Some examples of Bayesian thinking in policy analysis (with an emphasis on applications in health)

Alan M. Zaslavsky

Harvard Medical School

1

Bayes for evidence synthesis

- "… Bayesian inference readily offers a formal process for synthesizing data from multiple sources, in keeping with the principles of evidence-based medicine" *(McMahon et al. 2006)*

2

Meta-analysis

- Multiple studies give estimates of a treatment effect
  - Sampling variation within each study
  - Possibly: variation across studies
- "Random effects meta-analysis": hierarchical modeling to estimate:
  - Mean and variation of treatment effects
  - Inference about treatment effect in each study

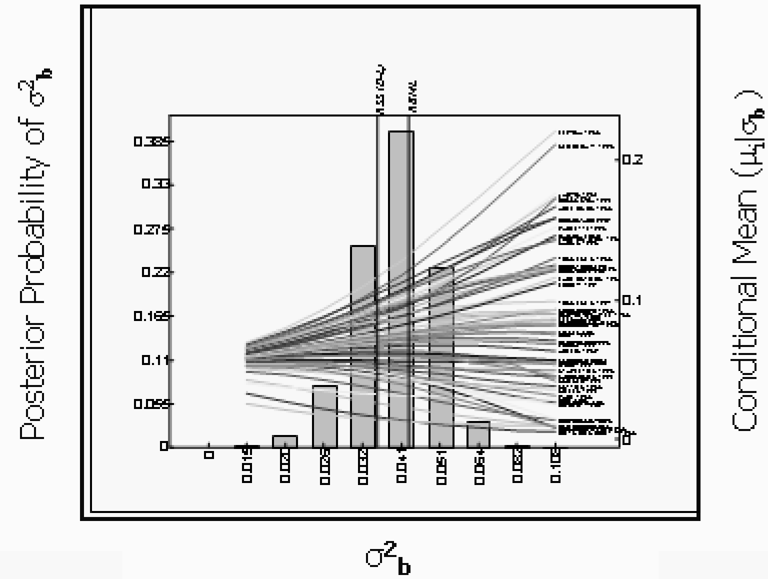
3

Bayesian profiling in health care

- Measures of "quality" in a number of units
  - Hospitals, doctors, health plans, etc.
  - Measures subject to sampling error
- Posit an exchangable distribution of "true" quality
  - Or: exchangeable errors from regression on covariates
- Inferences for "units" as probabilities of comparisons, orderings, exceeding thresholds, etc.
- Biggest payoff with unequal information for different units

4

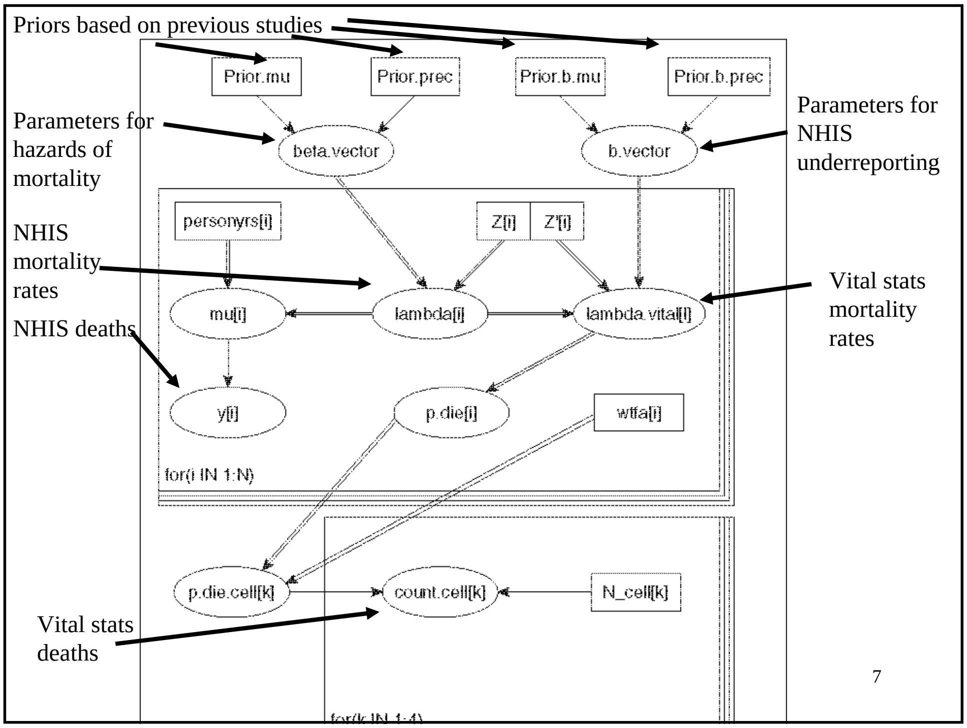
TRACE PLOT: Change in Random Effects as a function of σ_b



