

RESEARCH

Estimating treatment effects for individual patients based on the results of randomised clinical trials

 OPEN ACCESS

Johannes A N Dorresteijn *epidemiologist and medical doctor*¹, Frank L J Visseren *professor of vascular medicine, epidemiologist, and internist*¹, Paul M Ridker *Eugene Braunwald professor of medicine, epidemiologist, and cardiologist*², Annemarie M J Wassink *internist and postdoctoral researcher*¹, Nina P Paynter *assistant professor of epidemiology*², Ewout W Steyerberg *professor of medical decision making, and methodologist*³, Yolanda van der Graaf *professor of epidemiology and imaging*⁴, Nancy R Cook *associate professor of biostatistics and epidemiology*²

¹Department of Vascular Medicine, University Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht, Netherlands; ²Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; ³Department of Public Health, Erasmus Medical Center, Rotterdam, Netherlands; ⁴Julius Center for Health Sciences and Primary Care, Utrecht, Netherlands

Abstract

Objectives To predict treatment effects for individual patients based on data from randomised trials, taking rosuvastatin treatment in the primary prevention of cardiovascular disease as an example, and to evaluate the net benefit of making treatment decisions for individual patients based on a predicted absolute treatment effect.

Setting As an example, data were used from the Justification for the Use of Statins in Prevention (JUPITER) trial, a randomised controlled trial evaluating the effect of rosuvastatin 20 mg daily versus placebo on the occurrence of cardiovascular events (myocardial infarction, stroke, arterial revascularisation, admission to hospital for unstable angina, or death from cardiovascular causes).

Population 17 802 healthy men and women who had low density lipoprotein cholesterol levels of less than 3.4 mmol/L and high sensitivity C reactive protein levels of 2.0 mg/L or more.

Methods Data from the Justification for the Use of Statins in Prevention trial were used to predict rosuvastatin treatment effect for individual patients based on existing risk scores (Framingham and Reynolds) and on a newly developed prediction model. We compared the net benefit of prediction based rosuvastatin treatment (selective treatment of patients whose predicted treatment effect exceeds a decision threshold) with the net benefit of treating either everyone or no one.

Results The median predicted 10 year absolute risk reduction for cardiovascular events was 4.4% (interquartile range 2.6-7.0%) based on the Framingham risk score, 4.2% (2.5-7.1%) based on the Reynolds

score, and 3.9% (2.5-6.1%) based on the newly developed model (optimal fit model). Prediction based treatment was associated with more net benefit than treating everyone or no one, provided that the decision threshold was between 2% and 7%, and thus that the number willing to treat (NWT) to prevent one cardiovascular event over 10 years was between 15 and 50.

Conclusions Data from randomised trials can be used to predict treatment effect in terms of absolute risk reduction for individual patients, based on a newly developed model or, if available, existing risk scores. The value of such prediction of treatment effect for medical decision making is conditional on the NWT to prevent one outcome event.

Trial registration number Clinicaltrials.gov NCT00239681.

Introduction

Usually the results of trials are implemented in clinical practice by either treating all patients (in the case of a positive trial result) or treating no one (in the case of a negative trial result), expecting the treatment effect for every patient to be similar to the average treatment effect in the original trial. Clinicians intuitively know that this idea is oversimplified because in reality some patients benefit more than average from treatment, whereas others do not or may even be harmed.¹⁻⁶

The direct translation of trial results to individual patients in clinical practice is, however, complicated by some important limitations. The treatment effects of randomised trials are

Correspondence to: F L J Visseren F.L.J.Visseren@umcutrecht.nl

Extra material supplied by the author (see <http://www.bmj.com/content/343/bmj.d5888/suppl/DC1>)

Appendix 1: development of optimal fit model

Appendix 2: example calculation of net benefit assessment method

Appendix 3: application of treatment effect prediction methods to other datasets

typically expressed in terms of relative risks or hazard ratios at a group level—that is, treatment versus control. Yet treatment that is associated with a considerable reduction in relative risk will still result in a modest absolute effect when the incidence rate of the disease is low. Absolute risk reduction is usually more informative because it combines the relative risk reduction and the incidence rate of the disease outcome.^{1,2} The absolute risk reduction is sometimes expressed in trial reports as the number needed to treat (NNT). Still, implicit in the use of estimates at group level is that all patients are at average risk and all have the same likelihood of response to treatment. Usually at least one of these two assumptions is untrue because the expected absolute risk reduction resulting from treatment often depends on the characteristics of individual patients.¹⁻⁶

Although prespecified subgroup analyses take a step towards identifying those characteristics of patients that modify the treatment effect, some important limitations are retained. In subgroup analyses the study cohort is typically divided according to the presence or absence of a single patient characteristic such as diabetes, age (below or above a certain limit), or sex, and the effect of the intervention is presented accordingly. However, these univariable analyses do not fully incorporate all available patient characteristics and are less well powered but still return relative, rather than absolute, average effect measures at a group level.²⁻⁴

A more comprehensive approach towards making well informed decisions about treatment is to predict the treatment effect for individual patients based on all relevant characteristics together.^{1-5,7} Although not yet widely appreciated, data from randomised controlled trials usually provide an opportunity to develop models for the prediction of a treatment effect on the basis of individual patient characteristics.^{1,5} Such models can enable clinicians to estimate a treatment effect for individual patients in terms of absolute risk reduction for the disease of interest. This can be done before the start of intended treatment, and therefore decisions about treatment can be based on such predictions.⁴ Moreover, individualised predictions of treatment effect provide an opportunity to determine which implications the results of randomised trials should have in clinical practice.⁶ Making treatment decisions on the basis of a predicted treatment effect for individual patients may in some situations result in more net benefit on a group level than treating all patients (in the case of a positive trial result) or treating no one (in the case of a negative trial result). Although this approach is occasionally used in the research of cancer⁸⁻¹⁰ and cardiovascular disease,¹¹⁻¹³ the full potential has yet to be recognised by both researchers and clinicians.

