Empirical Bayes Methods for Estimation of Adverse Event Rates in Clinical Trials and Active Surveillance

William DuMouchel, PhD
Lincoln Technologies, Inc.

New Jersey Chapter ASA Symposium

June 7, 2005 (revised 6/9/2005)

Copyright © 2005 Lincoln Technologies, Inc.
All rights reserved

Data Mining of Adverse Drug Reactions

- Databases of Spontaneous Adverse Drug Reaction Reports
  - Objectives and Limitations
- Drug – Event Counts as a Two-Way Table
  - Empirical Bayes Approach to Disproportionality Analysis
- Software Implementation
  - In Use at FDA and Several Pharmaceutical Companies
- Logistic Regression for Spontaneous Reports
  - Helps Avoid Masking and Confounding
- Data Mining of Clinical Trial Drug ADR Data
  - Similarities and differences with post-market situation
- Bayesian Models for Collections of Clinical Trial 2x2 Tables
- Example

Clinical Trial Safety Data

- Similarities with spontaneous report data
  - Lots of 2 x 2 Tables
  - Must cope with multiple comparisons
- Differences
  - Smaller sample sizes
  - Usually just two treatments vs. thousands in database
  - Prospective study
  - Randomized allocation of treatment
  - More valid comparison group
- Active surveillance studies
  - Large longitudinal database of medical records
  - Attempt to match users of two drugs with propensity scores
  - Maybe closer to clinical trial data than to spontaneous reports

Bayesian Shrinkage Models

- Statistical validity of searching for extreme differences
- Classical approach to post-hoc interval estimates
  - Maintain centers of CI at observed differences
  - Expand widths of every CI
  - Expansion is greater the more differences you look at
  - If you look at too many, the CI’s are too wide to be useful
- Bayesian approach
  - Requires a prior distribution for differences
    - Can estimate it from the multiple observed differences available
  - Centers of CI’s are “shrunk” toward average or null difference
    - High-variance differences shrink the most
  - Widths of CI’s usually shrink a little too
  - The more you look at, the better you can model the prior dist.

Example Data (Berry & Berry 2004)

	BodySystem	a	b	c	d
Astenia/fatigue	1	57	40	91	92
Fever	1	34	26	114	106
Infection, fungal	1	2	0	146	132
Infection, viral	1	3	1	145	131
Malaise	1	27	20	121	112
Anorexia	2	7	2	141	130
Cendisiasis, oral	2	2	0	146	132
Constipation	2	2	0	146	132
Diarrhea	2	24	10	124	122
Gastroenteritis	2	3	1	145	131
Nausea	2	2	7	146	125
Vomiting	2	19	19	129	113
Lymphadenopathy	3	3	2	145	130
Dehydration	4	0	2	148	130
Crying	8	2	0	146	132
Insomnia	8	2	2	146	130
Irritability	8	75	43	73	89
Bronchitis	9	4	1	144	131
Congestion, nasal	9	4	2	144	130
Congestion, respiratory	9	1	2	147	130
Cough	9	13	8	135	124
Infection, upper respir.	9	28	20	120	112
Laryngotracheobronchitis	9	2	1	146	131
Pharyngitis	9	13	8	135	124
Rhinorrhea	9	15	14	133	118
Sinusitis	9	3	1	145	131
Tonsillitis	9	2	1	146	131
Wheezing	9	3	1	145	131

PATIENT COUNTS	Treatment Group	Comparator Group
Adverse Event	*a*	*b*
No Adverse Event	*c*	*d*

	BodySystem	a	b	c	d
Bite/Sting	10	4	0	144	132
Eczema	10	2	0	146	132
Pruritis	10	2	1	146	131
Rash	10	13	3	135	129
Rash, diaper	10	6	2	142	130
Rash, measles/rub.-like	10	8	1	140	131
Rash, varicella-like	10	4	2	144	130
Urticaria	10	0	2	148	130
Viral exanthema	10	1	2	147	130
Conjunctivitis	11	0	2	148	130
Otitis media	11	18	14	130	118
Otorrhea	11	2	1	146	131

