Trial designs fully integrating biomarker information for the evaluation of treatment-effect mechanisms in stratified medicine

Dr Richard Emsley

Centre for Biostatistics, Institute of Population Health,
The University of Manchester

www.medicine.manchester.ac.uk/staff/RichardEmsley
richard.emsley@manchester.ac.uk

Joint work with **Prof Graham Dunn**, **Dr Hanhua Liu** and **Dr Sabine Landau**

Presentation to Centre for Statistical Methodology, LSHTM
Friday 5th October 2012

Example: efficacy and mechanisms evaluation and stratified medicine

- A β-blocker may be effective at reducing risk of stroke in hypertensive patients, but *its effect might be greater in some patients than in others*.

- Similarly, it is likely to reduce average systolic blood pressure and, again, its effect on average blood pressure is *likely to vary from one patient to another*.

- We might expect that if one individual's blood pressure has been lowered considerably more than that of another individual then the risk of stroke is likely to have been reduced more in the first person than in the second.

  ➢ **What proportion of the β-blocker's effect on stroke is *explained* by its effect on average blood pressure?**
  ➢ **Who are β-blocker's effective for?**

Example of stratified medicine: biomarkers and depression

Open

Candidate Genes Expression Profile Associated with Antidepressants Response in the GENDEP Study: Differentiating between Baseline 'Predictors' and Longitudinal 'Targets'

Annamaria Cattaneo[1], Massimo Gennarelli[1,2], Rudolf Uher[3], Gerome Breen[3], Anne Farmer[3], Katherine J Aitchison[3,4], Ian W Craig[3], Christoph Anacker[5], Patricia A Zunsztain[5], Peter McGuffin[3] and Carmine M Pariante*[5]

[1]Department of Biomedical Sciences and Biotechnology, Genetic and Biology Section, University of Brescia, Brescia, Italy; [2]Genetic Unit, IRCCS San Giovanni di Dio, Fatebenefratelli Centre, Brescia, Italy; [3]Institute of Psychiatry, MRC Social, Genetic and Developmental Psychiatry, King's College London, London, UK; [4]Department of Psychiatry, University of Alberta, Edmonton, Canada; [5]Department of Psychological Medicine, Institute of Psychiatry, Section of Perinatal Psychiatry and Stress, Psychiatry and Immunology (SPI-lab), King's College London, London, UK

Example of stratified medicine: biomarkers and depression

- "To improve the 'personalized-medicine' approach to the treatment of depression, we need to identify biomarkers that, assessed before starting treatment, predict future response to antidepressants ('predictors'), as well as biomarkers that are targeted by antidepressants and change longitudinally during the treatment('targets')."

- Focus on markers in biologic systems described as abnormal in depression:
  - The glucocorticoid receptor complex (link to HPA, cortisol)
  - Inflammation: interleukin and TNF-α
  - Neuroplasticity.

- GENDEP: trial with two active pharmacological treatment arms.

- Analysis: is a good outcome predicted by biomarkers?

Contents

Treatment effects on outcome

- Consider a randomised controlled trial with two arms: treatment (T) versus control (C) and a continuous outcome Y

- Prior to randomisation to one of two competing treatment arms we can envisage two potential outcomes for each participant in the trial:
  - the outcome after an active treatment, Y_T
  - the outcome after receiving the placebo, Y_C

- For a given individual, the effect of treatment is the difference:
$$ITE(Y) = Y_T - Y_C$$

- The average treatment effect ATE is:
$$Average[ITE(Y)] = Average[Y_T - Y_C]$$

Treatment effects on mediator

- Similarly for a continuous mediator M, we can define:
  - ➤ the mediator after an active treatment, M_T
  - ➤ the outcome after receiving the placebo, M_C

- For a given individual, the effect of treatment is the difference:
$$ITE(M) = M_T - M_C$$

- The average treatment effect ATE is:
$$Average[ITE(M)] = Average[M_T - M_C]$$

Personalised medicine and treatment effect heterogeneity

- Treatment effect heterogeneity, whereby a given treatment will be more efficacious for some patients than for others, is the underlying foundation of personalised medicine.
  - Stratified/predictive/targeted medicine
  - Genomic medicine
  - Pharmacogenomics

- If a treatment is effective, we are interested in knowing who is it (most) effective for, in advance of treatment allocation/decisions to treat?

- We need access to pre-treatment characteristics that predict treatment-effect heterogeneity
  - **Not just predict outcome**

Critique of depression/biomarker paper

- Given an additive treatment effect, the outcome of treatment is:

$$Y_T = Y_C + \text{ITE}(Y)$$

  (i.e. their treatment-free outcome plus the effect of treatment)

- Now let's introduce a baseline marker, X.

- If we correlate X with treatment outcome Y_T then this can arise from two sources:
  - Y_C is correlated with M (prognosis), or
  - ITE(Y) is correlated with M (prediction)

- If X is prognostic then you can get a correlation between Y_T and M even when the ITE(Y) is **ZERO** for everyone in the study.

Contents

Biomarkers definition

- **PROGNOSTIC BIOMARKERS (RISK FACTORS)**
  - ➤ Predict outcomes (both intermediate or final), independently of treatment receipt.

- **PREDICTIVE BIOMARKERS**
  - ➤ Although they may have direct predictive effects on both intermediate and final outcomes, their essential characteristic is that they **moderate** (influence) treatment effects.

- **SURROGATE BIOMARKERS (not covered further)**
  - ➤ These replace clinical endpoints in clinical trials to evaluate the effects of treatment.

Prognostic biomarkers (risk factors)

- Predict outcomes (both intermediate or final), independently of treatment receipt (note that this is a statement concerning **association** only).

