BMJ Open

Prediction model study focusing on eHealth in the management of urinary incontinence: the Personalised Advantage Index as a decision-making aid

Anne Martina Maria Loohuis ,[1] Huibert Burger ,[1] Nienke Wessels,[1] Janny Dekker,[1] Alec GGA Malmberg,[2] Marjolein Y Berger,[1] Marco H Blanker ,[1] Henk van der Worp [1]

To cite: Loohuis AMM, Burger H, Wessels N, *et al*. Prediction model study focusing on eHealth in the management of urinary incontinence: the Personalised Advantage Index as a decision-making aid. *BMJ Open* 2022;**12**:e051827. doi:10.1136/bmjopen-2021-051827

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-051827).

Received 02 April 2021
Accepted 22 June 2022

[1]Department of General Practice and Elderly Care medicine, University Medical Center Groningen, Groningen, The Netherlands
[2]Department of Obstetrics and Gynaecology, University Medical Centre Groningen, Groningen, The Netherlands

Correspondence to
Dr Anne Martina Maria Loohuis; a.m.m.loohuis@umcg.nl

ABSTRACT

Objective To develop a prediction model and illustrate the practical potential of personalisation of treatment decisions between app-based treatment and care as usual for urinary incontinence (UI).

Design A prediction model study using data from a pragmatic, randomised controlled, non-inferiority trial.

Setting Dutch primary care from 2015, with social media included from 2017. Enrolment ended on July 2018.

Participants Adult women were eligible if they had ≥2 episodes of UI per week, access to mobile apps and wanted treatment. Of the 350 screened women, 262 were eligible and randomised to app-based treatment or care as usual; 195 (74%) attended follow-up.

Predictors Literature review and expert opinion identified 13 candidate predictors, categorised into two groups: Prognostic factors (independent of treatment type), such as UI severity, postmenopausal state, vaginal births, general physical health status, pelvic floor muscle function and body mass index; and modifiers (dependent on treatment type), such as age, UI type and duration, impact on quality of life, previous physical therapy, recruitment method and educational level.

Main outcome measure Primary outcome was symptom severity after a 4-month follow-up period, measured by the International Consultation on Incontinence Questionnaire the Urinary Incontinence Short Form. Prognostic factors and modifiers were combined into a final prediction model. For each participant, we then predicted treatment outcomes and calculated a Personalised Advantage Index (PAI).

Results Baseline UI severity (prognostic) and age, educational level and impact on quality of life (modifiers) independently affected treatment effect of eHealth. The mean PAI was 0.99±0.79 points, being of clinical relevance in 21% of individuals. Applying the PAI also significantly improved treatment outcomes at the group level.

Conclusions Personalising treatment choice can support treatment decision making between eHealth and care as usual through the practical application of prediction modelling. Concerning eHealth for UI, this could facilitate the choice between app-based treatment and care as usual.

Trial registration number NL4948t.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This is the first study to demonstrate the practical potential of prediction modelling to support decisions to personalise treatment when choosing between eHealth and care as usual.
⇒ The modifiers of treatment effect of eHealth for urinary incontinence in our sample (age, educational level and impact on quality of life) can be easily reproduced in clinical practice.
⇒ This study is based on data from a pragmatic randomised controlled trial with a representative first-line population, which makes the data well suited to developing a prediction model for personalising treatment decisions.
⇒ Despite a thorough search for candidate predictors, those interacting with eHealth treatment could have been missed, especially given that literature on this topic is scarce.
⇒ Our final model and the Personalised Advantage Index show moderate predictive performance and still requires further development and validation in a large primary care sample.

INTRODUCTION

Randomised controlled trials (RCTs) provide evidence of treatment effects at a group level, but they fail to provide the individual-level predictive information needed to optimise treatment in a given patient. This is especially relevant when two treatments show only marginal differences in effect at the group level, as occurs in a non-inferiority design, where the added value of personalised treatment decision might be greater.[1] A prediction model for treatment outcome,

based on a patient's individual characteristics, may facilitate the personalisation of treatment decisions. Different approaches to the development of clinical decision support tools informed by prediction models have been published in epidemiological and statistical literature, being developed for various disorders.[2–4] In mental healthcare, where treatment options for depression often show comparable effectiveness and marked individual variability, the Personalised Advantage Index (PAI) has shown utility.[4 5] This method predicts individualised outcomes for the treatment received (factual) and its alternatives (counterfactual), with the difference between these called the PAI. In this way, the optimal treatment and the magnitude of its predicted advantage can be quantified for a given patient. The PAI model accounts for patient characteristics that predict outcomes both irrespective of, and interacting with, the type of treatment.

