

Multilevel modeling and model averaging

by Sander Greenland, DrPH¹

Greenland S. Multilevel modeling and model averaging. *Scand J Work Environ Health* 1999;25 suppl 4:43–48.

Multilevel modeling, also known as hierarchical regression, generalizes ordinary regression modeling to allow explicit and flexible compromises between simple and complex models. This article provides an elementary introduction to multilevel modeling as a model-averaging technique. Model averaging provides an alternative to model selection, and it emphasizes the role of prior information in finding good models.

Key terms Bayesian statistics, empirical-Bayes estimation, hierarchical regression, random-coefficient regression, ridge regression, risk assessment, Stein estimation.

The vast increase in computing power over recent decades has led to the emergence of multilevel modeling as a practical and powerful analysis tool (1–5). This approach involves specifying 2 or more levels, or stages, of relationships among study variables and parameters. These levels are arranged in a hierarchy; hence the approach is also commonly known as hierarchical modeling or hierarchical regression. Ordinary regression techniques represent a special case in which there is only 1 level of the hierarchy.

Multilevel modeling has appeared and reappeared in many forms over the last 50 years. For example, the following techniques are all special cases of hierarchical regression: Bayesian, empirical-Bayes (EB), Stein, penalized likelihood, ridge, and random-coefficient regression and variance-components analysis (1–32). Conventional estimation methods, such as least squares and maximum likelihood, are also special cases in which there is only 1 level in the underlying model. The hierarchical approach unifies these methods and clarifies their meaning. Most importantly, it unifies the seemingly disparate methods of frequentist (“classical”) and Bayesian analysis. This unification is a major benefit of the approach, for it leads to methods more accurate than either classical frequentist or Bayesian methods. This benefit is especially important and obvious in analyses that screen a large number of associations for further investigation, as is common in occupational, environmental, and dietary

studies. [See the examples in Greenland (30), Greenland & Poole (22), and Witte et al (28).]

Despite several decades of published applications and methodologic studies demonstrating its advantages, multilevel modeling is not widely understood, taught, or employed in the health sciences. The lack of affordable and user-friendly software has no doubt been an obstacle, although conventional software is not that hard to adapt to fitting basic 2-level models; for example, an SAS procedure for fitting such models is now available on the World-Wide Web at <http://darwin.cwru.edu/~witte/hm.html> (6).

The present article focuses on a by-product of multilevel modeling — hierarchical estimation. I hope that readers unfamiliar with multilevel models will come away with at least a fairly clear intuition of what constitutes a 2-level model and estimator. An introduction to the fundamental principles behind such estimators will appear elsewhere (17).

To fix ideas, suppose we count $A_1, \dots, A_K$ spontaneous abortions among $N_1, \dots, N_K$ first planned pregnancies in K subcohorts numbered 1, ..., K , as we might obtain if we stratified a single first-planned pregnancy cohort on the level of exposure to a drinking-water contaminant and on potential confounders such as gravidity and age. Our objective might be to estimate the average risks $\theta_1, \dots, \theta_K$ in these subcohorts; more often, it would be to estimate the relation of these risk parameters to

¹ Department of Epidemiology, UCLA School of Public Health, Los Angeles, California, United States.

Correspondence to: Dr S Greenland, Department of Epidemiology, UCLA School of Public Health, 22333 Swenson Drive, Topanga, CA 90290, USA.

exposure and, perhaps, to gravidity and age. In other words, we want to estimate the *regression* of risk on exposure, gravidity, and age. For simplicity of the age relations, let us assume all the pregnancies occur after the age of 20 years.