Bayesian model for lung cancer mortality and competing risks

- Data sources:
  - NHIS: demographics, smoking history, cause of death; modest sample size, undercoverage of deaths
  - Vital stats: demographics, cause of death; population coverage
- Regression model with piecewise (by age range) constant hazards
- Objective: synthesize to obtain estimates of hazards of death from various causes, by smoking status

6



Lung cancer model: results

- Obtain posterior inferences about mortality models
- Vary assumptions about prior information, etc.
  (McMahon et al. 2006)

8

Synthesizing data sources: imputing underreported adjuvant therapy for cancer

- QOCC="Quality of Cancer Care" study
- California Cancer Registry data
  - Stage/site appropriate for adjuvant therapy
  - 13,878 incident cases in 1994-1998
  - Indicators for adjuvant chemotherapy
- Physician followback survey
  - Check medical record for adjuvant therapies
  - 1449 survey responses obtained

9

Reporting of chemotherapy

- 82% of adjuvant therapy was reported to Registry (among "respondents")
  - Substantial underestimation if Registry alone used
  - More complete in teaching hospitals, HMO affiliates, high volume hospitals, younger and rectal patients

Cress et al., *Medical Care* 2003

10

Adjuvant chemotherapy: “Naïve” survey analyses

- Analysis based only on “gold standard” survey + Registry data
- Strong variation by patient characteristics
- Substantial unexplained hospital variation

Ayanian et al., *J Clinical Oncology* 2003

11

Limitations of “naïve” analyses

- Survey respondents alone:
  - Small portion of available California data (1449 / 13,878)
  - Unrepresentative due to survey nonresponse, limited scope
  - Confounding of survey response, reporting, treatment variation
- Registry data alone:
  - Underreporting of chemotherapy
  - Reporting is nonuniform (by patient/hospital)

12

Combining Registry and survey data

- Combine
  - power of large Registry data
  - correction for underreporting based on survey
- Simple correction based on:

P(reported chemo) =

P(chemo) × P(report | chemo)

Therefore: P(chemo) =

P(reported chemo) / P(report | chemo)

13

Model-based methodology

(Yucel and Zaslavsky, *JASA* 2005)

- Disaggregated model
  - Take into account individual effects on both chemotherapy and reporting
  - Take into account hospital variation in both chemotherapy and reporting
- Imputation of chemo for individual cases
  - Allow fitting of any desired models
  - Multiple imputation to obtain proper measures of uncertainty with imputed data

14

Basic modeling approach

- Data consists of:
  - $Y_{(O)}$= true outcome (therapy given)
    - Observed for survey cases
  - $Y_{(R)}$= reported outcome (from Registry)
  - X = covariates
- Factor into outcome and reporting models

$$P(Y_{(O)}, Y_{(R)} \mid X, \theta) = P(Y_{(O)} \mid X, \theta_{(O)}) \cdot P(Y_{(R)} \mid Y_{(O)}, X, \theta_{(R)})$$

Each represents distinct scientific process; can build in different assumptions about predictors, “transportability” across areas and times, etc.

15

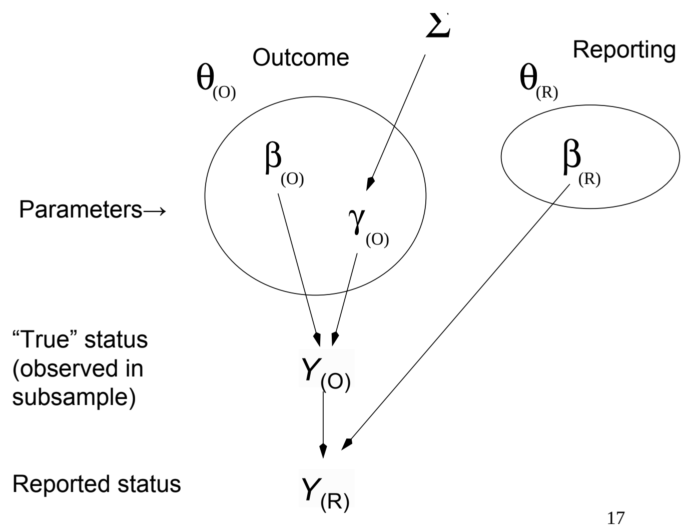
Models for reporting and therapy

- Logit or Probit regression for therapy (<u>outcome</u>)
  - Patient p has characteristics x_{hp}
  - Hospital h has characteristics z_h
  - Random effect γ_h for hospital h

  $\text{logit } P(\text{chemo}_{hp}) = \beta x_{hp} + \lambda z_h + \gamma_h$
- Similar model for <u>reporting</u> given therapy
- Fully Bayesian specification and model fitting

16

Hierarchical model structure



17

Direct interpretation of fitted model

- Probability of chemo given covariates
- Covariate effects broadly similar to those in survey-only analyses.
  - Hospital volume effects substantially different
- Substantial hospital random effects in both reporting and therapy rates
  - Indication of substantial unexplained variation – a problem (from health services standpoint)!