We developed and evaluated methods for predicting treatment effect using rosuvastatin in individual patients in a primary prevention setting based on data from the Justification for the Use of Statins in Prevention trial.¹⁴ This study was a randomised, double blind, placebo controlled, multicentre trial that showed on average a 44% relative risk reduction in major vascular events in those treated with rosuvastatin. As the trial was carried out in a primary prevention cohort at moderate absolute risk for cardiovascular disease, the overall treatment effect was modest for average absolute risk reduction. Therefore the trial represents a typical situation in which the prediction of treatment effect can be used to identify those who will benefit from treatment. We predicted treatment effects for individual patients based on data from randomised trials, taking rosuvastatin treatment in primary prevention of cardiovascular disease as an example, and evaluated the net benefit of making treatment decisions for individual patients based on predicted absolute treatment effect.

Methods

The design, rationale, and outcomes of the Justification for the Use of Statins in Prevention trial are described in detail elsewhere.¹⁴⁻¹⁶ Briefly, the trial evaluated the effect of rosuvastatin 20 mg daily compared with placebo on the occurrence of myocardial infarction, stroke, arterial revascularisation, admission to hospital for unstable angina, or death from cardiovascular causes among 17 802 apparently healthy men and women who had low density lipoprotein cholesterol levels of less than 3.4 mmol/L (130 mg/dL) and high sensitivity C reactive protein levels of 2.0 mg/L or more. After a median follow-up of 1.9 years the hazard ratio for occurrence of the primary end point was 0.56 (95% confidence interval 0.46 to 0.69), favouring rosuvastatin.¹⁴ Univariable subgroup analyses (for example, for age, sex, smoking status, ethnicity, and Framingham risk score) showed no significant deviations from this effect size.

Estimating treatment effects for individual patients

We estimated the baseline 10 year risk for cardiovascular events (myocardial infarction, stroke, arterial revascularisation, admission to hospital for unstable angina, or death from cardiovascular causes) for individual patients if untreated, using the existing Framingham risk score¹⁷ and the Reynolds risk score, without prior updating or refitting of the coefficients (see box).^{18,19} Based on the assumption that treatment effect increases linearly with baseline risk (fig 1), we estimated the patient's residual risk when given treatment by multiplying baseline risk by the overall relative effect measure (relative risk or hazard ratio) from the original trial report. Consequently the estimated absolute risk reduction achieved by treatment with rosuvastatin for 10 years (10 year treatment effect) is equal to the difference between these two [individual treatment effect=(1–overall relative effect measure from trial)×baseline risk derived from an existing prediction model].

Alternatively we developed a new prediction model (optimal fit model) based on trial data only (see web extra appendix 1). A theoretical advantage of this strategy over using existing risk scores is that the model may be better calibrated to the population of interest. Furthermore, such a model is not based on the assumption that treatment effect increases linearly with baseline risk: modification of the treatment effect by patient characteristics can be tested and, if significant, included in the model. Importantly, even in the absence of subgroup effects defined by univariable characteristics, a multivariable adjusted prediction model may contain such modifications of treatment effect. In situations where no existing prediction models are available, developing a new prediction model may be the only option.

Performance of the prediction models

We assessed the calibration of the predictions based on the Framingham risk score, the Reynolds risk score, and the optimal fit model. To do this we plotted the observed Kaplan-Meier survival for cardiovascular events at two years within 10ths of the predicted survival against the mean predicted two year survival of each 10th and by the P value derived from the Hosmer-Lemeshow test. Based on the assumption that the hazard rate is constant and thus survival is exponential over time we derived two year risk estimates of the Framingham risk score and the Reynolds risk score from the 10 year predicted risks. Discrimination was assessed by calculation of the C statistic.

Prediction of treatment effect for individual patients

10 year treatment effect (absolute risk reduction)=baseline risk without treatment–residual risk with rosuvastatin treatment

Baseline 10 year absolute risk for cardiovascular events (%) without treatment

Framingham risk score based method: risk as calculated from the Framingham risk score, published in the Adult Treatment Panel III guidelines¹⁷

Reynolds risk score based method: risk as calculated from the Reynolds risk score, derived for women from the Women's Health Study¹⁸ and for men from the Physicians Health Study¹⁹

Optimal fit model*:

$(1 - 0.985433^{(5 \times \exp[B])}) \times 100\%$, where:

$B = 0.09379363 \times \text{age in years} + 3.34656382 \text{ (if male)} - 0.03698750 \times \text{age in years (if male)} + 0.81823698 \text{ (if current smoker)} + 0.54045383 \text{ (if using blood pressure lowering drugs)} + 0.60281674 \text{ (if family history of premature coronary heart disease)} - 6.9932$

Residual 10 year absolute risk for cardiovascular events (%) with rosuvastatin treatment

Framingham risk score: $0.56 \times \text{baseline 10 year absolute risk for cardiovascular events (%) without treatment}$

Reynolds risk score: $0.56 \times \text{baseline 10 year absolute risk for cardiovascular events (%) without treatment}$

Optimal fit model*:

$(1 - 0.985433^{(5 \times \exp[B])}) \times 100\%$, where:

$B = 0.09379363 \times \text{age in years} + 3.34656382 \text{ (if male)} - 0.03698750 \times \text{age in years (if male)} + 0.81823698 \text{ (if current smoker)} + 0.54045383 \text{ (if using blood pressure lowering drugs)} + 0.00932154 \text{ (if family history of premature coronary heart disease)} - 7.484613$

*Treatment effect of rosuvastatin expressed in optimal fit model for residual risk as different coefficient for family history of premature coronary heart disease and different constant subtraction factor

Assessment of net benefit

We determined the value of individualised predictions of treatment effect for medical decision making using the previously described net benefit assessment method.⁵ This method calculates the impact of different treatment strategies using the event rates and the treatment rates in study participants. We considered the following approaches to rosuvastatin treatment of patients without previous vascular disease or diabetes mellitus and low levels of low density lipoprotein cholesterol: treat all, treat no one, or treat based on prediction—that is, the selective treatment of patients whose predicted treatment effect exceeds a decision threshold. To facilitate clinical interpretation we extrapolated the observed event rates at two years to 10 years (see web extra appendix 2 for an explanation of the net benefit assessment method, with a sample calculation).

