Medical and Pharmaceutical Statistics Research Unit

PSI Journal Club, 1 December 2009

Action following the discovery of a global association between the whole genome and adverse event risk in a clinical drug-development programme

John Whitehead, Patrick Kelly, Yinghui Zhou, Nigel Stallard, Helene Thygesen and Clive Bowman

Medical and Pharmaceutical Statistics Research Unit
Director: Professor John Whitehead
Tel: +44 1524 592350
Fax: +44 1524 592681
E-mail: j.whitehead@lancaster.ac.uk

MPS Research Unit
Department of Mathematics and Statistics
Fylde College
Lancaster University
Lancaster LA1 4YF, UK

Context

A major drug development programme, in which Drug A is to be compared with placebo

It is known that a particular form of ADR is a potential risk for patients receiving this drug

If ADRs are observed on Drug A with excessive frequency, what actions are possible?

(a) continue drug development without change
(b) abandon drug development
(c) continue drug development, excluding high risk patients

Which patients are at high risk?

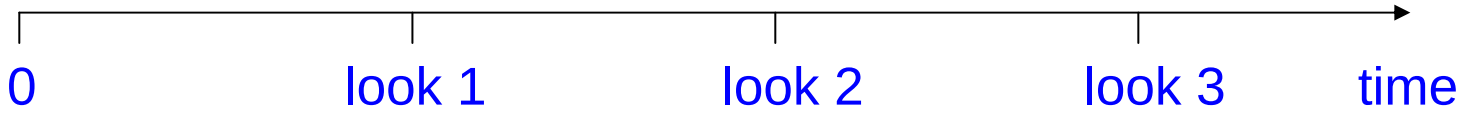
Those who are genetically pre-disposed to react badly to Drug A

If all patients with ADRs are genotyped, and some or all of the others, or genetic data are available from a comparable reference population, then we can seek a high risk subgroup

Should we exclude high risk patients?

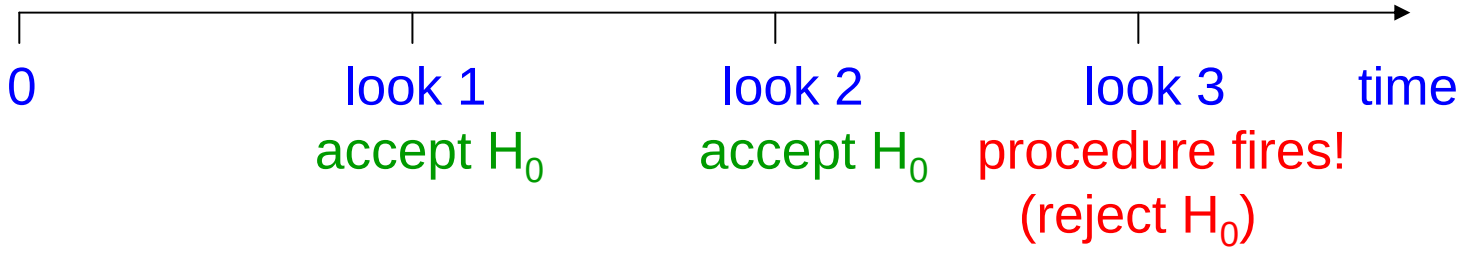
Yes: if the cost of suffering the ADR exceeds the cost of leaving the disease untreated

The search for a genetic effect on ADRs



- Test the hypothesis H_0: no SNP influences P(ADR)
- Test once or many times
- Allow for multiplicity of SNPs (permutation tests)
- Control the type I error rate (Kelly et al., 2006)

The search for a genetic effect on ADRs



- P(procedure ever fires $|$ H_0) = α
- When it fires, we know that there is a genetic effect on ADRs
- We don't know where the effect lies (at which SNPs)

Finding where the genetic effect lies

Method 1: The LASSO

Fit a logistic regression model:

$$\log\left(\frac{P(ADR)}{1-P(ADR)}\right)=\beta_0+\beta_1 x_1+\ldots+\beta_p x_p$$

where $x_{j(a)}$ and $x_{j(b)}$ represent the j^{th} of q SNPs (p = 2q)

