TITLE PAGE

Mixed models showed no need for initial response monitoring after starting anti-hypertensive therapy.

Authors:

Katy JL Bell[1] MBChB, MMed (Clin Epi)

Andrew Hayen[1] BA (Hons), MBiostat, PhD

Petra Macaskill[1] BA, MAppStat, PhD

Jonathan C Craig[1,2] MBChB, FRACP, MMed (Clin Epi), PhD

Bruce C Neal[3] MBChB, PhD

Les Irwig[1] MBBCh, PhD

Author Affiliations:

Screening and Test Evaluation Program (STEP), School of Public Health, Building A27, The University of Sydney, New South Wales 2006, Australia (Katy JL Bell, Andrew Hayen, Petra Macaskill, Jonathan C Craig, Les Irwig); Department of Nephrology, The Children's Hospital at Westmead, New South Wales 2145, Australia (Jonathan C Craig); The George Institute for International Health, Royal Prince Alfred Hospital, Building C39, The University of Sydney, NSW 2006 Australia (Bruce C Neal)

Text word Count: 3284 words

ABSTRACT

Objective: Demonstrate how mixed models may be used to estimate treatment effects and inform decisions on the need for monitoring initial response.

Study Design and Setting: Mixed models were used to analyse data from the Perindopril Protection Against recurrent Stroke Study (PROGRESS) which examined the effects of perindopril and indapamide in 6105 patients at high risk of a cerebrovascular event.

Results: The mean effect of perindopril was to lower blood pressure (systolic/diastolic) by 6/3 mmHg. The mean effects of perindopril/indapamide varied according to baseline blood pressure, and lowering of blood pressure ranged from 9/5 to 14/5 mmHg (for individuals with a baseline systolic blood pressure <140 and >150 mmHg respectively). We found no variation in the effects of treatment on blood pressure for either perindopril alone or in combination with indapamide. The effects of treatment on the individual can be predicted from the mean effect of treatment for the group (perindopril) or baseline systolic blood pressure subgroup (perindopril/indapamide).

Conclusion: Monitoring initial treatment response is unnecessary for anti-hypertensives similar to those examined in this study. To address this issue for other therapies we suggest trials report estimates of treatment effects from mixed models, and the CONSORT statement be expanded to include this item.

Abstract word count: 203

Key words: blood pressure, vascular diseases, chronic disease, linear models, statistical models

Running title: Mixed models showed no need for initial response monitoring

Clinical care commonly involves monitoring patients with chronic disease. Although monitoring is used in nearly every chronic disease, it is uncertain whether monitoring does more good than harm, and valid methods remain poorly defined. Monitoring may be divided into the following phases: pre-treatment, initial response, maintenance, re-establish control and post-treatment(1). Initial response monitoring uses repeated measurements soon after a new therapy is started to check a patient's response is within a range that maximises the benefits while minimising the harms. Monitoring initial response to drug treatment may be done using patient centred outcomes, intermediate outcomes or adherence measures.(2) In this article we limit our discussion to initial response monitoring of intermediate outcomes.

Intermediate outcomes (such as blood pressure and cholesterol) are usually used for initial response monitoring in patients with chronic conditions. These outcomes are used to predict patient-relevant, long-term endpoints like a patient's risk of stroke or myocardial infarction. These 'hard' end points are unsuitable for monitoring purposes as they may occur many years after the patient was first diagnosed, be irreversible or carry a substantial mortality risk. The intermediate outcomes are often responsive to therapy, and by using therapy to alter the value of an intermediate outcome early on in the disease process, the clinician hopes to change the patient's risk of developing later clinically important outcomes.

An intermediate outcome should only be considered for monitoring if a change in this outcome is known to predict the effect of treatment on the risk of the clinical outcome. Such evidence usually comes from population level meta-analyses of randomized controlled trials where change in intermediate outcome is related to change in risk of clinical outcome for patients on active treatment relative to those on placebo. However, although the population average treatment effect on an intermediate outcome may predict the population average treatment effect on the risk of a clinical outcome, the intermediate outcome might not be useful for monitoring the treatment effect in an

individual. Measurement variation can cause random change in the intermediate outcome in an individual. Failure to recognize non-treatment related variation in the intermediate may lead clinicians to make inappropriate changes to therapy or conversely to delay taking action when they should intervene(3).

Because of the potential for misinterpretation of changes observed in the intermediate outcome within an individual, unnecessary initial response monitoring is best avoided. Population data from randomized trials can be used to decide when initial response monitoring is unnecessary for the individual. For instance, monitoring is unnecessary (and best avoided) if the treatment effect on the intermediate outcome is the same for everyone. However, monitoring may be necessary if the treatment effect on the intermediate outcome differs between individuals. In this case the need for monitoring will usually depend on the probability of meeting defined treatment targets. If there is a high probability that a patient will meet a pre-determined target level with treatment, then there will be no need for monitoring. Conversely, if there is uncertainty whether the patient will meet the target level, then initial response monitoring will be needed.

