Dose-Response Modeling of Gene Expression Data in pre-clinical Microarray Experiments

Setia Pramana

Workshop on Multiplicity and Microarray Analysis

Diepenbeek, 19 February 2010

universiteit hasselt
I-BioStat
Interuniversity Institute for Statistics and statistical Bioinformatics
KATHOLIEKE UNIVERSITEIT LEUVEN

universiteit hasselt
I-BioStat
Interuniversity Institute for Statistics and statistical Bioinformatics
KATHOLIEKE UNIVERSITEIT LEUVEN

Outline

- Introduction
  - Dose-response Studies
  - Dose-response in Microarray Experiments
- Modeling Dose-response study in Microarray setting
  - Dose-response Modeling
  - Testing for Trend
  - Model Based
  - Model Averaging
- Application
  - Antipsychotic Study
  - Results
- Discussion

Dose-response studies

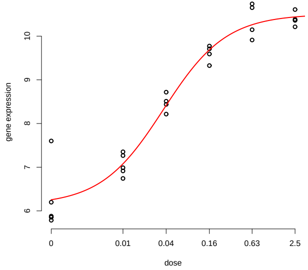
- The fundamental study in drug developments.
- Too high dose can result in an unacceptable toxicity profile.
- Too low dose decreases the chance of it showing effectiveness.
- Main aim: find dose or range of dose that is:
  - efficacious (for improving or curing the intended disease condition)
  - safe (with acceptable risk of adverse effects)

Dose-response studies

- Dose-response study investigates the dependence of the response on doses.
- Is there any dose-response relationship?
- What doses exhibit a response different from the control?
- What is the shape of the relationship?
- Estimates the target dose: minimum effective dose (MED), maximally tolerated dose (MTD) or half maximal effective concentration/dose (EC50), Ruberg,1995.

Dose-response in Microarray Experiments

- This research focuses on a dose-response study within a microarray setting.
- Monitoring of gene expression with respect to increasing dose of a compound.



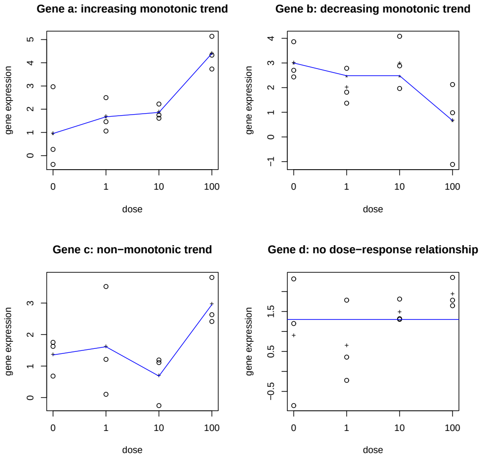
- No prior info about the dose-response shape.
- Genes have different shapes.
- Many ”noisy” genes hence need for initial filtering.

Steps

- Remove genes that do not show a monotone trend, using the monotonic trend test statistics, i.e., Likelihood Ratio Test (E^2_{01}).
- Fit several dose-response models in genes with a monotone trend.
- Estimate the target dose, e.g., EC_{50}.
- Apply model averaging.

Testing for Trend

- Dose-response relationship



Testing for Monotonic Trend

- For gene i ($i = 1, \cdots, m$) with K doses ($j = 0, \cdots, K$)

$$H_0: \quad \mu(d_0) = \mu(d_1) = \cdots = \mu(d_K)$$

$$H_1^{Up}: \quad \mu(d_0) \leq \mu(d_1) \leq \cdots \leq \mu(d_K)$$

or

$$H_1^{Down}: \quad \mu(d_0) \geq \mu(d_1) \geq \cdots \geq \mu(d_K)$$

with at least one inequality.

Test Statistics for Trend Test

Test statistic	Formula
Likelihood Ratio Test (LRT)	$\bar{E}^2_{01} = \frac{\sum_{ij}(y_{ij}-\hat{\mu})^2 - \sum_{ij}(y_{ij}-\hat{\mu}_i^\star)^2}{\sum_{ij}(y_{ij}-\hat{\mu})^2}$
Williams	$t = (\hat{\mu}_K^\star - \bar{y}_0)/\sqrt{2 \times \sum_{i=0}^{K}\sum_{j=1}^{n_i}(y_{ij}-\hat{\mu}_i)^2/(n_i(n-K))}$
Marcus	$t = (\hat{\mu}_K^\star - \hat{\mu}_0^\star)/\sqrt{2 \times \sum_{i=0}^{K}\sum_{j=1}^{n_i}(y_{ij}-\hat{\mu}_i)^2/(n_i(n-K))}$
M	$M = (\hat{\mu}_K^\star - \hat{\mu}_0^\star)/\sqrt{\sum_{i=0}^{K}\sum_{j=1}^{n_i}(y_{ij}-\hat{\mu}_i^\star)^2/(n-K)}$
Modified M (M’)	$M' = (\hat{\mu}_K^\star - \hat{\mu}_0^\star)\sqrt{\sum_{i=0}^{K}\sum_{j=1}^{n_i}(y_{ij}-\hat{\mu}_i^\star)^2/(n-I)}$

- More detail see Lin et.al 2007.
- In this case the LRT $\bar{E}^2_{01}$ is used.