- No statistical interaction in the model for outcome.
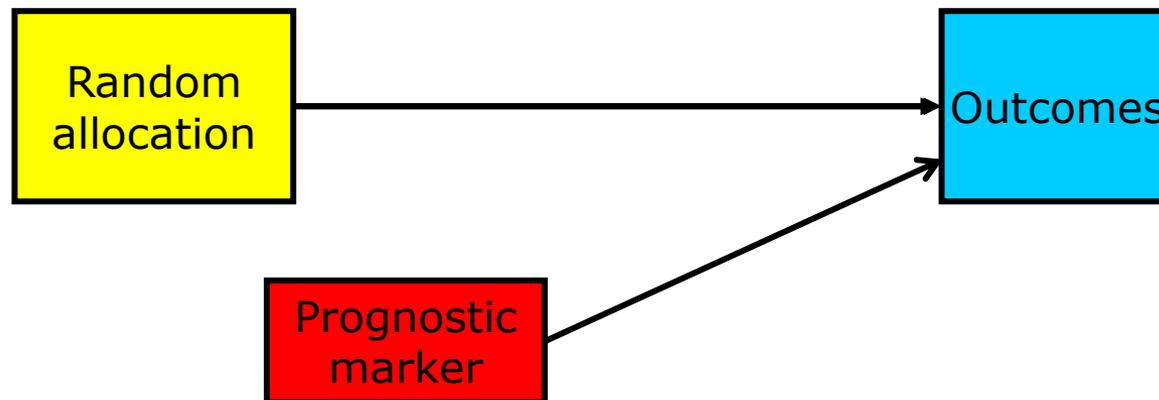
Prognostic biomarkers
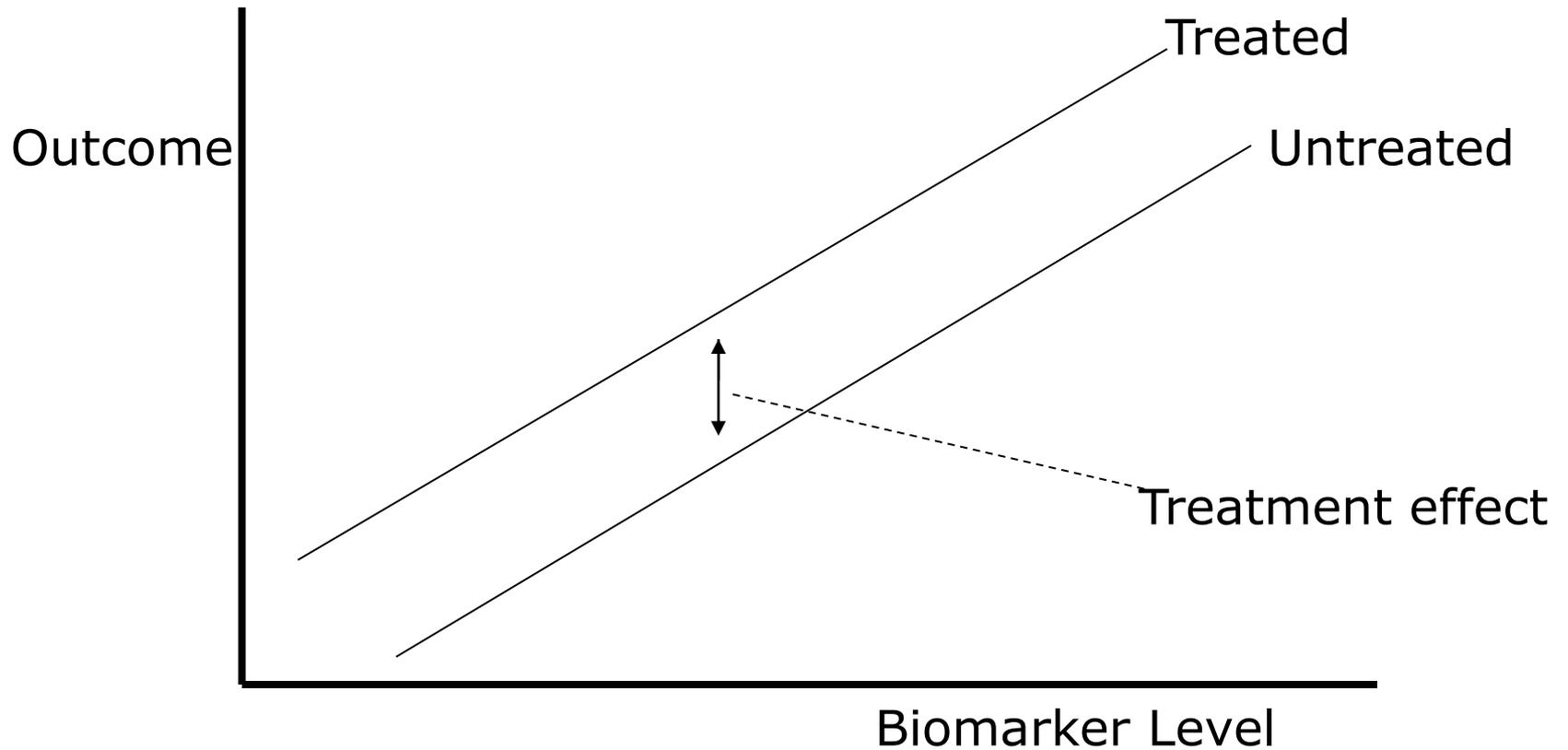
Predictive biomarkers – the basis for personalised medicine

- Although they may have direct predictive effects on both intermediate and final outcomes, their essential characteristic is that they **moderate** (influence) treatment effects.
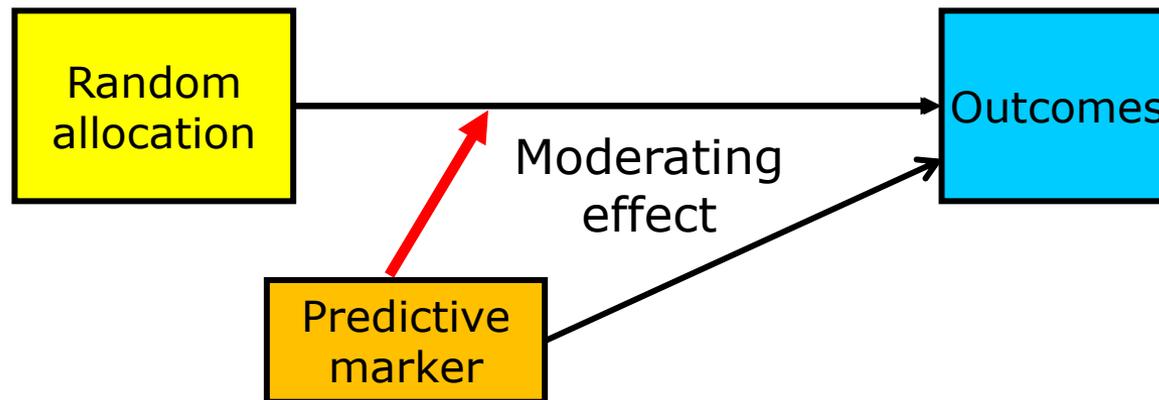
Predictive biomarkers
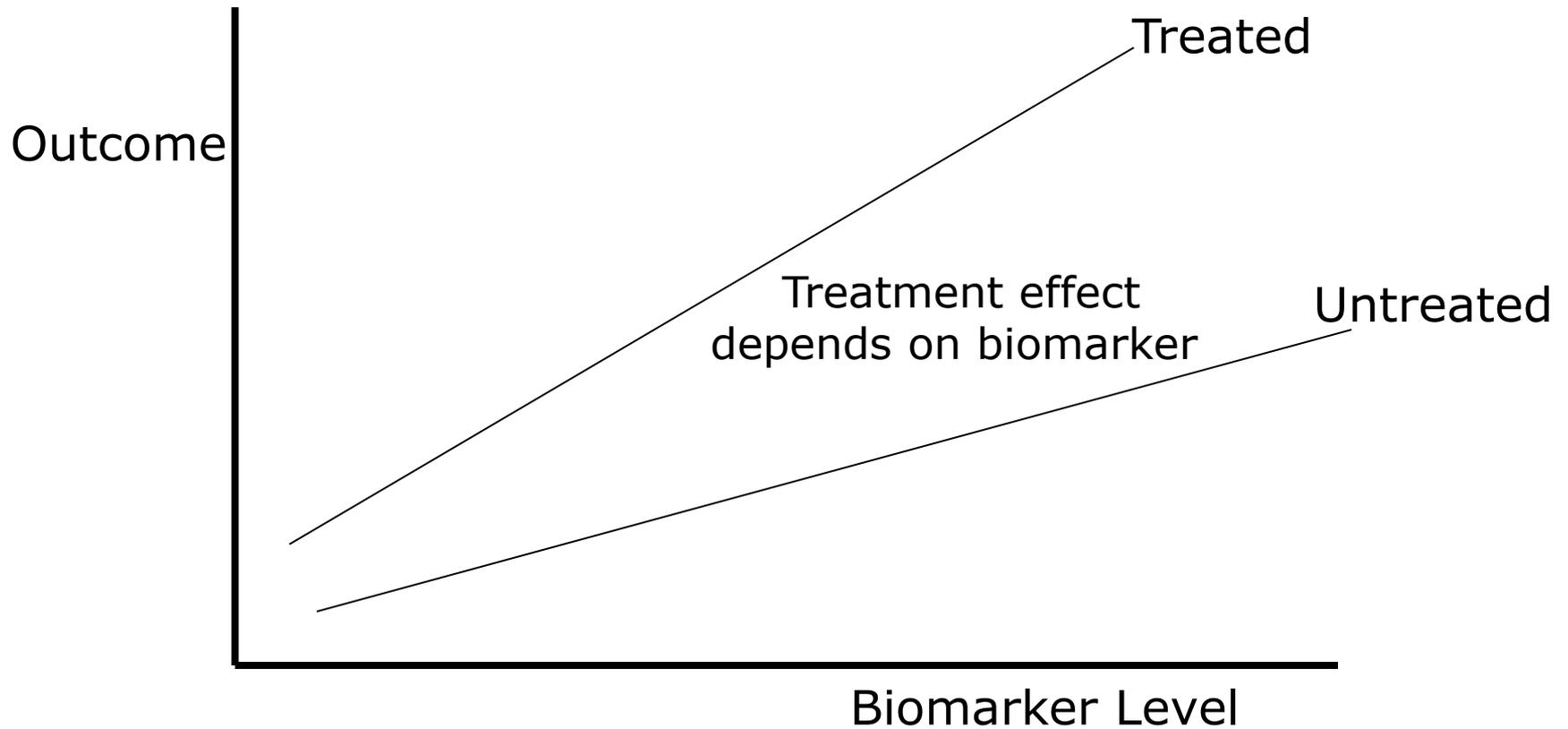
What happens when we have a predictive biomarker?