The decision threshold used for prediction based treatment represents the estimated harms of treatment, such as excess risk for adverse reactions, monetary costs, and the discomfort of sustaining treatment (fig 1). Notably, estimation of the harms of treatment is also needed to calculate and interpret the net benefit of one treatment strategy over another. One research team proposed to estimate the decision threshold by weighing the harms of treatment against the harms of an outcome event.^{5 20} For example, if the harms of a cardiovascular event are assumed to be 20 times worse than those of rosuvastatin treatment for 10 years, the appropriate decision threshold is 5% (1 divided by 20), and only those individuals whose predicted 10 year absolute treatment effect exceeds 5% should be advised to start rosuvastatin treatment. Usually, however, the level of the decision threshold is not discussed, but rather the maximum acceptable number needed to treat (NNT). For this purpose we propose to rename the NNT that is associated with clinical equipoise as the number willing to treat (NWT). If rosuvastatin treatment of 20 people for 10 years is assumed to be exactly as harmful as one outcome event (for example, a case of myocardial infarction), doctors would be willing to treat up to 20 patients to prevent one event, therefore the NWT is 20. The NWT is the

inverse of the decision threshold but generally more intuitive to clinicians.

The main harms resulting from rosuvastatin treatment include monetary costs and the discomfort of taking the drug daily, since multiple trials, including the Justification for the Use of Statins in Prevention trial, show that treatment with rosuvastatin 20 mg daily is not associated with an increased risk of adverse reactions, except for a small increase in the probability of newly diagnosed diabetes. Moreover, particularly among those with impaired fasting glucose (the group most likely to develop diabetes), large risk reductions in macrovascular disease are observed. None the less, the appropriate decision threshold is subjective and may differ between countries and over time. For this reason we did not make any assumptions about the severity of the harms associated with treatment but calculated the net benefit for a range of values of NWT. To graphically represent the net benefit assessment results for this range of values of NWT in a decision curve, we applied locally weighted scatter plot smoothing.^{5 20}

Analyses were done using open source statistical software, R version 2.10.0 (R Foundation for Statistical Computing, www.R-project.org).

Results

Table 1↓ shows the baseline clinical characteristics of the Justification for the Use of Statins in Prevention cohort. Overall, 140 events were observed in the rosuvastatin treated group (8853 participants) and 251 in the placebo treated group (8857 participants). Data related to 92 participants were excluded from the analyses owing to missing data for one or more predictor variables. The Kaplan-Meier survival curves of both treatment groups had been published before and did not show any remarkable aberrations at the two year follow-up.¹⁴

The box shows the models used for the prediction of the treatment effect from using rosuvastatin. The final optimal fit model contains terms for age, sex, age-sex interaction, smoking, blood pressure lowering drugs, and family history of premature

myocardial infarction. Importantly, the study population was selected to have low density lipoprotein cholesterol levels of less than 3.4 mmol/L (130 mg/dL) and high sensitivity C reactive protein levels of 2.0 mg/L or more. This might have contributed to the fact that neither lipids nor high sensitivity C reactive protein were selected in the final optimal fit model. The model contained one treatment-covariate interaction: a family history of premature coronary heart disease was the only patient characteristic that affected the rosuvastatin treatment effect.

Calibration and discrimination of all three prediction methods were moderate. The C statistic of the Framingham model based predictions in the Justification for the Use of Statins in Prevention cohort was 0.65 (95% confidence interval 0.62 to 0.68), almost equal to the C statistic of the predictions based on the Reynolds model (0.66, 0.63 to 0.69). As expected, the optimal fit model performed a little better because discrimination was tested in the same cohort from which it was developed. The C statistic was 0.71 (0.68 to 0.74). The Reynolds risk score somewhat overestimated risk for cardiovascular events within the highest 10th of predicted risk, resulting in some lack of fit as evidenced by a significant Hosmer-Lemeshow statistic (fig 2).

The Framingham risk score, the Reynolds risk score, and the optimal fit model can be applied to calculate the predicted 10 year treatment effect of using rosuvastatin for two patient scenarios (table 2). Likewise, the 10 year treatment effect was predicted for every individual within the Justification for the Use of Statins in Prevention cohort, the distributions of which are presented in figure 3. Coloured bars indicate how the predicted 10 year treatment effect of the two patient scenarios relate to that of all participants within the study cohort. The median predicted 10 year absolute risk reduction for all participants in the Justification for the Use of Statins in Prevention trial according to the Framingham based model was 4.4% (interquartile range 2.6-7.0%), the Reynolds based model was 4.2% (2.5-7.1%), and the optimal fit model was 3.9% (2.5-6.1%).

Net benefit assessment

Web extra appendix 2 shows an example of a net benefit calculation. In this example the net benefit of prediction based treatment using the Framingham risk score is compared with the net benefit of treating all patients, assuming that 20 is the appropriate NWT. Similar calculations were carried out for a range of values for NWT and also for prediction based treatment using the Reynolds risk score and optimal fit model. The net benefit of treating no one serves as a reference and is equal to zero. The net benefit of the other strategies represents the resulting decrease in the event rate minus the cost of treatment.