Model Based

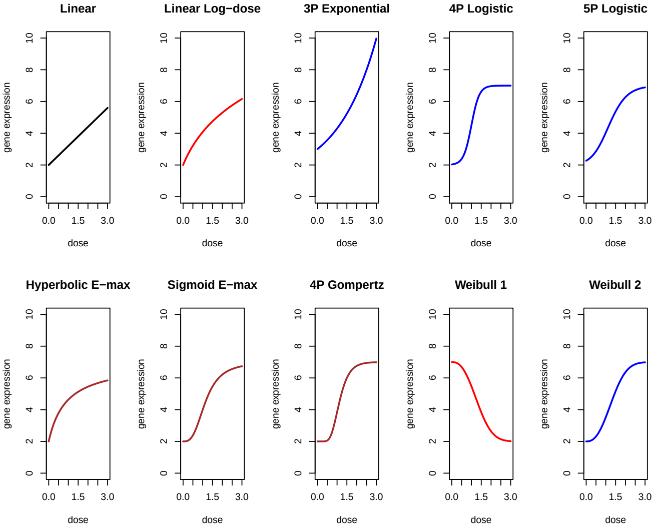
- Assumes a functional relationship between the response and the dose, taken as a quantitative factor, according to a pre-specified parametric model.
- Provides flexibility in investigating the effect of doses not used in the actual study
- Its result validity depends on the correct choice of the dose response model, which is a priori unknown.

Dose-response Models

Model Name	Function
Linear	$f(x) = E_0 + \delta x$
Linear Log-dose	$f(x) = E_0 + \beta log(x + c)$
Exponential	$f(x) = E_0 + E_1(e^{x/\vartheta} - 1)$
Four parameter logistic	$f(x) = E_0 + \frac{E_{max} - E_0}{1 + exp[(EC_{50} - x)/\phi]}$
Five parameter logistic	$f(x) = E_0 + \frac{E_{max} - E_0}{(1 + exp[(EC_{50} - x)/\phi])^{\gamma}}$
Hyperbolic E_{max}	$f(x) = E_0 + \frac{x \times (E_{Max} - E_0)}{x + EC_{50}}$
Sigmoidal E_{max}	$f(x) = E_0 + \frac{x^N \times (E_{Max} - E_0)}{x^N + EC_{50}^N}$
Gompertz	$f(x) = E_0 + (E_{Max} - E_0)e^{-exp(\varphi(EC_{50} - x))}$
Weibull 1	$f(x) = E_0 + (E_{max} - E_0)e^{-exp(b(log(x) - log(EC_{50})))}$
Weibull 2	$f(x) = E_0 + (E_{max} - E_0)(1 - e^{-exp(b(log(x) - log(EC_{50})))})$

Dose-response Profiles

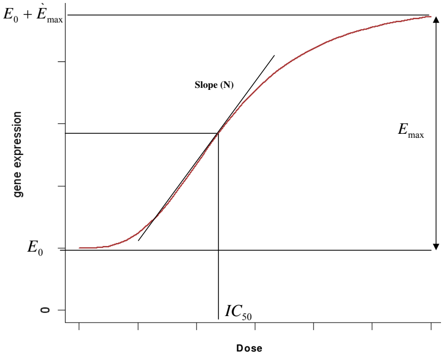
Dose-response profiles for each model



Target Dose: EC_{50}

The EC_{50}: dose/concentration which induces a response halfway between the baseline and maximum.

$$Y_{EC_{50}} = E_0 + \frac{E_0 + E_{max}}{2} \quad (1)$$



Model Averaging

- Combines results from different models.
- Account for model uncertainty.
- All fits are taken into consideration.
- Poor fits receive a small weights.
- Let θ be a quantity in which we are interested in and we can estimate θ from R models, the model averaged θ is defined as:

$$\hat{\theta} = \sum_{i}^{R} \omega_i \theta_i,$$

where θ_i is the value of θ from model i and ω_i the data-driven weights that sum to one assigned to model i.