- To date, there has been considerable investment in translational development and the use of complex data mining/bioinformatics methods for predictive biomarker validation, but very little methodological research on qualification of these biomarkers (i.e. evaluation of their clinical utility).

- "For a predictive biomarker that has met the necessary development milestones, it is necessary to evaluate its clinical utility through a confirmatory trial of its predictive value. Few markers have reached this status, making the design of such a trial, known as a validation trial, a relatively unexplored area."

  Simon R. (2010). Clinical trials for predictive medicine: new challenges and paradigms. *Clinical Trials*;7:516-524

Biomarker stratified design

1. Stratify patients according to marker status, and randomize to treatments (T and C) within each marker stratum.

2. Two parallel randomized clinical trials are conducted to compare the treatments within each marker stratum.

3. We assess the predictive value of the marker by formally testing whether the treatment effect is the same in each of the marker strata; that is, we assess the marker–treatment interaction

- These tells us whether the intention-to-treat effect differs in each strata
  - ➢ Treatment effect moderation;
  - ➢ Subgroup analysis.

- Says nothing about WHY there might be differences…

MRC Call for Stratified Medicine

"Patient response to drug treatments and therapeutic interventions varies markedly across the population as a result of differing underlying **mechanisms** of disease and patient responses to both disease and treatment.

Stratified medicine can be described as identifying the different strata within a disease and the deeper understanding of the **mechanisms** underpinning these strata.

Stratification will allow targeting of treatments to specific **disease pathways**, identification of treatments effective for particular groups of patients, and co-development of diagnostics to ensure the right patient gets the right treatment at the right time."

Contents

Efficacy and mechanisms evaluation

- The aim here is to go beyond evaluating whether an intervention is effective and to explain why it might be effective:
  - ➤ What are the putative mechanisms through which the treatment acts?
  - ➤ Do these mechanisms explain treatment effect heterogeneity?

- Usual analysis methods dominated by decomposing total effects into direct and indirect effects:
  - ➤ Mental health, psychology has been concerned with this idea for decades.
  - ➤ Widely cited Baron and Kenny paper for mediation analysis in social sciences.
  - ➤ Makes implicit assumptions which are unlikely to hold.

Simple mediation/mechanism diagram
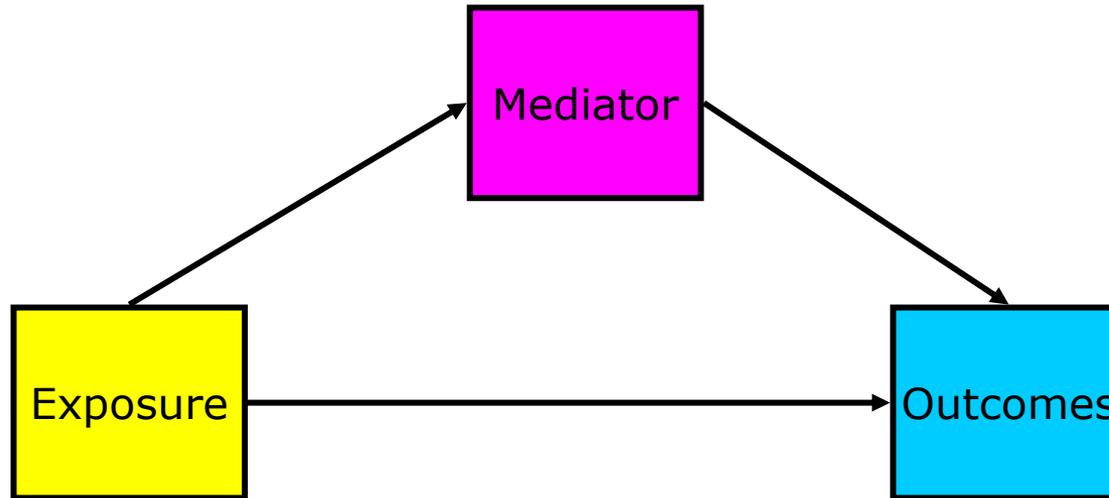
Total effect = direct effect + indirect effect

The basic underlying problem: estimating valid causal effects
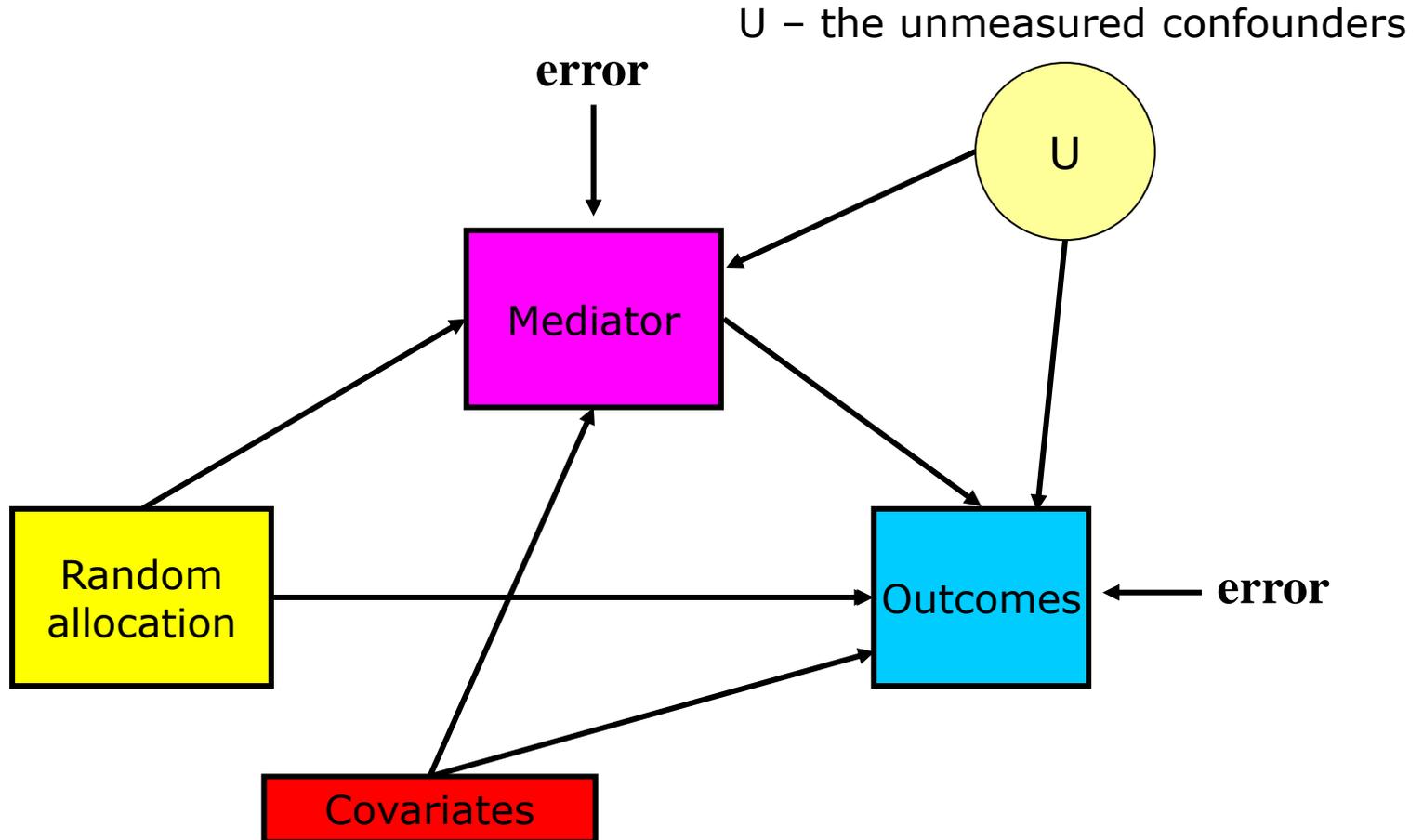
Solutions to unmeasured confounding in mechanisms evaluation

- We've proposed three solutions to analyse mediation allowing for unmeasured confounding:

1. Measure and adjust for potential confounders (sounds obvious, not always done);

2. **Instrumental variables;**

3. Principal stratification.

Explained in detail in:

Emsley RA, Dunn G & White IR (2010). Modelling mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research*, 19(3), pp.237-270.