The effectiveness of eHealth is often demonstrated as 'non-inferior' to a traditional treatment option, which is considered acceptable because of potential advantages unrelated to effectiveness, such as improved accessibility, privacy or cost savings.[6] However, treatment responses can vary widely at the individual level even when there is non-inferiority at the group level. For example, we demonstrated that an app-based treatment for female urinary incontinence (UI) was non-inferior to care as usual at a group level, but we equally found that individual outcomes at follow-up varied from 'much worse' to 'very much better' in both treatment groups.[7] Previously, higher age, treatment expectations and disease severity were reported to predict better outcomes for UI when using eHealth.[8 9] Although a given patient and caregiver could weigh these separate characteristics when making a treatment decision, it would be much more informative to know what specific outcomes one can expect from the available treatment options. We are unaware of the PAI having been applied to treatment decisions concerning eHealth.

In this study, we used existing RCT data to develop a prediction model and illustrate how the personalisation of treatment decisions affects the choice between app-based treatment and care as usual.[10] We also studied the practical potential of this approach in women with stress, urgency or mixed UI. First, we built a prediction model for the outcomes of app-based treatment and care as usual for UI, and we used this to predict outcomes given the actual treatment received (factual) and the hypothetical outcome of the treatment that was not received (counterfactual). Second, we used the PAI to identify the optimal treatment and to quantify its added benefit in individual participants. Third, we assessed the clinical relevance of any benefit and whether using the PAI improved treatment outcomes at the group level.

METHODS

Data source and study design

We used data from the URinControl-trial, a pragmatic, non-inferiority,RCT of women with stress, urgency or mixed UI who received either app-based treatment or care as usual via their general practitioner (GP). The trial design, the development and content of the app, and the clinical results have been published previously.[7 10] The original trial reported the non-inferiority of app-based treatment to care as usual at a group level. Baseline characteristics and outcome measures were based on data collected through validated questionnaires and a physical examination by a GP trainee. In this study, we use these data to build a prediction model, predict treatment outcomes at an individual level and calculate the PAI.

Participants

Participant enrollment took place from July 2015 to July 2018, with follow-up ending on 20 December 2018. We recruited participants in the north of the Netherlands via 88 GPs from 31 practices, and through social media and the lay press. Adult women were eligible if they had ≥2 episodes of self-reported stress, urgency or mixed UI per week, a wish to be treated, and access to a smartphone or tablet. The exclusion criteria were as follows: urinary tract infection, overflow or continuous UI, indwelling urinary catheter, urogenital malignancy, pregnancy or recent childbirth (<6 months ago), treatment for UI in the previous year, previous surgery for UI, terminal or serious illness, and cognitive impairment, psychiatric illness or the inability to complete a questionnaire in Dutch. The present analyses used the pretreatment data and the outcome data at 4 months for all women included in the original study.

Treatments

App-based treatment consisted of a step-by-step programme for the self-management of UI, with content based on relevant Dutch GP and international guidelines.[11 12] Care as usual comprised referral to the participant's GP, who was then free to engage in the following routine care: discussion of treatment options, such as pelvic floor muscle training and/or bladder training; prescribing of a pessary, drugs or absorbent products; and referral to a continence nurse, a pelvic physical therapist or secondary care.[12]

Outcome

The outcome predicted by the model was UI severity after 4 months of treatment, which we labelled the end-Urinary Incontinence Short Form (UISF) score. This continuous score was measured by the International Consultation on Incontinence Questionnaire, Urinary Incontinence Short Form (ICIQ-UISF),[13] a questionnaire measuring the self-reported frequency, severity and impact on daily life of UI. Scores ranged from 0 to 21, with higher scores indicating worse incontinence. Data analysts were blinded to the treatment arm at the time of analysis.

Predictors

We identified candidate predictors based on literature search and expert opinion. PubMed was searched for predictors of conservative UI treatment and eHealth

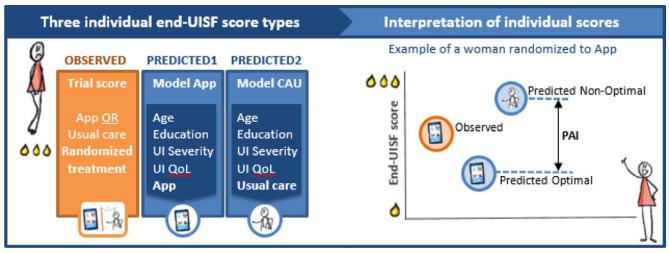

Figure 1 Calculating the Personal Advantage Index (PAI) from individual predicted scores legend: three individual outcome scores are possible for each patient (images, left): one observed and two predicted by the model. Optimal treatment is that with the lowest predicted outcome score (graph, right). The PAI is the difference between the optimal and non-optimal treatments. UISF, Urinary Incontinence Short Form; CAU, care as usual; UI, urinary incontinence; QoL, Quality of Life.

treatment (for UI and other conditions) (online supplemental table 1).[8 12 14–23] We also asked independent experts in eHealth and primary care (one pelvic floor physical therapist, two eHealth researchers, one GP with practical eHealth-experience and one GP/eHealth-researcher urogynecology) to list factors they considered relevant to the success or failure of app-based and usual treatment in women with UI, as well as to comment on the factors identified by literature search. This process identified 30 candidate predictors, as summarised in online supplemental table 2, from among which we selected 13 based on availability in our dataset and usability in clinical practice.