Treatment of all patients is more beneficial than treating no one if the NWT is high (little harm, treat even at low risk) but not if the NWT is low (considerable harm, treat at high risk only; table 3 and fig 4). If the NWT is about 20, then the benefits of treating all patients and treating no one are equivalent (zero). Prediction based treatment is associated with equal net benefit as treating all patients for high values of NWT and the net benefit curves of prediction based treatment converge to zero (treat no one) for lower values of NWT (fig 4). For a range of NWT (between about 15 and 50), prediction based treatment is the preferred strategy of choice. Notably, the net benefits of prediction based treatment based on the optimal fit model and the Framingham or Reynolds risk score were similar. Therefore the assumption that treatment effect increases linearly with

baseline cardiovascular risk (fig 1) appears to be true in this example.

Interpreting the size of the net benefit advantage of one strategy over another is complex. One study proposed to imagine that the same net benefit value was achieved by an infallible prediction model that identifies a certain percentage of people as being not at risk for the outcome and thus not in need of treatment. Such a fictitious infallible prediction model reduces the treatment rate without increasing the event rate.⁵ If this method is applied to the present data this means that for a NWT of 30, the net benefit advantage of prediction based treatment (mean net benefit over all three methods is 0.0228) over treating all patients (net benefit is 0.0165), is equivalent to that of treatment by a fictitious infallible prediction model that reduces the treatment rate by 19% without increasing the event rate. Likewise, if the NWT is 20, the mean advantage of prediction based treatment over treating all patients is equal to a 16% reduction of the treatment rate.

Translation to clinical practice

Figure 5 illustrates how the findings could be translated to clinical practice. Treatment of all patients is the strategy of choice if the 10 year NWT is 50 or more. Treat none is preferable if the 10 year NWT is 15 or fewer. If the NWT is between 15 and 50, prediction based treatment results in most net benefit. Because the three prediction methods resulted in similar net benefit, treatment prediction based on existing risk scores is most appropriate in clinical practice. These risk scores are already externally validated and more easily implemented. This means that if, for example, the 10 year NWT to prevent one cardiovascular event is 20, patients with a baseline (for example, Framingham score or Reynolds score) risk of 11.4% or more (95% confidence interval 9.3% to 16.1%) benefit from treatment. Likewise, if the 10 year NWT is 30, patients with a baseline risk of 7.6% or more (6.2% to 10.8%) benefit from treatment. These findings do not contradict the current guidelines, which also recommend treating those whose risk for cardiovascular events exceeds a certain threshold.²¹ However, our findings do suggest that the optimal treatment threshold may be lower than is often assumed.

Discussion

The direct translation of results of trials to individual patients in clinical practice is often difficult because not all respond to treatment similar to the average patient enrolled in a trial. This is because the effect of treatment often depends on the characteristics of individual patients. In the present study we have shown how data from randomised clinical trials can be used to predict absolute treatment effects for individual patients, taking patient characteristics into account. In addition, we have assessed the added value of such predictions for medical decision making.

Implementation of an individualised prediction of treatment effect in clinical practice is not necessarily complicated. Several prediction rules are already available for estimating baseline risk for vascular events in primary prevention—for example, the Framingham risk score and Reynolds risk score. The example from the Justification for the Use of Statins in Prevention trial shows that estimation of an individual treatment effect can be as easy as multiplying the individual baseline risk, as estimated from the Framingham risk score or the Reynolds risk score, by the average relative treatment effect from the trial report. If, however, risk scores are not yet available in a certain area of medicine, a new prediction model to estimate individual

treatment effect can be developed from the trial data. The methods described in this paper can thus be applied to various medical specialties. Online calculators and integration of prediction models in electronic patient record systems could facilitate the widespread use of prediction of treatment effect in clinical practice. The trial example used in this article also shows that even when discrimination and calibration of a prediction model are moderate, the net benefit of treatment assignment according to prediction can still be superior to both treating all patients within the study domain and treating no one for a certain range of NWT (in this example between about 15 and 50).

Prediction of treatment effect for individual patients may enable doctors to practise individualised medicine in an evidence based manner. It could help to make better informed treatment decisions and perhaps motivate patients to adhere to treatment. Presentation of the net benefit of all possible strategies of treatment assignment for a spectrum of NWT is useful in this respect because the NWT possibly varies with patient and provider preferences. This is especially true when treatment is associated with important adverse reactions. For example, treatment with tissue plasminogen activator for acute myocardial infarction is associated with an increased risk for intracranial haemorrhage that also varies according to individual patient characteristics.^{6 13} If patients have difficulties understanding the concept of risk, the predicted individual treatment effect (expressed in terms of absolute risk reduction) can be expressed as a NNT (the number of similar patients that needs to be treated to prevent one outcome event; table 2), which might be more intuitive, and this can be compared to the appropriate NWT.

Prediction of treatment effect for individual patients might also facilitate the work of practice guideline committees that aim to make well informed decisions about indications for treatment on a group level. When the trial results are presented using the methods presented in this paper, the remaining issue that guideline committees need to focus on is the appropriate NWT. The NWT is estimated by weighing the total harms of treatment (for example, adverse reactions, monetary costs, discomfort of sustaining treatment) against the harms of the outcome event of interest (cardiovascular event). For any given NWT three possible treatment strategies must be considered: treat everyone, treat no one, or treat based on prediction (selective treatment of patients whose predicted treatment effect exceeds a decision threshold). When the NWT is agreed on, the trial results can be used to estimate the net benefit of each strategy (table 3 and fig 4) and to determine the optimal treatment strategy (fig 5). The treatment strategy with the highest net benefit for the appropriate value of NWT results in the most favourable trade-off between treatment rate and event rate. Applying this strategy in clinical practice leads to more selective treatment of patients who will benefit from treatment.