Instrumental variables

- An instrumental variable is:
  1. (Strongly) predictive of the mediating variable;
  2. Has no direct effect on the outcome, except through the intermediate;
  3. Does not share common causes with the outcome.

- Randomisation, where available, often satisfies this criteria when accounting for departures from randomised treatment.

- If we consider this at the design stage of the complex intervention trial, we can measure variables that <u>MIGHT</u> meet these requirements.

Mediation diagram with instrumental variables
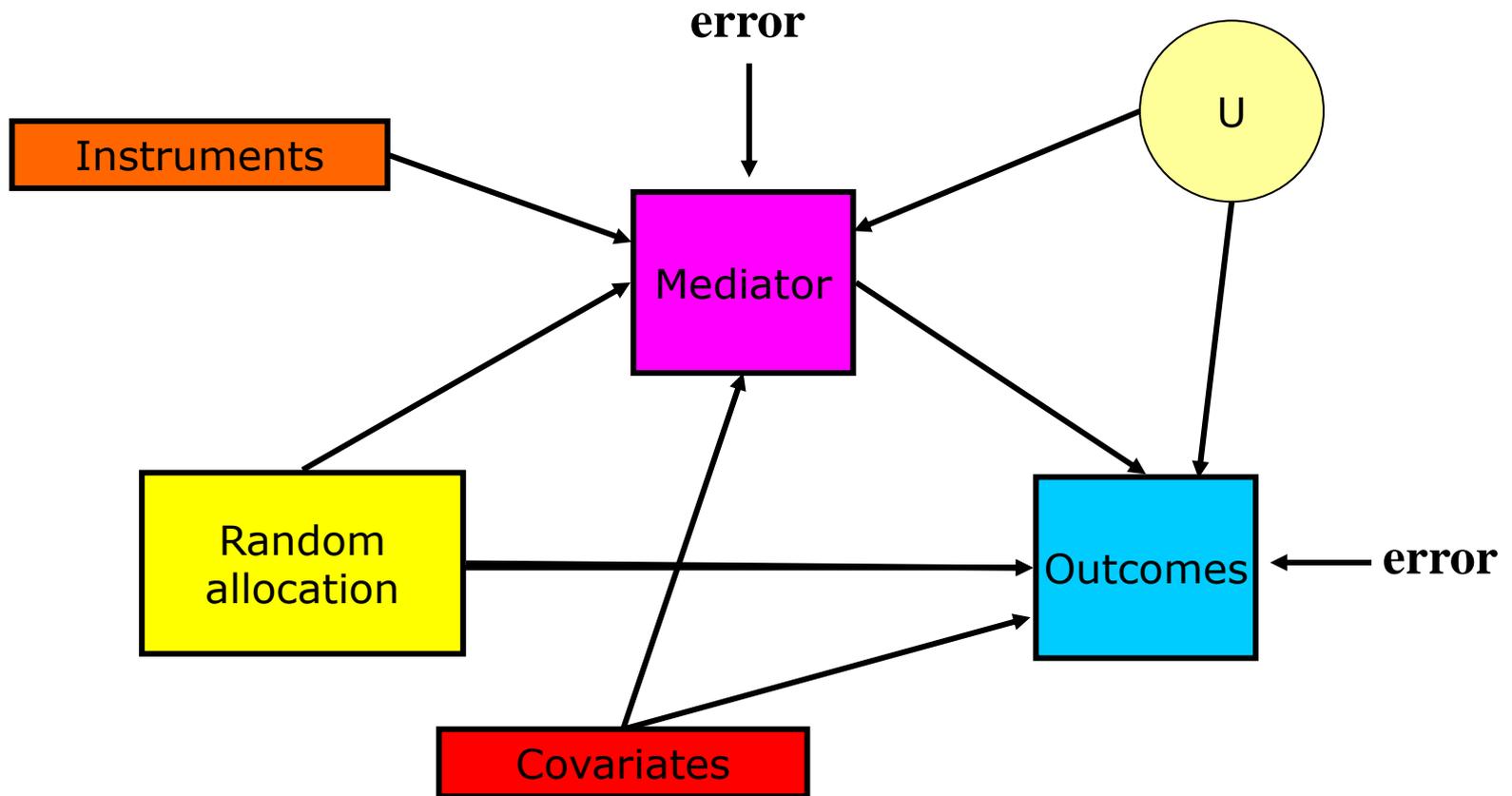
We need more information to make progress

- Genetic and phenotypic markers
- Clinical history
- Past environmental exposures, lifestyle, etc.

- Advantage of genetic markers is that they are essentially randomised and, in particular, (in a conventional RCT) independent of treatment allocation. And, of course, they are not influenced by treatment.

- Can we use **markers** (prognostic and predictive markers, e.g. biological or biomarkers, social and psychological markers) as this extra information?
  - ➤ How we do this depends on the assumptions we make about relationships between markers and outcomes.

Prognostic biomarker as a confounder

- If the prognostic markers have a **causal influence** on both intermediate and final outcomes then they are **confounders** of the effect of the intermediate on final outcome, and of the direct effect of treatment receipt on final outcome.
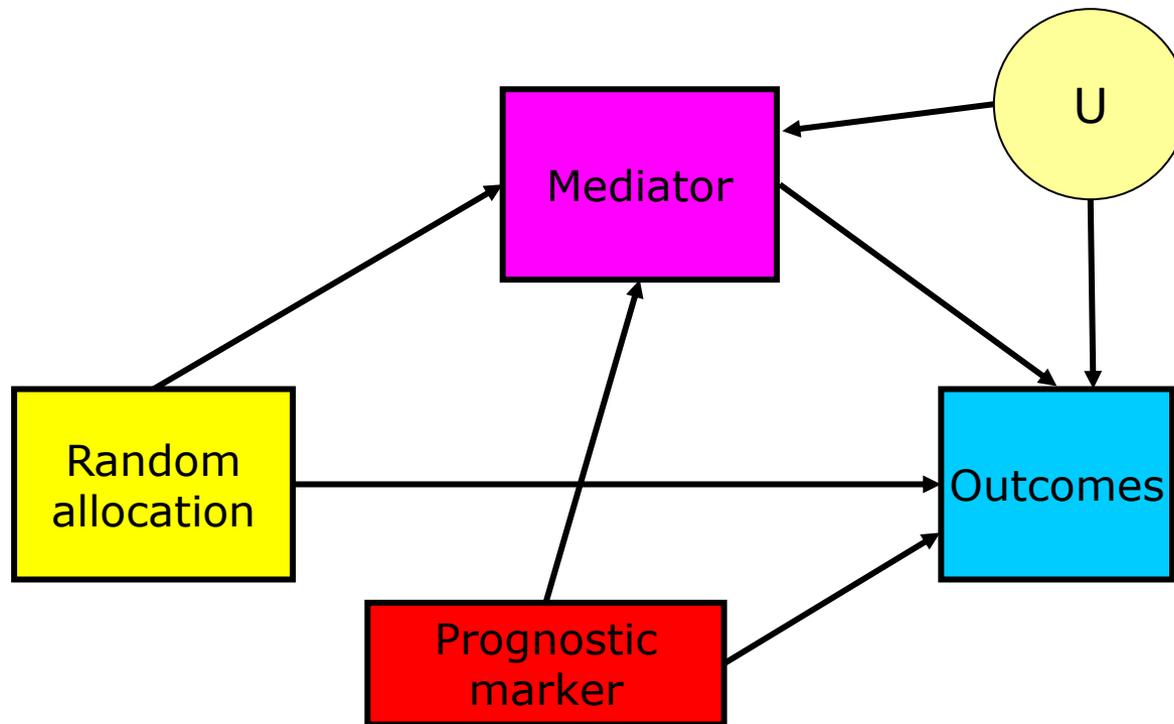
Instrumental variables or instruments

- If the **causal influence** of the prognostic marker on the final outcome can be fully explained by its influence on the intermediate, then the marker can be used as an **instrumental variable** (or **instrument**, for short).

