A comparison of approaches for combining predictive markers for personalised treatment recommendations

Matthias Pierce and Richard Emsley

Abstract

**Background:** In the presence of heterogeneous treatment effects, it is desirable to divide patients into subgroups based on their expected response to treatment. This is formalised via a personalised treatment recommendation: an algorithm that uses biomarker measurements to select treatments. It could be that multiple, rather than single, biomarkers better predict these subgroups. However, finding the optimal combination of multiple biomarkers can be a difficult prediction problem.

**Methods:** We described three parametric methods for finding the optimal combination of biomarkers in a personalised treatment recommendation, using randomised trial data: a regression approach that models outcome using treatment by biomarker interactions; an approach proposed by Kraemer that forms a combined measure from individual biomarker weights, calculated on all treated and control pairs; and a novel modification of Kraemer’s approach that utilizes a prognostic score to sample matched treated and control subjects. Using Monte Carlo simulations under multiple data-generating models, we compare these approaches and draw conclusions based on a measure of improvement under a personalised treatment recommendation compared to a standard treatment. The three methods are applied to data from a randomised trial of home-delivered pragmatic rehabilitation versus treatment as usual for patients with chronic fatigue syndrome (the FINE trial). Prior analysis of this data indicated some treatment effect heterogeneity from multiple, correlated biomarkers.

**Results:** The regression approach outperformed Kraemer’s approach across all data-generating scenarios. The modification of Kraemer’s approach leads to improved treatment recommendations, except in the case where there was a strong unobserved prognostic biomarker. In the FINE example, the regression method indicated a weak improvement under its personalised treatment recommendation algorithm.

**Conclusions:** The method proposed by Kraemer does not perform better than a regression approach for combining multiple biomarkers. All methods are sensitive to misspecification of the parametric models.

**Keywords:** Personalised medicine, Stratified medicine, Precision medicine, Personalised treatment recommendations, Predictive biomarkers, Moderators
Background

One of the primary aims of the modern paradigm of stratified medicine is to move beyond a one-size-fits-all approach that allocates treatment based on population average responses, towards identifying patient subgroups for whom a given treatment is beneficial and those for whom it is not. Given a patient population with heterogeneous treatment response, it might be possible to produce an algorithm for clinical use that provides a recommendation for treatment based on measurable traits (biomarkers). For these purposes, it is necessary to separate biomarkers into those that predict treatment response (moderating biomarkers) and those that predict the outcome, regardless of treatment (prognostic biomarkers). When the treatment choice is binary (the situation considered in this paper), the algorithm may recommend a treatment over an alternative for values of a single moderating biomarker, or a weighted combination of multiple moderating biomarkers. Such an algorithm is referred to as a personalised treatment recommendation (PTR).

In many disease areas, it might be that a combination of multiple biomarkers is more effective in identifying subgroups with a beneficial treatment outcome than any single biomarker [1]. Finding the optimal combination of biomarkers in a PTR algorithm is a challenging prediction problem. In order to avoid the confounding between treatment assignment and outcome, it is considered optimal that PTR's are estimated from randomised controlled trial (RCT) data. A method used to estimate a PTR is to fit a regression model with treatment by biomarker interaction terms [2, 3]. Kraemer [4] proposes an alternative method that uses a parametric model fitted to all pairwise combinations of treated and control subjects.

The method assigns a weight to each moderator from the correlation between pair of treated and control subjects. The method assigns a weight to each moderator from the correlation between pair of treated and control subjects. The method assigns a weight to each moderator from the correlation between pair of treated and control subjects. The method assigns a weight to each moderator from the correlation between pair of treated and control subjects. The method assigns a weight to each moderator from the correlation between pair of treated and control subjects. The method assigns a weight to each moderator from the correlation between pair of treated and control subjects. The method assigns a weight to each moderator from the correlation between pair of treated and control subjects.
A combined moderator is obtained by the sum of the individual moderator weights: \( Z_\ast = \sum_{k=1}^{K} w_k Z_{ik} \). This combined moderator can be used to separate subgroups based on their expected treatment response (see below) [6].

**Modified Kraemer approach: matching on a prognostic score**

We modify Kraemer’s approach to determine whether considering only treated and control pairs that have similar values of a ‘prognostic score’ improves its performance. A prognostic score is an estimate of the treatment-free outcome. This is calculated by fitting a regression model to the subjects in the control group only (thus excluding any moderators):

\[
\mu(Y, A = -1/2) = a_0 + aX^\top
\]

The resulting model can be used to predict the outcome under the control condition, for subjects in both the control and treatment groups.