18

Application: MI analysis of effect of chemotherapy on survival

- Fit underreporting model including *2-year survival* as predictor of *chemotherapy*
- Multiple imputation = device for representing a joint predictive distribution
- Using imputed corrected chemotherapy, fit model with *chemotherapy* (and other variables) as predictor of *survival*
  - Correct variances with multiple imputation
  - Missing info ≈70% for chemo, 1-4% for other variables
- Significant positive effect (OR=1.26) of chemo on survival

19

Bayesian approach to characterizing uncertainty in a policy microsimulation model

- Focus here on synthesizing information about ***uncertainty***

20

Elements of microsimulation models

1. Base data set (survey or administrative records).
2. Merge with information from other files (statistical matching).
3. Simulation of individual behaviors:
   - over time (dynamic aging),
   - under changed policies/conditions (behavioral response).

   (Simulation is at the level of the individual unit, unlike cell-based or regression models.)
4. Simulate outcomes (costs and benefits) under policy alternatives, at alternative conditions, or over time.
5. Summarize aggregates and calculate contrasts.

Sources of uncertainty in microsimulation models

1. **Sampling variability** in the base data set.
2. **Stochastic variability** in
   - completion of simulation data set, and
   - stochastically simulated outcomes.
3. **Model parameter uncertainty:** model parameters are obtained from studies external to the microsimulation and/or "expert opinion".
4. **Model specification uncertainty** arises because there are alternative model specifications and methods.

   * *Models are required for:*
   - combining files (matching), imputing missing items, and correcting for undercoverage and response error (calibration);
   - imputing behavioral responses;
   - imputing changes in individual status over time;
   - representing contextual ("macro") effects.
5. **Choice of welfare measures:** lack of consensus.

Why do we model uncertainty?

- What do we really know? (Don't make policy decisions based on random noise!)
- Sensitivity to alternative assumptions.
- Information for design of future data collection, modeling, and simulation efforts.

What do we do with uncertainties?

- *Stochastic simulation uncertainty:* Estimate its contribution to variance; reduce through additional simulation effort.
- *Sampling error; parameter estimation:* Estimate their contribution to variance; reduce through further data collection.
- *Model specification and assumed model parameters; macro effects; welfare measures:* Sensitivity analysis to varying assumptions, summarize by estimating their contribution to posterior variance. Seek better understanding and consensus.

The MATH-CPS model

- Based on Current Population Survey (CPS) March Supplement: 60,000 households, complex sampling design with varying weights.
- Data "aged" from 1993 to 1996
- Imputations of monthly income, monthly employment status, financial and vehicular assets, medical and childcare expenses
- Corrections (calibration) to national totals for:
  - CPS totals to match census totals (via CPS weights)
  - AFDC, SSI, and GA payments and participation rates to administrative totals
  - FSP recipiency (in general, among AFDC recipients) to administrative totals

MATH-CPS model outputs

- Models participation and benefits in the food stamp program under base plan and reform plan
- Typical (and important) estimand:
  % change in benefits $= \frac{\text{reform benefits - baselaw benefits}}{\text{baselaw benefits}} \times 100\%$
- Tables contain many other estimands corresponding to effects on particular regions, income groups, etc.

Uncertainties in the MATH-CPS model

- Sampling in the Current Population Survey *(sampling error)*
- Stochastic model for behavioral response to change in potential benefits. *(Stochastic simulation error.)*
- Stochastic model for underreporting of participation. *(Stochastic simulation error.)*
- Calibration of AFDC totals to administrative records. *(Model specification uncertainty.)*
  Note: calibration both *reduces* and *increases* uncertainty.
- Allocation of annual income to months. *(Model specification uncertainty.)*
- Assumptions regarding unemployment rates ("macro effect"). *(Model parameter uncertainty.)*

Error analysis techniques for MATH-PC

- *Stochastic simulation error:*
  - Repeated simulations with different random number *seeds*.
  - Variance reduction using antithetic *streams* of random numbers (implemented in other models, not MATH-CPS).
  - Could be broken down into components corresponding to different stochastic features of model.
- *Sampling error:* Jackknife variance estimation.
  - Grouping for computational convenience.
  - "Rotation groups" are essentially replications of the CPS sample design: gives valid estimates of survey error.
- *Model parameter and specification uncertainty:* repeated simulation with alternative values of parameters or alternative specification of models.
  (Levels of these effects have identifiable meanings.)
  Select levels as ± 1 SD range of parameter values, or equally plausible alternative models.