Conventional (single-level) regression models

One typical single-stage regression analysis would fit the following logistic risk model:

$$\theta_k = \text{expit}(\alpha + \beta x_k + \gamma g_k + \delta z_k), \quad (\text{equation 1})$$

where $\text{expit}(\bullet)$ stands for the logistic transform

$$\text{expit}(\bullet) \equiv \frac{\exp(\bullet)}{1 + \exp(\bullet)},$$

and x_k , g_k , z_k are the exposure level, gravidity, and age for pregnancies in stratum k ; x_k , g_k , and z_k are called the model *regressors* (33). Model 1 can be rewritten in terms of the odds $\theta_k/(1 - \theta_k)$ of spontaneous abortion:

$$\begin{aligned} \frac{\theta_k}{1 - \theta_k} &= \exp(\alpha + \beta x_k + \gamma g_k + \delta z_k) \\ &= e^{\alpha} e^{\beta x_k} e^{\gamma g_k} e^{\delta z_k}. \end{aligned} \quad (\text{equation 2})$$

This equation makes it easier to see the unrealism of model 1, for it shows that the model imposes an exponential relation of the odds to exposure, gravidity, and age; this relation is called the exponential-odds assumption (33). Equation 2 also shows that the model imposes a multiplicative relation among the odds ratios for exposure, gravidity, and age; this relation is called the multiplicative-odds assumption (33). The extremity of these impositions is often ignored, in part because model 1 is often presented as a linear model for the *logits*, or log odds, of the average risks θ_k ; taking the natural logarithms of both sides of equation 2, we get

$$\text{logit}(\theta_k) \equiv \ln \left(\frac{\theta_k}{1 - \theta_k} \right) = \alpha + \beta x_k + \gamma g_k + \delta z_k. \quad (\text{equation 3})$$

Because of this relation, model 1—3 is sometimes called a “log-linear” model (34). It is important to bear in mind, however, that equations 1, 2, and 3 all impose exactly the same exponential-multiplicative relations among the odds; that is, they all represent the same model.

A major problem with model 1—3 is that it rigidly assumes far more than we know about the relation of spontaneous abortion to exposure, gravidity, and age. Exponential trends in odds imply very steep risk trends when risks are not high. Although such trends are sometimes reasonable approximations of the true trends, we cannot be sure that they occur in this example; in fact, many other possibilities are reasonable. Similarly,

multiplicative relations among odds ratios (often called “multiplicative interactions”) can imply dramatic variations in risk trends even when only modest trends are present for all regressors; again, although such trend variation sometimes occurs, we cannot be sure that it occurs here.

If any model assumption is wrong, the modeling results may give a misleading impression of risk patterns. Analysts typically address this problem by testing the fit of their model against the observed data; unfortunately, the power of such tests is so low that they usually fail to reject the model, even if it is very wrong. More creatively, analysts may modify the coding of the regressors to allow more realistic trends and occasionally add product terms to allow more variety in trend variation (33). Such modifications can reduce bias, but the resulting model remains dogmatic in its assertions, however plausible the model might be, and tends to produce misleading results due to an underestimation of variability (8).

At the other extreme from the oversimplification of model 1—3, consider the the following model:

$$\theta_k = \text{expit}(\eta_k), \quad (\text{equation 4})$$

where $\eta_1, \dots, \eta_K$ are stratum-specific coefficients (parameters). Model 4 has a regression coefficient for each observed proportion and imposes no assumption whatsoever about how the risk parameters are related to exposure, gravidity, or age. It is sometimes called a *saturated model* because it is “saturated” with parameters, one for every combination of the regressors (exposure, gravidity, age) (33, 34). The η_k are just the logits (log odds) of spontaneous abortion. Solving equation 4 for η_k we get:

$$\text{logit}(\theta_k) \equiv \ln \left(\frac{\theta_k}{1 - \theta_k} \right) = \eta_k. \quad (\text{equation 5})$$

Under model 4, the ordinary (maximum-likelihood) estimators of the θ_k turn out to be A_k/N_k , so that using model 4 is the same as examining the subcohort-specific proportions A_k/N_k and their standard errors.

Although direct examination of subcohort-specific proportions or rates can be very helpful, it often runs up against the roadblock of small numbers. Most of the subcohorts may be so small that the A_k/N_k are highly unstable. As a result, the A_k/N_k may resemble random noise when plotted and reveal few or no general relations; that is, patterns of risk (signals) in observed proportions can be drowned out by random fluctuations (noise). Analysts often try to address this problem by making the stratification very crude, for example, by using only a few categories to represent each of the regressors. This strategy helps keep the A_k large, but unfortunately can itself obscure patterns by combining regressor levels that have disparate risks and by introducing confounding (35, 36).