Much of the often considerable variation observed between patients on treatment may be explained by pre-treatment differences between patients, short-term variability and measurement error(4). Temporal variation in an intermediate outcome in the placebo arm of a randomized trial represents variation from all non treatment sources. This variation includes within person measurement variability due to both measurement error and short term biological fluctuations. It also includes between person variation in baseline level of the intermediate outcome and change in the intermediate outcome over time; this variation may arise because of differences in underlying physiology and the effect of co-interventions besides the trial medication (these co-interventions may be non-pharmacological such as diet and exercise, or they may be other non-trial medications). If there is no variation in the treatment effect between patients then variability in the treated group

should be equivalent to that in the placebo group. Analysis of the difference in variability between placebo and active treatment groups provides insight into whether the treatment effect differs between individuals. The main sources of variability in randomized controlled trials are summarized in Table 1 (adapted from (4)).

A common example of initial response monitoring is blood pressure monitoring after starting a new anti-hypertensive agent. Individuals are started on anti-hypertensive treatment if they are judged to be at increased risk of a vascular event. A recent population based prospective study found 1.8% of an English population suffered one or more vascular events over a three year period, with a steep increase in risk with increasing age(5). The proportion of individuals at increased risk of a vascular event (for whom anti-hypertensives are prescribed) is considerably greater than this and will grow further as populations age. As nearly all individuals commenced on anti-hypertensive treatment are monitored for their initial response, there are a large number of individuals having potentially unnecessary initial response monitoring of blood pressure.

In this paper, we use data from the Perindopril Protection Against recurrent Stroke Study (PROGRESS)(6), a randomized trial of perindopril and indapamide in patients at high risk of stroke, to demonstrate how mixed models may be used to inform decisions on initial response monitoring. Monitoring questions potentially answered by this trial include: Should we monitor blood pressure after starting perindopril in patients at high risk of a vascular event? Should we monitor after starting perindopril and indapamide together?

METHODS

Study Design and Population

We analysed data from PROGRESS(6). This trial evaluated the effects of perindopril alone compared to single placebo or perindopril together with indapamide compared to double placebo on

the risk of stroke in 6,105 high risk patients. The decision on whether patients would be allocated to single therapy (single placebo or perindopril) or dual therapy (double placebo or perindopril/indapamide) was decided on clinical grounds before randomization; for this reason single therapy and dual therapy groups are considered separately.

Outcome Measures

We fitted mixed models using systolic and diastolic blood pressure as outcomes. Measurements were made to the nearest 2 mmHg, with a standard mercury sphygmomanometer. We used the average of two measurements made at each clinic visit for each individual. Blood pressure data from the first three clinics post randomization were used: one month, three months and six months after treatment was started. Baseline blood pressure (measured prior to starting the trial) was included as one of the predictors. We limited analysis to the first six months post randomization to reflect a period of time that might be used clinically for monitoring initial response after starting a new therapy.

Statistical Analysis

A series of mixed models were fitted where there was the same treatment effect for everyone (treatment had a fixed effect) or between person differences in treatment effect (treatment had random effects). We outline the alternative models in order of increasing complexity. Model 0 (random intercept, no treatment effect) included all significant predictors except treatment: baseline blood pressure, age, gender, other anti-hypertension treatment and presence of left ventricular hypertrophy on baseline electrocardiogram. To correct for measurement error in individuals' baseline blood pressure measurements, we adjusted blood pressure measurements made at the pre-randomization visit to reflect average blood pressure levels found during follow up in the placebo group. This was done by regressing the mean of three follow up measurements on the pre-treatment measurement for individuals on placebo and using the resulting regression equation to adjust all

individuals' pre-treatment measurements. This method is adapted from previously described methods of correction for measurement error(7). The predictors were fitted to have the same effect for everyone (i.e. they were fixed effects). This model also included a term for between person variation in the first measurement on treatment (random intercept). In Model 1 treatment was fitted to have the same effect for everyone (Model 1: random intercept, fixed treatment effect). Model 1 was also used to investigate effect modifiers of treatment apart from time. The following interaction terms were assessed for significance: treatment×baseline blood pressure, treatment×age, treatment×gender, treatment×other treatment, treatment×LVH. In this way we identified factors that predict the effect of treatment. Model 1 was then extended to include a term for between person variation in change over time to account for the fact that individuals may differ in how their measurements change over time (Model 2: random intercept and random time effects, fixed treatment effect). This model was further extended to include a term for between person variation in treatment effect on the first measurement (Model 3: random intercept, random time effects and random treatment effects). Two further models were fitted which included a treatment-by-time interaction where the effect was the same for everyone (Model 4: random intercept, random time effects, fixed treatment effect and fixed treatment time interaction), and where the effects differed between individuals (Model 5: random intercept, random time effects and random treatment time interactions).

The alternative models were compared using likelihood ratio tests to see which model provided the best fit for the data. Figure 1 provides a schematic overview of the model fitting process. We started with the least complex model (Model 0) and used forward selection to add the parameters that were needed until the final model was reached.

As the primary analysis we used systolic blood pressure as our outcome (univariate response models). Systolic blood pressure is the best type of blood pressure measurement for predicting

macrovascular disease(8, 9). However many clinicians use both systolic and diastolic blood pressure measurements to guide decisions about therapy and monitor treatment effects. These two types of measurement are highly correlated when taken at the same time on the same individual(10). As an extension to the analysis outlined above, we also built bivariate response models which enabled the two outcomes to be considered simultaneously while allowing for their correlation.

