Development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ)

Polly Duncan,1 Mairead Murphy,1 Mei-See Man,2 Katherine Chaplin,2 Daisy Gaunt,3 Chris Salisbury

ABSTRACT

Objective To develop and validate a new scale to assess treatment burden (the effort of looking after one’s health) for patients with multimorbidity.

Design Mixed-methods.

Setting UK primary care.

Participants Content of the Multimorbidity Treatment Burden Questionnaire (MTBQ) was based on a literature review and views from a patient and public involvement group. Face validity was assessed through cognitive interviews. The scale was piloted and the final version was tested in 1546 adults with multimorbidity (mean age 71 years) who took part in the 3D Study, a cluster randomised controlled trial. For each question, we examined the proportion of missing data and the distribution of responses. Factor analysis, Cronbach’s alpha, Spearman’s rank correlations and longitudinal regression assessed dimensional structure, internal consistency reliability, construct validity and responsiveness, respectively. We assessed interpretability by grouping the global MTBQ scores into zero and tertiles (>0) and comparing participant characteristics across these categories.

Results Cognitive interviews found good acceptability and content validity. Factor analysis supported a one-factor dimensional structure, internal consistency reliability, construct validity and responsiveness, respectively. The MTBQ score had a positive association with a comparator treatment burden scale ($r_s = 0.58, P<0.0001$) and with self-reported disease burden ($r_s = 0.43, P<0.0001$), and a negative association with quality of life ($r_s = -0.36, P<0.0001$) and self-rated health ($r_s = -0.36, P<0.0001$). Female participants, younger participants and participants with mental health conditions were more likely to have high treatment burden scores. Changes in MTBQ score over 9-month follow-up were associated, as expected, with changes in measures of quality of life (EuroQol five dimensions, five level questionnaire) and patient-centred care (Patient Assessment of Chronic Illness Care).

Conclusion The MTBQ is a 10-item measure of treatment burden for patients with multimorbidity that has demonstrated good content validity, construct validity, reliability and responsiveness. It is a useful research tool for assessing the impact of interventions on treatment burden.

Trial registration number ISRCTN06180958.

INTRODUCTION

Treatment burden is a patient’s perception of the effort required to self-manage their medical conditions and the impact that this has on their general well-being.1 This includes complex medication regimens, coordinating healthcare appointments, making lifestyle changes and self-monitoring.

This is particularly relevant to patients with multimorbidity (having multiple long-term conditions). Associated with the ageing population, multimorbidity has become the norm, affecting over two-thirds of adults attending general practice.2 Current health policy envisages greater support for patients to self-manage their chronic medical conditions. However, the time and energy this requires of patients can be overwhelming.

In order to understand the impact of treatment burden, and particularly to assess the effects of interventions that might increase or decrease burden, a valid patient-reported outcome measure (PROM) is essential. There are four existing PROMs that measure aspects of treatment burden for patients with multimorbidity,4–8 all of which have important limitations. The 13-question Treatment Burden Questionnaire (TBQ) was originally developed in French, and subsequently a revised 15-question English version was tested.15 Some of the content is healthcare system-specific and the wording...
is relatively complex, perhaps reflecting the fact that the
English version was tested in a relatively young and highly
educated population of volunteers recruited from the
‘Patients like me’ website (mean age 51 years, 78% with
college education), not all of whom had multimorbidity.1

The Patient Experience with Treatment and Self-man-
agement (PETS) PROM was recently developed in the
USA and includes 48 questions grouped under nine sepa-
rate domains of treatment burden.5 While this measure
is comprehensive, its length is a limitation. The Multi-
morbidity Illness Perceptions Scale (MULTIPLes) was
developed and validated in elderly patients (mean age
70 years) with multimorbidity and includes a six-question
treatment burden subscale and a three-question activity
limitation subscale.7 This measure is brief but omits
several important aspects of treatment burden. Similarly,
the 11-question Healthcare Task Difficulty (HCTD) ques-
tionnaire was designed to measure only one aspect of
treatment burden.6

The purpose of this study was to develop and validate
a new concise measure of treatment burden for patients
with multimorbidity.