More generally, problems of biased patterns can adversely affect results from a model that is too constrained

(simplified) to approximate all reasonable possibilities, while problems of instability (noise) and estimate inflation can adversely affect results from a model that imposes too few constraints on the possibilities (8, 17, 35, 36). Analysts sometimes try to cope with these problems by searching for suitable intermediate models, but, in doing so, they often use methods known to be misleading (such as stepwise regression) (8, 17, 27, 37). Multilevel modeling provides a rational and sometimes optimal compromise between the extremes of overconstraint (such as model 1) and its attendant biases and underconstraint (such as model 4) and its attendant variability (37).

Two-level regression

There are several ways to extend regression methods in the multilevel framework. Suppose first that we are interested in examining the risks directly. One of the oldest and simplest approximate methods estimates the risks by taking a weighted average of the fitted risks $\hat{\theta}_{sk}$ from a possible but simple model, such as model 1—3, and the fitted risks $\hat{\theta}_{Gk}$ from a model general enough to approximate most reasonable possibilities (34, 37). That is, to estimate risk θ_k ($k = 1, \dots, K$), we could use the weighted average:

$$\hat{w}_k \hat{\theta}_{Gk} + (1 - \hat{w}_k) \hat{\theta}_{sk}, \quad (\text{equation 6})$$

where $\hat{\theta}_{Gk}$ is the estimate of θ_k from the chosen general model (model 4 in our example), $\hat{\theta}_{sk}$ is the estimate of θ_k from the chosen simple model (model 1 in our example), and $\hat{w}_k$ are estimated weights attached to the general model estimates. (Various weight formulas can be found in references 22, 24, 26, 30, and 34.) In the present example, $\hat{\theta}_{Gk}$ could be the estimate A_k/N_k from model 4, while $\hat{\theta}_{sk}$ could be the estimate from the exponential multiplicative model 1; that is

$$\hat{\theta}_{sk} = \text{expit}(\hat{\alpha} + \hat{\beta}x_k + \hat{\gamma}g_k + \hat{\delta}z_k), \quad (\text{equation 7})$$

where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\delta}$ are the coefficient estimates obtained by fitting model 1 to the observed data using ordinary maximum likelihood.

The hierarchical structure underlying the estimator in equation 6 is as follows. The first stage is a model for the proportions A_k/N_k as probabilistic functions of the parameters η_k ; it assumes that A_k/N_k deviates randomly about $\text{expit}(\eta_k)$. The second stage is a model for the first-stage parameters η_k as probabilistic functions of the second-stage logits in equation 3; it assumes that η_k deviates randomly around the linear term on the right-hand side of equation 3 (which comes from model 1). In this approach, model 1 has been moved up the hierarchy to serve as a second-stage model. More generally, the

simple model used to derive $\hat{\theta}_{sk}$ is the prior model; hence, the $\hat{\theta}_{sk}$ are called the *estimated prior risks*, and the resulting weighted estimators are called the *estimated posterior risks*, the *empirical-Bayes (EB) estimators*, or the *smoothed risks* (although the latter term is sometimes reserved for situations in which the general model is saturated).

As scientific matter, model 4 is too general for most purposes — it excludes no possibility, even an absurd one such as that the odds of spontaneous abortion varies 10^{10} -fold across the observed subcohorts. If all the case counts (A_k) are very large (over 10 or so), this unparsimonious aspect of model 4 should cause no practical problems. Otherwise, to minimize errors, we should use a simpler general model, one general enough to approximate the trends reasonably and interactions we think are reasonably possible. “Reasonably” here depends on context, both on the accuracy we think we need and on what previous information we have. Good acquaintance with the subject is essential here. A 10^{10} -fold variation is patently absurd in most applications; on the other hand, $10^2 = 100$ -fold variation may be absurd (as in the variation in breast-cancer risk across levels of alcohol use in a birth cohort of Finnish women) or may not be absurd (as in the variation in HIV risk across levels of sexual activities).