Assessing the power of mixed models to detect random effects is a challenging and understudied area. To minimise the possibility that a failure to detect random effects of treatment reflected insufficient analytic power, we chose a large dataset for our example. Models were fitted using data from 6105 individuals, with four measurement occasions for each individual. The power was further increased by using bivariate response models that incorporated both systolic and diastolic blood pressure.

A detailed description on how mixed models inform decisions on initial response monitoring is provided in the Appendix. Analysis was done using MLwiN (using iterative generalised least squares).

RESULTS

For both single and dual therapy groups, there was an improvement in model fit for Model 1 (where treatment has a fixed effect and time has no effect) compared to Model 0 (where both treatment and time have no effect), indicating a significant treatment effect on mean systolic blood pressure at one month. For single therapy, there was no evidence that treatment effect was modified by any of the other predictors included in the model. For dual therapy there was very strong evidence that treatment effect was modified by baseline blood pressure (all other interactions were not statistically significant).

There was further improvement in model fit for Model 2 (where treatment has a fixed effect and time has random effects) compared with Model 1, indicating a significant change in mean systolic blood pressure over the six months as well as between person variation in change. There was no further improvement in model fit for Model 3 (where both treatment and time have random effects) compared with Model 2. This means there was no evidence of between person variation in treatment effect on blood pressure at one month for either perindopril alone or in combination with indapamide (likelihood ratio test for single therapy: χ^2=1.64, 2df, p=0.44, likelihood ratio test for dual therapy: χ^2=3.8, 2df, p=0.15). There was also no improvement in model fit for Model 4 and Model 5 compared with Model 2 which means there was no evidence that the treatment effect changed over time.

We adopted the second model as our final model for both single and dual therapy groups. The parameter estimates for fixed effects and random effects in the final model are presented in Table 1 along with results of significance testing (single therapy estimates are on the left-hand side and dual therapy estimates are on the right-hand side).

The estimated effect of perindopril alone (relative to single placebo) was to lower systolic blood pressure by 6 mmHg. The estimated effects of perindopril and indapamide combined (relative to double placebo) varied according to baseline blood pressure, with systolic blood pressure lowered by 9, 11, 13 and 14 mmHg for individuals with a baseline systolic blood pressure <140, 140-145, 145-150 and >150 mmHg respectively.

The bivariate response analyses (where systolic and diastolic blood pressures were considered simultaneously) yielded similar conclusions to the analyses of systolic blood pressure alone. There remained no evidence of variation in treatment effect between patients (likelihood ratio test for single therapy: χ^2=6.9, 7df, p=0.44, likelihood ratio test for dual therapy: χ^2=9.6, 7df, p=0.21). The

estimated effect of perindopril alone (relative to single placebo) was to lower (systolic/diastolic) blood pressure by 6/3 mmHg. The estimated effects of perindopril and indapamide combined (relative to double placebo) varied according to baseline blood pressure, with blood pressure lowered by 9/5, 11/5, 12/5 and 14/5 mmHg for individuals with a baseline systolic blood pressure <140, 140-145, 145-150 and >150 mmHg respectively (Treatment effect on diastolic blood pressure did not vary according to baseline diastolic blood pressure).

DISCUSSION

Clinical medicine has a long history of initial response monitoring, aimed at ensuring therapy is tailored to meet the specific needs of the individual. Where initial response monitoring achieves this, patients are likely to benefit, but current monitoring practices are rarely based on empirical research and may actually cause harm(11). In this paper we have illustrated how mixed models on individual patient data from clinical trials can be used to inform decisions on the need for initial response monitoring.

Our example used trial data to examine the need for blood pressure monitoring of antihypertensive therapy in patients at high risk of a cerebrovascular event. Although there was substantial variability in blood pressure both between and within patients, we found no evidence of variation in treatment effect between patients after allowing for background variability and other significant predictive factors. The effect of perindopril alone was assumed to be the same for everyone, and the effect for the individual could be predicted from the group's mean of 6/3 mmHg reduction in blood pressure. The effects of perindopril and indapamide combined differed depending on the patient's baseline blood pressure and ranged from a lowering of 9/5 mmHg (for individuals with baseline systolic blood pressure below 140 mmHg) to 14/5 mmHg (for individuals with baseline systolic blood pressure greater than 150 mmHg). For this patient group there is no need for initial response monitoring of perindopril alone (4mg), or of perindopril in combination with indapamide (4mg/2-

2.5mg). Decisions to alter therapy made on the basis of monitoring blood pressure in this early period after treatment initiation would be misguided as any variation in intermediate outcome is likely to be due to background random variation that is not related to treatment.

Current guidelines for the management of hypertension recommend a treatment target of below 130/80 for a number of groups of high risk patients(12, 13). On this basis, for individuals with a true baseline BP <136/83 (known through averaging a large number of pre-treatment measurements) perindopril (4mg) alone will be sufficient therapy to meet the target. For patients with baseline blood pressure levels above this but below 141/85, the combination of perindopril (4mg) and indapamide (2-2.5mg) will usually achieve the target but for patients with higher baseline blood pressure levels more intensive therapy would be required. Treatment decisions about which therapy should be commenced to achieve the desired target can be made before any blood pressure lowering agents are begun. Treatment targets will be met more quickly, patients will be saved unnecessary clinic visits, clinical benefit will be maximised and inappropriate cessation of treatment will be averted.