METHODS

Study setting
This questionnaire was developed and validated as part of
the 3D Study, a multicentre, cluster randomised control
trial that aims to improve the management of patients
with multimorbidity within primary care.9 Participants
aged 18 years or older with three or more of the long-
term conditions included in the 2014 UK Quality and
Outcomes Framework were recruited from 33 general
practices in three areas of the UK.

Development of the questionnaire
We reviewed the literature on the concept and measure-
ment of treatment burden in multimorbidity using PubMed in July 2014. We identified a number of rele-
vant qualitative studies10-12 and three relevant existing
PROMs that were not specific to a particular medical
condition. These were the TBQ,4 5 the MULTIPLes,7
and the HCTD questionnaires.6 A further measure, the
PETS scale, was published later.8 We identified relevant
domains for the PROM by reviewing the three existing
PROMs against a framework of treatment burden which
had been developed following qualitative interviews and
focus groups.1 We then sought the views from a patient
and public involvement (PPI) group of eight patients
with multimorbidity formed for the purpose of the 3D
Study, discussing the concept of treatment burden, the
existing measures, the treatment burden framework
and the domains of treatment burden to be included
in the questionnaire. We then developed a draft ques-
tionnaire with 12 questions and undertook two rounds
of cognitive interviews with eight PPI group members to
improve the face and content validity of the scale (on-line
supplementary appendix A).13 Participants were asked
to ‘think aloud’13 as they completed the questionnaire
commenting on the reasoning behind their ratings;
and perceived question meaning, the layout, title, intro-
duction and general wording. They also gave their own
examples of treatment burden and reflected on whether
these would be captured by the questionnaire. Modifi-
cations to the questionnaire were made between the
two rounds and an additional question was added about
accessing healthcare during the evenings and weekends
(see the Results section). Following written consent,
the interviews were audio-taped and field notes were
taken. The second round of cognitive interviews led to
only minor changes to the questionnaire with no new
insights emerging. A debriefing meeting was held with
PPI members and final changes to the 13-question ques-
tionnaire were made.

Recruitment, data collection and measures
Data were collected in two related studies, the cross-sect-
ional 3D pilot study and the longitudinal main 3D Study,
a cluster randomised controlled trial. The 13 candidate
questions were included in a questionnaire, named
the Multimorbidity Treatment Burden Questionnaire
(MTBQ). Sociodemographic information (see table 1)
was collected at baseline in both the pilot and main
studies. Details of participants’ medical conditions were
collected from their family practice computer records.
Measures of health-related quality of life (EQ-5D-5L),14
self-rated health (single question item), self-reported
disease burden (Bayliss)15 and patient-centred care
(Patient Assessment of Chronic Illness Care (PACIC))16
were collected at baseline and 9 months in both the
pilot and main 3D studies. Following a review of existing
measures and discussion with the PPI group, the HCTD
questionnaire was included in the pilot study question-
naire as the best comparator for the MTBQ. A key reason
for choosing this measure was the simple wording and
brevity. This was felt to be important because many of the
participants of the 3D Study were older people and some
had low literacy levels.

The questionnaire was sent to participants by post. For
non-responders, a reminder letter was sent 10–14 days
later, and a second reminder phone call was made 10–14
days after this.

Analysis
Data were analysed using STATA V.14. We generated
descriptive statistics of participant characteristics for
the pilot and main studies. The pilot study data were used to
test the prespecified hypothesis of a positive association
between global MTBQ score and HCTD score. The main
study data were used for the remainder of the analysis.

We tested the psychometric properties of the question-
naire against the minimum standards set out by the Inter-
national Society for Quality of Life Research (ISOQOL).17
The analysis plan and results are described in relation to
ISOQOL’s six recommended standards.
1. Conceptual and measurement model

Conceptual framework
See the ‘Development of the questionnaire’ section.

Question properties
To assess the properties of the questions, we examined the proportion of missing data and ‘does not apply’ responses and the distribution of responses. Responses of ‘not difficult’ or ‘does not apply’ were scored as 0. Floor and ceiling effects of the MTBQ were compared with the HCTD. Questions with a proportion of ‘does not apply’ responses greater than 40% were removed and excluded from the analysis.