The resulting estimates are referred to as *prognostic scores* and have been used in observational research to control for confounding [7].

We propose modifying Kraemer’s approach so that the composite moderator is derived using a sample of treated and control pairs that only have similar values of their prognostic scores. In our implementation, we use a single nearest neighbour matching algorithm, with replacement, and with a caliper such that, for each treated subject, a control is sought that is within 0.1 times the standard deviation of the prognostic score. The caliper width is arbitrary and was set prior to any implementation.

The rationale for this modification is to minimise the contribution to \( \Delta(Y) \) by variables that are irrelevant to treatment effect modification. There is a trade-off between minimising this variance and losing treated and control pairs who do not fit the matching criteria. We investigate whether this modification results in an improvement on the Kraemer approach by applying these techniques to simulated datasets. First, we establish how to measure improvement in the context of determining a personalised treatment recommendation.

**Constructing a personalised treatment recommendation (PTR)**

A PTR uses biomarker values to recommend whether a patient should be treated or not. For example, statins are recommended in the UK if a person is aged over 40 and if their estimated CVD risk is at least 10% over a 10-year period [8]. Formally, this can be represented as: PTR = \( I \{ \text{age} \geq 40 \& \text{CVD risk} \geq 10\% \} \), where \( I \) indicates that treatment (statins) is recommended over the alternative (no statins) when the bracketed expression is true.

Assuming higher values of a continuous outcome are advantageous, an optimal PTR is one that recommends treatment when, conditional on a set of moderating biomarkers \( Z \), the mean outcome under treatment \( \mu(A = 1/2, Z) \) is greater than the mean outcome under control \( \mu(A = -1/2, Z) \):

\[
\text{PTR} = I\{\mu(A = 1/2, Z) - \mu(A = -1/2, Z) > 0\}
\]

Under the parametrisation in Eq. (1):

\[
I\{\mu(A = 1/2, Z) - \mu(A = -1/2, Z) > 0\} = I\{\beta_0 + \beta Z^* > 0\}
\]

(2)

The parameters \( \beta_0 \) and \( \beta \) can be estimated using ordinary least squares regression with treatment by moderator interaction terms [2, 3]. This we refer to as the regression approach.

A PTR can be constructed using the Kraemer or modified Kraemer method using the following steps: (1) calculate the combined moderator; (2) regress the outcome on a model with treatment, the combined moderator and their interaction with treatment: \( \mu(A, Z^*) = a_0 + aZ^* + A(\beta_0^* + \beta^* Z^*) \) (where \( Z^* \) is the combined moderator and \( \beta_0^* + \beta^* \) are the average effect of treatment and the moderated effect, respectively); and (3) use these parameters to calculate the PTR: \( \text{PTR} = I\{\beta_0 + \beta Z^* > 0\} \).

**Measuring the performance of a personalised treatment recommendations**

Under randomisation, an unbiased estimate of the population mean outcome under a PTR is provided by the mean of the observed outcome in those receiving the treatment they were recommended weighted by the probability of being randomised to their respective group (\( n \)) [9–11]:

\[
\mu(\text{PTR}) = \frac{1}{n} \sum_{n} \left( \frac{(A + 1/2) \cdot (\text{PTR} + 1/2)}{\pi} Y + \frac{(1/2 - A) \cdot (1/2 - \text{PTR})}{1 - \pi} Y \right)
\]

This can be contrasted with the average outcome under treatment \( \mu(A = 1/2) = \frac{1}{n} \sum_{n}(A + 1/2) Y \) or control \( \mu(A = -1/2) = \frac{1}{n} \sum_{n}(1/2 - A) Y \) to get the parameters: \( \theta_T = \mu(\text{PTR}) - \mu(A = 1/2) \) and \( \theta_C = \mu(\text{PTR}) - \mu(A = -1/2) \). These are interpreted as the expected change in outcome under a PTR compared to a policy where everybody receives treatment or everybody receives control.