There are problems with using blood pressure targets to guide therapy. There is a risk of under treating high risk individuals such as stroke patients who happen to have a low or normal level of blood pressure at baseline. For this reason it may be preferable to use a target change in blood pressure rather than a target level. For example, the target change might be a decrease in systolic/diastolic blood pressure of at least 10/5 mmHg (which corresponds to a relative risk reduction of about 33% for stroke and of 20% for ischaemic heart disease(14-16)). On the basis of our results, only when perindopril (4mg) and indapamide (2-2.5mg) are given together will the treatment effect usually exceed this minimum target change. Some advocate even larger reductions in blood pressure for high risk patient(14). So far studies have failed to identify a "threshold" (or “J-shape”) below which a lower blood pressure is not associated with a lower risk of stroke or

coronary heart disease. An intensive approach to therapy is to continue escalation of anti-hypertensive agents as far as the patient will tolerate. Adopting this approach there is even less reason to monitor blood pressure after starting therapy – monitoring would only be considered if the patient began to experience adverse effects from therapy that might be due to hypotension.

Methods were used in PROGRESS to minimise measurement error in blood pressure estimation (for example blood pressure was measured to the nearest 2 mm Hg with a mercury sphygmomanometer and the average of two measurements taken at each clinic was used in analysis). Measurement error may have been further reduced with the use of 'out of office' monitoring techniques such as 24 hour ambulatory blood pressure monitoring and patient self monitoring (both use the average of a large number of blood pressure measurements). Although this may have resulted in smaller estimates of within person variation in the mixed models, the estimates of between person variation would remain unchanged, and the finding of no variation in treatment effects would still hold.

We have focused attention on a continuous outcome because these are by far the most common form of intermediate outcome (blood pressure, cholesterol, bone mineral density). However the methods described can be adapted to other forms of intermediate outcome such as binary outcomes (e.g. presence of retinopathy on fundoscopy) and ordinal outcomes (e.g. number of plaques on MRI in multiple sclerosis, amount of protein on urinalysis).

The type of analysis we have presented in this paper requires individual patient data which is only available to clinicians if reported by trialists. For intermediate outcomes used for initial response monitoring, it is most useful if both the mean treatment effect and the between person variability in treatment effect are reported from a mixed model analysis. These results will often be context specific. This means that not only do we need this information for trials of different treatments, we

also need it for trials of the same treatment with different patient populations, different doses of treatment or different intermediate outcomes. For example, while it is reasonable to generalise our results to other angiotensin converting enzyme inhibitors and thiazide diuretics at equivalent doses, further research is needed for other classes of antihypertensive drugs such as calcium channel blockers, beta-blockers and angiotensin-II receptor antagonists. In this way not only would trials inform clinical decisions on which treatment to choose (treatments that are proven to reduce the risk of adverse clinical outcomes) but also whether initial response monitoring is needed (when there is clinically relevant between person variation in the treatment effect on intermediate outcomes). We would recommend these data should be requested by journals in the next revision of the CONSORT statement(17).

ACKNOWLEDGEMENTS

Authors' Contributions:

KJLB conceived the study, did the analysis and wrote the paper. AH helped with the analysis and helped write the paper. JCC and BCN helped write the paper. PM and LI supervised the analysis and helped write the paper. KJLB is guarantor.

Access to Data:

KJLB had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Potential Conflicts of interests:

Consultancies or honoraria

BCN has received honoraria from a range of pharmaceutical companies, including the sponsor of PROGRESS, for presentations about the effects of blood pressure lowering made at scientific meetings.

Grants

BCN has received research grants from a range of pharmaceutical companies, including the sponsor of PROGRESS, for research investigating the effects of blood pressure lowering agents.

Financial Support information: The authors have received funding from the Australian National Health and Medical Research Council Program Grants (No.s 402764 & 358395). Bruce Neal is also supported by a Career Development Award from the National Heart Foundation of Australia. PROGRESS was funded by grants from Servier, the Health Research Council of New Zealand, and the National Health and Medical Research Council of Australia. Sponsors did not influence: design and conduct of the study; collection, management, analysis and interpretation of data; preparation and approval of manuscript.

Thank-you: We thank Dr Nicholas Cross and Dr Peter Grimison for their helpful comments on earlier drafts of this paper, the PROGRESS collaboration for providing data from the PROGRESS study, and Dr Toshiharu Niomiya for helping with the transfer of data.