Dimensionality
To examine the dimensionality of the scale, we performed factor analysis. This is a statistical technique used to reduce a larger number of items into a smaller number of common factors that reflect shared variance. Items that share a lot of variance should have high ‘loadings’ (correlation between the item and the factor) and low uniqueness (variance that is unique to the item, not common to the factor). Loading of at least 0.4 and uniqueness of less than 0.6 are acceptable. The number of factors extracted was decided by a combination of Kaiser’s rule (eigenvalues greater than 1),20 the scree plot18 and by interpretability of domains.

2. Reliability
To test internal consistency reliability, we examined the interitem correlation matrix and calculated Cronbach’s alpha, a measure of consistency between the items in a scale. Interitem correlations between 0.2 and 0.4 were deemed ideal.21 A Cronbach’s alpha of 0.7–0.9 was acceptable.22

3. Validity
Content validity
The content validity of the questionnaire was tested iteratively using cognitive interviews (see the ‘Development of the questionnaire’ section).

Construct validity
Each question was scored as follows: 0 (not difficult/does not apply), 1 (a little difficult), 2 (quite difficult), 3 (very difficult) and 4 (extremely difficult). Participants were excluded if more than 50% of their responses were missing. To calculate a global score, each participant’s average score was calculated from the questions answered and multiplied by 25 to give a score from 0 to 100.

Construct validity was examined by testing five prespecified hypotheses: first, a positive association between global MTBQ score and global HCTD score; second, a negative association between global MTBQ score and health-related quality of life (EQ-5D-5L); third, a positive association between global MTBQ score and self-reported disease burden score; fourth, a positive association between global MTBQ score and number of self-reported comorbidities; and fifth, a negative association between global MTBQ and self-rated health (single question item). We applied Spearman’s rank correlation to test these hypotheses.

Responsiveness
According to the ISOQOL guidelines, responsiveness to change should be assessed.22 Due to the non-normal distribution of the global MTBQ score, standard methods to assess responsiveness to change such as calculating an effect size22 were not possible. We therefore tested the responsiveness of the global MTBQ score by assessing whether changes over time in measures of quality of life (EQ-5D-5L) and patient-centred care (PACIC) were inversely associated with changes in MTBQ as anticipated. We used a linear regression model of the standardised change in quality of life (EQ-5D-5L) score between baseline and 9 months on the standardised change in MTBQ between baseline and 9 months. These standardised change scores were calculated at the participant level by dividing the individual difference in 9-month and baseline MTBQ (or EQ-5D-5L) score by the SD of the overall MTBQ (or EQ-5D-5L) change score for all individuals. We then further adjusted this linear regression model in a subsequent analysis by age, gender, number of long-term conditions and individual participant deprivation level. All participants who died prior to the 9-month follow-up were given an EQ-5D-5L follow-up score of 0.

We then used the same model for MTBQ specified as above but included the standardised change in PACIC scores between baseline and 9-month follow-up, defined as previously, and subsequently further adjusted this model by the additional covariates as specified.

4. Interpretability of scores
The distribution of global MTBQ scores was examined and compared with the distribution of HCTD scores.

We assessed interpretability of the questionnaire by grouping the global MTBQ scores greater than 0 into tertiles. Four categories were generated: no burden (score 0), low burden (score 1–10), medium burden (10–22) and high burden (≥22). Participant characteristics and key outcome variables, including EQ-5D-5L, Bayliss disease burden score and self-rated health, were compared across these four categories. To test for associations between treatment burden score category and participant characteristics, we performed ordinal logistic regression of MTBQ group (four treatment burden categories) on each participant characteristic. We then further adjusted these ordinal logistic regression models by age, gender, number of comorbidities, age left full-time education and individual deprivation score.