- This is the theoretical rationale in the use of so-called '**Mendelian Randomisation**'.
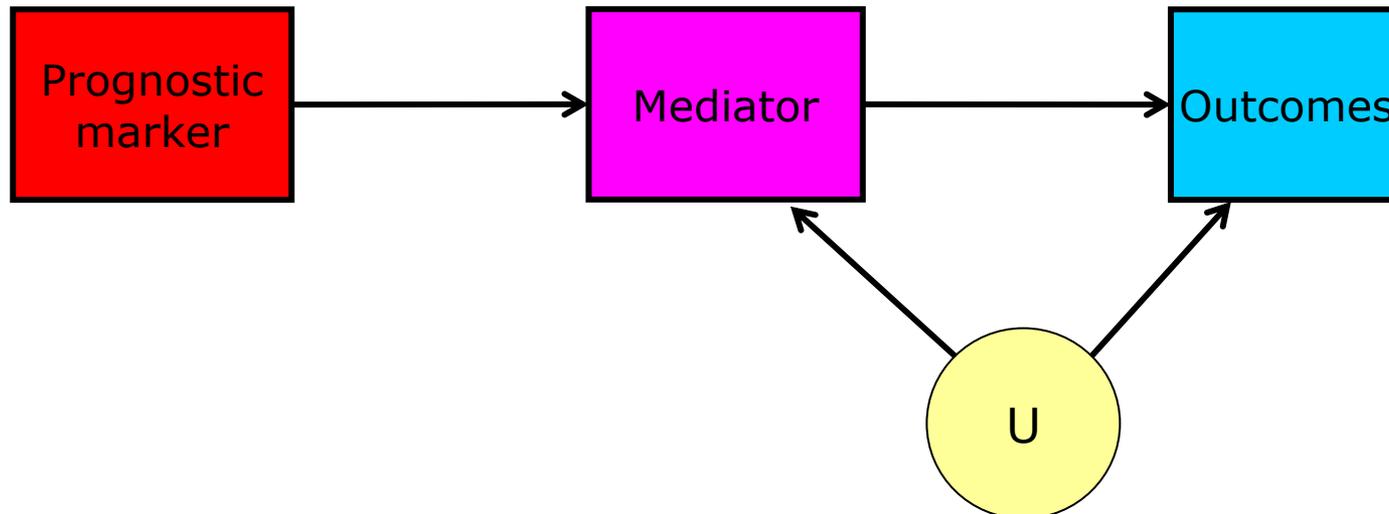
Prognostic biomarkers as instrumental variables

No direct effect of the prognostic marker on outcome
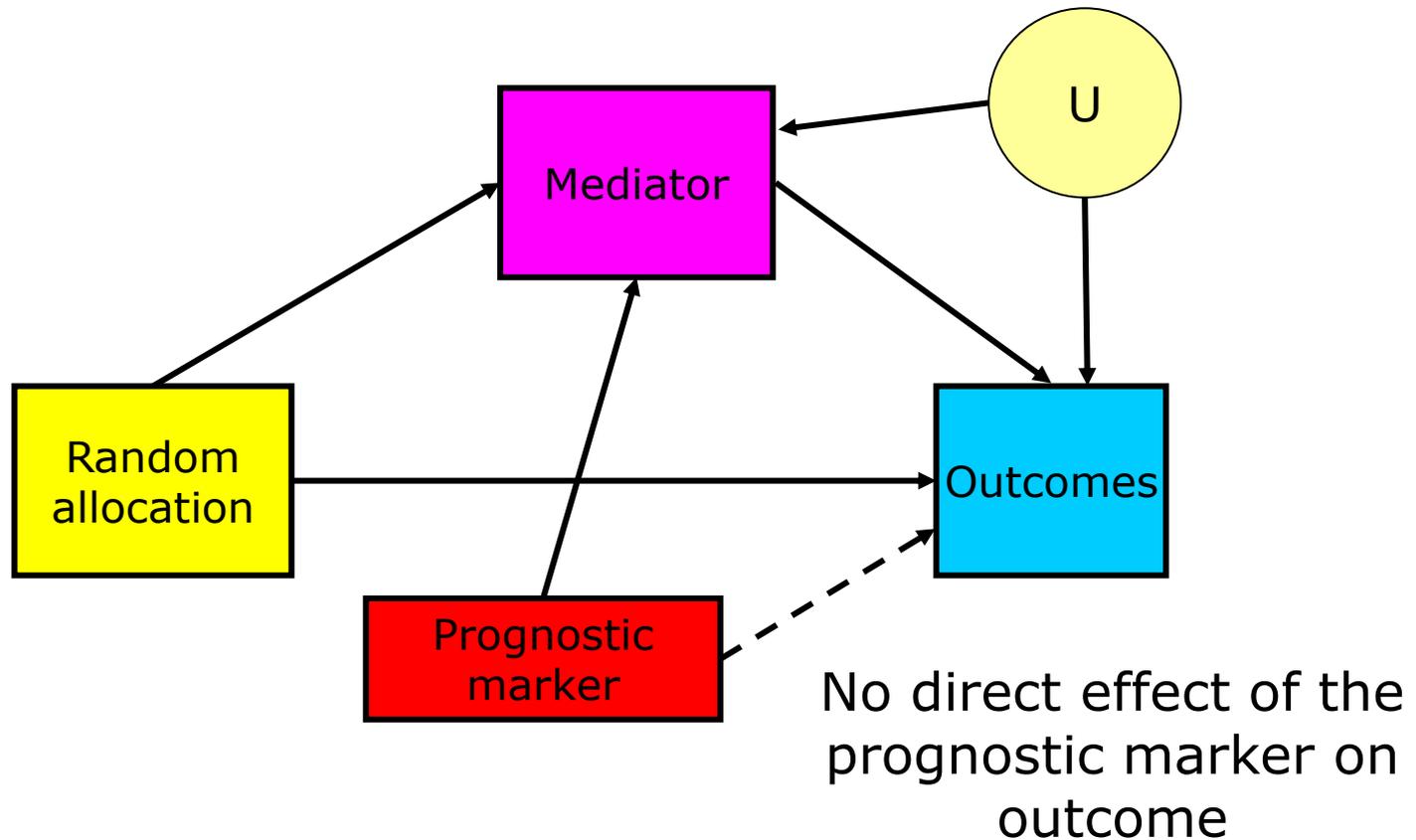
Predictive markers: mechanisms evaluation in personalised medicine

- If the treatment-effect moderation on final outcome is wholly explained by the moderation of the effect of treatment on the intermediate outcome, then the latter (i.e. a treatment by marker interaction) can be used as an instrument.