Choices of Bayesian Model

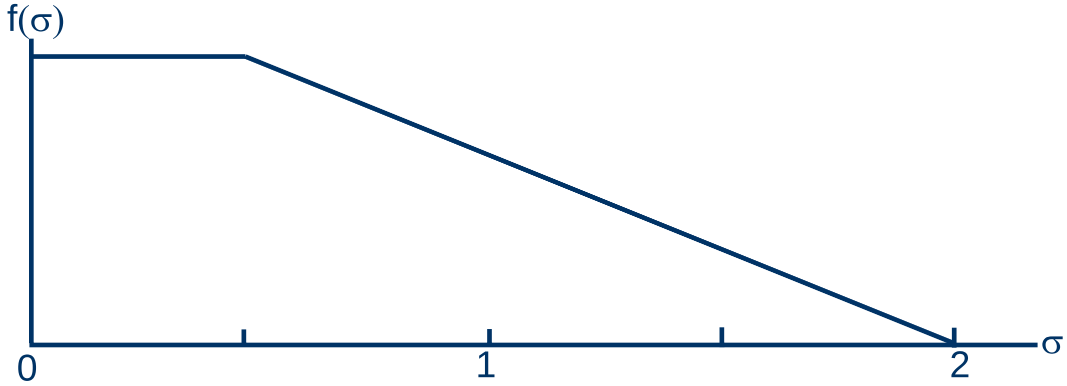
- MGPS as used for spontaneous reports
  - Gamma-Poisson model assumed baseline E’s were known
  - In 2 x 2 table from database, a << b << d and a << c << d
    - This made the Poisson assumption work
  - In clinical trial data, a, b are of similar magnitude
    - Model needs two binomial distributions, one for each arm
- Bayesian model of Berry & Berry (Biometrics 2004)
  - Three stage hierarchical model
    - Body systems, AE’s within body system, data in 2 x 2 tables
    - Use binomial likelihood function for each group for each AE
    - Logistic transform of control group probabilities has normal dist.
    - Log odds has mixture dist. of normal and point mass at 0
    - Dozens of parameters define the complete hierarchy
  - Markov chain Monte Carlo estimation procedure
    - Complicated special programs required

Simple Empirical Bayes Model

- Normal approximation to distribution of log odds ratio
  - Notation: 2 x 2 counts are (a_k, b_k, c_k, d_k) for $1 \le k \le K$
  - $y_k = \log[(a_k+.5)(d_k+.5)/ ((b_k+.5)(c_k+.5))]$
  - $v_k = 1/(a_k+.5) + 1/(b_k+.5) + 1/(c_k+.5) + 1/(d_k+.5)$
  - Assume $y_k \sim N(\theta_k, v_k)$ (θ_k is "true" log odds ratio for table k)
  - $\theta_k \sim N(\mu, \sigma^2)$ (prior distribution for θ_k -- must estimate σ^2)
- Assume μ is known
  - Usually take null hypothesis $\mu = 0$
  - μ could be based on a higher level analysis
    - y_k based on patient subgroups where μ is log OR of full sample
- Posterior distribution for each subtable
  - $\theta_k|y_k \sim N(\mu + (y_k - \mu)\, \sigma^2/(\sigma^2 + v_k),\ \ v_k\, \sigma^2/(\sigma^2 + v_k))$
    - Note simple shrinkage formula for both mean and variance
    - Take anti-logs to get CI for odds ratios

Estimating Prior Variance of Log Odds

- Bayesian estimate preferred to maximum likelihood
  - Likelihood for σ defined by $y_k \sim N(\mu, v_k + \sigma^2)$
    - $\text{logLike}(\sigma) = - \Sigma_k [\log(v_k + \sigma^2) + (y_k - \mu)^2/(v_k + \sigma^2)]/2$
    - Very simple to maximize, but don't believe it if MLE of $\sigma = 0$
  - Better to define a prior for σ and compute posterior mean
    - Uniform prior for $0 < \sigma < .5$, triangular for $.5 < \sigma < 2$
    - Prior mean for σ of 0.7 corresponds to factor of 2 for odds ratio
    - Multiply likelihood by prior to get posterior distribution
    - Compute mean by one-dimensional numerical integration