Previously, risk stratified reporting of trial results was proposed as a method for presenting heterogeneity of treatment effects in trials.^{3 7} In line with this, the relative risk and NNT for participants of the Justification for the Use of Statins in Prevention trial within subgroups of estimated baseline risk were published earlier.²²⁻²⁴ Stratified analysis of treatment effects in subgroups of the total study cohort may, however, lead to imprecision owing to loss of statistical power. Moreover, existing risk scores are not available for many diseases, invalidating the risk stratified approach. Also, risk based stratification may still obscure important modification of relative treatment effect that can be discovered or excluded (as in the Justification for the Use of Statins in Prevention trial example) by a multivariate model for predicting treatment effect based

on trial data. Also, the cut-off values for defining subgroups of estimated baseline risk are usually predefined, whereas the methods shown in this paper allow searching for the treatment threshold that is associated with maximum net benefit.

Although data from clinical trials have been used before to predict treatment effects for individual patients, evidence supporting the added value of individualised prediction of treatment effect for clinical practice has been sparse.⁸⁻¹³ Expensive and long lasting impact trials were needed to show the benefit of prediction based treatment.²⁵ In this article we show that the net benefit assessment methods, described previously, provide a more efficient and readily available opportunity for evaluating the potential net benefit of prediction based treatment and for determining implications of contemporary trial results for clinical practice.⁵ This report also shows that the added value of individualised prediction of treatment effect for medical decision making may not be universal but instead is conditional on the NWT.

Limitations and challenges of the study

Limitations of using trial data for individualised predictions of treatment effect generally include short and variable follow-up times, whereas meaningful predictions of cardiovascular event risk usually comprise a 10 year period. This is particularly true for the Justification for the Use of Statins in Prevention trial because the study was discontinued early, but few clinical trials have a follow-up period as long as 10 years either. Thus the predictions and observations usually need to be extrapolated. Furthermore, similar to conventional trial reports, generalisability of the results may be problematic. Trial participants were often selected on the basis of strict eligibility criteria and are healthier and more compliant to treatment than are patients in clinical practice.⁶ In this example, the results apply to patients without manifest vascular disease or diabetes, but additional eligibility criteria of the trial were low levels of low density lipoprotein cholesterol and increased levels of high sensitivity C reactive protein. Hence application of trial based predictions of treatment effect to the general population may be suboptimal. This is especially true for newly fit models because important risk factors (such as low density lipoprotein cholesterol and high sensitivity C reactive protein in our example) may not be included in the prediction model if all trial participants had similar characteristics.

Apart from these practical constraints, many feel reluctant to interpret the implications of subgroup analyses let alone multivariate prediction of treatment effect, because over-accuracy and chance findings may occur.²⁶ Predictions of treatment effect should therefore be based on existing risk scores developed in external data when possible.^{2 7} Yet even when validated risk scores are available, as in our example, developing a new prediction model fit to the trial data can help to confirm the assumption that treatment effect increases linearly with baseline risk (fig 1). Moreover, it should be stressed that the estimated treatment effects in prediction models originating from randomised trials are not subject to confounding bias, because treatment was allocated randomly in the study population. Over-fitting can be minimised by careful and preferably prespecified selection of candidate predictors and shrinkage of the model coefficients when needed. Web extra appendix 3 summarises considerations that need to be taken into account when applying the methods described in this paper to other trial datasets.

Conclusions

Data from randomised trials can be used to predict treatment effect in terms of absolute risk reduction for individual patients before the start of intended treatment. Predictions could be based on existing risk scores, if available, or a newly developed model. The value of such prediction of treatment effect for medical decision making is conditional on the NWT to prevent one outcome event. Prediction based treatment may result in positive net benefit for a range of NWT, even when model calibration and discrimination are moderate. The methods shown in this paper could therefore become a routine part of reporting clinical trials and be used in everyday clinical practice.

Contributors: JAND designed and carried out the data analyses, interpreted the results, and drafted the manuscript. FLJV conceived the research question, designed the data analyses, interpreted the results, and revised the manuscript for important intellectual content. PMR conceived the research question, collected the data, designed the data analyses, interpreted the results, revised the manuscript for important intellectual content, and is guarantor for the validity of the data and analyses. AMJW and NPP designed the data analyses, interpreted the results, and revised the manuscript for important intellectual content. EWS and YvdG conceived the research question, designed the data analyses, interpreted the results, and revised the manuscript for important intellectual content. NRC conceived the research question, collected the data, designed the data analyses, interpreted the results, and revised the manuscript for important intellectual content.

Funding: The Justification for the Use of Statins in Prevention was an investigator initiated trial. The sponsor of the study collected the trial data and monitored the study sites but had no role in the conduct of the analyses or drafting of the report. All statistical analyses were done by the investigators.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: PMR is the principal investigator of the investigator initiated Justification for the Use of Statins in Prevention trial, which was funded by AstraZeneca (Wilmington, Delaware). PMR received grant support from Novartis and Roche; consulting fees from Siemens Medical Systems, ISIS, and Vascular Biogenetics; and is listed as a co-inventor on patents held by the Brigham and Women's Hospital that relate to the use of inflammatory biomarkers in cardiovascular disease that have been licensed to Siemens Medical Systems (Erlangen, Germany) and AstraZeneca. FLJV's department receives grant support from Merck, the Netherlands Organisation for Health Research and Development, and the Catharijne Foundation Utrecht; and speaker fees from Merck and AstraZeneca. JAND, AMJW, NPP, EWS, YvdG, and NRC have no relationships with industry that might have an interest in the submitted work in the previous three years. All authors have no non-financial interests that may be relevant to the submitted work.