Based on the literature review and expert opinion, we prespecified the baseline characteristics either as potential prognostic factors or as potential modifiers. Prognostic factors predicted the outcome irrespective of treatment type, while modifiers predicted the outcome depending on the treatment received (the modifiers accounted for the difference in treatment effect in the counterfactual analysis).

Six baseline characteristics were selected as potential prognostic factors: UI severity, based on the ICIQ-UISF questionnaire (range 0–21 for low–high severity); postmenopausal state (yes or no); vaginal births (yes or no); general physical health, based on the EQ-5D-5L-VAS questionnaire (range 0–100 for low–high physical health); pelvic floor muscle function (normal, overactive or underactive) and body mass index. Seven baseline characteristics were selected as potential modifiers, or prescriptive factors, as described by DeRubeis *et al*[4]: age (years); UI type (stress or urgency), duration (years) and impact on quality of life (ICIQ-LUTS-QoL questionnaire, range 19–76 for low–high impact); previous physical therapy (yes or no); recruitment method (through GP or media) and educational level (iMTA-MCQ-PCQ questionnaire, rated as higher or lower). Predictors were measured at baseline and educational level was assessed during a follow-up.

Statistical analysis

We calculated the maximum number of parameters needed to build the model according to the guidance of Riley *et al*, based on a clinical prediction model with a continuous outcome and a known sample size of 262 participants.[24] Given a mean 9.9±3.3-point UI severity score from our trial population and an anticipated R^2 (0.6) from Nyström *et al*,[25] we calculated that a maximum of 28 parameters were needed to build the model.

Data were missing for the outcome measure and one predictor, which we accommodated by multiple imputation under the assumption of data being missing at random. We assessed the missing data mechanism by looking at patterns and predictors of missingness to substantiate assumptions of being missing at random or not at random.[26] All variables that predicted missingness of a certain variable were included in the imputation model together with all variables from the analyses.

All statistical analysis was performed using IBM SPSS for Windows, V.26.0 (IBM) and R. We performed multiple imputation in R, using the MICE package and constructed 50 imputed datasets.[27]

Development and validation

We developed a model to predict the treatment outcome based on prognostic factors and modifiers, assessed overoptimism by internal validation and applied this as a correction. This model was used to predict two outcomes for each participant: (1) for the actual treatment the patient received and (2) for the counterfactual treatment to which the patient was not allocated. We used these to construct and calculate the PAI, before assessing its benefits at individual and group levels.

Step 1: development of the prediction model

We investigated multicollinearity in the non-imputed dataset by correlation matrix, which revealed that none of the candidate predictors correlated highly with another (r>0.8).[28] As described by Kraemer *et al*, continuous predictors were centred by subtracting the median and dichotomous variables were set at 0.5 and −0.5.[29] The end-UISF score was predicted by linear regression, with the potential prognostic factors entered as main effects and the potential modifiers entered as both main effects and terms representing their interactions with treatment. We used a stepwise, backward elimination strategy, excluding variables from the model based on an alpha of 0.25.[30] Predictors selected in at least 50% of the imputed datasets were included in the final model.[28] We forced treatment type and the main effects of every included interaction into the final model irrespective of their significance.[31] Model performance was assessed by R^2, goodness-of-fit and calibration slope. The 95% CIs are reported as appropriate.

Step 2: internal validation of the prediction model

Stability of the regression coefficients, inclusion percentages and the mean adjusted R^2 was assessed across 500

Table 1 Baseline characteristics of all participants with urinary incontinence

	Characteristic	Total (n=262)	App treatment (n=131)	Care as usual (n=131)
Prognostic factors	Severity UI at baseline*	9.9±3.3	9.5±3.2	10.3±3.4
	Body mass index (kg/m^2)*	27.8±5.3	27.6±5.5	28.0±5.2
	Postmenopausal status, yes	123 (47.1%)	64 (49.2%)	59 (45.0%)
	Vaginal births,≥1	216 (82.8%)	111 (85.4%)	105 (80.2%)
	Pelvic floor muscle function			
	Normal activity	84 (32.1%)	44 (33.6%)	40 (30.5%)
	Overactive	44 (16.8%)	18 (13.7%)	26 (19.8%)
	Underactive	134 (50.8%)	69 (52.7%)	65 (49.6%)
	General physical health status*	74±20	73±20	75±21
Modifiers	Age, (years)	52.2±11.6	53.2±12.8	51.3±10.3
	Educational level, higher	107 (52.7%)	58 (54.2%)	49 (51.0%)
	Duration of UI (years)*	7 (4–14)	7 (4–15)	8 (4–13)
	UI impact on quality of life*	33.6±8.0	33.9±8.3	33.4±7.8
	Type of UI			
	Stress	180 (68.7%)	87 (66.4%)	93 (71.0%)
	Urgency	82 (31.3%)	44 (33.6%)	38 (29.0%)
	Previous physical therapy for UI, yes	66 (25.3%)	31 (23.8%)	35 (26.7%)
	Recruitment type			
	General practitioner	152 (58%)	76 (58.0%)	76 (58.0%)
	Lay press or social media	110 (42%)	55 (42.0%)	55 (42.0%)