Refinements of the Simple Model

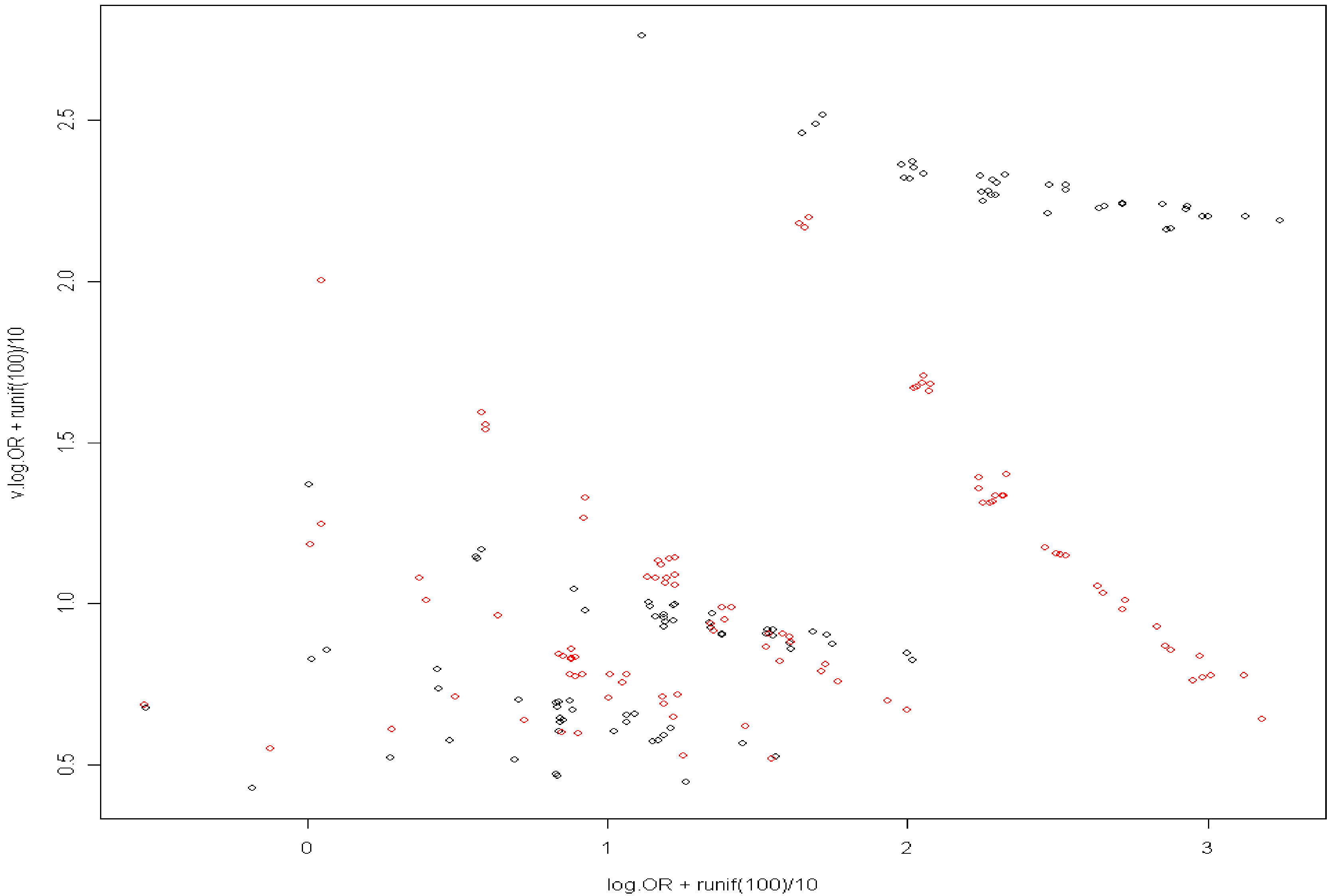
- Mixture prior distribution for θ_k
  - Assume effect *sparsity*: not much difference for most AEs
    - $\theta_k \sim N(\mu_0, \sigma^2/10)$ with probability P_0
    - $\theta_k \sim N(\mu_1, \sigma^2)$ with probability $(1 - P_0)$
    - Changing factor of 10 above would lead to different P_0
  - Must estimate μ_1, σ and P_0
    - Use posterior mean for σ and P_0
    - MLE for μ_1 using E-M algorithm
  - Posterior distribution for each θ_k is a mixture of normals
    - Usually well approximated by a single normal distribution
- Although this refinement is intuitively appealing, and doesn't involve much extra computational complexity, it is hard to find examples where it makes much of a difference
  - Should prevent over-shrinkage in presence of very large effects

Refinement (2)

- Definition of v_k produces correlation between y_k & v_k
  - $y_k = \log[(a_k+.5)(d_k+.5)/ ((b_k+.5)(c_k+.5))]$
  - $v_k = 1/(a_k+.5) + 1/(b_k+.5) + 1/(c_k+.5) + 1/(d_k+.5)$
  - Suppose a's, c's and d's are usually 3 or more, and the b's are randomly 0, 1, or 2
    - b = 0 increases both y and v, b = 2 decreases both y and v
    - Results in smaller values of y getting more weight than large y
- Iterative reweighting can reduce this effect
  - First get shrinkage estimates of every OR_k using original v_k
  - Find e_k: $(a_k+.5+e_k)(d_k+.5+e_k)/ ((b_k+.5–e_k)(c_k+.5–e_k)) = OR_k$
  - Set $v_k = 1/(a_k+.5+e_k) + 1/(b_k+.5–e_k) + 1/(c_k+.5–e_k) + 1/(d_k+.5+e_k)$
  - Refit using new v's and repeat these steps until convergence
    - Variances match fitted OR's and are less noisy than original v_k