Ethical approval: The protocol for the Justification for the Use of Statins in Prevention trial was approved by the local institutional review boards at each participating centre. All study participants provided written informed consent before taking part.

Data sharing: No additional data available.

- 1 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356-9.

- 2 Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol* 2006;6:18.
- 3 Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA* 2007;298:1209-12.
- 4 Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995;345:1616-9.
- 5 Vickers AJ, Kattan MW, Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. *Trials* 2007;8:14.
- 6 Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 2004;82:661-87.
- 7 Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;11:85.
- 8 Gill S, Loprinzi CL, Sargent DJ, Thome SD, Alberts SR, Haller DG, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol* 2004;22:1797-806.
- 9 Loprinzi CL, Thome SD. Understanding the utility of adjuvant systemic therapy for primary breast cancer. *J Clin Oncol* 2001;19:972-9.
- 10 Steyerberg EW, Homs MY, Stokvis A, Essink-Bot ML, Siersema PD. Stent placement or brachytherapy for palliation of dysphagia from esophageal cancer: a prognostic model to guide treatment selection. *Gastrointest Endosc* 2005;62:333-40.
- 11 Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. European Carotid Surgery Trialists' Collaborative Group. *Lancet* 1999;353:2105-10.
- 12 Kent DM, Hayward RA, Griffith JL, Vijan S, Beshansky JR, Califf RM, et al. An independently derived and validated predictive model for selecting patients with myocardial infarction who are likely to benefit from tissue plasminogen activator compared with streptokinase. *Am J Med* 2002;113:104-11.
- 13 Califf RM, Woodlief LH, Harrell FE Jr, Lee KL, White HD, Guerci A, et al. Selection of thrombolytic therapy for individual patients: development of a clinical model. GUSTO-I Investigators. *Am Heart J* 1997;133:630-9.
- 14 Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM Jr, Kastelein JJ, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med* 2008;359:2195-207.
- 15 Ridker PM. Rosuvastatin in the primary prevention of cardiovascular disease among patients with low levels of low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive protein: rationale and design of the JUPITER trial. *Circulation* 2003;108:2292-7.
- 16 Ridker PM, Fonseca FA, Genest J, Gotto AM, Kastelein JJ, Khurmi NS, et al. Baseline characteristics of participants in the JUPITER trial, a randomized placebo-controlled primary prevention trial of statin therapy among individuals with low low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive protein. *Am J Cardiol* 2007;100:1659-64.
- 17 National Cholesterol Education Program. Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA* 2001;285:2486-97.
- 18 Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds risk score. *JAMA* 2007;297:611-9.
- 19 Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds risk score for men. *Circulation* 2008;118:2243-51.
- 20 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565-74.
- 21 National Institute for Health and Clinical Excellence. Technology appraisal 94: statins for the prevention of cardiovascular events. 2006. www.nice.org.uk/TA094.
- 22 Ridker PM, Macfadyen JG, Nordestgaard BG, Koenig W, Kastelein JJ, Genest J, et al. Rosuvastatin for primary prevention among individuals with elevated high-sensitivity c-reactive protein and 5% to 10% and 10% to 20% 10-year risk. Implications of the Justification for Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial for "intermediate risk." *Circ Cardiovasc Qual Outcomes* 2010;3:447-52.
- 23 Ridker PM, MacFadyen JG, Fonseca FA, Genest J, Gotto AM, Kastelein JJ, et al. Number needed to treat with rosuvastatin to prevent first cardiovascular events and death among men and women with low low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive protein: justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin (JUPITER). *Circ Cardiovasc Qual Outcomes* 2009;2:616-23.
- 24 Koenig W, Ridker PM. Rosuvastatin for primary prevention in patients with European systematic coronary risk evaluation risk $\geq 5\%$ or Framingham risk $>20\%$: post hoc analyses of the JUPITER trial requested by European health authorities. *Eur Heart J* 2011;32:75-83.
- 25 Selker HP, Beshansky JR, Griffith JL. Use of the electrocardiograph-based thrombolytic predictive instrument to assist thrombolytic and reperfusion therapy for acute myocardial infarction. A multicenter, randomized, controlled, clinical effectiveness trial. *Ann Intern Med* 2002;137:87-95.
- 26 Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117.

Accepted: 12 August 2011

Cite this as: *BMJ* 2011;343:d5888

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: <http://creativecommons.org/licenses/by-nc/2.0/> and <http://creativecommons.org/licenses/by-nc/2.0/legalcode>.

What is already known on this topic

In clinical practice some patients benefit more than average from treatment, whereas others do not or may even be harmed

Implementing trial results by treating all or no patients, expecting the treatment effect for everyone to be similar to the average treatment effect in the original trial, may not lead to optimal benefit

What this study adds

Data from randomised trials can be used to predict treatment effect in terms of absolute risk reduction for individual patients

Predictions could be based on existing validated risk scores, if available, or a new prediction model fit to the trial data

The value of such prediction of treatment effect for medical decision making is conditional on the number willing to treat (NWT) to prevent one outcome event

Tables

Table 1 | Baseline characteristics of participants in Justification for the Use of Statins in Prevention trial. Values are medians (interquartile ranges) unless specified otherwise

Characteristic	Population (n=17 710)
Age (years)	66 (60-71)
Men (%)	61.8
White ethnicity (%)	71.3
Current smoker (%)	15.8
Family history of premature coronary heart disease (%)	11.5
High density lipoprotein cholesterol (mmol/L)	1.3 (1.0-1.6)
Low density lipoprotein cholesterol (mmol/L)	2.8 (2.4-3.1)
Total cholesterol (mmol/L)	4.8 (4.4-5.2)
High sensitivity C reactive protein (mg/L)	4.3 (2.9-7.1)
Systolic blood pressure (mm Hg)	134 (124-145)
Blood pressure lowering drug use (%)	49.5
Body mass index	28.4 (25.3-32.0)