Table 1 Sources of variation in measured outcomes in randomised controlled trials

Source	Description
Between treatments	Variation caused by average effect of treatment
Between patients	Variation caused by differences between patients unrelated to treatment
Within patients	Variation caused by measurement variability
Between patients in treatment effect	Variation caused by differences between patients in effect of treatment

Adapted from (4)

Table 2: Predictors of Systolic Blood Pressure in Final Models*

		Perindopril vs Single placebo		Perindopril/Indapamide vs Double Placebo		
Variable	**Parameter**	**Estimate of effect (Standard Error)**	**Likelihood Ratio test†**	**Estimate of effect (Standard Error)**		**Likelihood Ratio test†**
Fixed Effect						
Treatment (active vs placebo)	Mean (mmHg)	-5.9 (0.5)	χ^2=154.6, 1df, p<0.001	Baseline SBP‡	Treatment Estimate	$\chi^2_{treatment}$ =684.1, 1df, p<0.001
				<140	-9.5	
				140-145	-10.9	$\chi^2_{treatment\times baseline}$ =29.5, 1df, p<0.001
				145-150	-12.2	
				>150	-13.6	
Random Effects¤						
Intercept (one month after treatment started)	Mean (mmHg)	137.6 (0.5)	N/A	142.4 (0.4)		N/A
	Between person Variance ($mmHg^2$)	103.7 (5.8)	χ^2= 1053.3, 2df, p<0.001	119.2 (5.2)		χ^2=1810.2, 2df, p<0.001
Time (months)	Mean (mmHg)	0.15 (0.06)	χ^2=5.8,1df, p=0.02	0.22 (0.1)		χ^2=15.6,1df, p=<0.001
	Between person Variance ($mmHg^2$)	1.7 (0.4)	χ^2=25.1, 2df, p<0.001	2.3 (0.3)		χ^2=71.2, 2df, p<0.001
Residual						
	Within person Variance ($mmHg^2$)	109.8 (3.1)	N/A	105.7 (2.5)		N/A

* All models adjusted for baseline blood pressure, age, gender, other anti-hypertension treatment and presence of left ventricular hypertrophy on baseline electrocardiogram.
†Likelihood ratios calculated by comparing the model that included all covariates to a model without the covariate in question
‡Baseline blood pressure adjusted to reflect average blood pressure levels found during follow up in the placebo group. Baseline blood pressure and its interaction with treatment were fitted as continuous variables for significance testing. Estimates of treatment effects by baseline blood pressure are from bivariate response model.
¤ Mean and variance refers to distribution of random effects

REFERENCES

1. Glasziou PP, Irwig L, Mant D. Monitoring in chronic disease: a rational approach. BMJ. 2005;330:644-8.
2. Bell KJL, Irwig L, Craig JC, Macaskill P. Use of randomized trials to decide when to monitor response to new treatment. BMJ. 2008;336:361-5.
3. Morton V, Torgenson D. Regression to the mean: treatment effect without the intervention. Journal of Evaluation in Clinical Practice. 2004;11(1):59.
4. Senn S. Individual therapy: New dawn or false dawn? Drug Information Journal. 2001;35(4):1479-94.
5. Rothwell PM, Coull AJ, Silver LE, Fairhead JF, Giles MF, Lovelock CE, et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). Lancet. 2005;366:1773–83.
6. PROGRESS, Collaboration, Group. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. Lancet. 2001;358:1033-41.
7. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Statistics in Medicine. 1989;8:1051-69.
8. Kannel WB, Dawber TR, Sorlie P, Wolf PA. Components of Blood Pressure and Risk of Atherothrombotic Brain Infarction: The Framingham Study. Stroke. 1976;7:327-31.
9. Stamler S, Neaton JD, Wentworth DN. Blood Pressure (Systolic and Diastolic) and Risk of Fatal Coronary Heart Disease. Hypertension. 1989 1989;13(Suppl 1):2-12.
10. Flack JM, Gardin JM, Yunis C, Liu K. Static and pulsatile blood pressure correlates of left ventricular structure and function in black and white young adults: The CARDIA study Am Heart J 1999;138(5):856-64.
11. Morton V, Torgenson D. Effect of regression to the mean on decision making in health care. BMJ. 2003;326:1083-4.
12. Task Force for the Management of Arterial Hypertension of the ESH and of the ESC. 2007 Guidelines for the Management of Arterial Hypertension. Journal of Hypertension. 2007;25:1105-87.
13. Chobanian A, Bakris G, Black H, Cushman W, Green L, Izzo J, et al. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension. 2003;42:1206-52.
14. Law MR, Wald NJ. Risk factor thresholds: their existence under scrutiny. BMJ. 2002;324:1570-6.
15. The Asia Pacific cohort Studies Collaboration. Blood pressure and cardiovascular disease in the Asia Pacific region. Journal of Hypertension 2003;21:707–16.
16. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, et al. Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335:827-38.
17. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. 2001;357(9263):1191.
18. Goldstein H, Browne W, Rabash J. Multilevel modelling of medical data. Statistics in medicine. 2002;21:3291-315.
19. Singer J, Willet J. Exploring Longitudinal Data on change. Applied Longitudinal Data Analysis. New York: Oxford University Press, Inc.; 2003. p. 42-4.

APPENDIX

Using Mixed Models to decide on need for initial response monitoring

Mixed models are a powerful method for analysing data from longitudinal studies, in which there are multiple measurements on each subject (18, 19). This approach allows explicit modelling of the within and between person variation in the intermediate outcome, while also taking into account the correlation between measurements taken on the same individual. In this type of analysis, predictive factors may be fitted to have the same effect for everyone, in which case they are said to be 'fixed effects' (these effects are equivalent to the estimated effects in ordinary least squares regression models). Alternatively, the predictive factors may be fitted to have effects which differ between individuals, in which case they are said to have 'random effects' (these effects are not allowed for in ordinary least squares regression models).