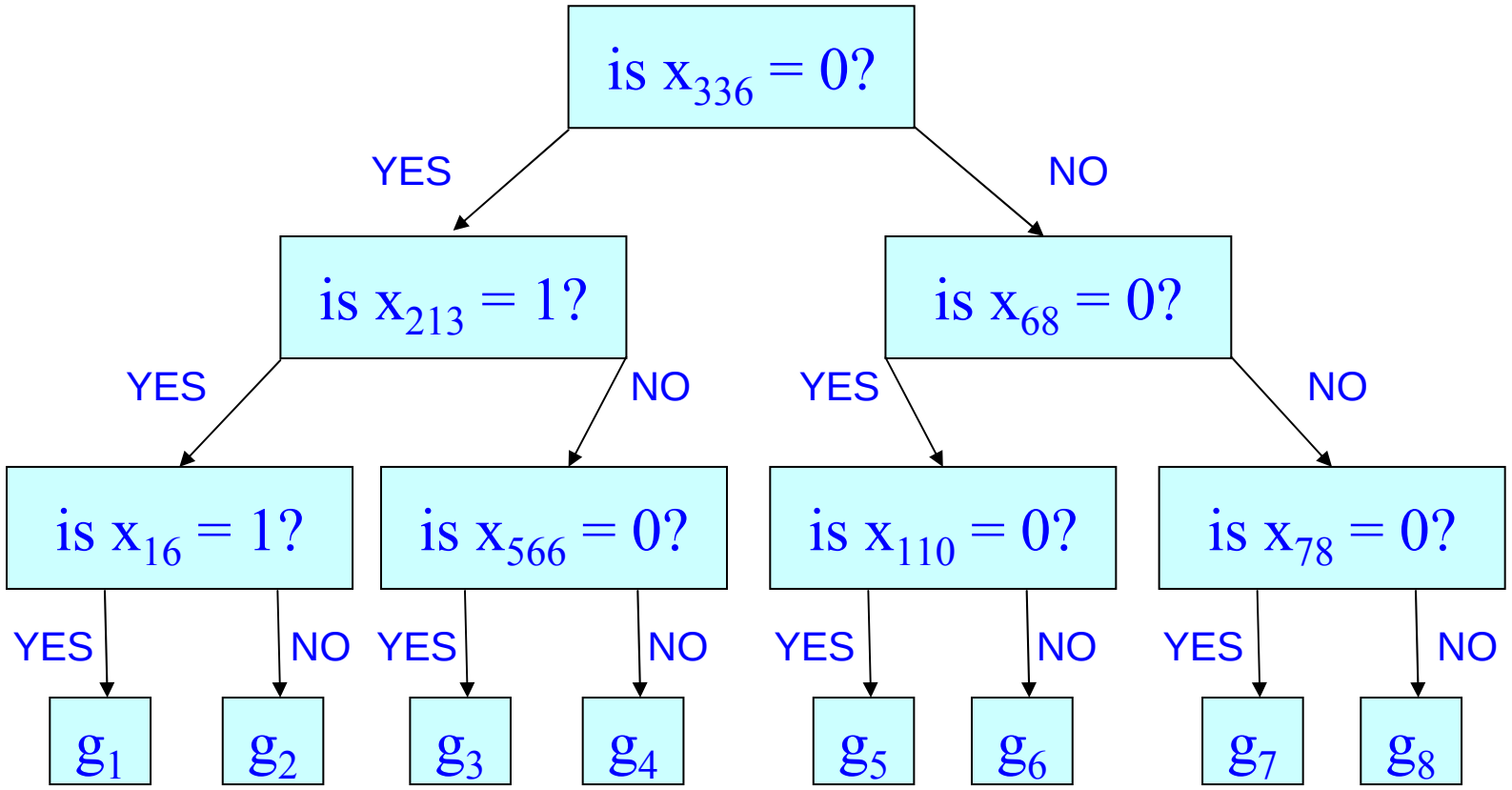
Find log-likelihood, ℓ, and maximise

$$\ell(\beta_1,\ldots,\beta_p)-\lambda\sum_{j=1}^{q}\left\{\left|\beta_{j(a)}\right|+\left|\beta_{j(b)}\right|\right\}$$