Table 2| Calculation example of predicted 10 year treatment effect for two patient scenarios

Variables	Scenario 1*			Scenario 2†		
	Framingham based	Reynolds based	Optimal fit model	Framingham based	Reynolds based	Optimal fit model
Baseline 10 year risk for cardiovascular disease (%)	16	13.9	16.6	2	4.3	2.6
Residual risk if treated with rosuvastatin for 10 years (%)	9	7.8	5.9	1.1	2.4	1.6
Predicted absolute risk reduction (%)	7	6.1	10.6	0.9	1.9	1.0
NNT (patients with similar characteristics) (%)	14	16	9	111	53	100

NNT=number needed to treat.

Following values were the same in both scenarios: total cholesterol level 4.8 mmol/L, high density lipoprotein cholesterol level 1.3 mmol/L, low density lipoprotein cholesterol level 2.8 mmol/L, high sensitivity C reactive protein level 4.3 mg/L, and 134 mm Hg systolic blood pressure.

*Male non-smoker aged 60 years with family history of premature coronary heart disease and taking blood pressure lowering drugs.

†Female smoker aged 55 years with no family history of premature coronary heart disease and not taking blood pressure lowering drugs.

Table 3| Results of net benefit assessment

	NWT	Decision threshold (%)	Treat all	Prediction based treatment net benefit (% treatment rate)		
				Framingham score	Reynolds score	Optimal fit model
Little harm: treat even at low risk	Infinity	≥0	0.0499	0.0499 (100)	0.0499 (100)	0.0499 (100)
	100	≥1	0.0399	0.0398 (93)	0.0388 (97)	0.0407 (98)
	75	≥1	0.0365	0.0371 (89)	0.0376 (94)	0.0376 (96)
	50	≥2	0.0299	0.0320 (84)	0.0307 (83)	0.0316 (84)
	30	≥3	0.0165	0.0233 (70)	0.0180 (62)	0.0271 (58)
	20	≥5	-0.0001	0.0054 (46)	0.0080 (42)	0.0106 (34)
	15	≥7	-0.0168	0.0030 (32)	0.0081 (28)	0.0058 (19)
Considerable harm: treat at high risk only	12	≥8	-0.0335	-0.0038 (18)	0.0003 (18)	-0.0017 (10)
	10	≥10	-0.0501	-0.0011 (8)	0.0040 (12)	-0.0032 (6)
	0	Infinity	-Infinity	0 (0)	0 (0)	0 (0)

NWT=number willing to treat.

The net benefit of treating no one serves as a reference and is equal to zero. Positive values imply that treatment of all patients or prediction based treatment was superior to treating no one, given the corresponding NWT.

Figures

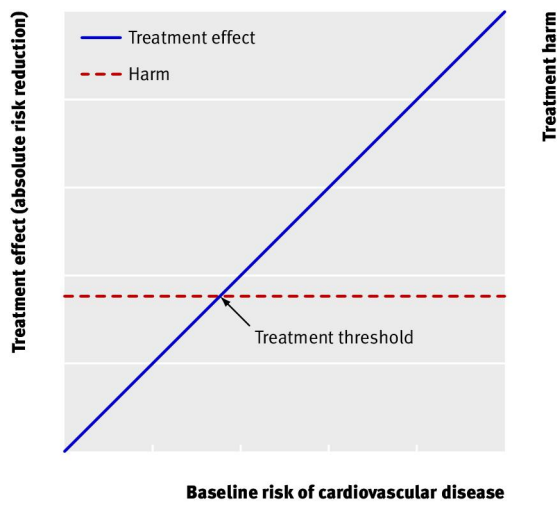


Fig 1 Basic concept for weighing treatment effect against harm. Treatment effect usually increases with baseline risk, whereas harm is relatively constant for all patients. Those whose treatment effect exceeds treatment related harm (reflected by decision threshold) benefit from treatment¹

