Appendix 3. Application of treatment-effect prediction methods to other datasets

Prediction of treatment effect is not restrained to cholesterol lowering for prevention of cardiovascular events. In general prediction-based treatment is especially useful when:

- The effect of treatment is limited, because if a therapy is known to be effective in everyone (e.g. antibiotics for pneumonia), all patients will benefit from treatment.
- Treatment is associated with considerable harm, e.g. adverse reactions or high costs. In such situations benefit will outweigh harm only when predicted treatment effect is high.
- The outcome of interest can be predicted accurately on the basis of clinical patient characteristics. For example, the mortality risk within 30-days after an acute myocardial infarction can be predicted accurately on the basis of age, location of infarction, heart rate, Killip class, systolic blood pressure, and history of prior myocardial infarction.[1]

Not all situations where prediction-based treatment is useful meet all three criteria. In the present JUPITER-trial example, treatment was not associated with major harm and model-accuracy was moderate. Yet, prediction-based treatment was associated with optimal net benefit for a range of NWT.

A trial dataset is required to develop and test treatment effect prediction models. As a rule of thumb, this trial dataset needs to contain at least ten outcome events for every candidate risk factor that is considered as a candidate predictor for the treatment effect prediction model. Because trials are often powered to include only as many endpoints as needed to demonstrate a significant treatment effect, this may sometimes hamper the development of a prediction model.[2] [3] Still, the number of endpoints that is required to develop a prediction model not necessarily exceeds the number that is needed to demonstrate a significant treatment effect. Moreover, suitable trials for developing treatment-effect prediction models preferably have few eligibility criteria. Especially when trials have selected participants on the basis of important risk factors for the outcome event, this may limit generalisability to the general population. Finally, duration of follow-up should preferably be long enough to provide meaningful predictions. When this is not the case (like in the JUPITER-trial example), survival estimates can often be extrapolated.

Methods for fitting treatment-effect prediction models are not different from fitting prediction models in general. Candidate predictors and candidate treatment-covariate interactions need to be selected carefully on the basis of prior literature to avoid chance-findings. Cox proportional hazards modelling is appropriate for time-to event data and logistic regression analysis is appropriate when duration of follow-up is fixed.[1] In smaller datasets, models may need to be adjusted for 'optimism', for example during selection of the coefficients using the Lasso-method or after selection of the coefficients by multiplying them by a shrinkage factor. Moreover, like is customary with any newly developed prediction model, treatment effect prediction models can be validated in external datasets.

Presentation of treatment-effect prediction models should preferably include the following:

- Model structure and coefficients (Box 1). These are needed to build online treatment-effect calculators and implement the model in electronic patient record systems.
- Measures of model performance, such as a calibration curve (Figure 2) and Hosmer-Lemeshow test to assess calibration and a c-statistic to assess discrimination. Other performance measures may also be used.
- A histogram of the distribution of predicted treatment effect in the trial population (Figure 3). This could help to determine whether the predicted treatment effect for an individual patient is lower, equal to or higher than average.
- A decision-curve covering a clinically relevant range of NWT (Figure 4). The methods for calculating net benefit and plotting a decision curve are described in detail by Vickers et al.[4] [5]

Finally, when treatment is associated with adverse reactions, additional information on excess risk might be helpful for patients and physicians to determine the appropriate level of NWT. This could be presented as the excess number of adverse events resulting from treatment of certain numbers of patients. For example, how many gastrointestinal bleeding events result from aspirin treatment of 30, 50 or 100 patients? When the occurrence of adverse reactions is related to certain patient characteristic, such as age, excess adverse event risk could be presented for different age strata. When the occurrence of adverse reaction is related to more than one patient characteristic, patient-specific adverse event risk could be modelled also.

1 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.

2 Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995;345:1616-9.

3 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.

4 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.

5 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565-74.