- The LASSO will usually set most β_is to zero
- The other β_is will be shrunken relative to standard estimates
- The magnitude of the *tuning parameter* λ governs the number of non-zero β_is
- The result is a manageable model of the genetic effect on ADRs

Finding where the genetic effect lies

Method 2: CART (Classification and Regression Trees)



- Each split maximises the difference between the two groups in terms of ADR incidence
  – *for example in terms of deviance*
- Keep the group with more ADR risk on the right
- Stop when each group is sufficiently homogeneous or sufficiently small
- The result is a series of genetic groups with differing ADR risks

Deciding who should be excluded

E: Receiving the experimental therapy (Drug A)

D: Failing to control the disease – *cost d*

A: Suffering an ADR – *cost a*

Assume that the cost of both is d + a

The *utility* of excluding patients in set S is

$$U = \sum_{g \in S} h(g)\left\{ dP\left(D \middle| g, E\right) + aP\left(A \middle| g, E\right) - dP\left(D \middle| g, \bar{E}\right) - aP\left(A \middle| g, \bar{E}\right)\right\}$$

This is the difference, per patient, between the expected cost of administering the drug and the expected cost of withholding it

Deciding who should be excluded

The exclusion set S will be chosen to maximise U

If D is independent of genetic factors, then a patient in genetic group g should be excluded if

$$P\left(A\middle|g,E\right)-P\left(A\middle|g,\overline{E}\right)>\left(d/a\right)\left\{P\left(D\middle|\overline{E}\right)-P\left(D\middle|E\right)\right\}$$

Deciding who should be excluded

$P\left(D\middle\vert\overline{E}\right)-P\left(D\middle\vert E\right)$	is found from external data or from design assumptions
$P\left(A\middle\vert g,\overline{E}\right)$	will often be zero
$P\left(A\middle\vert g,E\right)$	is found from LASSO, CART or similar
a/d	represents the cost of the ADR relative to failure to control the disease

h(g) = P(patient $\in$ g) can be estimated from the data to allow calculation of P(exclusion) and P(ADR $\vert$ included)

Examples

DATA:

from two GSK trials of abacavir in HIV infection

ADR:

hypersensitivity

SCENARIOS:

artificially constructed, using real genotypes and ADRs

External information

$$P\left(D\middle|\overline{E}\right) - P\left(D\middle|E\right) = 0.5$$

– *US package inserts for abacavir cite response rates of 49% - 69%*

$$P\left(A\middle|g,\overline{E}\right) = 0$$

– *no hypersensitivity in the placebo group*

$$a/d = 5$$

– *hypersensitivity is 5 times worse than failure to control the HIV*

So patients should be excluded if

$$P\left(A\middle|g,E\right) > 0.1$$

Scenario 1

The following data cause concern:

	ADR	No ADR
Drug A	52	593
Placebo	0	650

The risk of an ADR on Drug A is 8.8%

All patients have been genotyped

A significant association between ADR and genotype is found

The LASSO

Applied to the 52 cases and 593 fixed controls

Choose the tuning parameter $\lambda = 9$, and then we find

$$\log\left(\frac{P(ADR)}{1-P(ADR)}\right) =$$

$$-2.691 + 2.166x_{336a} + 0.018x_{410a} + 0.014x_{825b} - 0.079x_{847a}$$

leading to 16 genetic groups

Excluded genetic groups from the LASSO method

genetic group, g_i	x_{336a}	x_{410a}	x_{825b}	x_{847a}	$h(g_i)$	$P(A\|g_i, E)$	cases excluded[1]	controls excluded[2]
g_1	1	0	0	0	0.012	0.372	4 (28)	3
g_2	1	0	0	1	0.010	0.353	1 (19)	5
g_3	1	0	1	0	0.008	0.375	3 (32)	2
g_4	1	0	1	1	0.002	0.357	0 (27)	1
g_5	1	1	0	0	0.007	0.376	4 (14)	0
g_6	1	1	0	1	0.003	0.358	0 (19)	2
g_7	1	1	1	0	0.012	0.379	7 (20)	0
g_8	1	1	1	1	0.007	0.361	3 (15)	1

[1]Cases excluded amongst the 52 used in the example
(and amongst the remaining 478 in the dataset)