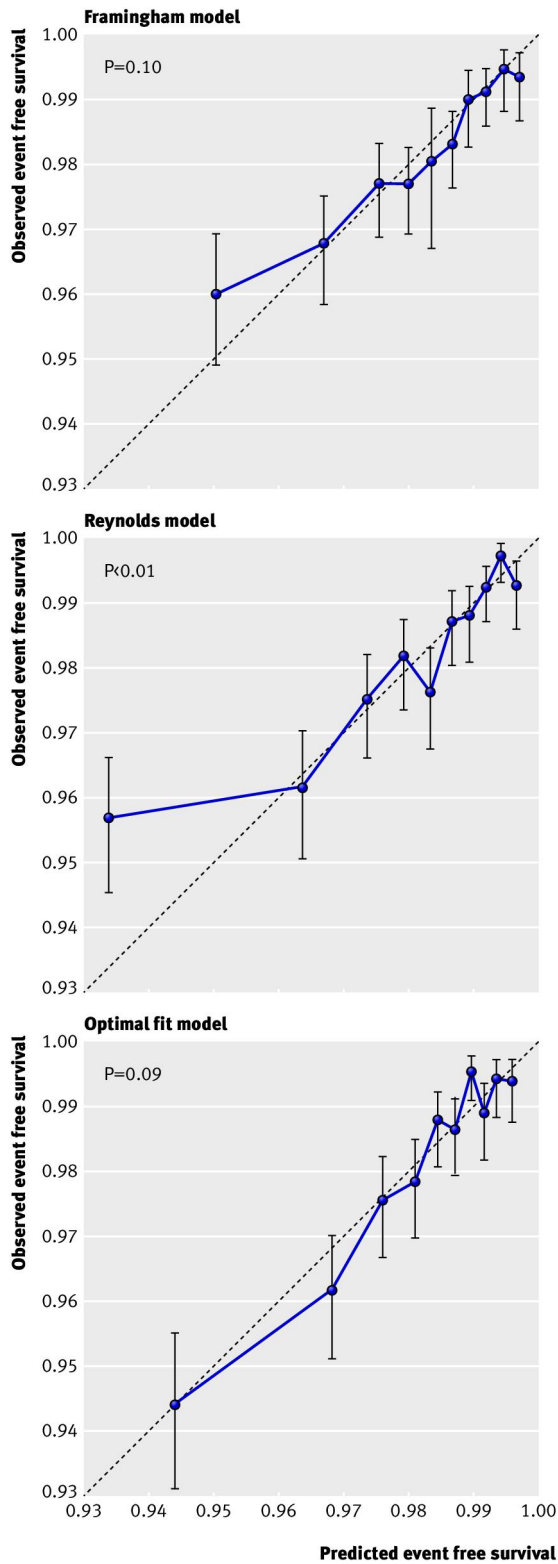


Fig 2 Calibration plots. Predicted and observed two year event free survival for cardiovascular events within 10ths of predicted survival using three models. P values derived from the Hosmer-Lemeshow test

- Female smoker aged 55 years with no family history of premature coronary heart disease and not taking blood pressure lowering drugs
- Male non-smoker aged 60 years with family history of premature coronary heart disease and taking blood pressure lowering drugs
- Total JUPITER population

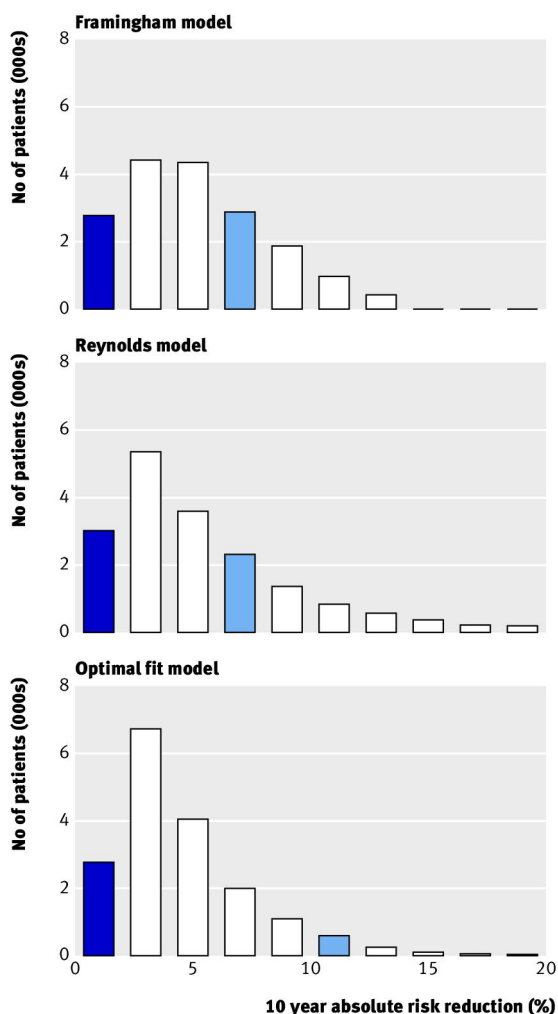


Fig 3 Distribution of predicted 10 year absolute treatment effect (absolute risk reduction) based on Framingham risk score, Reynolds risk score, and optimal fit model, with coloured bars indicating predicted treatment effects for two different patient scenarios. JUPITER=the Justification for the Use of Statins in Prevention trial

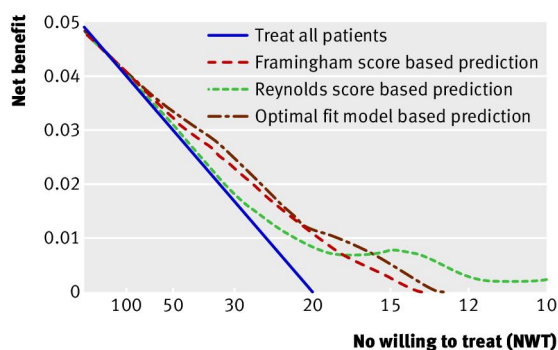


Fig 4 Decision curve: graphical representation of net benefit. For large values of numbers willing to treat (NWT), the net benefit of treating all patients is about equal to the net benefit of prediction based treatment. The net benefit of treating all patients becomes negative if the NWT is less than 20, whereas the net benefit of prediction based treatment is still positive for a NWT of 20 and converges to zero for smaller values of NWT

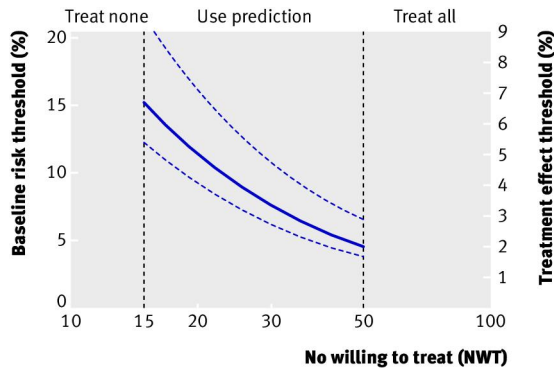


Fig 5 Implications for clinical practice. Justification for the Use of Statins in Prevention trial shows that treatment of all patients is the strategy of choice if the 10 year number willing to treat (NWT) is 50 or more. Treating no one is preferable if the 10 year NWT is 15 or fewer. If the NWT is between 15 and 50, prediction based treatment results in most net benefit