Prognostic factors predict outcomes irrespective of treatment type. Modifiers predict outcomes dependent on the treatment (modifier). Values are presented as means±SD deviation, percentages or medians (IQR).
*N was lower: missing data of one baseline assessment, three baseline questionnaires and educational level were assessed at follow-up.
UI, urinary incontinence.

bootstrapped samples. We examined precision with the true error (mean observed score minus mean predicted score) and the SE. To correct for overoptimism, we equally applied uniform shrinkage to the final model coefficients.[32]

Step 3: construction of the PAI

Having determined the predictors of differential response, a model can be constructed to generate treatment recommendations, making use of the PAI.[4] Given our aim to study the practical potential of this index, we focused on clinical utility over technical detail (figure 1).[4 5]

Prediction of individual outcomes

For each patient, we predicted the end-UISF score for app-based treatment and for care as usual by completing the model twice with the patient's observed values: once with the value of app-based treatment (−0.5) and once with the value of care as usual (0.5). This predicted the end-UISF score for the treatments the patient received (factual score) and did not receive (counterfactual score). To predict individual end scores, we split the sample into five equal groups and used a linear regression model with sampling weights based on data from four groups to predict end scores in the targeted group. This fivefold cross-validation reduced the risk of overfitting by avoiding the inclusion of an individual's own data when estimating the relevant regression coefficients.

Interpretation of individual outcomes

Three end-UISF scores were documented for each patient: (1) the observed score after receiving the randomised treatment in the trial, (2) the predicted score for app-based treatment and (3) the predicted score for care as usual. The lowest of the two predicted scores was the optimal treatment (figure 1).

Step 4: assessment of the PAI

Assessment of individual benefit

For each patient, we then calculated the PAI as a measure of the benefit of one treatment over the other and assessed its magnitude (ie, clinical relevance). The PAI was the difference between the highest and lowest predicted score. Based on a difference of 1.58 points having previously been defined as the minimum clinically important difference for the ICIQ-UISF,[25] optimal treatment with a PAI higher than this was expected to have a noticeable benefit for the patient.

Assessment of improvement at the group level

Finally, we assessed whether treatment personalisation using the PAI significantly and substantially improved

Table 2 Final model predicting UI severity after 4 months (end-UISF score)

Variable	Unstandardised beta*	Inclusion frequency (%) 50 imputed sets	95% CI
Intercept	7.55		7.18 to 7.94
Treatment type, App (–0.5) or CAU (0.5)†	–0.07	100	–0.82 to 0.68
Age, years†	–0.01	2	–0.04 to 0.02
Educational level, lower (–0.5) or higher (0.5)†	0.01	0	–0.72 to 0.75
UI severity $_{\text{at baseline}}$	0.56	100	0.42 to 0.74
Impact of UI on quality of life†	0.08	100	0.02 to 0.15
Age×treatment type	0.06	92	–0.01 to 0.12
Educational level*treatment type	1.59	96	0.18 to 3.08
Impact on quality of life*treatment type	0.07	58	–0.02 to 0.17

Median centering of continuous values: age, 51.45; UI severity, 10; impact of UI on quality of life, 32.
*Uniform shrinkage was applied on beta with factor 0.98.
†Treatment type and the main interaction effects (age, educational level, impact on quality of life) were forced into the backward selection procedure irrespective of significance.
CAU, care as usual; UI, urinary incontinence; UISF, Urinary Incontinence Short Form.

treatment outcomes at a group level (ie, the usefulness of the PAI as a tool to improve treatment selection and thereby effectiveness). Using the observed outcome scores from the trial, we compared patients who randomly received an optimal treatment with those who randomly received a non-optimal treatment. Randomisation for this comparison allowed causal interpretation at the group level because we built the model on a separate selection of participants (using fivefold cross-validation) and because the predicted end-UISF scores were not tied to the randomisation or treatment received.

Patient and public involvement

We involved patients, the public and professionals from the start of the study.[10] Each group provided feedback on the study design, assessed the app in the development phase, provided feedback during the trial phase and have been informed of previous results by email. They have been informed of previous results by email and will be informed of the current results after publication. To facilitate the process of informing the public, we have produced plain-language summaries in text, illustration and video formats for dissemination on social media and our website (www.urincontrol.nl).