We can conceptualise a series of these models where there is the same treatment effect for everyone (treatment has a fixed effect) or between person differences in treatment effect (treatment has random effects). In order of increasing complexity, these alternative models are outlined below and graphically illustrated in Appendix Figure using hypothetical data for ten individuals. In this figure, time point 1 is the first measurement taken after treatment is initiated. Dashed lines are used for patients on placebo and solid lines are used for those on active treatment. The two thick lines are the mean changes in intermediate outcome over time for placebo and active treatment groups, while the thin lines represent individuals' changes in intermediate outcome over time.

In the basic mixed model (Model 0: random intercept, no treatment effect), the intermediate outcome is predicted by a number of factors (such as time, age, gender, pre-trial intermediate outcome measurement) but not by treatment: treatment has no effect. The other predictors have the same effect for everyone (i.e. they are fixed effects). This model also includes a term for between person variation in the first measurement (random intercept). The random intercept represents

variation in the first measurement between individuals beyond that explained by the fixed effect predictors.

Model 0, Figure 1 shows variation between the individuals in the intermediate outcome at Time 1. There is a fixed effect of time on the intermediate outcome, so the trajectories are all parallel. The distribution for individuals on active treatment is the same as that for individuals on placebo reflecting the absence of any treatment effect.

The random intercept allows for a more realistic representation of the data than conventional models where there is only one value of intercept for everyone. There is also a residual error term which represents the random variation seen within an individual over time. It is important to include all statistically significant fixed effects in the basic mixed model so as to minimise the variability in outcome that is unexplained. This increases the power of analysis to detect random effects added in subsequent models and gives better estimates of the random effects' variance(s).

Model 1 is identical to Model 0 except that treatment now has an effect assumed to be the same for everyone (Model 1: random intercept, fixed treatment effect). Model 1, Figure 1 shows a downward shift in the distribution of the intermediate outcome at Time 1 for individuals on active treatment compared to those on placebo. The variances of the two distributions are the same, reflecting the fact that the treatment effect is the same for everyone. There remains a fixed effect of time on the intermediate outcome for both treatment groups, and all trajectories are again parallel.

Model 1 can be extended to include a term for between person variation in change over time to account for the fact that individuals may differ in how their measurements change over time (Model 2: random intercept and random time effects, fixed treatment effect). In Model 2, Figure 1 there is again a downward shift in the Time 1 distribution for those on active treatment compared with those

on placebo. Again both groups have the same variance. However whereas in the previous models time had a fixed effect, in this model individuals differ in how the intermediate outcome changes with time. On average there is the same upward drift in the intermediate outcome between Time 1 and Time 3 as there was in Model 0 and Model 1, but there is now some variation in change between individuals (for example two people in each group have downward sloping trajectories). The effects of time are the same for both active treatment and placebo groups. Treatment effect is again the same for everyone.

This model can be further extended to include a term for between person variation in treatment effect on the first measurement (Model 3: random intercept, random time effects and random treatment effects). Model 3, Figure 1 shows an increased variance in the intermediate outcome distribution at Time 1 for those on active treatment compared to those on placebo. This reflects between person differences in the effects of treatment. Once the effects of treatment at Time 1 are accounted for, there are no further differences between individuals on active treatment and individuals on placebo over time.

If the effect of treatment is thought to change over time in a way that is the same for everyone, a treatment time interaction may be added as a fixed effect (Model 4: random intercept, random time effects, fixed treatment effect and fixed treatment time interaction). In Model 4, Figure 1 the distance between the mean intermediate outcome level in the placebo group and the active treatment group is larger at Time 3 compared to Time 1. This reflects an increasing effect of treatment over time. The variability between individuals around the two mean trajectories is the same. This means that at any particular time point the effect of treatment is the same for everyone.

If the effect of treatment is thought to change over time in a way that differs between individuals, a treatment time interaction may be added as random effects (Model 5: random intercept, random

time effects and random treatment time interactions). In Model 5, Figure 1 there is again an increasing effect of treatment over time. However, there is now more variability in individuals' changes in the active treatment group than there is in the placebo group. This means there are differences between individuals in how the effects of treatment change over time.
The interactions in Models 4 and 5 can also be thought of as effect modification of treatment effect by time.

The alternative models can be compared using likelihood ratio tests to see which model provides the best fit for the data. Figure 2 provides a schematic overview of the model fitting process. The usual strategy for mixed model fitting is to start with the least complex model (Model 0) and use forward selection to add parameters that are needed until a final model is reached.

As can be seen in Figure 2, the model comparisons for our theoretical Models 1-5 do not follow a simple numerical progression. Rather, there are two branches to the model building. If treatment effect stays constant over time, the model comparisons are between Models 0, 1, 2 and 3. If treatment effect changes over time, the model comparisons are between Models 0, 1, 2, 4 and 5.

If Model 0 has the best fit, then treatment can be assumed to have no effect.
If Models 1,2 or 4 have the best fit then treatment can be assumed to have the same effect for everyone. For models 1 and 2 the initial treatment effect is maintained over time (with an underlying change in the intermediate outcome over time that is uniform for Model 1 and variable for Model 2). If Model 4 has the best fit then the treatment effect changes over time, but at any one time the effect is the same for everyone.

