QUINTILES

R vs. SAS in model based drug development

Michael O'Kelly

Quintiles Centre for Statistics in Drug Development

UseR!
2009

Overview

QUINTILES

1. Introduction to model based drug development
2. An example
3. Current practice in the pharmaceutical industry
4. Planning how to model the development of a treatment
5. Coding and running the simulations: R vs. SAS
6. (Results of our simulation of a development program)
7. Summary

Model based drug development

QUINTILES

- "Model based drug development " a very fuzzy term. Can mean...
  - using a regression model to explore the dose-effect profile of a treatment (as opposed to comparing results for each dose vs. results from placebo group)
  - modelling the biological mechanism of the treatment– "pharmacometrics"
  - modelling a study or studies
    - simulating pseudo-subjects in a drug development program
      - » (early dose-finding phase IIa -> dose selection phase IIb -> pivotal final phase III)

Model based drug development

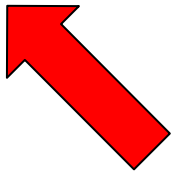
QUINTILES

- "Model based drug development " a very fuzzy term. Can mean...
  - **using a regression model to explore the dose-effect profile of a treatment** (as opposed to comparing results for each dose vs. results from placebo group)
  - **modelling the biological mechanism of the treatment**– "pharmacometrics"
  - **modelling a study or studies**
    - simulating pseudo-subjects in a drug development program
      - » (early dose-finding phase IIa -> dose selection phase IIb -> pivotal final phase III)
  - **a combination of any of the above three**

Model based drug development

QUINTILES

- "Model based drug development " a very fuzzy term. Can mean...
  - **using a regression model to explore the dose-effect profile of a treatment** (as opposed to comparing results for each dose vs. results from placebo group)
  - **modelling the biological mechanism of the treatment**– "pharmacometrics"
  - **modelling a study or studies**
    - simulating pseudo-subjects in a drug development program
      - » (early dose-finding phase IIa -> dose selection phase IIb -> pivotal final phase III)
  - **a combination of any of the above three**



This presentation

An example

QUINTILES

- Challenging program: use model based drug development
  - model/simulate the working of the three phases, automating a mechanism for choosing doses at end of each phase
    - phase IIa -> phase IIb: choose highest safe dose
    - phase IIb -> phase III: choose safe dose with highest model-predicted efficacy
  - little data available about efficacy and safety in humans needing to be treated, so **under a number of possible scenarios**, evaluate
    - probability of finding viable dose
    - probability of finding the best dose (most efficacious dose that is safe)
  - "**scenario**": profile of safety + profile for efficacy across candidate doses of treatment, including zero dose (placebo)

Current practice in the pharma industry?

QUINTILES

- Many large pharmaceutical companies talk of their model based drug development initiatives
  - in most cases this may describe work modelling the mechanism of action of the drug (simulating biology), not clear that it includes modelling a development program (simulating a chain of clinical trials)
- Business Week: “Pfizer, GlaxoSmithKline and others are using simulation...Novartis is far ahead of the rest of the industry” (again unclear what kind of simulation to which Business Week is referring)

/Current practice in the pharma industry?

QUINTILES

- Many large pharmaceutical companies talk of their model based drug development initiatives
  - in most cases this may describe work modelling the mechanism of action of the drug (simulating biology), not clear that it includes modelling a development program (simulating a chain of clinical trials)
- Business Week: “Pfizer, GlaxoSmithKline and others are using simulation...Novartis is far ahead of the rest of the industry” (again unclear what kind of simulation to which Business Week is referring)

(Emphasis on knowledge of how the drug works.)

/Current practice in the pharma industry?

QUINTILES

- Many large pharmaceutical companies talk of their model based drug development initiatives
  - in most cases this may describe work modelling the mechanism of action of the drug (simulating biology), not clear that it includes modelling a development program (simulating a chain of clinical trials)
- Business Week: “Pfizer, GlaxoSmithKline and others are using simulation...Novartis is far ahead of the rest of the industry” (again unclear what kind of simulation to which Business Week is referring)

(Emphasis on knowledge of how the drug works.)

R library MSToolkit simulates single studies and dose selection, such as we are describing here. May become open source.

/Current practice in the pharma industry?

QUINTILES

- Many large pharmaceutical companies talk of their model based drug development initiatives
  - in most cases this may describe work modelling the mechanism of action of the drug (simulating biology), not clear that it includes modelling a development program (simulating a chain of clinical trials)
- Business Week: “Pfizer, GlaxoSmithKline and others are using simulation...Novartis is far ahead of the rest of the industry” (again unclear what kind of simulation to which Business Week is referring)

(Emphasis on knowledge of how the drug works.)

R library MSToolkit simulates single studies and dose selection, such as we are describing here. May become open source.