In a simulation study, an additional measure of the performance of a PTR is the rate of misclassification; that is, the proportion of subjects whose PTR conflicts with their known optimal treatment: \( \mathbb{P}(\text{PTR}(X) \neq \mathbb{P}(\text{PTR}(X) = \text{OPT}(X)) \)


PTR\textsuperscript{opt}, where PTR\textsuperscript{opt} indicates treatment if the simulated outcome under treatment is greater than the simulated outcome under control.

**Simulations comparing approaches**
Monte Carlo simulations were constructed to compare the regression, Kraemer and modified Kraemer approaches to estimating a PTR. Training datasets were simulated with sample sizes 75, 200 and 300 with 1:1 randomisation. These datasets were generated under a range of scenarios (shown in Table 1). PTRs were estimated using all three approaches and applied to a test dataset of the same size and generated using the same specifications as the training dataset. For each PTR, we use the test dataset to calculate the change under PTR (\(\theta_T\)) and the misclassification rate. For each data-generating scenario, 5000 simulations were carried out and we evaluate each method by averaging \(\theta_T\) and the misclassification across simulations.

**Application to randomised trial data**
The three approaches to constructing a PTR were applied to data from the Fatigue Intervention Nurses Evaluation (FINE) randomised trial [12]. This trial randomised 296 patients diagnosed with Chronic Fatigue Syndrome to three groups: home-delivered pragmatic rehabilitation, supportive listening or treatment as usual. It included in the PTR: baseline fatigue score (with positive values indicating an improvement). Variables were designated moderators if they had a univariate \(p\) value in the initial analysis of less than 0.10. Moderating variables were then excluded then if their \(p\) value was greater than 0.3, in a multivariate model that included all moderating variables. Three variables remained to include in the PTR: baseline fatigue score (\(p\) value for interaction = 0.24), EQ-5D mobility (no problems, some problems, severe problems, \(p = 0.16\)) and score on the Oslo Social Support scale, relating to concern (\(p = 0.15\)). The data was randomly split in half between training and test datasets (size \(n = 98\) and \(n = 97\) respectively) and then PTRs were evaluated on the test dataset using the parameter \(\theta_C\) outlined above. Inference for this parameter was determined by drawing 1000 bootstrap samples and using the normal approximation.

**Results**

**Simulations**
Across all data-generating scenarios and sample sizes, the regression method was superior to both the Kraemer and the modified Kraemer methods: on average, it was

**Table 1** Different scenarios for simulations, comparing approaches to combining multiple biomarkers to construct personalised treatment recommendations

| #  | Scenario                        | Data generation model                                                                 | Variables in the prediction model                          |
|----|---------------------------------|--------------------------------------------------------------------------------------|-------------------------------------------------------------|
| 1  | Simple linear                   | \(Y = 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 2Z_1 - 1.5Z_3 - 3Z_3,_{\nu}) + \epsilon\) | All variables                                               |
| 2  | Linear with weak moderators     | \(Y = 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 0.5Z_1 - 0.5Z_2 - 0.25Z_3,_{\nu}) + \epsilon\) | All variables                                               |
| 3  | Strong unobserved prognostic marker | \(Y = 12U_1 + 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 2Z_1 - 1.5Z_3 - 3Z_3,_{\nu}) + \epsilon\) | Excluding \(U_1\)                                            |
| 4  | Strong unobserved moderator variables | \(Y = 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 10U_2 + 2Z_1 - 1.5Z_2 - 3Z_3,_{\nu}) + \epsilon\) | Excluding \(U_2\)                                            |
| 5  | Misspecified prognostic part    | \(Y = 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 10U_2 + 2Z_1 - 1.5Z_2 - 3Z_3,_{\nu}) + \epsilon\) | All linear terms, excluding \(U_1\)                           |
| 6  | Misspecified moderator part     | \(Y = 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 2Z_1 - 1.5Z_2 - 3Z_3,_{\nu}) + \epsilon\) | All linear terms, excluding \(U_2\)                           |
| 7  | Non-linear model                | \(Y = 3X_1 + 2Z_1 - 0.5Z_3,_{\omega} - 1.5Z_3,_{\nu} + a(2 + 2Z_1 - 1.5Z_2 - 3Z_3,_{\nu}) + \epsilon\) | All as linear terms                                          |
Table 2: Results of simulation studies showing mean values of theta and misclassification rate across 5000 simulations under a range of data-generating scenarios.

