bayesian posterior mean and larger standard error $\beta_9 = -1.86$, se = 1.04 than the corresponding NPW estimate of $w_9 = -1.64$, se = 0.94. This discrepancy is most likely due to the inaccuracy of the NPW asymptotic estimator in a small sample of $n_{9,0} + n_{9,1} = 27$ subjects.

More important than the numerical similarity between the NPW and Bayesian results is that the interpretation is completely different. The credible intervals give the fixed interval in which the random treatment effect probably lies, in sharp contrast to the NPW interpretation of the random interval that probably covers the fixed treatment effect.

Figure 2 ***Forest Plot of Logit Effect Sizes for the Bayesian Analyses***



Note: Circles with horizontal bars represent independent logit effects and their respective 95% credible intervals. The solid vertical line is the fixed logit effect and the dashed vertical lines are its 95% credible interval.

Fixed-Effect Analysis

The fixed-effect analysis again assumes that there is a unique treatment effect for all the studies and that each study displays this unique effect accompanied by random error. The fixed-effect analysis merely replaces each study-specific β_i in the independent-effects analysis with a single β for all studies in the binomial-logistic regression. Thus the treatment

effect is, instead, a logistic function of $\alpha_i + \beta$. The prior distributions for the 12 α_i's and the one β are the same normal distributions as in the independent-effects case.

Again, the resulting posterior distribution is analytically intractable and MCMC simulation was used. The second row of Table 3 presents the Bayesian fixed-effect result. The numerical values are identical to those of the NPW analysis but the interpretation in the Bayesian case is that $\Pr(-0.48 \leq \beta \leq -0.21) = .95$. Furthermore, because β is a random variable, probabilistic hypotheses can be entertained. Of interest here is whether the fall-prevention programs reduced the proportion of falls, which translates into the hypothesis whether β is negative. The probability is that $\beta < 0$ can be readily evaluated from its relevant posterior distributions. The rightmost column of Table 3 shows that $\Pr(\beta < 0) > .999$.

Bayesian Analyses II

Random-Effects Analysis

The purpose of the Bayesian fixed-effect analysis was to demonstrate that the Bayesian approach can replicate the numerical results of the NPW approach, even though the conceptual interpretations of the results are different and incompatible between the two traditions. The purpose of the random-effects analysis is to show how the Bayesian approach can easily accommodate more realistic assumptions in the statistical model.

The assumption of a unique, fixed treatment effect for each study is excessively restrictive. It implies that all differences among study effects are due solely to random error. It ignores possible differences in settings and implementations of the prevention programs that would have caused different effects in addition to random fluctuations. A more realistic and plausible assumption is that the population effects themselves arise from a distribution of treatment effects. The crucial assumption required for such an analysis, called *exchangeability*, is that the magnitudes of the program effects are equally as likely to appear in one study as in another (Gelman et al., 2004). The assumption of exchangeability allows for heterogeneous but related effects in place of the assumption of a homogeneous effect.

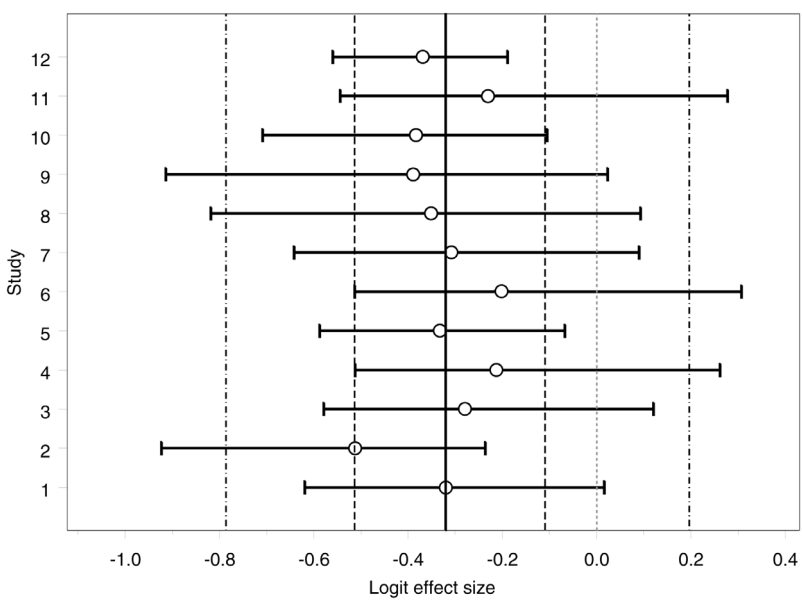
As in the previous analyses, the binomial-logistic regression model is assumed. The baseline propensity to experience a fall is assumed to be specific to each study, implying that the α_i's are again assumed to be fixed. The β_i's, however, are not assumed to be fixed for each study but assumed to be part of an overall effect of fall prevention on all the studies. To accommodate heterogeneity of effects among the studies, the β_i's are themselves assumed to be independently distributed random variables from a normal distribution with unknown mean μ_β and unknown