Model based drug development: planning

QUINTILES

- Clear and detailed specification
  - In Quintiles: specification is followed independently by originator and by quality-control (QC) statistician
    - two simulation programs produced
    - the two programs produce same or similar outputs
    - Originator used R
    - QC used SAS
  - opportunity to compare R vs. SAS in this work
  - since SAS and R cannot produce simulation of identical random variables, QC of the generation of data is best separate from QC of the processing of those variables
    - 1. QC the generation of random variables
    - 2. share the random data
    - 3. QC the processing of that data in the simulated studies

/Model based drug development: planning

QUINTILES

- /Clear and detailed specification
  - facilitate using same programs for other treatments
    - self-contained functions in R for data generation, simulated analysis, dose selection
    - self-contained macros in SAS
  - rules for selection of doses from a dose-effect curve
    - many simulations -> some unexpected outcomes – rules must cater for these
    - e. g. unsafe dose between two safe doses – assess the higher dose as unsafe?
  - agree how to deal with simulations where development is discontinued mid-program
    - “empty” simulations passed to next stage of simulation?
    - “make up the numbers” by drawing at random from the simulations where the development was not stopped?

Coding

QUINTILES

R vs SAS

- **rv** package in R: nice compact generation of simulated data
  - its objects (=lists) not intended for use in modelling or other complex statistics
- Bulk of code: 972 lines (R) vs. 1323 lines
- Example

```
#simulate a placebo binary var and a group of active-group binary vars
ppe      <- 0.15
pevents<- rvbinom(n=1, size=controls, prob=ppe)
pae      <- max(.01, min(rvnorm(1, mean=activerate, sd=0.0001), 1))
aevents <- rvbinom(n=1, size=actives, prob=pae)
pnon     <- controls - pevents
anon     <- actives - aevents
```

R

```
do d=1 to maxdose;
                    cohort=d;
*** randomly purturb the fixed SAE rate for each simulation ***;
                    rate_sp=max(.01, rand('NORMAL', saerate_dose(d), 0.0001));

                    do pt=1 to ptN;
                                        subjid=d*1000+pt;
                                        dose=1;
                                        sae=rand('BERNULLI', rate_sp);
                                        if sae=0 then sae=2;
                                        output;
                    end;

                    rate_sp=max(.01, rand('NORMAL', saerate_dose(0), 0.0001));
                    do pt=1 to ptN/2;
                                        subjid=pt;
                                        dose=0;
                                        sae=rand('BERNULLI', rate_sp);
                                        if sae=0 then sae=2;
                                        output;
                    end;
end;
```

SAS

/Coding

QUINTILES

R vs SAS

- **rv** package in R: nice compact generation of simulated data
  - its objects (=lists) not intended for use in modelling or other complex statistics
- Bulk of code: 972 lines (R) vs. 1323 lines
- Example

```
# simulate fishers exact test for placebo
(AE/death rate 15% vs. the active dose (rate specified)
for specified numbers of subjects
fishersim<-function(activerate=0.15, controls, actives) {
binfish         <- sim.event(activerate=activerate, controls=controls, actives=actives)

binfish$pnon<- controls - binfish$pevents
binfish$anon<- actives - binfish$aevents

rvfish<-rvmapply(function(x, y, a, b)
  fisher.test(
       matrix(c(x, a, y, b), nrow=2), alternative="l"
                                                ),
          x=binfish$pevents,
          a=binfish$pnon,
          y=binfish$aevents,
          b=binfish$anon
                                  )
list(pvalue=rvfish$p.value)
}
```

R

```
proc freq data = IIa noprint ;
  tables dose*sae / nowarn fisher out=IIaSAEfreq(keep=scenario
sim cohort dose sae count where=(sae=1))  ;
  by scenario sim cohort;
  output out = IIaFish(keep=scenario sim cohort xpl_fish) exact ;
run;
```

SAS

/Coding

QUINTILES

R vs SAS

- Readability?
  - R more compact, but its density may render it less readable for some users

Coding and running

QUINTILES

R vs SAS

- Repeated processing of the same analysis required for simulation:
  - R, "for (i in 1:n)";
    - alternative: **rv** [=**r**andom **v**ariable] objects: compact simulation-based representation of random variables, useful for simple statistics, e.g.

```
» > setnsims(2500)
» > sim.binom <- rvbinom(n=1, size=200, prob=0.20)an(sim.binom)
» > sim.binom
»      mean          sd 1% 2.5% 25% 50% 75% 97.5% 99% sims
» [1]   40          5.6 28   29   36   40   44   51    53   2500
» > sim.binom**2
»      mean          sd  1%   2.5%  25%  50%  75%   97.5%  99% sims
» [1] 1625          458 729  841 1296  1600  1936   2655   2916 2500
```

