Strategic roadmap for the IT support of the analysis and interpretation of data in drug discovery

Rudi Verbeeck, Johnson & Johnson Pharmaceutical Research & Development
Franky De Cooman, Business & Decision Benelux
Overview

• The process of developing new drugs
• Vision for data analysis for drug discovery
• Roadmap to realize the vision
  – Basic analyses
  – Expert level analyses
• ANOVA example
• Conclusions
Business & Decision: Group Profile

- European group, headquarters in Paris
- Founded in 1992
- Listed on Euronext (French “New Market”)
- Core business:
  - Business Intelligence
  - Customer Relationship Management and
  - Management Consulting
  - E-business
- + 650 employees
- Turnover of 52.7 Million Euro (2001/2002)
- Offices in France (7 regions), Belgium, Spain, Switzerland and the United Kingdom
Strong presence in Banking, Insurance Government and Manufacturing

Core business: Business Intelligence, Business Engineering, Risk, Management Solutions

Strong presence in Pharmaceutical, Government and Distribution

Core business: Business Intelligence, Supply Chain Optimization, Bio-Statistics

2003: Business&Decision Benelux NV/SA

Solutions: CRM, Life Sciences, Risk, Performance

Core competences: Business Intelligence and Business Engineering

Sectors: Banking & Insurance, Distribution, Government, Manufacturing ...

+ 60 consultants

+ 20 years experience in consulting services in Belgium

Turnover of + 5.5 Million Euro
Janssen Pharmaceutica

- Founded in 1953
- A Johnson & Johnson company since 1961
- More than 80 new compounds
- 5 products on WHO list of essential drugs (~200), more than any other company
- Johnson & Johnson is a family of > 200 companies with 110,600 employees
- Active in 3 sectors: consumer, professional & pharmaceutical
Janssen in the world
How are new drugs developed?

Drug discovery:
- new disease targets
- new molecular entities

Drug development:
- PK, tox
- process chemistry, formulation
- clinical studies

Registration

Post marketing surveillance

Drug on market
Vision for data analysis in drug discovery

• The researcher is the subject matter expert. They know their data best.
• Therefore bring the analysis to the researcher instead of the research data to the analyst.
A diverse set of needs

- Consumers of analysis methods
- Project teams
- Labs
- Specialised analysis (bix, cix…)
- IT department
- Biometrics
- Experts
- Literature
- Universities
- Competitors
- History

Support for routine analyses:
- Training
- Standard analysis environment

Deployment

Providers of analysis methods
Training path

New scientist discovery

Communicating numerical data

Smart research design

Applied statistics for non-statisticians

Biostatistics for non-statisticians

Excel
Spotfire

Xantipe
SAS/JMP

SAS for life scientists

Recommended
Optional
A diverse set of needs
Application assembly line

1. Request for analysis → Discovery
2. Discovery → Biometrics
3. Biometrics → External collaborations
4. External collaborations → Analysis method
5. Analysis method → IT - reporting
6. IT - reporting → Algorithm + reporting
7. Algorithm + reporting → IT - web deployment
8. IT - web deployment → Web based application
9. Web based application → IT - helpdesk support
10. IT - helpdesk support → Application support
11. Application support → Discovery

Keywords:
- Biometrics
- External collaborations
- IT - helpdesk support
- IT - web deployment
- IT - reporting
- Web based application
- Algorithm + reporting
- Application support
- Discovery
- Analysis method
- Request for analysis
Assembly line for statistical services

- Discovery site 1
- Discovery site 2
- Biometrics site 1
- Biometrics site 2
- External collaborations
- IT - reporting
- IT - helpdesk support
- IT - web deployment
- IT - application dvl
- Web based application
- Web service
- Application support
- Deployment in any application

- Request for analysis
- Ownership of algorithms @ biometrics
- Leverage of expertise, not applications
- Analysis method
- Algorithm + reporting
- Web service
Anova example: sketch

- $Y = f(\text{categorical variables}) + \text{Error}$
- Analysis by scientists was done using SPSS
- Did not check for assumptions
  - Normal distribution of errors
  - Homogeneity of variances
- Danger of drawing false conclusions
Anova example: Statistical Solution

• Apply Box-Cox transformation
• Use PROC MIXED for ANOVA but with GROUP statement
• The group statement allows for different variances within levels of categorical variable(s)
  e.g. GROUP = a
• Automatic model selection
Anova example: algorithm

Overview
Anova example: algorithm

Transform data

START
Rename original variables to A B C
Generate combinations A B C AB AC BC ABC
Calculate DF and maximum allowable depth of interactions
Determine Optimal Box-Cox transformation variable lambda & transform
Construction of GLM macro
Construction of lilmmodel_group macro
STOP
Anova example: algorithm

Forward group modeling

START
- loop = 1
- Group model empty

Try to add variable to group model using Log Likelihood Ratio

One significant variable found
- Yes
  - Add most significant variable to Group model

No
- Use current Group for group model

STOP

START
- MIXED on current model with SELECTED group variables

Try adding one-by-one all not-yet modeled variables

Calculate Log-Likelihood Ratio Test for all added effects

STOP
Anova example: algorithm

Backward effect modeling

START

Construct parameter set for starting model effects

Model MIXED with known Group variables and model effects

Look for least significant effect not being mandatory neither appearing in Group

Effect significant

Yes

STOP

No

Take least significant effect out of modeling parameter set not being mandatory neither appearing in Group

All possible combinations of effects take part in the parameter set if enough degrees of freedom, Group variables and mandatory variables should always be part of the parameter set.
Anova example: algorithm

Back transform data
Anova example: algorithm

Exploit final model

START → Calculate LSMEANS on mandatory variable → Calculate all pairwise comparisons on mandatory variable + significance → QQ plot on residuals of final model → STOP
Conclusions

• Experimental data in drug discovery should be analysed by the subject matter experts
• Support the scientists for routine analyses by training and standard software packages
• Develop expert solutions using an assembly line
• In a distributed organisation, use web services to leverage expertise
• Elaborated an example for robust ANOVA
Special Thanks To

- Luc Bijnens (J&J PRD)
- Luc Wouters (Barrier Therapeutics)
- Luc Bernolet (LACO NV)
- Limburgs Universitair Centrum