RESULTS

Participants

We included data for 262 women who participated in the trial (table 1). The only remarkable difference was a lower severity of UI in the app-based treatment group. At 4 months, 195 women (74.4%) had reported end-UISF scores, resulting in 67 cases of missing data for the end-UISF score and educational level. The outcome variable was missing at random and predicted by a younger age, a higher body mass index and no prior treatment, but not by severity of incontinence.[7]

Development of the prediction model

Table 2 shows the variables included in the final model. The model explained 46% of the variance in predicting the outcome measure (R^2 0.46; 95% CI 0.36 to 0.55)). The mean difference between observed and predicted outcomes in the original data—that is, the goodness-of-fit—was 0.015 (95% CI −0.308 to 0.278), showing a calibration intercept at −0.06 and a calibration slope of 1.01 (online supplemental figure 1). A lower end-UISF score, indicating a better treatment outcome, was predicted by care as usual, lower baseline severity of UI, and lower impact of UI on quality of life. Success of app-based treatment was associated with higher age, higher impact of UI on quality of life and higher educational level. Success of care as usual was associated with lower educational level.

Internal validation

Regression coefficients and inclusion percentages across the bootstrapped samples were stable. The mean R^2 (% explained variance) was 0.455 (95% CI 0.357 to 0.547) after bootstrapping (online supplemental table 3). The uniform shrinkage factor calculated by bootstrapping was small with a factor of 0.98 (table 2). The true error was 1.85 (values plotted in online supplemental figure 2) and the SE was 0.15.

Personalised Advantage Index

At the group level, the mean change in the unimputed observed UISF score after 4 months indicated a symptoms improvement of −2.35±3.05 points. The change of symptom score varied from −15 to 6 points among patients.

Individual observed and predicted outcome scores

The observed scores showed a mean end-UISF of 7.58±3.46 points (range 0–18). The mean predicted optimal and non-optimal scores per patient were 7.15±2.46 points

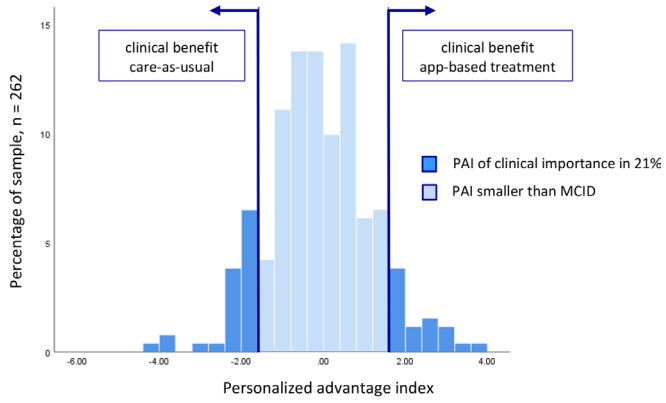

Figure 2 Individual PAI scores and their clinical relevance legend: the figure shows the individual variability of treatment response above and below the minimum clinical important difference of 1.58. PAI, Personalised Advantage Index; MCID, minimum clinically important difference.

(range 0–13) and 8.14±2.52 points (range 2–16), respectively.

Individual benefit

The PAI showed a mean benefit of 0.99±0.79 points for the optimal treatment over the non-optimal treatment, which ranged from 0.02 to 4.21 points at the individual level (figure 2). This difference was clinically relevant at ≥1.58 points for 55 patients (21%).[25] Online supplemental table 4 shows a comparison of baseline characteristics between 55 patients with a clinically relevant PAI and the other patients.

Improvement on group level

Finally, we compared the observed trial outcomes of patients receiving optimal (n=135; 51%) and non-optimal (n=127; 49%) treatments, which had mean scores of 7.01±3.33 points and 8.20±3.51 points, respectively. The observed difference in means between the randomised optimal and non-optimal treatment was statistically significant with a mean difference of 1.19 points (95% CI 0.355 to 2.021).

DISCUSSION

Statement of principal findings

We illustrated a method for predicting optimal treatment and for quantifying its benefit compared with non-optimal treatment at the individual patient level when eHealth is being considered for UI management. Four baseline characteristics, namely UI severity (a prognostic factor) and age, educational level and impact of UI on quality of life (modifiers), were identified as suitable for helping with decisions in this model. The mean advantage according to the PAI was 0.99 points, and it exceeded the threshold for clinical relevance of 1.58 in 21% of individuals. Applying the PAI to facilitate decision making also significantly improved treatment outcomes at the group level, which may be relevant when considering other measures of quality with this treatment, such as cost-effectiveness. To our knowledge, this is the first study to have translated established methods for predicting treatment outcomes from mental health and somatic disease settings[2–4] to an eHealth setting.