To decide on whether the effect of treatment differs between individuals, Model 3 is compared with Model 2 or Model 5 is compared with Model 4. If Model 3 has a better fit than Model 2, then the

immediate effects of treatment differ between individuals. If Model 5 has a better fit than Model 4, then the change in treatment effect over time differs between individuals.

Likelihood Ratio Testing using mixed models

The likelihood ratio test compares the difference in -2Log likelihood for two models to a chi-squared distribution. The degrees of freedom for the test reflects the number of additional parameters in the more complex model compared with the more simple one. The test is only valid if the models are 'nested' – i.e. the variables being tested in one model are a subset of those in the other model.

Likelihood ratio test to compare nested models for variance components may be based on maximum likelihood (or iterative generalised least squares) or restricted maximum likelihood (or restricted generalised least squares). Likelihood ratio tests to compare nested models for fixed effects should only be based on maximum likelihood (or iterative generalised least squares)(19).

Assumptions for using Mixed Models

There are three main assumptions invoked for using mixed models in the standard form

1.Functional Form. In the simplest type of mixed model (as presented in this paper) one assumes that there is a linear relationship between the intercept for the intermediate outcome and each of the individual level predictors and between the change over time for the intermediate outcome and each of the individual level predictors. One also assumes a linear relationship between the level of intermediate outcome and time for each individual.

2. Normality. One assumes that the distribution of the Random Effects, as well as the within person residuals is normal.

3. Homoscedasticity. One assumes that the distribution of the Random Effects do not vary across different individual level predictors. One also assumes that the distribution of within person residuals do not vary over time.

If the assumptions are violated there are a number of possible solutions. Transforming either the intermediate outcome or the time variable is the simplest solution (e.g. logarithmic transformation). If this fails to resolve the problem(s) then another option may be to add higher order terms for the time variable and adopt a 'polynomial growth model' (e.g. add a quadratic term). If there is still concern assumptions are being violated then adopting a nonlinear model may be the solution (e.g. 'logistic individual growth curve') or a model with a more complex error structure (e.g. 'Autoregressive').

In the blood pressure example we used in this paper, all the assumptions were met with one minor exception. When we assessed the homoscedasticity of random intercepts by quartiles of baseline

blood pressure we found similar variances for the first three quartiles and a moderately increased variance for the highest quartile (for both single and dual therapy groups).

Applying estimates of treatment effect(s) to patients

In our example, there was no variation in treatment effect for either perindopril or perindopril/indapamide which meant the application of the trial population results to individual patients is relatively straight forward. In other situations, we may find evidence of between person variation in treatment effect. In this case, the ease of application to individual patients will vary depending on the extent of any correlation between treatment effect and either baseline level or background change in intermediate outcome. In the simpler case where there is no substantial correlation, the estimate of variance for treatment effect can be used to calculate a 95% distribution of treatment effect. Individuals with a baseline level of intermediate outcome that is lower than [treatment target level + the 2.5^{th} percentile of treatment effect] are very likely to meet the target, and initial response monitoring is not needed. Individuals with a baseline level higher than [treatment target level + the 97.5^{th} percentile of treatment effect] are unlikely to meet the target without additional therapy, and initial response monitoring is not needed. It is only when individuals' baseline levels of intermediate outcome lie between these two points that there is uncertainty whether the treatment target level will be met and initial response monitoring is needed.

Even if information on variability of treatment effect on the intermediate outcome is not available, randomized trials may help inform decisions on the need for initial response monitoring. If trialists present the mean change and variance of change in intermediate outcome from baseline, then inferences can be made about the treatment effect on the intermediate outcome using a simple method of analysis which compares summary variances of the placebo and active treatment groups(2). This method makes the following important assumptions: data are normally distributed

with no relationship between variability and level of intermediate outcome, and there is no substantial correlation between the treatment effect and background change in intermediate outcome.

For example, comparing the standard deviations presented in the Appendix Table for change in systolic blood pressure between baseline and one month post-randomization in the single therapy group, we find weak evidence of between person variation in the effects of perindopril alone (17.0 vs 16.2 mmHg, $F_{1227,1244}=1.09$, p=0.06). The estimated 95% distribution of treatment effect ranges from a decrease in systolic blood pressure of 3.7 mmHg to a decrease of 7.3 mmHg. Comparing variances for change in systolic blood pressure between baseline and one month post-randomization in the dual therapy group, we find no evidence of between person variation in the effects of perindopril in combination with indapamide ($F_{1735,1744}=1.04$, p=0.20) and the treatment effect for all patients is estimated as a decrease in systolic blood pressure of 11.5 mmHg (the mean treatment effect).

Because the simple variance comparison approach only uses data from two time points (and there are multiple potential comparisons that could be made using different pairs of time points), it may be more likely to find spurious variation in treatment effect than the more complete analysis possible with mixed models. Other weaknesses of this method are that the assumptions may not always be reasonable (although they were found to hold for the PROGRESS data) and there is no allowance made for other predictors (and importantly, treatment effect modifiers) when estimating the treatment effect. For these reasons, inferences about treatment effects on intermediate outcomes made on the basis of a mixed model analysis on individual patient data are preferred whenever possible.