| Scenario | Sample size | Parameter | Method          | Regression (SD) | Kraemer (SD) | Modified Kraemer (SD) |
|----------|-------------|-----------|-----------------|----------------|--------------|------------------------|
| 1a. Simple linear | 50, 200, 300 | θ (SD) | 0.72 (0.676), 0.71 (0.473), 0.72 (0.337) | 0.46 (0.810), 0.58 (0.555), 0.65 (0.397) | 0.62 (0.679), 0.68 (0.468), 0.70 (0.335) |
|           |             | Misclass. rate | 0.05, 0.04, 0.02 | 0.21, 0.15, 0.10 | 0.14, 0.09, 0.06 |
| 1b. Simple linear with correlated biomarkers | 50, 200, 300 | θ (SD) | 0.73 (0.629), 0.61 (0.422), 0.62 (0.276) | 0.50 (0.695), 0.48 (0.483), 0.57 (0.331) | 0.59 (0.616), 0.56 (0.423), 0.61 (0.278) |
|           |             | Misclass. rate | 0.05, 0.05, 0.02 | 0.21, 0.17, 0.1 | 0.17, 0.11, 0.06 |
| 2. Linear with weak moderators | 50, 200, 300 | θ (SD) | 0.71 (0.665), 0.71 (0.482), 0.72 (0.336) | 0.45 (0.799), 0.56 (0.567), 0.66 (0.401) | 0.61 (0.667), 0.66 (0.481), 0.71 (0.336) |
|           |             | Misclass. rate | 0.05, 0.04, 0.02 | 0.21, 0.15, 0.1 | 0.13, 0.09, 0.06 |
| 3. Strong unobserved prognostic marker | 50, 200, 300 | θ (SD) | 0.23 (3.150), 0.23 (3.150), 0.07 (1.556) | −0.24 (3.149), −0.24 (3.149), 0.06 (0.401) | −0.31 (3.115), −0.31 (3.115), 0.01 (1.535) |
|           |             | Misclass. rate | 0.41, 0.41, 0.32 | 0.41, 0.41, 0.32 | 0.42, 0.42, 0.34 |
| 4. Strong unobserved moderator variables | 50, 200, 300 | θ (SD) | 0.22 (1.328), 0.43 (0.893), 0.56 (0.606) | 0.16 (1.405), 0.35 (0.946), 0.51 (0.654) | 0.14 (1.323), 0.35 (0.911), 0.52 (0.620) |
|           |             | Misclass. rate | 0.42, 0.42, 0.32 | 0.45, 0.43, 0.42 | 0.45, 0.44, 0.34 |
| 5. Misspecified prognostic part | 50, 200, 300 | θ (SD) | 0.53 (0.866), 0.62 (0.582), 0.67 (0.406) | 0.34 (1.010), 0.49 (0.682), 0.60 (0.495) | 0.40 (0.882), 0.57 (0.588), 0.65 (0.407) |
|           |             | Misclass. rate | 0.17, 0.12, 0.09 | 0.25, 0.18, 0.13 | 0.22, 0.15, 0.11 |
| 6. Misspecified moderator part | 50, 200, 300 | θ (SD) | 0.70 (0.773), 0.78 (0.559), 0.80 (0.381) | 0.44 (0.872), 0.60 (0.651), 0.72 (0.443) | 0.58 (0.775), 0.72 (0.558), 0.77 (0.382) |
|           |             | Misclass. rate | 0.16, 0.13, 0.12 | 0.25, 0.2, 0.16 | 0.21, 0.16, 0.14 |
both associated with higher values of $\theta_T$ (the expected benefit under PTR compared to treating everyone) and the lowest misclassification rate (Table 2). The modification of the Kraemer method, where treated and control subjects are matched on their prognostic score, improved on the Kraemer method across most data-generating scenarios. The exception is the scenario with a strong prognostic variable that is not included in the regression model. This suggests that the prognostic score is useful only when it captures sufficient variation in the prognostic effects. Post hoc, we changed the size of the calliper distance but this did not make any noticeable difference until it was $< 0.05$ SD or $> 1.5$ SD of the prognostic score (Table 2).

In the scenario with a non-linear data-generating model, no method, on average, constructed a PTR where subjects had a better outcome compared to a policy where everybody was treated. It is worth noting that, in this scenario, if each simulated subject were allocated the treatment they should have received, then their expected outcome would be, on average, 1.40 higher than if everybody were treated.