[2]Controls excluded amongst the 593 used in the example

Action:

Exclude patients in genetic groups g_1, …, g_8

*– that is if x_{336a} = 1 (the genotype at locus 336 is **not** aa)*

- This excludes patients if $P(A|g_i, E) \geq 0.10$
- From cross-validation:
  - 24 cases from 52 excluded (46% sensitivity)
  - 19 controls from 593 excluded (97% specificity)
- Of the 478 cases not used in the model, 174 (36%) excluded
- Leads to the inclusion of 93.9% of patients, with an ADR risk of 6.2% (rather than 8.8%)
- The estimated utility of the policy is U = 0.0164a per patient

Scenario 2

The monitoring method of Kelly et al. (2006) is used

This compares the genotypes of emerging cases of ADR with those of a reference population of fixed controls

A sequential plan with up to 20 looks at the data was devised: type I error was controlled at 0.05

After 18 ADRs on drug A, the procedure stopped

Excluded genetic groups from CART

genetic group	specification of g_i	$h(g_i)$	$P(A\|g_i, \Pi)$	$P(A\|g_i, E)$	cases excluded[1]	controls excluded[2]
g_4	$x_{61}=1$; $x_{556}=2$; $x_{751}=1$; $x_{821}=2$	0.009	0.60	0.913	3 (5)	2
g_7	$x_{336}=0$; $x_{437}=1$; $x_{586}=0,1$; $x_{809}=1$	0.065	0.20	0.639	1 (2)	4
g_8	$x_{336}=0$; $x_{437}=1$; $x_{568}=2$	0.009	0.80	0.965	4 (3)	1
g_{10}	$x_{64}=0,2$; $x_{169}=1,2$; $x_{336}=1,2$	0.012	0.29	0.739	2 (40)	5
g_{11}	$x_{64}=0,2$; $x_{169}=0$; $x_{336}=1,2$	0.012	0.99	1.000	7 (45)	0

[1]Cases excluded amongst the 18 used in the example (and amongst the remaining 512 in the dataset)

[2]Controls excluded amongst the 593 used in the example

The table features

$P(A|g,\Pi)$ – *probability of an ADR in group g based on the 18 cases and the 593 fixed controls*

and

$P(A|g,E)$ – *probability of an ADR in group g for a patient on Drug A*

To get from the first to the second we use $P(A|E) = 0.08$

from the US prescribing information for abacavir,
and an argument used in matched case-control studies

Action:

Exclude patients in genetic groups g_4, g_7, g_8, g_{10}, g_{11}

- This excludes patients if $P(A|g_i, E) \geq 0.10$
- From cross-validation:
  - 4 cases from 18 excluded (22% sensitivity)
  - 44 controls from 593 excluded (93% specificity)
- Of the 512 cases not used in the model, 174 (36%) excluded
- Leads to the inclusion of 90.1% of patients, with an ADR risk of 2.1% (rather than 8.8%)
- The estimated utility of the policy is U = 0.0893a per patient

Conclusions

- Associations between ADR risk and genotype can be found with controlled risk of false discovery
- A reaction is then needed in terms of conduct of studies
- Exclusion of patients must be based on estimated risks and the relative costs of ADRs and failure to control the disease
- The utility approach can be combined with various methods for estimating ADR risks

References

Kelly P, Stallard N, Zhou Y, Whitehead J, Bowman C. Sequential genomewide association studies for monitoring adverse events in clinical evaluation of new drugs. *Statistics in Medicine* 2006; 25: 3081-3092.

Kelly P, Zhou Y, Whitehead J, Stallard, N, Bowman, C. Sequentially testing for a gene-drug interaction in a genomewide analysis. *Statistics in Medicine* 2007 (to appear).

Roses AD. Pharmacogenetics and Drug Development: The path to safer and more effective drugs. *Nature Reviews Genetics* 2004; **5**: 643-655.

Roses AD. Genome-wide screening for drug discovery and companion diagnostics. *Expert Opinion in Drug Discovery* 2007; **2**: 489-501.

Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of Royal Statistical Society Series B* 1996; **58:** 267-288.

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botsterin D and Altman R. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001; **17:** 520-525.