variance σ^2_β. This likelihood is equivalent to a two-level logistic regression (Rice, 2001).

To complete the Bayesian specification, prior distributions are required for the 12 α_i's, μ_β, and σ^2_β. Again, diffuse priors are chosen so that prior beliefs have little influence on the observed data. The fixed α_i's are once again given a normal prior with a mean of zero and a standard deviation of 1,000. The priors for the parameters of the distribution of the random treatment effects are similarly diffuse. The prior for μ_β is a normal distribution with a mean of zero and a standard deviation of 1,000, and the prior for σ^2_β is a diffuse gamma distribution. Again, the formula for the posterior distribution is analytically intractable, so that MCMC methods are used to simulate the posterior distributions.

Figure 3 presents the posterior mean logit effect for each of the 12 studies together with their 95% credible intervals. The first feature to note is that all the logit effects are negative, in contrast to the independent-effects analysis in which the effects of studies 4, 6, and 11 were positive.

Figure 3 ***Forest Plot of Random Logit Effect Sizes***



Note: Circles with horizontal bars denote the posterior mean logit effect sizes and their respective 95% credible intervals. The solid vertical line is the posterior mean logit effect size for the distribution of effects. The dashed vertical lines contain its 95% credible interval. The dot-dashed vertical lines contain the 95% credible interval for the predictive posterior mean effect size.

The second feature to note, related to the first, is that the logit effects tend to be closer together than in the independent-effects analysis, with no extreme cases. (The x-axis ranges only from -1 to 0.4 rather than from -5 to 2 for the independent-effects analysis.) These two features demonstrate the well-known phenomena of "shrinkage towards the mean" and "studies borrowing strength from each other" found in random-effects analyses.

The mean posterior mean treatment effect is $\mu_\beta = -0.32$ with $Pr(-0.51 \leq \mu_\beta \leq -0.11) = .95$. This mean and credible interval is also displayed in Figure 3. As shown in Table 3, the mean posterior mean and its standard error are slightly larger than that of the fixed logit effect. This shrinkage towards zero accompanied by a larger standard error is also common in random-effects models. The probability that μ_β is negative is .996.

Predictive Analysis

The random-effects analysis also yields the posterior distribution of the logit treatment effects. The distribution of treatment effects is a normal distribution (by assumption) with a posterior mean for $\mu_\beta = -0.32$ and a posterior mean for $\sigma_\beta = 0.17$. Thus it would be reasonable to predict that any treatment effect not yet observed would on average be -0.32 with a 95% credible interval of [-0.67, + 0.02]. However, a logical defect in making this particular out-of-sample prediction is that only the means of the distributions of μ_β and σ_β are used. A more realistic out-of-sample prediction would include the imprecision of μ_β and σ_β as well. Let *B* denote the out-of-sample treatment effect. A Bayesian *predictive analysis* obtains the distribution of *B* by sampling it from a normal distribution with mean μ_β and variance σ^2_β. However, μ_β and σ_β are not fixed points but random variables. Thus, the posterior distribution of *B* will have the same mean as the posterior distribution of μ_β but will have a larger variance than σ^2_β, comprising the variance of the normal distribution, the posterior variance of μ_β, and the posterior variance of σ^2_β.