**Discussion**

This paper reported on a comparison of three methods for constructing personalised treatment recommendations from randomised controlled trial data: the regression method that models outcome using treatment by moderator biomarker interactions; a method proposed by Kraemer that forms a combined moderator from individual moderator weights, calculated on all treated and control pairs; and a modification of Kraemer’s approach that utilises a prognostic score to sample pairs of treated and control subjects. Across all simulations, the regression approach outperformed Kraemer’s approach. The modification of Kraemer’s approach appeared to indicate higher values of $\theta$, except in the case where there was a strong unobserved prognostic marker. The superiority of the regression approach was replicated using real-world data from a randomised trial of home-delivered pragmatic rehabilitation for chronic fatigue patients; however, for this example, no method conclusively demonstrated that a PTR does better than a policy of ‘treatment as usual’ despite there being several individual moderators of treatment effect [13]. Therefore, in this case, we conclude that forming a PTR is more difficult than finding individual treatment effect moderators.

**Trial data**

The results of the PTR algorithms and the estimated change under a PTR, compared to a treatment-as-usual approach, are shown in Table 3. There is weak evidence that the regression method results in a PTR that results in a greater reduction in chronic fatigue symptoms compared to an approach where everybody receives treatment as usual ($\theta = 1.92$, 95% CI $−0.65$ to $4.49$). There was little evidence that a PTR estimated using the Kraemer method or the modified Kraemer method results in an improvement compared to treatment as usual ($\theta = 0.47$ and $p = 0.13$ respectively). Eight subjects were excluded when implementing the modified Kraemer approach because they did not have a match within the set calliper distance of the prognostic score.

**Table 2** Results of simulation studies showing mean values of theta and misclassification rate across 5000 simulations under a range of data-generating scenarios (Continued)

| Scenario       | Sample size | Parameter | Method          | Regression | Kraemer | Modified Kraemer |
|----------------|-------------|-----------|-----------------|------------|---------|------------------|
| 7. Non-linear model | 50          | $\theta$ (SD) | 2.11 (8.414)   | 1.94 (8.606)   | 2.10 (8.566)   |
|                |             | Misclass. rate | 0.45 | 0.45 | 0.44 |
|                | 200         | $\theta$ (SD) | 2.89 (9.242)   | 2.84 (10.822)  | 1.62 (10.261)  |
|                |             | Misclass. rate | 0.44 | 0.44 | 0.43 |
|                | 300         | $\theta$ (SD) | 0.31 (10.085)  | 0.06 (10.831)  | $−0.30$ (10.797) |
|                |             | Misclass. rate | 0.44 | 0.44 | 0.43 |

**Table 3** Results of the analysis of FINE data

| Method       | PTR algorithm | $\theta$ | $p$ value | 95% CI         |
|--------------|---------------|----------|-----------|----------------|
| Regression   | $−2.020 + 1.709$. baselinefat $+ 1.705$. EQ5D $+ 1.685$. concern $< 0$ | 1.92 | 0.07 | $−0.65$ to $4.49$ |
| Kraemer      | $−2.342 + 0.633$. baselinefat $+ 2.253$. EQ5D $+ 2.069$. concern $< 0$ | $−0.10$ | 0.47 | $−2.26$ to $2.06$ |
| Modified Kraemer | $−2.365 + 1.278$. baselinefat $+ 1.568$. EQ5D $+ 2.101$. concern $< 0$ | 0.69 | 0.13 | $−0.49$ to $1.87$ |
the likelihood of correctly specifying the model might be low. For example, none of the approaches, on average, indicated an improvement under PTR in the scenario with a non-linear data-generating model. These simulations were limited because they did not include any variable selection or transformations of variables based on model fit. Such processes require care and often require knowledge of the variables at hand. In practice, researchers will be insuring against model underfitting by testing for non-linear terms and higher-order interactions. Overfitting of these models could be counterbalanced using regularisation techniques, such as Lasso regression. Additionally, methods exist that are more robust to model misspecification, for example methods that seek to maximise the expected outcome under a PTR using classification techniques [14–16].

The application to the FINE randomised control trial showed that a PTR, as estimated using the regression approach, might result in an improvement over a recommendation where everybody is provided treatment as usual; however, the 95% confidence interval for all approaches included estimates where the PTR strategy is associated with a small amount of harm. These results should not be over-interpreted: the 95% confidence intervals were wide, which indicates insufficient power to detect a change. Data from another trial testing the use of the PTR would be needed to confirm whether any PTR results in an overall benefit. Another aspect that should be considered is whether including cost information in the measure of benefit has an effect on the decision to adopt a PTR strategy.