Model Averaging

- Model averaging uses all (or most) of the candidate models whereas model selection selects the best model (the model with the highest value of ω_i).
- Akaike's weights:

$$\omega_i(AIC) = \frac{exp(-\frac{1}{2}\Delta AIC_i)}{\sum_{i=1}^{R} exp(-\frac{1}{2}\Delta AIC_i)} \tag{2}$$

where $\Delta AIC_i = AIC_i - AIC_{min}$

Model Averaged $EC50$

The model-averaged EC_{50} is defined as:

$$\widehat{\overline{EC}}_{50} = \sum_{i=1}^{R} \omega_i(AIC)\widehat{EC}_{50,i}, \quad (3)$$

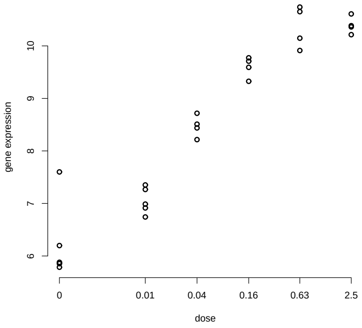
where ω_i: the Akaike's weight and $\widehat{EC}_{50,i}$: the EC_{50} of model i.
The estimator for variance of $\overline{EC}_{50}$ is defined as:

$$\widehat{var}(\overline{EC}_{50}) = \left[\sum_{i=1}^{R} \omega_i(AIC)\sqrt{\widehat{var}(\widehat{EC}_{50,i}|M_i) + (\widehat{EC}_{50,i} - \widehat{\overline{EC}}_{50})^2}\right]^2, \quad (4)$$

where $\widehat{var}(\widehat{EC}_{50,i}|M_i)$ is the variance of EC_{50} in model i.

Antipsychotic Study

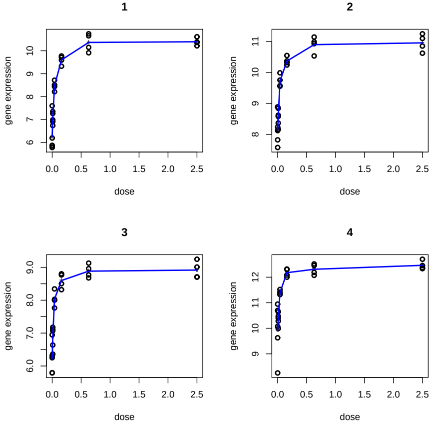
- Case study: a study focuses on an antipsychotic compound.
- 6 dose levels with 4-5 samples at each dose level.
- Each array consists of 11,565 genes.



Results

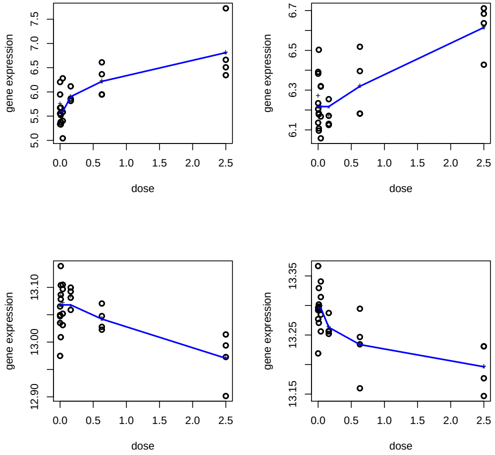
- 250 genes have a significant monotonic trend (FDR=0.05) using E^2 test statistic with 1000 permutations.

Data and isotonic trend of four best genes



Results

Data and isotonic trend of four other genes



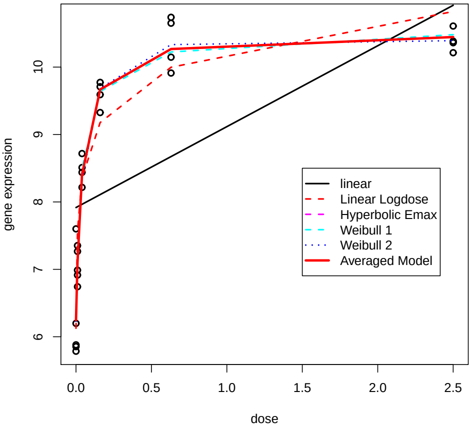
Results

- Number of models that converged and number of times as the best model:

Model	Number of models converge	Number selected as the best model
Linear	250	18
Linear log-dose	250	31
Three parm exponential	45	15
Four parm logistic	199	4
Five parm logistic	135	0
Sigmoidal E_{max}	25	1
Hyperbolic E_{max}	250	153
Four parm Gompertz	8	0
Weibull 1	213	22
Weibull 2	213	6