Figure 3 presents the posterior predictive mean $B = -0.32$ with a 95% credible interval ranging from -0.79 to +0.20, considerably wider than the point-based credible interval mentioned in the previous paragraph. The hypothesis test shows that *B* only has probability of .93 of being negative.

Discussion

The Bayesian statistical tradition offers a comprehensive view of the nature of probability and statistical inference. Probability is interpreted as a logic of judgement regarding degrees of belief, and statistical inference is the revision of those subjective judgements according to observational

evidence. Unobserved quantities are random variables, known observations are fixed, inferences are based only on the observations, hypotheses can be probabilistically compared one with another, and credible intervals give the interval that probably contains the unobserved quantity. The NPW tradition also offers a comprehensive view of the nature of probability and statistical inference. Probability is interpreted as relative frequency, and statistical inference is the evaluation of the performance of decision rules for inductive behaviour. Unknown parameters are fixed, known observations are random, inferences are not based solely on the observations, statistical hypotheses are choices between inductive behaviours, and $1 - \alpha$-level confidence intervals give the interval that contains the point-null hypotheses that would not have been rejected by an α-level statistical test. Comparisons between the two traditions cannot be made on statistical grounds because each tradition contains its own, internal standards of evaluation, standards that need not be applicable to the other tradition. Choosing between traditions must instead be based on higher-order cognitive values regarding the goals and norms of scientific inference (Laudan, 1984). Many statisticians and philosophers of science believe that the Bayesian tradition offers a natural and commonsense interpretation of probability and statistical inference. These same authors often remark on the persistent tendencies of investigators to interpret *p* values as evidential support and confidence intervals as fixed intervals of probability as indications of the inferential perversity of the NPW tradition.

The demonstration here shows that a Bayesian statistical analysis can replicate an NPW analysis by yielding the same numerical results. The fixed-effect meta-analyses addressed whether the fall-prevention programs reduced the proportion of falls — that is, had an effect in the negative direction. The NPW approach yielded a logit effect size of $w = -0.35$ with a 95% confidence interval ranging from -0.48 to -0.21. The Bayesian approach likewise yielded a logit effect size of $\beta = -0.35$ with a 95% credible interval ranging from -0.48 to -0.21. These two identical numerical results have incommensurable interpretations. The NPW confidence interval holds that the set of point-null hypotheses from -0.48 to -0.21 would not have been rejected by a two-tailed test at the 5% significance level. The Bayesian credible interval holds that β falls between -0.48 and -0.21 with probability .95.

The random-effects analysis further showed how a Bayesian analysis could go beyond the assumption of a fixed effect by allowing a distribution of effects. By assuming a distribution of effects, the Bayesian analysis accommodated possible heterogeneity among study-specific effects and yielded the overall mean effect. The Bayesian analysis also yielded the entire distribution of effects along with its mean $\mu_\beta = -0.32$ and standard

deviation σ_β= -0.17. The Bayesian predictive analysis also yielded the distribution of an out-of-sample effect that included not only the variability in effects but also the variability in estimating the distribution of the effects. This final result indicates that future fall-prevention programs can expect, with 95% confidence, to have a logit effect size between -0.8 and +0.2. Thus the predictive analysis indicates that the effect of fall-prevention programs is less definitive than might be presupposed from the either the fixed-effect or random-effects analyses.

To begin using Bayesian methods, one must be prepared to undertake substantial changes in one's ideas about statistics. This demonstration, though simple, was sufficient to uncover some of the advantages of Bayesian statistical methods. The Bayesian approach is known for its ability to handle complicated and unusual situations (Best, Spiegelhalter, Thomas, & Brayne, 1996). Thus, the Bayesian statistical tradition appears well-suited for the difficult inferential problems found in nursing science.

References

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (Eds.). (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: BMJ Books.

Bayes, T., & Price, R. (1763). An essay toward solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions, 53,* 370–418.

Beard, M. T., Curry, E. L., Edwards, K., & Adams, B. N. (1997). Advances in meta-analysis as a research method. *ABNF Journal, 8*(5), 92–97.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association, 95*(452), 1269–1276.