Combining different levels of information

Suppose now that we are interested in estimating the coefficients in a “general-enough” model, or some simple functions of the coefficients, such as their antilogarithms (which often serve as odds-ratio estimates). In these situations, it is easier to use a more general approach to multilevel estimation in which we average coefficient estimates instead of risk estimates. More precisely, we estimate the coefficients of interest by averaging the coefficient estimates from the general first-stage model with the coefficients predicted from the simpler second-stage (prior) model (27—30, 32, 37—39).

Returning to the example, suppose that study subjects come from Finland, Sweden, Denmark, and Norway; denote these countries by F, S, D, N. Model 1 has no term for country of pregnancy; thus, it assumes that neither risk nor the relations of risk to the regressors vary by country. Suppose we take model 1 as our second-stage model, and as our first-stage model we want a more general model, in which risks and relations to risk may vary among countries. We now must include country in our stratification, so that θ_k represents the average risk for a specific combination of country, exposure, gravidity, and age. Let $C = F, S, D$ or N be the variable standing for country; C is an example of an *ecologic* or *group-level*

variable. One first-stage model that incorporates country effects is

$$\theta_k = \text{expit}(\alpha_c + \beta_c x_k + \gamma_c g_k + \delta_c z_k). \quad (\text{equation 8})$$

This model has 16 coefficients, 4 for each country: $\alpha_F, \beta_F, \gamma_F, \delta_F$ for Finland; $\alpha_S, \beta_S, \gamma_S, \delta_S$ for Sweden; $\alpha_D, \beta_D, \gamma_D, \delta_D$ for Denmark; $\alpha_N, \beta_N, \gamma_N, \delta_N$ for Norway.

Model 1, with only 4 coefficients, predicts that risks are the same across countries, given the same exposure, gravity, and age. This prediction translates to $\alpha_F, \alpha_S, \alpha_D, \alpha_N$ all equal a common value α ; $\beta_F, \beta_S, \beta_D, \beta_N$ all equal a common value β ; $\gamma_F, \gamma_S, \gamma_D, \gamma_N$ all equal a common value γ ; $\delta_F, \delta_S, \delta_D, \delta_N$ all equal a common value δ .

Intuitively, we can imagine an EB estimator for (say) the Finnish exposure coefficient β_F as a weighted average of the estimator $\hat{\beta}_F$ obtained by fitting model 8 and the estimator $\hat{\beta}$ obtained by fitting model 1, which is the model-1 prediction of β_F . Actual estimators are a bit more complicated, as they involve a process of simultaneous *matrix*-weighted averaging of all the coefficient estimators. (See references 27, 30, and 39 for examples.)

Regardless of such technicalities, the underlying 2-level structure can be described as follows: In the first-stage model, A_k/N_k is assumed to vary randomly about the θ_k given by model 8, which allows essentially unlimited systematic variation of incidence across countries (ie, at the ecologic level). In the second-stage model, however, this cross-country (ecologic) variation is constrained to depart only randomly from a common underlying structure for average risks, given by model 1. More precisely, it is assumed that the exposure coefficients $\beta_F, \beta_S, \beta_D, \beta_N$ vary only randomly around an underlying average value β , and so on for the other sets of coefficients.

Of course, we would not really think this variation in true effects is completely random, for there may be some uncontrolled environmental or perhaps genetic differences among the countries that affect risk. But the underlying philosophy in use of the second-stage model 6 is that we have not yet identified any such differences; therefore, we can do no better than treat these differences as unknown and hence as random deviations. Such an approach is certainly more realistic than pretending no such differences exist, which is what using model 1 in a single-stage analysis assumes. It also produces much more accurate results than does an analysis that pretends such differences can be arbitrarily large, which is what using model 8 in a single-stage analysis does.

Semi-Bayes estimation and three-level models

If sample sizes are small, one can do even better than the EB estimators already given if one can specify in advance a probabilistic range for the departures of first-

stage associations from second-stage predictions. In the last example, this strategy would specify a probable range of variation of coefficients across countries. For details of this "semi-Bayes" approach, see reference 22, 27, 28, 32, or 39; for an SAS procedure to perform it, see reference 6.