Strengths and weaknesses of the study

We developed a model for predicting the treatment option most likely to improve UI symptoms in individuals and assessed the clinical relevance of that prediction. The use of data from a pragmatic RCT with a representative first-line population make that data well suited to developing a prediction model for personalising treatment decisions.[2] The model is also usable because the predictors are both easily reproduced (age and answers to validated questions) and readily available in clinical practice.[1] Other strengths are the use of a patient-centred outcome measure, the selection of predictors based on literature and expert opinion, the inclusion of both prognostic factors (treatment independent) and modifiers (treatment dependent), the power of the prediction study and the minimal overfitting of the model (shrinkage factor=0.98).[28]

There are several important limitations that should also be considered. First, the prediction model required external validation via a sample comparing app-based treatment to care as usual; however, such a sample does not exist for UI. Internal validation only confirmed the stability of development and performance of the model. Second, the explained variance of 46% was moderate, being similar to that reported for other eHealth models for UI (range 30%–61.4%).[8 9] Third, the true error was 1.85 points in our sample, which is larger than the mean PAI (0.99 points) and threshold for clinical relevance (1.58 points), possibly indicating low precision for personalised predictions and probably affecting the estimation of the magnitude of an individual's advantage. Performance and precision could be increased by adding stronger predictors that interact with eHealth treatment to the model. Despite a thorough search for candidate predictors, those interacting with eHealth treatment could have been missed, especially given that literature on this topic is scarce. Finally, we could not include some variables identified by literature search and expert opinion, such as the eHealth literacy and treatment expectations of participants, because these were missing from our dataset.

Strengths and weaknesses in relation to other studies and key differences

Our approach of predicting the best treatment option for an individual shows important features and challenges of predictive heterogeneity of treatment effect analysis. Our methods fall under the larger 'effect modelling' approach (as opposed to risk modelling), meaning that a term for treatment assignment and interactions between treatment and baseline covariates are included in the prediction model, as described in a recent State of the art review.[33] Disaggregation of the overall results in our

pragmatic trial appeared to improve treatment effects on the individual and population level. However, we also encountered barriers linked to effect modelling described in the review; for example, a priori modifiers that were not yet well established, limited statistical power, multiple testing problems and the evaluation whether a particular prediction-decision strategy (in our trial for example clinical relevance threshold) would optimise the net benefit in the general population. The PAI predicted clinically relevant improvement in 60% of patients with depression for the choice between antidepressant medication and cognitive behavioural therapy.[4 5] Compared with the mental health setting,[5] clinically relevant improvement was only predicted in 21% of our cohort, possibly because there is less existing knowledge or less identifiable variation in treatment effect for UI. To date, however, we are unaware of any other studies having assessed and compared the interaction of predictors for an eHealth treatment and care as usual. We believe this type of assessment is essential to strengthen treatment-specific outcome predictions and to optimise clinical decision making for personalised medicine. Lindh *et al*, for example, showed that higher age predicted greater treatment success with eHealth for UI,[8] but this only considered their total sample (internet-based treatment and controls) and did not assess treatment interactions. In our study, the interaction of age with treatment type implies that a higher age may favour app-based treatment over care as usual. This new information is relevant to both researchers and clinicians because it runs counter the general expectation that eHealth is better suited to younger patients.

Possible mechanisms and explanations for findings

The predictors in our model should not be interpreted as strict causal or etiological factors for UI symptoms. The present data analysis was designed specifically to identify a set of variables that had high predictive accuracy in combination, rather than to unravel the causal factors influencing UI symptom severity at follow-up. However, the predictors that remained in the model had high a priori predictive value and are plausible causal factors.

In the developed model, increased age, educational level and impact of UI on quality of life predicted a better treatment outcome for app-based treatment compared with care as usual. Educational level had the greatest modifying effect, with a higher level associated with benefit from app-based treatment and a lower level associated with care as usual. This is likely to reflect differences in health, eHealth literacy and self-efficacy, but it could also reflect the app's design (eg, there are lengthy sections of text or instructions that may be too difficult to understand) or better adaptability by a given healthcare professional to a patient's need for support.

Other studies indicate that lower health literacy is associated with poorer health outcomes and with difficulties using eHealth effectively.[34 35] A mobile app has the potential to be tailored to specific users, such as those with low literacy and may bridge this gap.[36] Given that the content of our app was not tailored to users with low literacy, we will develop it further to improve its availability, readability and usability. Furthermore, we plan to add improved technological and practical support, specifically targeting users with low literacy.

Potential implications for clinicians or policy-makers

Prediction modelling at a group level only allows patients and caregivers to guess how a given characteristic influences treatment outcomes at an individual level. The PAI helps to correct this by quantifying the expected outcomes and benefits of an optimal treatment over its alternative given an individual's characteristics. The results are easy to interpret and can inform decisions immediately.

Our model requires further development and validation, but in the meantime, we believe it can be of use in clinical practice. Indeed, using the tool is certainly superior to the current situation where no support is available and where its use will pose little risk to the patient if the prediction is wrong (ie, the options are non-inferior, but its use could improve outcomes).[1] The PAI could also be implemented in clinical practice with ease, either within the app itself or on a patient information website, where the necessary prognostic factors and modifiers can be entered by users to predict the option most likely to be of benefit. This approach could be especially helpful for shared decision making and could be used to guide patients who wish to consider using a freely accessible eHealth intervention when they experience barriers seeking help from a caregiver. Currently, these patients often start to use an available app with no knowledge of what to expect.