Berlowitz, D. R., Christiansen, C. L., Brandeis, G. H., Ash, A. S., Kader, B., Morris, J. N., & Moskowitz, M. A. (2002). Profiling nursing homes using Bayesian hierarchical modeling. *Journal of the American Geriatrics Society, 50*(6), 1126–1130.

Best, N. G., Spiegelhalter, D. J., Thomas, A., & Brayne, C. E. G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 159*(2), 323–342.

Brophy, J. M., Belisle, P., & Joseph, L. (2003). Evidence for use of coronary stents: A hierarchical Bayesian meta-analysis. *Annals of Internal Medicine, 138*(10), 777–786.

Brophy, J. M., & Joseph, L. (2000). A Bayesian meta-analysis of randomized megtrials for the choice of thrombolytic agents in acute myocardial infarction. In D. K. Stangl & D. A. Berry (Eds.), *Meta-analysis in medicine and health policy.* New York: Marcel Dekker.

Brophy, J. M., Joseph, L., & Rouleau, J. L. (2001). β-blockers in congestive heart failure: A Bayesian meta-analysis. *Annals of Internal Medicine, 134*(7), 550–560.

Burr, D., Doss, H., Cooke, G. E., & Goldschmidt-Clermont, P. J. (2003). A meta-analysis of studies on the association of the platelet p1a polymorphism of glycoprotein iiia and risk of coronary heart disease. *Statistics in Medicine, 22,* 1741–1760.

DuMouchel, W. (1990). Bayesian metaanalysis. In D. A. Berry (Ed.), *Statistical methodology in the pharmaceutical sciences* (pp. 509–529). New York: Marcel Dekker.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Fleiss, J. F. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman & Hall.

Hacking, I. (2001). *An introduction to probability and inductive logic.* Cambridge: Cambridge University Press.

Higgins, J. P., & Spiegelhalter, D. J. (2002). Being sceptical about meta-analyses: A Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology, 31*(1), 96–104.

Hill-Westmoreland, E. E., Soeken, K., & Spellbring, A. M. (2002). A meta-analysis of fall prevention programs for the elderly: How effective are they? *Nursing Research, 51*(1), 1–8.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago: Open Court.

Insightful Corporation. (2001). *S-PLUS 6 for Windows user's guide.* Seattle: Author.

Kadane, J. B. (1995). Prime time for Bayes. *Controlled Clinical Trials, 16,* 313–318.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Laudan, L. (1981). A problem-solving approach to scientific progress. In I. Hacking (Ed.), *Scientific revolutions* (pp. 144–155). Oxford: Oxford University Press.

Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate.* Berkeley: University of California Press.

Louis, T. A., & Zelterman, D. (1994). Bayesian approaches to research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 411–422). New York: Russell Sage Foundation.

McCaullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Nam, I.-S., Mengersen, K., & Garthwaite, P. (2003). Multivariate meta-analysis. *Statistics in Medicine, 22,* 2309–2333.

Rice, N. (2001). Binomial regression. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics.* Chichester: John Wiley.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm.* London: Chapman & Hall.

Shimony, A. (1993). Scientific inference. In *Search for a naturalistic world view: Scientific method and epistemology* (Vol. 1). Cambridge: Cambridge University Press. (Original work published 1970)

Spiegelhalter, D. J., & Best, N. G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine, 22,* 3687–3709.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2002). *WinBUGS Version 1.4 user's manual.* MRC Biostatistics Unit. Available: http://www.mrc-bsu.cam.ac.uk/bugs/

Stangl, D. K., & Berry, D. A. (2000). *Meta-analysis in medicine and health policy.* New York: Marcel Dekker.

Warn, D. E., Thompson, S. G., & Spiegelhalter, D. J. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. *Statistics in Medicine, 21,* 1601–1623.

Author's Note

I thank Dr. Hill-Westmoreland for generously providing the data.

Joseph F. Lucke, PhD, is Assistant Professor and Statistical Scientist, School of Nursing, University of Texas Health Science Center at San Antonio, USA.