Results

Data and fitted value for the best gene



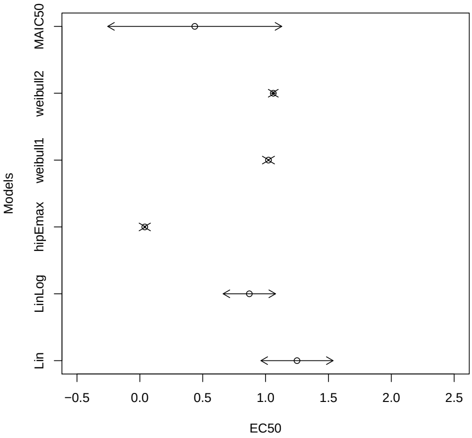
Results

AIC, EC_{50}, and Akaike's weight for the best gene

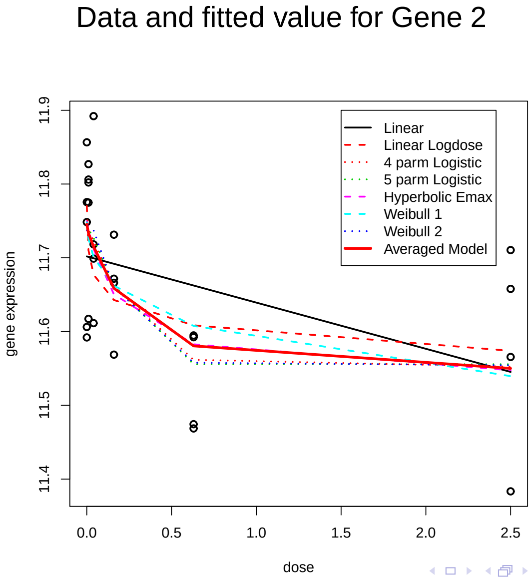
Model	AIC	EC_{50}	Weight
Linear	93.062	1.250	1.027e-14
Linear log-dose	93.062	0.871	1.212e-13
Hyperbolic E_{max}	29.656	0.039	0.603
Weibull 1	31.709	1.023	0.216
Weibull 2	32.064	1.061	0.181
Model Average	-	0.436	-

Results

Plot of Confidence Interval of EC_{50}



Results


Data and fitted value for Gene 2
gene expression
dose
11.4
11.5
11.6
11.7
11.8
11.9
0.0
0.5
1.0
1.5
2.0
2.5
Linear
Linear Logdose
4 parm Logistic
5 parm Logistic
Hyperbolic Emax
Weibull 1
Weibull 2
Averaged Model

Results

AIC, EC_{50}, and Akaike's weight for Gene 2

Model	AIC	EC_{50}	Weight
Linear	-35.03	1.25	0.021
Linear log-dose	-37.06	0.871	0.056
Four parameter logistic	-39.61	0.073	0.204
Five parameter logistic	-37.81	0.373	0.083
Hyperbolic E_{max}	-39.67	0.095	0.211
Weibull 1	-39.91	1.240	0.236
Weibull 2	-39.44	1.141	0.187
Model Average	-	0.706	-

Discussion

- In dose-response modeling in a microarray setting, fitting directly the proposed models to all genes (which can be tens thousands) can create problems, such as complexity and time consumption.
- We propose a three steps approach:
  - Select the genes with a monotone trend using the LRT (E^2) test statistic.
  - Fit the selected genes with the candidate models to estimate the target dose.
  - Average the target dose from the candidate models.
- Testing for trend
  - Assumes no specific dose-response relationship shape.
  - Filters genes with a non monotonic trend

Discussion

- Model-based approach:
  - Assumes a functional relationship.
  - Provides flexibility.
- Model averaging:
  - Accounts for uncertainty.
  - Takes in to account all the proposed models.
- Genes then can be ranked based on the Model Averaged EC_{50}
- Software: `IsoGene` and `IsoGeneGUI` packages for testing for trend.

Selected References

- Barlow, R.E., Bartholomew, D.J., Bremner, M.J. and Brunk, H.D. (1972) *Statistical Inference Under Order Restriction*, New York: Wiley.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Statist. Soc. B*, **57**, 289-300.
- Lin, Dan, Shkedy, Ziv, Yekutieli, Dani, Burzykowski, Tomasz, Göhlmann, Hinrich, De Bondt, An, Perera, Tim, Geerts, Tamara and Bijnens, Luc.(2007) Testing for Trends in Dose-Response Microarray Experiments: A Comparison of Several Testing Procedures, Multiplicity and Resampling-Based Inference, *Statistical Applications in Genetics and Molecular Biology: Vol. 6 : Iss. 1, Article 26*.
- Pinheiro, J.C. and Bates, D.M. (2000) *Mixed Effects Models in S and S-Plus.* Springer-Verlag, New York.

Thank You for Your Attention