Alternative model specifics I: Calibration of AFDC totals

MATH model: Simulate participants among eligible nonreporters.
Problems:

1. Targets are not met:

Benefits ($\times \$10^6$)	Target	MATH model
AFDC-Basic	1,850	1,499
AFDC-UP	180	73

2. Model assumption: truth for nonreporters is the way they look in CPS. Not necessarily true!

Alternate model: Upweight recipients to meet control totals; differential upweighting for those receiving larger benefits.
(Recalculate upweighting for each jackknife replication and random seed.)

1. Better fit to controls.
2. Compare to alternative assumption about response error.

Alternative model specifics II: Income allocation

MATH model: Allocate earned income evenly across months worked.

Alternate model: For low-income families, multiply income by a random number drawn from a Uniform(.5,1.5).

Alternative model specifics III: Unemployment rate

MATH model: Use 1993 unemployment rate (6.4%)

Alternate model: For 40% of people employed part of the year, make simulation month be one of employment. Result: 4.1% unemployment.

Design of simulation experiment

- Jackknife replicate (j): 8 levels
- Seed (s): 2 nested in each jackknife replicate (16 total)
- Calibration for AFDC ($c = \pm 1$): 2 levels (yes/no)
- Unemployment ($e = \pm 1$): 2 levels (6%, 4%)
- Income allocation ($i = \pm 1$): 2 levels (even, uneven)
- Altogether $8 \times 2 \times 2 \times 2 \times 2 = 128$ conditions
- Design (**seed** nested in **jackknife**) $\times$ **calibration** $\times$ **unemployment** $\times$ **income allocation**

Modeling population and posterior variances using a variance components model

1. Write a variance components model for the population.
2. Write expected mean squares in simulation experiment ANOVA in terms of these variance components.
3. Solve for variance components estimates.
4. Calculate posterior variances for quantities of interest, using the estimated variance components.

Prior distributions of "fixed effects" and model specifications

- Assume flat priors for

$$\beta, \beta_C, \beta_E, \beta_I, \beta_{CE}, \beta_{CI}, \beta_{EI}, \beta_{CEI}$$

- Assume prior equipoise regarding model specification

$$c = \begin{cases} 1 & \text{with probability } .5 \\ -1 & \text{with probability } .5 \end{cases}$$

$$e = \begin{cases} 1 & \text{with probability } .5 \\ -1 & \text{with probability } .5 \end{cases}$$

$$i = \begin{cases} 1 & \text{with probability } .5 \\ -1 & \text{with probability } .5 \end{cases}$$

Inference: variance components

Means squares in ANOVA for experimental outputs can be equated with expectations under the model. Solve for $\sigma_H^2, \sigma_S^2, \ldots$ ("method of moments").

Inference: "fixed effects"

With flat priors, posterior mean = unbiased point estimate:

$$\mathbf{E}\, \beta|\text{data} = \hat{\beta} = \hat{Y}_{.....}$$

and posterior variance = sampling variance

$$\text{Var}\, \beta|\text{data} = \text{Var}\, \hat{\beta}|\beta = \frac{\sigma_H^2}{n} + \frac{\sigma_S^2}{n(J-1)S}$$

(and similarly for $\beta_C, \beta_E, \beta_I, \ldots$)

Results for one reform
Estimand: percent change in benefits

Effects	
$\hat{\beta}$	8.47 (.69)
$\hat{\beta}_C$	−0.82 (.10)
$\hat{\beta}_E$	−0.08 (.36)
$\hat{\beta}_I$	0.04 (.13)
$\hat{\beta}_{CE}$	0.02 (.006)
...	
posterior SE	1.15
95% interval	(6.18, 10.76)

% var due to:		
sampling (H)	34.64	28.66
stochastic (S)	1.53	17.72
calibration (C)	51.43	42.56
employment (E)	0.50	0.41
income (I)	0.10	0.90
C × E	0.02	0.02
C × I	0.001	0.001
E × I	0.001	0.001
E × sampling	9.67	8.00
...		