- A more subtle (and more realistic?) version of Mendelian Randomisation.
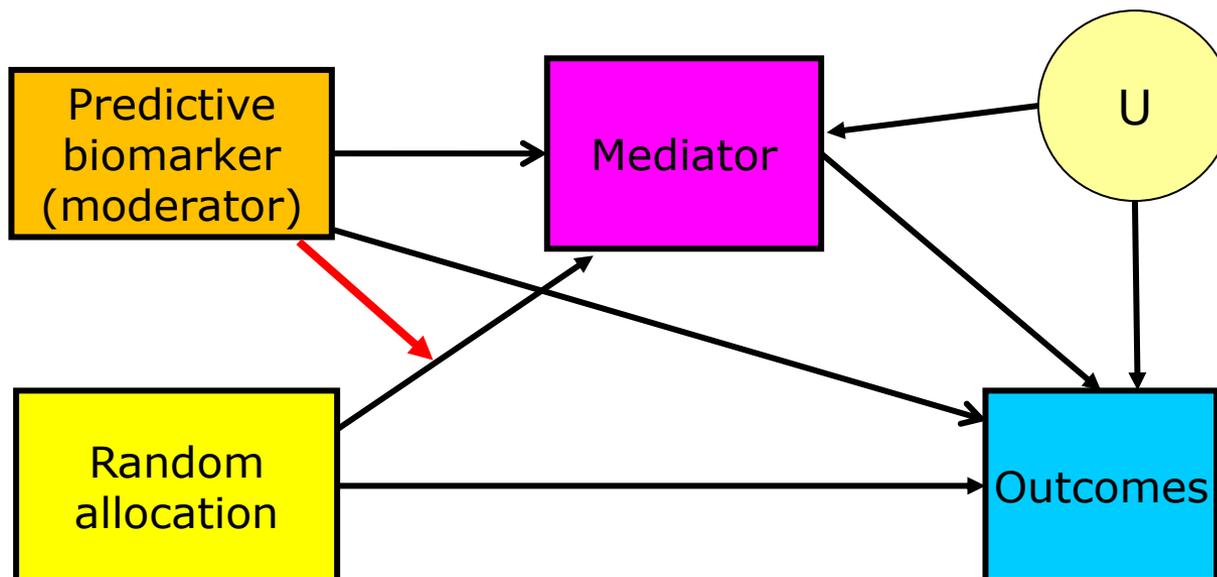
Contents

1. Stratified medicine: a causal inference perspective

2. The biomarker stratified (BS) trial design

3. Efficacy and mechanisms evaluation

4. Putting it all together: a BS-EME trial

5. Evaluation of the BS-EME trial design: simulation studies

6. Conclusions and future directions

Putting it all together: potential roles of genetic and other markers

Potential roles of prognostic biomarkers – measured confounder or instrument

U – unmeasured confounders
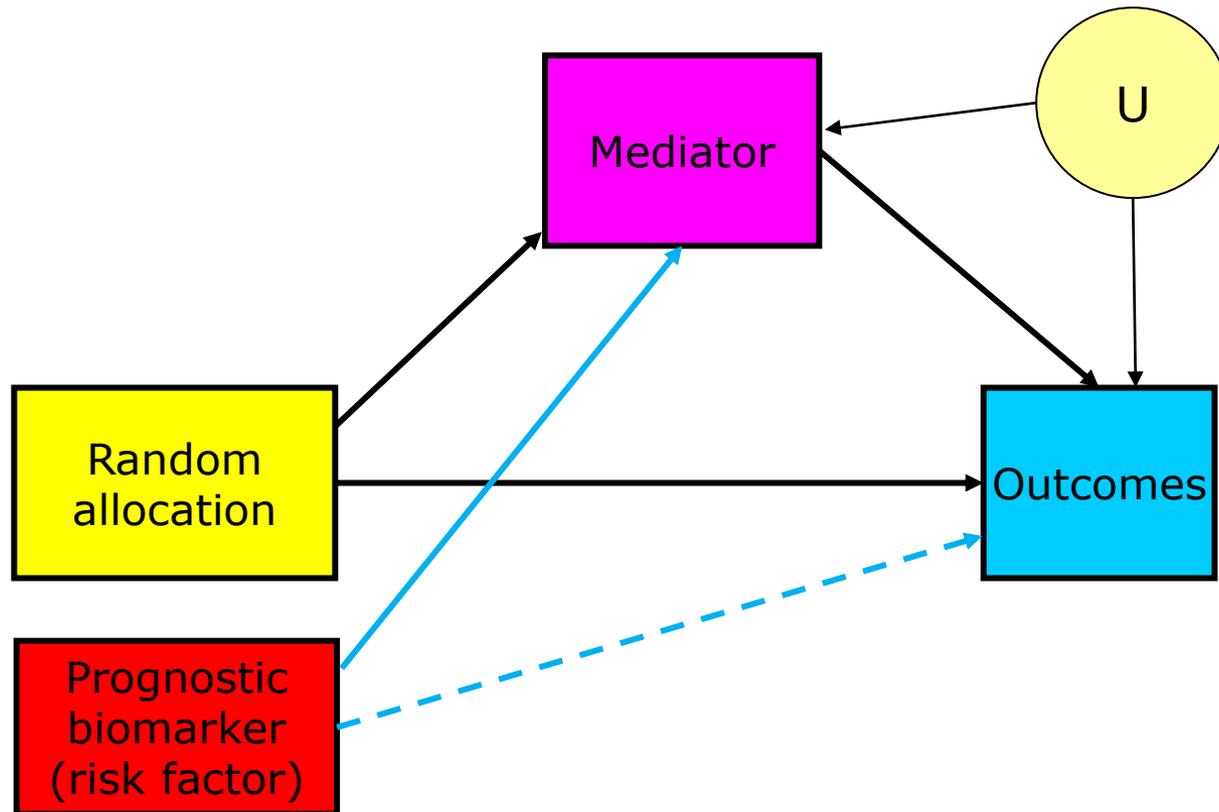
Blue dotted line – pathway we might assume are absent
Alternatively, we might assume that there are no longer any Us

Potential role of predictive biomarkers

U – unmeasured confounders
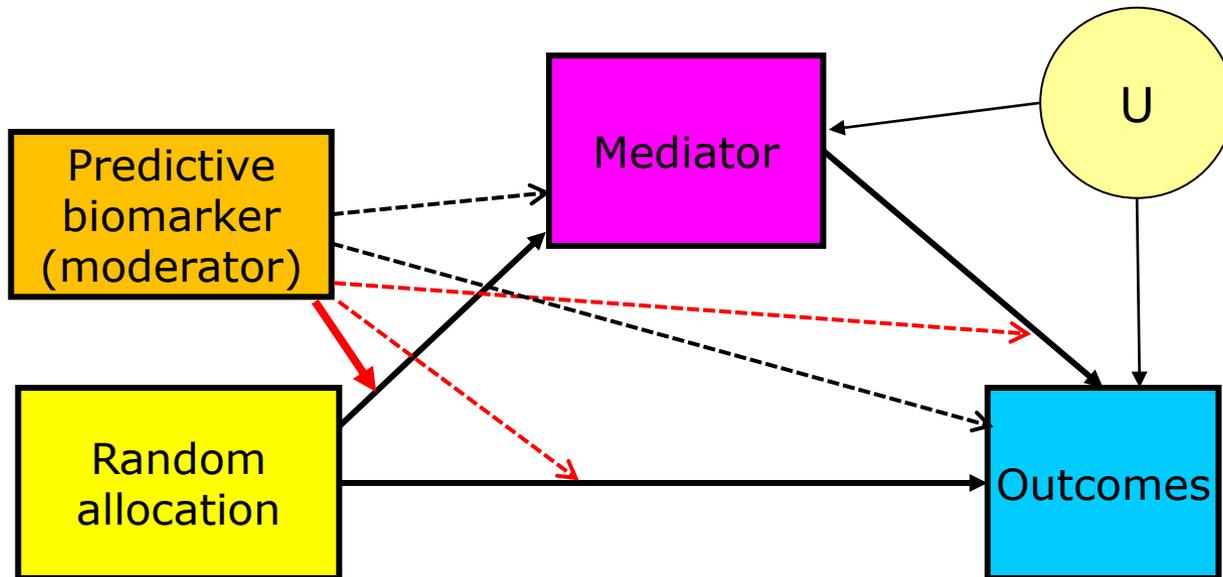
Red dotted lines – interaction pathways we might be justified in assuming are absent