In the discussion to the paper, Kraemer says: ‘an irrelevant baseline factor and a non-specific predictor can have no influence on making different choices between [treatments]’ We argue that this appears to be false, based on our findings: information from a prognostic score appears to result in a composite moderator that more effectively discriminates between those who should receive treatment and those that should not. Whilst the Kraemer approach does not appear to improve on the regression approach, it should be noted that the original paper provides a useful example of how to judge the relative effect sizes of multiple modifiers that would form a useful exploratory analysis before forming a PTR.

**Conclusion**

Our simulations demonstrate that the parametric method proposed by Kraemer does not result in a more effective personalised treatment recommendation than a method that uses a regression model. Utilising a prognostic score improves the Kraemer method, however not to an extent that it should be adopted over the regression method.

**Abbreviations**

PTR: Personalised treatment recommendation; RCT: Randomised controlled trial; FINE: Fatigue Intervention Nurses Evaluation randomised trial

**Acknowledgements**

We would like to thank Alison Wearden and the rest of the FINE team for the use of their trial data and Andrew Pickles and Jennifer Hellier for their helpful comments on this work. This work is dedicated to the memory of Professor Graham Dunn, who provided us with support and advice when developing this manuscript.

**Authors’ contributions**

RE conceived of the study and MP developed the methodology and carried out the analysis. All authors read and approved the final manuscript.

**Funding**

This work was supported by the MRC North West Hub for Trials Methodology Research (MR/K025635/1). This work represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

**Availability of data and materials**

Stata do-files used to construct simulations are available on request from the corresponding author.

**Ethics approval and consent to participate**

Ethical approval for the FINE trial was granted by the Eastern Multicentre Research Ethics Committee, reference 03/S/5/62.

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1Centre for Biostatistics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, 1st Floor, Jean McFarlane Building, Oxford Road, Manchester M13 9PL, UK. 2Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK.

**Received:** 20 February 2020 **Accepted:** 15 November 2020 **Published online:** 06 January 2021

**References**

1. The Academy of Medical Sciences. Realising the potential of stratified medicine: Academy of Medical Sciences; 2013. p. 105.
2. Byar DP. Assessing apparent treatment - covariate interactions in randomized clinical trials. Stat Med. 1985;4:255–63.
3. Gail AM, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. Int Biometric Soc. 2009;41:361–72.
4. Kraemer HC. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. Stat Med. 2013;32:1994–73.
5. Wearden AJ, Dowell C, Chew-Graham C, Bentall RP, Morris R, Peters S, et al. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. Br Med J. 2010; 340:c1777.
6. Wallace ML, Frank E, Kraemer HC. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. JAMA Psychiatry. 2013;70:1241–7.
7. Hansen BB. The prognostic analogue of the propensity score. Biometrika. 2008;95:481–8.
8. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction including lipid modification (NICE Guidelines
9. Song X, Pepe MS. Evaluating markers for selecting a patient’s treatment. Biometrics. 2004;60:874–83.
10. Janes H, Brown MD, Huang Y, Pepe MS. An approach to evaluating and comparing biomarkers for patient treatment selection. Int J Biostat. 2014;10:99–121.
11. Kapelner A, Bleich J, Cohen ZD, DeRubeis RJ, Berk R. Inference for treatment regime models in personalized medicine 2014: http://arxiv.org/abs/1404.7844.
12. Wearden AJ, Riste L, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, et al. Fatigue intervention by nurses evaluation—the FINE trial. A randomised controlled trial of nurse led self-help treatment for patients in primary care with chronic fatigue syndrome: study protocol. BMC Medicine. 2006;4:9.
13. Wearden AJ, Dunn G, Dowrick C, Morriss RK. Depressive symptoms and pragmatic rehabilitation for chronic fatigue syndrome. Br J Psychiatry. 2012;201:227–32.
14. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. J Am Stat Assoc. 2012;107:1106–18.
15. Rubin DB, van der Laan MJ. Statistical issues and limitations in personalized medicine research with clinical trials. Int J Biostat. 2012;8:18.
16. Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber E. Estimating optimal treatment regimes from a classification perspective. Stat. 2012;1:103–14.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.