Unanswered questions and future research

We missed important predictors by not anticipating the present analysis at the inception of our trial. Therefore, we recommend that eHealth researchers consider adding a method for personalising treatment decisions to allow them to consider and include all relevant predictors. This is especially relevant if researchers are conducting a (pragmatic) RCT, which otherwise provides the perfect foundation for this method.[1 37] If more eHealth researchers conducted similar research, we may see a large-scale improvement in clinical decision making, treatment outcomes and our knowledge of the predictors that interact with eHealth treatment.

External validation of the model in the present study is needed, but this is complicated by the lack of a suitable sample. More validation samples may be available for researchers applying this method to other eHealth settings where there is a greater body of research comparing eHealth to care as usual (eg, obesity and diabetes).[36] Finally, an impact study comparing treatment outcomes for groups with and without this decision support tool would be of interest.[1]

CONCLUSION

Prediction modelling can directly support decisions to personalise treatment when choosing between eHealth and care as usual. We applied this principle to an eHealth treatment for UI and, despite our model having only moderate predictive performance and still requiring external validation, we demonstrated its practical potential.

Twitter Anne Martina Maria Loohuis @loohuisanne and Marco H Blanker @Marco_Blanker

Acknowledgements We thank the participating general practices for their ongoing support, as well as all the participants for their invaluable contributions to this study. Special thanks is due to patients involved in the development of the app and to those at the Bekkenbodem4all patient organisation. Finally, we thank Dr Robert Sykes (www.doctored.org.uk) for providing editorial services.

Contributors AMML collected the data, did the analysis and wrote the paper; HB designed the study and contributed to the analysis and writing of the paper; NW collected the data and contributed to the writing of the paper; JD designed the study, acquired the funding, and contributed to the writing of the paper; AGGAM assisted in the study design, the content of the app and contributed to the writing; MYB assisted in the study design and contributed to the writing of the paper; MHB designed the study, acquired the funding, was project leader, contributed to the analysis and contributed to the writing of the paper; HvdW contributed to the analysis and contributed to the writing of the paper; and MHB is guarantor. The corresponding author attests that all listed authors meet the criteria for authorship and that no others meeting the criteria have been omitted.

Funding This work was supported by a grant from ZonMw, The Dutch Organisation for Health Research and Development (project number: 837001508) and subfunded by a grant from the P.W. Boer foundation. The study won the Professor Huygen award 2016 for best study proposal in general practice, which included additional funding.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Consent obtained directly from patient(s)

Ethics approval This study involves human participants and was approved by the Medical Ethical Review board of the University Medical Center Groningen (Netherlands) (METc-number: 2014/574) approved this study. All participants gave written informed consent. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Individual participant data that underlie the results reported in this article, after the identification, will be available including data dictionaries. Data are available to investigators who provide a methodologically sound proposal for analyses to achieve aims in the approved proposal. The available period begins 9 months after publication and ends 36 months following article publication and after approval of a proposal. Additional information available is the study protocol. There are no additional restrictions on the use of the data. The data are available from MHB, m.h.blanker@umcg.nl.

ORCID iDs

Anne Martina Maria Loohuis http://orcid.org/0000-0001-6582-1502
Huibert Burger http://orcid.org/0000-0001-6484-0349
Marco H Blanker http://orcid.org/0000-0002-1086-8730
Henk van der Worp http://orcid.org/0000-0001-5545-4155