Both EB and semi-Bayes estimators are derived from 2-level models, with EB estimators having less bias but larger variability than typical semi-Bayes estimators. Bayes empirical-Bayes (BEB) estimators, derived from 3-level models, can achieve greater accuracy over a broad range than either 2-level estimator. In the last example, a BEB estimator could be obtained by adding a third-stage (hyperprior) model for the parameters of the second-stage model (model 1) (9, 31). Unfortunately, published examples of BEB analysis remain rare, in part due to a lack of software for 3-level modeling in major statistical packages.

Discussion

The multilevel modeling or hierarchical regression approach is completely general, in the sense that the models can encompass all other parametric regression models and can be extended to nonparametric regression as well (via the relation of smoothers to roughness penalties) (40, 41). A separate issue is the estimation criteria and procedures used to fit the models. Nearly all discussions, including the present one, focus on minimizing the total expected squared error as an estimation criterion. This focus leads to natural extensions of least squares and likelihood estimation for multilevel models (1–3, 27, 30, 37, 39, 42). Except for normal (Gaussian) linear models, these are approximate methods with certain sample size requirements, although these requirements tend to be less restrictive than those for single-stage modeling (27, 30). Exact hierarchical estimation is also possible via Monte-Carlo methods (4, 5).

Multilevel estimation criteria can be modified to recognize qualitative differences in estimation errors. Such differences arise when model parameters can be divided into 2 groups, with 1 group containing parameters that represent effects of the exposure or exposures of interest (including any product terms involving an exposure) and the other containing parameters that represent purely confounding effects. Model 8 provides an example in which the β_c represent the exposure effects. There, accurate estimators for country, gravity, and age effects (α_c, γ_c , and δ_c) are important for the control of confounding, but errors in these estimators are important only insofar as they induce errors in estimation of the exposure effects. A simple way to account for this coefficient grouping is to use separate estimation criteria for the exposure and

confounder coefficients (43). Nonetheless, even greater accuracy for the exposure coefficients can be obtained by using accuracy in estimating the exposure coefficients as the sole criterion (44, 45). Such a criterion leads to shrinkage of both the exposure and the confounder coefficient estimates, but it sacrifices accuracy in the confounder estimates whenever such sacrifice appears to improve accuracy in the estimation of exposure effects.

The preceding presentation has emphasized potential benefits of the multilevel or hierarchical perspective: reduction in estimation error and the use of prior information in a relatively cautious but thorough fashion. The benefits of such use have been illustrated in many published applications (7, 25–28, 30). Further applications of multilevel modeling include the control of multiple-comparison artifacts (22, 27, 28), statistical decision analysis (46), and the meta-analysis of study-specific estimates (47). Multilevel modeling is, however, the most widely recognized as a method for entering ecologic (group-level) variables such as country and individual-level variables such as age into the same analysis, while keeping their levels distinct (1, 3, 38). The theory and examples of these applications are now found in textbooks (1, 3), as well as in many articles in education and related fields. In what may be a promising sign, these ideas are starting to appear in the health sciences as well (48).

Acknowledgments

I would like to thank Charles Poole and Ellen Eisen for their comments on the initial draft of this paper.