Stratification and mechanisms evaluation

Using the treatment by marker interaction as an instrument

U – unmeasured confounders
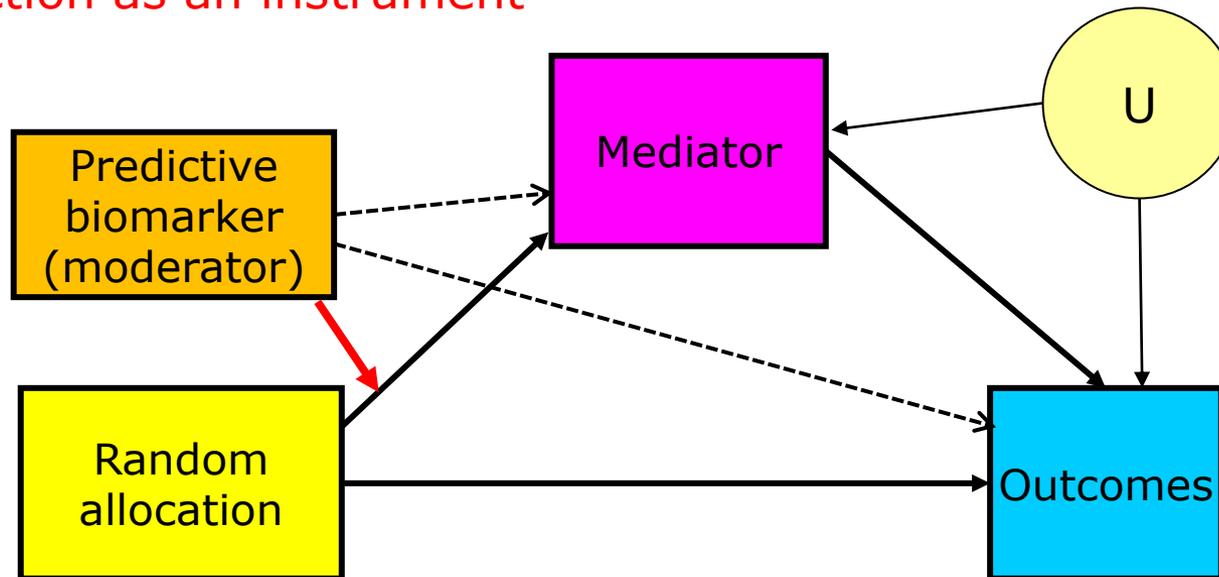
- Are we correct in assuming that there is no moderating effect on the other pathways?

- Dependent on prior knowledge of the biology/biochemistry of the system.

Predictive markers as instrumental variables

- Given a predictive biomarker *X10, M, Y, treat, X11=X10*treat*, our two causal structural models are:

  ➢ $M = \beta_0 + \beta_1 X10 + \beta_2 treat + \beta_3 X11 + \varepsilon_m$

  ➢ $Y = \psi_0 + \psi_1 X10 + \psi_2 treat + \psi_3 M + \varepsilon_y$

  ➢ $ATE(M) = E[M_T - M_C] = \beta_2 + \beta_3 X10$

  ➢ $ATE(Y) = E[(Y_T - Y_C)|X10] = \psi_2 + \psi_3 E[(M_T - M_C)|X10]$

  ➢ $Cov(\varepsilon_m, \varepsilon_y) \neq 0$

Treatment effects

- β_2 - effect of treatment on mediator when *X10=0*

- $\beta_2+\beta_3$ - effect of treatment on mediator when *X10=1*

- Ψ_2 – effect of treatment on outcome

- Ψ_3 – effect of mediator on outcome

- $\beta_2\Psi_3+\Psi_2$ – total effect of treatment on outcome when *X10=0*

- $(\beta_2+\beta_3)\Psi_3+\Psi_2$ – total effect of treatment on outcome when *X10=1*

Using strong theory and all available prognostic marker information

The BS-EME trial design

- Taking the biomarker stratified design as described previously we supplement the baseline information (i.e. $X10$ status) by:
  - ➤ measuring all previously-validated prognostic markers ($X1$ to $X9$, say)
  - ➤ baseline covariates (demographic information; clinical and treatment history; co-morbidity; social, psychological and cultural variables; etc.) thought to have prognostic value.
  - ➤ baseline measurement of the putative mediator.
  - ➤ baseline value for the final outcome measurement.

- The rationale for all of these measurements is

  (a) to allow for as much confounding of the effects of the mediator on final outcome as is feasible,

  (b) to assess sensitivity of the results to assumptions concerning residual hidden confounding and, perhaps more importantly,

  (c) increase the precision of the estimates of the important causal parameters.

Contents

1. Stratified medicine: a causal inference perspective

2. The biomarker stratified (BS) trial design

3. Efficacy and mechanisms evaluation

4. Putting it all together: a BS-EME trial

5. Evaluation of the BS-EME trial design: simulation studies

6. Conclusions and future directions

A simulated BS-EME trial

- Trial with 200 participants (100 treated, 100 controls).

- Quantitative outcome, Y.

- Binary predictive marker ($X10$): Moderating effect of $X10$ on outcome solely through the mediator ($X10$ known to be an IV).

- Variants of $X10$ equally probable (50:50).

- Nine prognostic (genetic) uncorrelated binary markers $X1$-$X9$ and all nine are confounders.

Data generating models

- Mediator (M):

    $M=5*(X1-X9)+5*X10+5*\text{treat}+20*X11+e12$

    ➤ $X11$ = treat*$X10$, e12 is a random 'error' term

- Outcome (Y):

    $Y=5*(X1-X9)+5*X10+2*M+10*\text{treat}+e13$

    ➤ e13 is a random 'error' term (uncorrelated with e12).

    ➤ There is no $X11$ (interaction) in the outcome model.

    ➤ THERE ARE NO UNMEASURED COMMON CAUSES (i.e. $X1-X9$, and $X10$, are all measured).

Simulated BS-EME trial
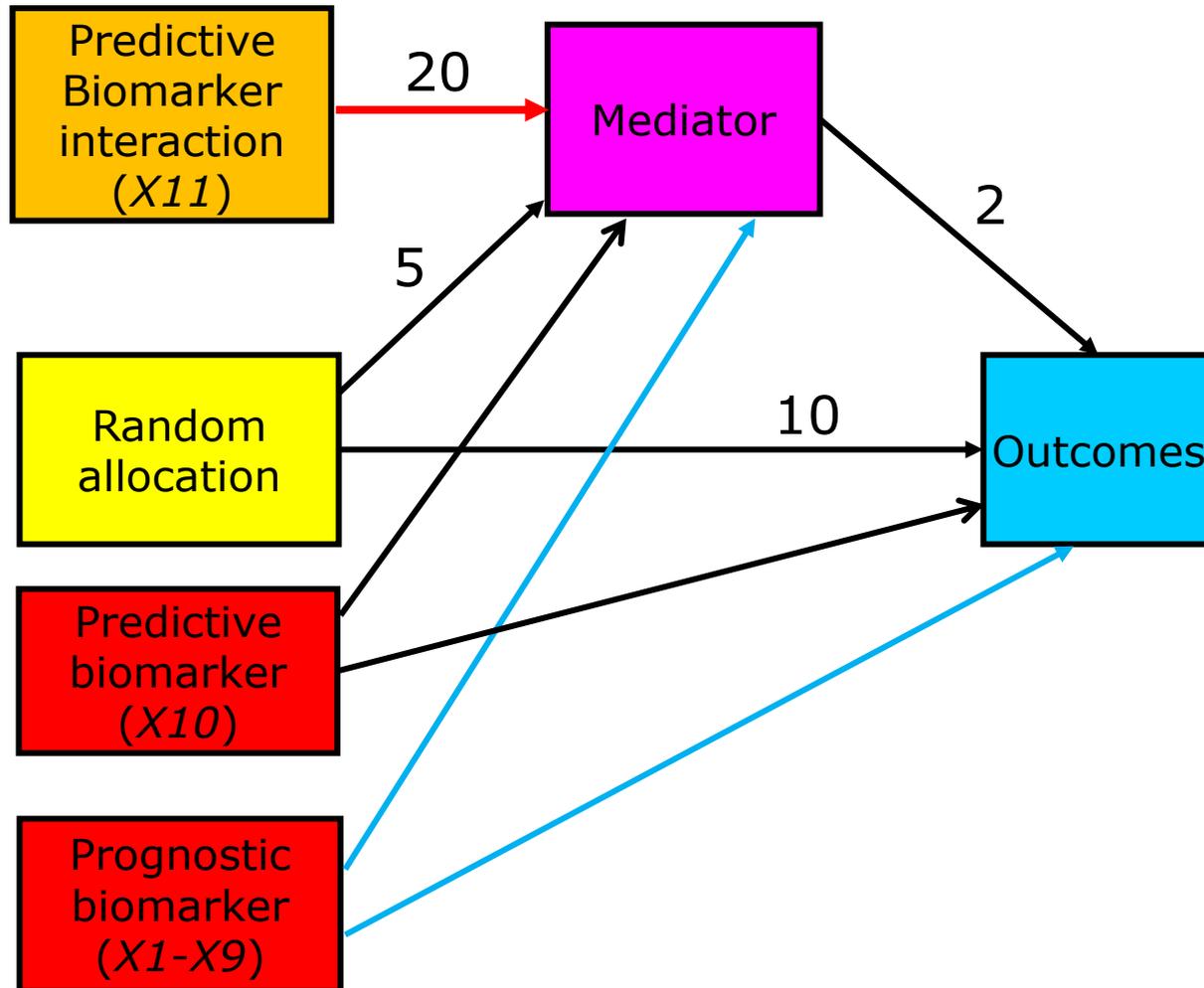
Instrumental variables analysis