REFERENCES

1 Moons KGM, Altman DG, Vergouwe Y, *et al*. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606–90.
2 van Diepen M, Ramspek CL, Jager KJ, *et al*. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrol Dial Transplant* 2017;32:ii1–5.
3 Farooq V, van Klaveren D, Steyerberg EW, *et al*. Anatomical and clinical characteristics to guide decision making between coronary artery bypass surgery and percutaneous coronary intervention for individual patients: development and validation of SYNTAX score II. *Lancet* 2013;381:639–50.
4 DeRubeis RJ, Cohen ZD, Forand NR, *et al*. The personalized advantage index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One* 2014;9:e83875–8.
5 van Bronswijk SC, DeRubeis RJ, Lemmens LHJM, *et al*. Precision medicine for long-term depression outcomes using the personalized advantage index approach: cognitive therapy or interpersonal psychotherapy? *Psychol Med* 2021;51:279–89.
6 Kummervold PE, Johnsen J-AK, Skrøvseth SO, *et al*. Using noninferiority tests to evaluate telemedicine and e-health services: systematic review. *J Med Internet Res* 2012;14:e132.
7 Loohuis AMM, Wessels NJ, Dekker JH, *et al*. App-Based treatment in primary care for urinary incontinence: a pragmatic, randomized controlled trial. *Ann Fam Med* 2021;19:102–9.
8 Lindh A, Sjöström M, Stenlund H, *et al*. Non-face-to-face treatment of stress urinary incontinence: predictors of success after 1 year. *Int Urogynecol J* 2016;27:1857–65.
9 Nyström E, Asklund I, Sjöström M, *et al*. Treatment of stress urinary incontinence with a mobile APP: factors associated with success. *Int Urogynecol J* 2018;29:1325–33.
10 Loohuis AMM, Wessels NJ, Jellema P, *et al*. The impact of a mobile application-based treatment for urinary incontinence in adult women: design of a mixed-methods randomized controlled trial in a primary care setting. *Neurourol Urodyn* 2018;37:2167–76.
11 Damen-van Beek Z, Teunissen D, Dekker JH, *et al*. [Practice guideline 'Urinary incontinence in women' from the Dutch College of General Practitioners]. *Ned Tijdschr Geneeskd* 2016;160:D674.
12 Abrams P, Andersson K-E, Apostolidis A, *et al*. 6Th International consultation on incontinence. recommendations of the International scientific Committee: evaluation and treatment of urinary incontinence, pelvic organ prolapse and faecal incontinence. *Neurourol Urodyn* 2018;37:2271–2.
13 Avery K, Donovan J, Peters TJ, *et al*. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol Urodyn* 2004;23:322–30.
14 Wyman J, Allen A, Hertsgaard L. Effect of smoking cessation on overactive bladder symptoms in adults: a pilot study. *Neurourol Urodyn* 2014;33.
15 Wells MJ, Jamieson K, Markham TCW, *et al*. The effect of caffeinated versus decaffeinated drinks on overactive bladder: a double-blind, randomized, crossover study. *J Wound Ostomy Continence Nurs* 2014;41:371–8.
16 Burgio KL, Goode PS, Locher JL, *et al*. Predictors of outcome in the behavioral treatment of urinary incontinence in women. *Obstet Gynecol* 2003;102:940–7.
17 Cammu H, Van Nylen M, Blockeel C, *et al*. Who will benefit from pelvic floor muscle training for stress urinary incontinence? *Am J Obstet Gynecol* 2004;191:1152–7.
18 Kim H, Yoshida H, Suzuki T. The effects of multidimensional exercise treatment on community-dwelling elderly Japanese women with stress, urge, and mixed urinary incontinence: a randomized controlled trial. *Int J Nurs Stud* 2011;48:1165–72.
19 Dumoulin C, Bourbonnais D, Morin M, *et al*. Predictors of success for physiotherapy treatment in women with persistent postpartum stress urinary incontinence. *Arch Phys Med Rehabil* 2010;91:1059–63.
20 Hendriks EJM, Kessels AGH, de Vet HCW, *et al*. Prognostic indicators of poor short-term outcome of physiotherapy intervention

in women with stress urinary incontinence. *Neurourol Urodyn* 2010;29:336–43.
21 Yoo E-H, Kim Y-M, Kim D. Factors predicting the response to biofeedback-assisted pelvic floor muscle training for urinary incontinence. *Int J Gynaecol Obstet* 2011;112:179–81.
22 Schaffer J, Nager CW, Xiang F, *et al*. Predictors of success and satisfaction of nonsurgical therapy for stress urinary incontinence. *Obstet Gynecol* 2012;120:91–7.
23 Vitacca M, Montini A, Comini L. How will telemedicine change clinical practice in chronic obstructive pulmonary disease? *Ther Adv Respir Dis* 2018;12:175346581875477–19.
24 Riley RD, Ensor J, Snell KIE, *et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:1–12.
25 Nyström E, Sjöström M, Stenlund H, *et al*. ICIQ symptom and quality of life instruments measure clinically relevant improvements in women with stress urinary incontinence. *Neurourol Urodyn* 2015;34:747–51.
26 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Belmont, California: Lifetime Learning Publications, 1982.
27 van Buuren S. *Package "mice": Multivariate Imputation by Chained Equations*. CRAN Repos, 2019.
28 Steyerberg. Clinical prediction models: a practical approach to development, validation and updating. In: *Kybernetes*. Vol. 38. New York, NY: Springer-Verlag, 2009.
29 Kraemer HC, Blasey CM. Centring in regression analyses: a strategy to prevent errors in statistical inference. *Int J Methods Psychiatr Res* 2003;13:2647500.
30 Walker E. Regression Modeling Strategies. In: *Technometrics*. , 2003: 45, 170.
31 Steyerberg EW. Clinical Prediction Models. In: *Statistics for biology and health*. 2nd edition, 2019.
32 Van Calster B, van Smeden M, De Cock B, *et al*. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: simulation study. *Stat Methods Med Res* 2020;29:3166–78.
33 Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 2018;363:k4245.
34 Berkman ND, Sheridan SL, Donahue KE, *et al*. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;155:97–107.
35 Jensen JD, King AJ, Davis LA, *et al*. Utilization of Internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. *J Aging Health* 2010;22:804–26.
36 Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. *Patient Educ Couns* 2017;100:1073–82.
37 Liu JLY, Wyatt JC. The case for randomized controlled trials to assess the impact of clinical information systems. *J Am Med Inform Assoc* 2011;18:173–80.