Simulate a=Bin(100, .05) vs. b=Bin(100, .01): Plot logOR vs V(logOR)

Original: cor=.76, fitted OR=2.9; Reweighted: cor=.07, fitted OR=4.1



Refinement (3)

- Suppose a trial has 200 patients in each arm
  - For an AE overall, counts are 8/200 and 4/200
  - Patients taking concomitant drug, counts are 5/8 and 0/2
    - Viewed alone, small subgroup difference is not significant
    - But outside this subgroup, fewer than 2% have the AE!
    - Isn't 5 out of 8 significant in light of the rest of the data?
- Empirical Bayes smoothing of AE proportions
  - Instead of adding .5 to every cell, estimate α, β, γ, δ
  - $y_k = \log[(a_k+\alpha)(d_k+\delta)/ ((b_k+\beta)(c_k+\gamma))]$, analogously for v_k
  - Beta prior distributions for prob(AE) in each arm
    - Prob(AE|Treatment) ~ Beta(α, γ)
    - Prob(AE|Control) ~ Beta(β, δ)
    - Can estimate α, β, γ, δ using well-known beta binomial model
  - If proportions are consistent across k: α, β, γ, δ become large
    - Distinguish consistency of proportions vs. consistency of odds ratios

Refinement (3, continued)

Example Data

	a	b	c	d
1	5	0	3	2
2	0	0	6	4
3	0	0	6	4
4	0	0	4	6
5	1	0	13	6
6	1	0	13	6
7	0	0	9	11
8	0	0	8	12
9	0	0	15	15
10	0	0	14	16
11	0	0	15	15
12	1	0	12	17
13	0	1	22	17
14	0	2	18	20
15	0	1	15	24
16	0	0	19	21
total	8	4	192	196

- Estimated hyperparameters
  - $\alpha = .15$, $\beta = .99$, $\gamma = 2.64$, $\delta = 48.99$
  - Estimation performed under constraints
    - α < min(1, mean(a_k)); γ < min(250, mean(c_k))
    - β < min(4, 4*mean(b_k)); δ < min(1000, 4*mean(d_k))
    - Prevents prior from overwhelming the data
    - Restrict prior for P(AE|Treatment) more severely because it is also affected by prior for Odds Ratio
- Estimate and 90% CI for group 1 OR
  - Also used iterative reweighted variances
  - Set μ_0 = 0.5 since overall data has OR = 2
  - $y_1 = 3.85$, $v_1 = 0.92$, $\sigma = 1.05$
  - $OR_1 = 10.9$, $3.5 < OR_1 < 34.1$