- Let's pretend we've not measured *X1-X9*: now there *are* 'unmeasured' common causes.

- An instrumental variable regression in *Stata*:

ivregress 2sls *Y* treat *X10* (*M = X11*), first

- This is a two-stage least-squares procedure which simultaneously estimates the effect of treatment on *M* (the first-stage regression), the effect of *M* on *Y*, and direct effect of treatment on *Y* (the second stage).

Simulated BS-EME trial

Multiple regression models for the joint effects of treatment and the mediator on outcome: No interactions (i.e. *x11*) in the analysis model.

		Naïve model regress y x10 treat m		Adjusted for all confounders regress y x10 treat m x1-x9	
Effect	**True**	**Estimate**	**s.e.**	**Estimate**	**s.e.**
treat	**+10**	**+3.16**	1.21	**+9.97**	1.04
M	**+2**	**+2.46**	0.05	**+2.00**	0.05

- On the left, assume no measure on any of *x1-x9*.

- On the right, have made adjustments for the effects of *x1-x9*.

- The results on the left are clearly biased. If we know all the confounders and if we make adjustments for them all then we can retrieve the correct treatment effects (the column on the right).

Simulated BS-EME trial

Multiple regression models for the joint effects of treatment and the mediator on outcome: Including a treatment by predictive marker interaction (the effect of *X11*) in the analysis model

Effect	True	Naïve model regress y x10 treat m x11		Adjusted for all confounders regress y x10 treat m x11 x1-x9	
		Estimate	s.e.	Estimate	s.e.
treat	+10	+6.80	1.26	+9.97	1.08
M	+2	+2.63	0.05	+2.00	0.07
X11	**0**	**-12.41**	2.08	**+0.03**	2.07

- The results of the naïve analysis reveal a highly statistically significant interaction, which is an artefact of confounding.

- Only if we correctly allow for all of the known confounders (the column on the right) then we obtain a small and statistically non-significant effect.

Simulated BS-EME trial: instrumental variable estimators (*x11* as the instrument)

First stage regressions (modelling *M* as the outcome)

Effect	True	No prognostic markers in the analysis model ivregress 2sls y x10 treat (m=x11)		All prognostic markers in the analysis model ivregress 2sls y x1-x9 x10 treat (m=x11)	
		Estimate	s.e.	Estimate	s.e.
treat	+5	+5.00	1.28	+5.02	0.80
X11	+20	+20.00	2.83	+20.00	1.76

Second stage regressions (modelling *Y* as the outcome)

Effect	True	Estimate	s.e.	Estimate	s.e.
treat	+10	+10.00	2.07	+9.97	1.31
M	+2	+2.00	0.12	+2.00	0.07

Some conclusions from the simulations

- Larger sample sizes than 200 give the same results with respect to bias, but are obviously more precise.

- Standard approach: If (but *only if*) we've measured all confounders then this is valid and it is the most precise method. But ...we never know, and it seems unlikely that all prognostic markers will be measured.

- IV approach is unbiased but less precise – we don't get something for nothing.

- However, there is a considerable gain in precision with prognostic markers:
  - ➢ Measurement of prognostic markers not essential, but it makes the design more efficient (i.e. get away with a smaller trial) – perhaps the difference between a viable trial and one that's just not feasible.

Changing the prevalence of the marker

- We assumed a 50:50 ratio of the predictive marker *X10* in the sample for the simulations.

- If there is a strong prior belief that one biomarker strata will not benefit from the treatment, it may be unethical to expose a large number of patients in that strata to the treatment (though in order to test that belief, some patients would need to be exposed to treatment).

- We need to examine the effect of different biomarker prevalence on our estimation approach
  - ➢ i.e. the implications of differential sampling of predictive biomarker prevalence for the trial design.

Simulated BS-EME trial: instrumental variable estimators (*x11* as the instrument)

- Changing the prevalence to 90% marker positive
  - ➢ Second stage of IV regressions (modelling *Y* as the outcome)

Effect	True	Estimate	s.e.	Estimate	s.e.
treat	+10	+10.48	6.01	+9.87	3.12
M	+2	+1.98	0.25	+2.01	0.13

- Slight increase in the bias of the estimates.

- The importance of measuring the prognostic markers for the increase in precision is clear.

Contents

1. Stratified medicine: a causal inference perspective

2. The biomarker stratified (BS) trial design

3. Efficacy and mechanisms evaluation

4. Putting it all together: a BS-EME trial

5. Evaluation of the BS-EME trial design: simulation studies

6. Conclusions and future directions

Some conclusions

1. Personalised (stratified) medicine and treatment-effect mechanisms evaluation are inextricably linked;
2. Stratification without corresponding mechanisms evaluation lacks credibility;
3. In the almost certain presence of mediator-outcome confounding, mechanisms evaluation is dependent on stratification for its validity;
4. Both stratification and treatment-effect mediation can be evaluated using a biomarker stratified trial design together with detailed baseline measurement of all known prognostic biomarkers and other prognostic covariates;
5. Direct and indirect (mediated) effects should be estimated through the use of instrumental variable methods together with adjustments for all known prognostic biomarkers (confounders) – the latter adjustments contributing to increased precision (as in a conventional analysis of treatment effects) rather than bias reduction.

Key references

- Dunn G, Emsley RA, Liu H & Landau S. (2012). Trial designs fully integrating biomarker information for the evaluation of treatment-effect mechanisms in personalised (stratified) medicine. Submitted to *Clinical Trials.*

- Emsley RA, Dunn G & White IR. (2010). "Mediation and moderation of treatment effects in randomised controlled trials of complex interventions", *Statistical Methods in Medical Research*, 19(3),237-270.

- Simon R. (2010). Clinical trials for predictive medicine: new challenges and paradigms. *Clinical Trials*;7:516-524.

- Young KY, Laird A, Zhou ZX. (2010). The efficiency of clinical trial designs for predictive biomarker validation. *Clinical Trials*;7:557-566.

- Freidlin B, McShane LM, Korn EL. (2010). Randomized clinical trials with biomarkers: design issues. *Journal of the National Cancer Institute*;102(3):152-160.

Research Programme:
Efficacy and Mechanisms Evaluation

Funded by MRC Methodology Research Programmes:

- **Estimation of causal effects of complex interventions in longitudinal studies with intermediate variables (2009-2012)**
  - ➤ Richard Emsley (MRC Fellow), Graham Dunn.

- **MRC Early Career Centenary Award (2012-13)**

- **Designs and analysis for the evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health (2010-12)**
  - ➤ Graham Dunn (PI), Richard Emsley, Linda Davies, Jonathan Green, Andrew Pickles, Chris Roberts, Ian White & Frank Windmeijer with **Hanhua Liu**.

- **Developing methods for understanding mechanism in complex interventions (2013-15)**
  - ➤ Sabine Landau (PI), Richard Emsley, Andrew Pickles, Graham Dunn, Ian White, Paul Clarke