References

1. Bryk AS, Raudenbush SW. Hierarchical linear models. Newbury Park (CA): Sage, 1992.
2. Searle SR, Casella G, McCulloch CE. Variance components. New York (NY): Wiley, 1992.
3. Goldstein H. Multilevel statistical models. 2nd ed. London: Edward Arnold, 1995.
4. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. New York (NY): Chapman and Hall, 1995.
5. Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis. New York (NY): Chapman and Hall, 1996.
6. Witte JS, Greenland S, Kim L-L. Software for hierarchical modeling of epidemiologic data. *Epidemiology* 1998;9:563–6.
7. Efron B. Biased versus unbiased estimation. *Adv Math* 1975;16:259–77.
8. Leamer EE. Specification Searches. New York (NY): Wiley, 1978.
9. Good IJ. Good thinking. Minneapolis (MI): University of Minneapolis Press, 1983:chapter 9.
10. DeFinetti B. The theory of probability, vol 1 and 2. New York (NY): Wiley, 1974.
11. Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *Am Sci* 1988;76:159–65.
12. Howson C, Urbach P. Scientific reasoning: the Bayesian approach. 2nd ed. LaSalle (IL): Open Court, 1993.
13. Greenland S. Probability logic and probabilistic induction. *Epidemiology* 1998;9:322–32.
14. Copas JB. Regression, prediction, and shrinkage. *J R Stat Soc [B]* 1983;45:311–54.
15. Good IJ. Hierarchical Bayesian and empirical Bayesian methods [letter]. *Am Stat* 1987;41:92.
16. Box GEP. An apology for ecumenism in statistics. In: Box GEP, Leonard T, Wu CF, editors. Scientific inference, data analysis, and robustness. New York (NY): Academic Press, 1983.
17. Greenland S. Principles of multilevel modeling. *Int J Epidemiol*. In press.
18. Hill JR. A general framework for model-based statistics. *Biometrika* 1990;77:115–26.
19. Efron B, Morris CN. Stein's estimation rule and its competitors — an empirical Bayes approach. *J Am Stat Assoc* 1973;68:117–30.
20. Brandwein AC, Strawderman WE. Stein estimation: the spherically symmetric case. *Stat Sci* 1990;5:356–69.
21. Casella G. An introduction to empirical-Bayes data analysis. *Am Stat* 1985;39:83–7.
22. Greenland S, Poole C. Empirical Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health* 1994;49:9–16.
23. Draper D, Hodges JS, Mallows CL, Pregibon D. Exchangeability and data analysis. *J R Stat Soc [A]* 1993;156:9–37.
24. Morris CN. Parametric empirical Bayes: theory and applications (with discussion). *J Am Stat Assoc* 1983;78:47–65.
25. Efron B, Morris CN. Stein's paradox in statistics. *Sci Am* 1977;236:119–27.
26. Efron B, Morris CN. Data analysis using Stein's estimator and its generalizations. *J Am Stat Assoc* 1975;70:311–9.
27. Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. *Stat Med* 1993;12:717–36.
28. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology* 1994;5:612–21.
29. Witte JS, Greenland S. Simulation study of hierarchical regression. *Stat Med* 1998;15:1161–70.
30. Greenland S. Second-stage least squares versus penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. *Stat Med* 1997;16:515–26.
31. Deely JJ, Lindley DV. Bayes empirical Bayes. *J Am Stat Assoc* 1981;76:833–41.
32. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ Health Perspect* 1994;102(suppl 8):33–9.
33. Greenland S. Introduction to regression models. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, 1998:chapter 20.
34. Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis: theory and practice. Cambridge (MA): MIT Press, 1975.
35. Greenland S. Dose-response and trend analysis: alternatives to category-indicator regression. *Epidemiology* 1995;6:356–

- 65.
36. Greenland S. Problems in the average-risk interpretation of categorical dose-response analysis. *Epidemiology* 1995; 6:563—5.
37. Greenland S. Introduction to regression modeling. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed. Philadelphia (MA): Lippincott-Raven, 1998:chapter 21.
38. Wong GY, Mason WM. The hierarchical logistic regression model for multilevel analysis. *J Am Stat Assoc* 1985;80:913—24.
39. Greenland S. A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupation cancer-mortality study. *Stat Med* 1992;11:219—230.
40. Hastie T, Tibshirani R. *Generalized additive models*. New York (NY): Chapman and Hall, 1990.
41. Green CT, Silverman BW. *Nonparametric regression and generalized linear models*. New York (NY): Chapman and Hall, 1994.
42. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993;88:9—25.
43. Lindsay BG. Using empirical partially Bayes inference for increased efficiency. *Ann Stat* 1985;13:914—31.
44. Kalish L. Reducing mean squared error in the analysis of pair-matched case-control studies. *Biometrics* 1990;46:493—9.
45. Greenland S. Reducing mean squared error in the analysis of stratified epidemiologic studies. *Biometrics* 1991;47:773—5.
46. Cox DR, Hinkley DV. *Theoretical statistics*. New York (NY): Chapman and Hall, 1974:chapter 11.
47. Raudenbush SW, Bryk AS. Empirical-Bayes meta-analysis. *J Educ Stat* 1985;10:75—98.
48. Von Korff M, Koepsell T, Curry S, Diehr P. Multilevel analysis in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol* 1992;135:1077—82.