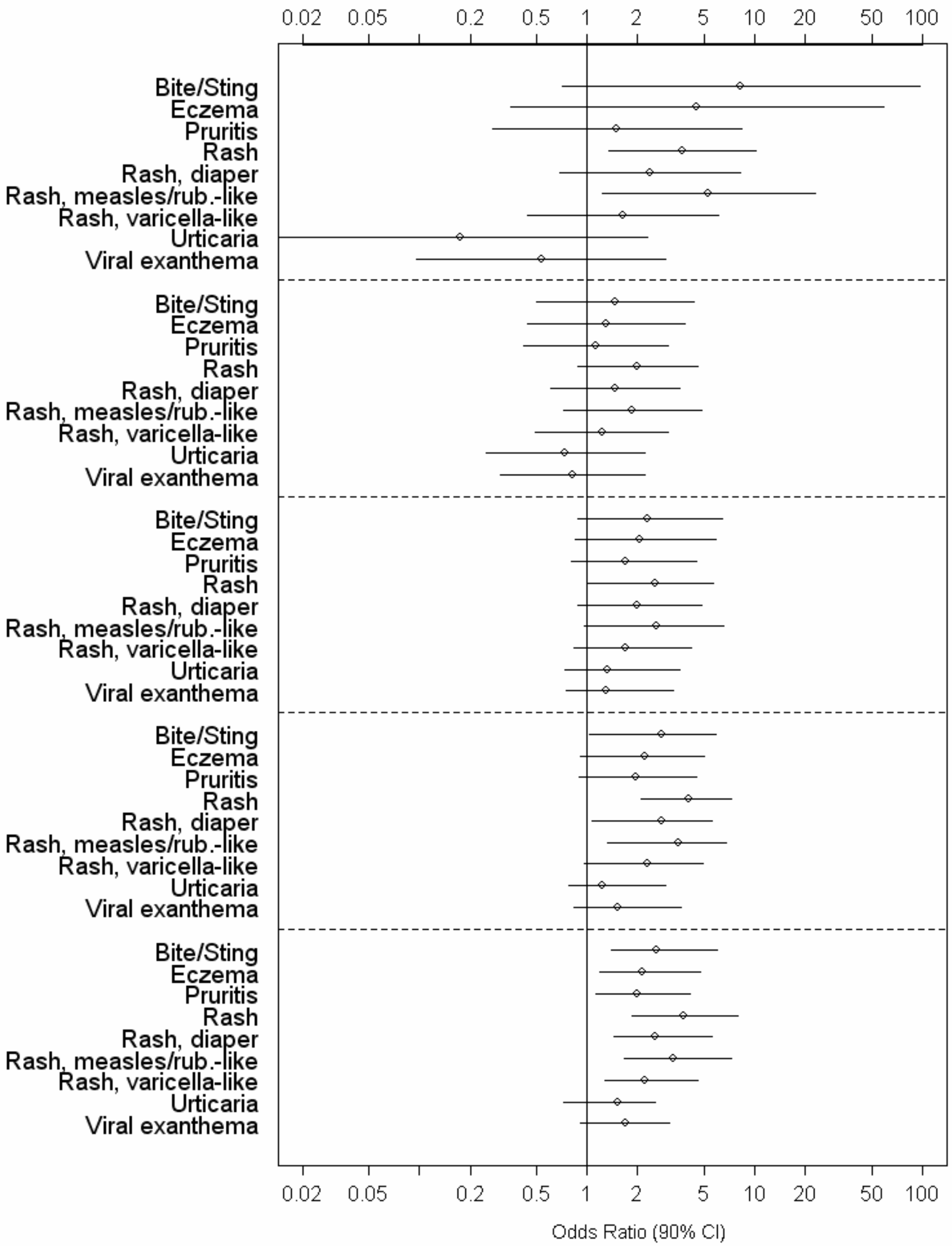
Hierarchical Approach

- A “Top Down” strategy
  - First fit a Bayesian shrinkage model where each table represents a higher level group
    - Obtain shrinkage estimates of odds ratios for each table
  - For each higher level group, fit a model to the set of subtables, setting μ_0 = shrinkage estimate of the group
  - Can even repeat at a third level, if data is available
    - E.g., Body system data, individual AEs, patient subgroups
- This may not be strictly proper, since the data are being reused at each level in a way that is hard to account for
  - But it allows for reuse of one simpler estimation procedure without requiring a very complex estimation algorithm

Example

	a	*b*	*c*	*d*
Bite/Sting	4	0	144	132
Eczema	2	0	146	132
Pruritis	2	1	146	131
Rash	13	3	135	129
Rash, diaper	6	2	142	130
Rash, measles/rub.-like	8	1	140	131
Rash, varicella-like	4	2	144	130
Urticaria	0	2	148	130
Viral exanthema	1	2	147	130

- Data from Berry and Berry (June 2004 Biometrics)
  - New vaccine formulation
  - a/(a+c) = proportion of AEs for treatment group
  - b/(b+d) = proportion of AEs for control group
- Fit several of the models we have discussed
  - Standard frequentist
  - Simplest Bayesian
  - Two component mixture
  - Modified y and v
  - Hierarchical determination of μ_0



Confidence Intervals from Different Models

Frequentist using y and v
(add .5 to every cell)
no multiplicity correction

Shrink toward log(OR) = 0
prior s.d. σ = 0.73

Mixture model: Shrink toward
36% N(0, $.58^2/10$)
64% N(0.90, $.58^2$)

Add (1.0, 4.0, 32.4, 361.3) to a,b,c,d
modify v's iteratively
27% N(0, $.48^2/10$) + 77% N(1.09, $.48^2$)

Add (1.0, 4.0, 32.4, 361.3) to a,b,c,d
modify v's iteratively
Hierarchical estimate of μ_0=0.6
43% N(0.6, $.65^2/10$) + 57% N(1.00, $.65^2$)

Conclusion

- Data mining of clinical trial safety data has many of the same challenges as analysis of spontaneous reports
  - Although the data will be cleaner, there will be less of it and the multiple comparisons issues are just as significant
- Bayesian models can be useful here too
- Many plausible Bayesian modeling strategies
  - Trade-off between plausibility of model assumptions and complexity and transparency of analysis
  - Simple normal approximations (perhaps with relatively simple refinements) may be good enough
  - More research and practice may eventually determine the best approach