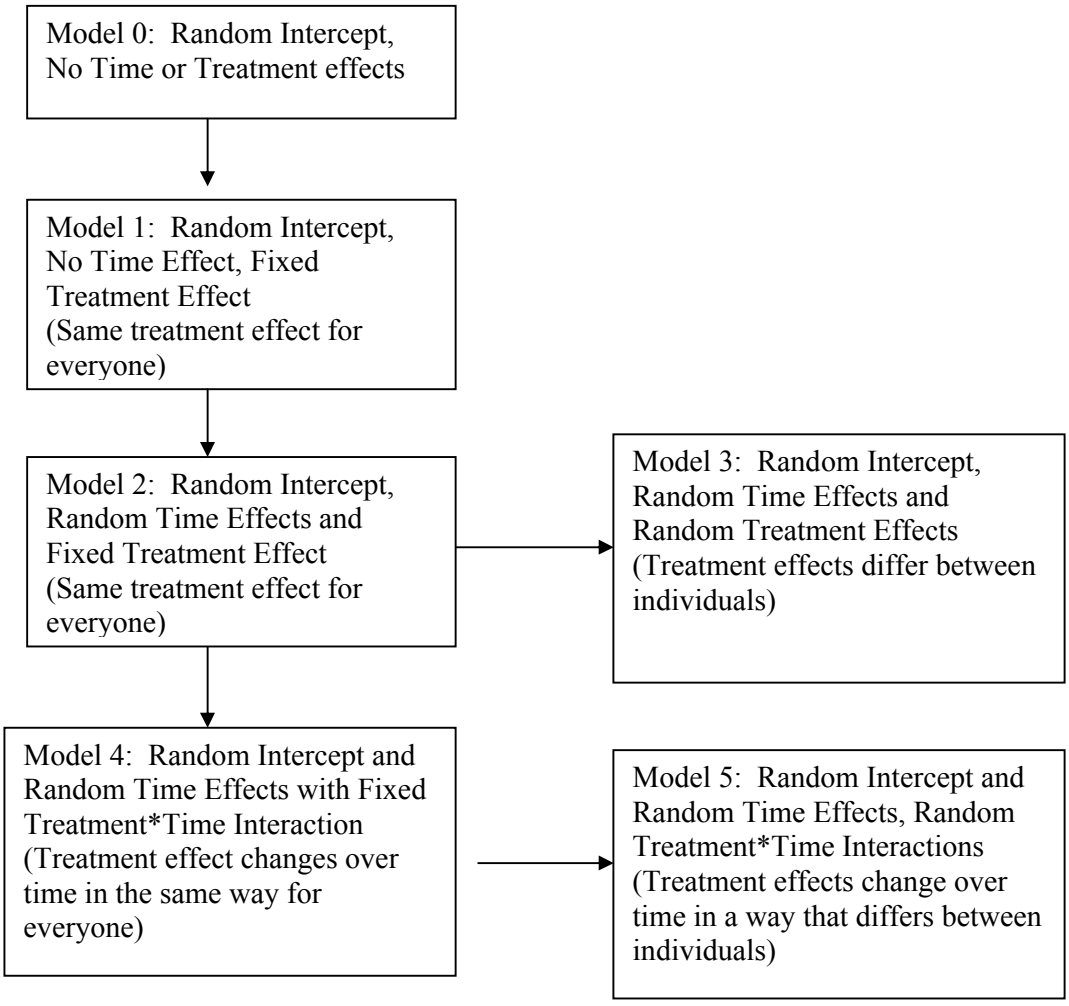
Appendix Table: Variation in Systolic Blood Pressure measurement during the first 6 months of PROGRESS trial.

			Systolic Blood Pressure Measurement (mm Hg)							
			1 month pre-randomization		1 month post-randomization		3 months post-randomization		6 months post-randomization	
			Level	***Change****	Level	***Change****	Level	***Change****	Level	***Change****
Single Therapy	Placebo N = 1245†	Mean	143.4	-	139.0	***4.4***	140.2	***3.2***	140.2	***3.1***
		Standard Deviation	18.2	-	18.0	***16.2***	18.0	***16.7***	17.5	***17.5***
	Perindopril N = 1228†	Mean	144.1	-	134.0	***9.9***	133.9	***9.9***	134.3	***9.5***
		Standard Deviation	19.1	-	18.3	***17.0***	17.9	***17.3***	17.8	***18.1***
Dual Therapy	Placebo N = 1745†	Mean	149.5	-	144.6	***4.8***	145.1	***4.2***	145.7	***3.5***
		Standard Deviation	19.2	-	18.3	***17.4***	18.2	***18.1***	18.4	***18.5***
	Perindopril & Indapamide N = 1735†	Mean	149.1	-	132.8	***16.3***	133.3	***15.8***	133.8	***15.2***
		Standard Deviation	18.6	-	16.3	***17.7***	16.6	***18.7***	16.7	***19.1***

* Change is from measurement one month prior to randomization
† N for change at one month, three months and six months differ slightly

Figure(s)

Figure 1: Schematic overview of Model fitting



Figure(s)

Appendix Figure: Key Scenarios for Initial Response to Treatment

Mixed models based on hypothetical data for ten patients