Results for one reform
Estimand: total benefits

Effects ($ million)	
$\hat{\beta}$	1874 (43.7)
$\hat{\beta}_C$	121 (6.2))
$\hat{\beta}_E$	−101 (9.7)
$\hat{\beta}_I$	6 (7.5)
$\hat{\beta}_{CE}$	− 8.3 (0.4)
...	
posterior SE	165
95% interval	(1544, 2204)

% var due to:		
sampling (H)	6.9	6.7
stochastic (S)	0.1	1.8
calibration (C)	54.2	53.2
employment (E)	37.7	37.1
income (I)	0.15	0.15
C × E	0.3	0.2
C × I	0.001	0.001
E × I	0.004	0.004
E × sampling	0.3	0.3
...		

Results for one reform
Estimand: total benefits

Effects ($ million)	
$\hat{\beta}$	1874 (43.7)
$\hat{\beta}_C$	121 (6.2))
$\hat{\beta}_E$	−101 (9.7)
$\hat{\beta}_I$	6 (7.5)
$\hat{\beta}_{CE}$	− 8.3 (0.4)
...	
posterior SE	165
95% interval	(1544, 2204)

% var due to:		
sampling (H)	6.9	6.7
stochastic (S)	0.1	1.8
calibration (C)	54.2	53.2
employment (E)	37.7	37.1
income (I)	0.15	0.15
C × E	0.3	0.2
C × I	0.001	0.001
E × I	0.004	0.004
E × sampling	0.3	0.3
...		

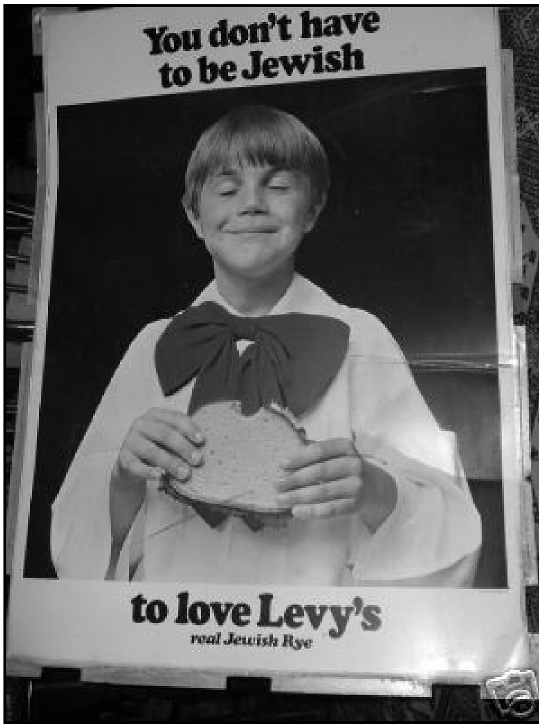
Percentages of posterior variance of % change in benefits, due to major sources of error, in six MATH model reforms.

	reform number					
	1	2	3	4	5	6
sampling (H)	34.638	0.000	9.133	0.000	48.481	0.000
seed (S)	1.529	0.789	5.192	10.857	0.477	4.265
calibration (C)	51.431	22.766	35.247	55.545	44.777	5.492
unemployment (E)	0.497	58.354	11.614	0.956	0.233	86.843
income (I)	0.103	16.337	7.683	0.153	0.005	1.696
C*E	0.022	0.062	0.272	0.021	0.004	0.055
C*I	0.001	0.065	0.037	0.000	0.000	0.016
E*I	0.001	0.099	0.023	0.001	0.000	0.002
C*E*I	0.000	0.001	0.000	0.000	0.000	0.000
E*H	9.671	0.000	28.994	22.516	4.335	0.000

Conclusions on microsimulation error

- Only from a (subjective) Bayesian perspective can we combine/compare these different types of error.
- Similar principles apply any time we try to make policy projections based on data and models

38



"You don't have to be Jewish to love Levy's real Jewish Rye"

39

You don't have to be a Bayesian to love …

- Bayes's theorem
- Hierarchical modeling
- Exchangeable distributions
- Coherent inference based on probability distributions
- Consistent computational approaches
- Consistent framework for all types of uncertainty
- Good frequency properties

40

Models or Bayes?

- More important decision might be whether to use models, not whether to use priors
- Subjective probability statements might better correspond to ordinary discourse about uncertainty
  - More comprehensible to policymakers

41

The iron discipline of Bayes

- Once models are fully specified, inferences flow “automatically”
  - Little room for kluges
  - If there is a problem, you have to fix the model
- Model everything?
- Statistical judgement – what is important?
- Still jobs for skilled and wise analysts!

42