  - SAS, BY statement – built-in economies in re-use of matrices

Coding and running

QUINTILES

R vs SAS, elapsed time for simulations of drug development program, same machine used

- caveat #1: different programming styles and skill levels for each program language
- caveat #2: no particular attempt to improve time-efficiency of R code
- caveat #3: elapsed timings dependent on workload of machine being used

Coding and running

QUINTILES

R vs SAS, elapsed time for simulations of drug development program, same machine used

- 2500 simulations, no modelling involved
  - R: 10.65 minutes
  - SAS: 5 minutes

Coding and running

QUINTILES

R vs SAS, elapsed time for simulations of drug development program, same machine used

- 2500 simulations, no modelling involved
  - R: 10.65 minutes
  - SAS: 5 minutes
- simulation involving 4 logistic regressions
  - 2500 simulations, 7 scenarios, (run at night)
    - SAS: 18 minutes
    - R: 7+ hours
  - logistic regression alone, (run during daytime)
    - R: 77.93 minutes for 2500 simulations of **single scenario**
    - SAS: 8.77 minutes for 2500 simulations of **7 scenarios**!

Summary

QUINTILES

- Thorough QC of simulations is recommended when modelling a "chain" of clinical trials
  - requires careful specification
  - QC the generation of random variable data, which may not be matched across languages like SAS and R, separately from the processing of that data
- Problem: how to simulate selection of dose where this is not based on quantitative rule, as in some dose-escalation studies?
- tricky points in programming
  - "empty" simulations (where there was a safety issue or there was no efficacy detected, so development stopped for that simulation)
  - specify the structure of the simulation as well as the algorithms and analyses
    - » facilitate QC
    - » facilitate re-use of code for other the development of other treatments

/Summary

QUINTILES

- R vs. SAS
  - R compact, sometimes less readable
    - R's rv package nice for generating random variables, but of limited use for e.g. repeated simulations of ANCOVA or logistic regression
  - SAS somewhat faster than R for simulations where no generalised linear modelling involved
  - SAS considerably faster than R where generalised linear modelling is involved in simulations (by a factor of c.30)

Extra slides

QUINTILES

An example

QUINTILES

- Challenging drug development program
  - mechanism of action of treatment not well understood
  - data only from animals and healthy subjects
  - broad options:
    - high risk, explore few doses
    - cautious approach, more doses
  - outline of program:
    - small phase IIa to identify highest safe dose (as often in oncology)
    - phase IIb to identify dose for further development
    - phase III to provide enough evidence of safety and efficacy for regulatory approval
  - model based drug development to help decide design of the phase IIb study

Current practice in the pharma industry?

QUINTILES

- US, EU regulators
  - Robert O'Neill, FDA: "Clinical scenario assessments are **much, much** more important than anything else that statisticians in drug development are doing at the moment, including adaptive design…"
  - Robert Hemmings, EMEA: "I had a dream; 10 years from now there will be no drug development programs without clinical scenario simulations"

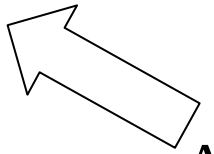
/Current practice in the pharma industry?

QUINTILES

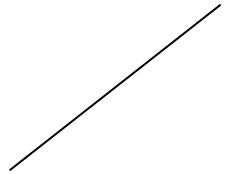
- Roche reported to work with Pharsight's "Trial simulator"

Simulates trials, "point and click", not based on a language

- (For most adaptive clinical trials, simulation or complex integration is required to estimate power of the trial.)



Addplan, Decimaker, East



User can add modules in R

Results of our simulation of a development program

QUINTILES

- Difficult to perform simulation of small IIa study to find highest safe dose
  - in real development the decision about safety would be based on the totality of the clinical data: labs, adverse events, baseline scores, etc.
  - our simulation used only assumed proportion of SAEs/deaths

/Results of our simulation of a development program

QUINTILES

Recall broad options for the sponsor:

- high risk, explore few doses

- cautious approach, more doses

/Results of our simulation of a development program

QUINTILES

Recall broad options for the sponsor:

- high risk, explore few doses

good success rate when no safety issues and all doses worked

- cautious approach, more doses

/Results of our simulation of a development program

QUINTILES

Recall broad options for the sponsor:

- high risk, explore few doses

good success rate when no safety issues and all doses worked

good at stopping early when there was no viable dose

- cautious approach, more doses

/Results of our simulation of a development program

QUINTILES

Recall broad options for the sponsor:

- high risk, explore few doses

good success rate when no safety issues and all doses worked

good at stopping early when there was no viable dose

- cautious approach, more doses

strength was in scenarios where some doses were unsafe or doses did not have efficacy