5. Translation
Not applicable.

6. Demands on patient respondents and investigators
The effort required of patient respondents to complete the questionnaire was assessed during the cognitive interviews, and by reviewing the proportion of missing
Table 1  Participant characteristics (main study n=1546, pilot study n=143)

<table>
<thead>
<tr>
<th></th>
<th>Pilot study n/N* (%)</th>
<th>Main study n/N* (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean age (SD)</strong></td>
<td>74 (10)</td>
<td>71 (12)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤50</td>
<td>3 (2)</td>
<td>79 (5)</td>
</tr>
<tr>
<td>51–60</td>
<td>9 (6)</td>
<td>196 (13)</td>
</tr>
<tr>
<td>61–70</td>
<td>27 (19)</td>
<td>420 (27)</td>
</tr>
<tr>
<td>71–80</td>
<td>67 (47)</td>
<td>510 (33)</td>
</tr>
<tr>
<td>81–90</td>
<td>33 (23)</td>
<td>315 (20)</td>
</tr>
<tr>
<td>≥90</td>
<td>4 (3)</td>
<td>26 (2)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>65 (45)</td>
<td>763 (49)</td>
</tr>
<tr>
<td><strong>Number of comorbidities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>109 (76)</td>
<td>1234 (80)</td>
</tr>
<tr>
<td>4</td>
<td>23 (16)</td>
<td>277 (18)</td>
</tr>
<tr>
<td>5</td>
<td>10 (7)</td>
<td>31 (2)</td>
</tr>
<tr>
<td>6</td>
<td>1 (&lt;1)</td>
<td>4 (&lt;1)</td>
</tr>
<tr>
<td><strong>Comorbidities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>138 (97)</td>
<td>1445 (97)</td>
</tr>
<tr>
<td>Stroke/Transit ischaemic attack</td>
<td>35 (25)</td>
<td>527 (34)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>63 (44)</td>
<td>811 (52)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>83 (58)</td>
<td>464 (30)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease or asthma</td>
<td>58 (41)</td>
<td>770 (50)</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>6 (4)</td>
<td>76 (5)</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>46 (32)</td>
<td>529 (34)</td>
</tr>
<tr>
<td>Severe mental health problems†</td>
<td>2 (1)</td>
<td>66 (4)</td>
</tr>
<tr>
<td>Depression</td>
<td>26 (18)</td>
<td>560 (36)</td>
</tr>
<tr>
<td>Dementia</td>
<td>6 (4)</td>
<td>60 (4)</td>
</tr>
<tr>
<td>Learning disability</td>
<td>3 (2)</td>
<td>14 (1)</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>9 (6)</td>
<td>103 (7)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>14 (10)</td>
<td>157 (10)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White British</td>
<td>135/136 (99)</td>
<td>1502/1519 (99)</td>
</tr>
<tr>
<td><strong>Age left full-time education (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤14</td>
<td>22 (15)</td>
<td>154/1541 (10)</td>
</tr>
<tr>
<td>15 or 16</td>
<td>74 (52)</td>
<td>907/1541 (59)</td>
</tr>
<tr>
<td>17 or 18</td>
<td>25 (17)</td>
<td>222/1541 (14)</td>
</tr>
<tr>
<td>≥19</td>
<td>22 (15)</td>
<td>258/1541 (17)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully retired from work</td>
<td>113/139 (81)</td>
<td>1044/1501 (70)</td>
</tr>
</tbody>
</table>

Continued

Table 1  Continued

<table>
<thead>
<tr>
<th></th>
<th>Pilot study n/N* (%)</th>
<th>Main study n/N* (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deprivation score quartile‡</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower quartile</td>
<td>99/143 (69)</td>
<td>445/1079 (41)</td>
</tr>
<tr>
<td>Middle lower quartile</td>
<td>44/143 (31)</td>
<td>304/1079 (28)</td>
</tr>
<tr>
<td>Middle upper quartile</td>
<td>0</td>
<td>196/1079 (18)</td>
</tr>
<tr>
<td>Upper quartile</td>
<td>0</td>
<td>134/1079 (12)</td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower quartile</td>
<td>105/467 (22)</td>
<td>46/467 (10)</td>
</tr>
<tr>
<td>Middle lower quartile</td>
<td>46/467 (10)</td>
<td>156/467 (33)</td>
</tr>
<tr>
<td>Middle upper quartile</td>
<td>160/467 (34)</td>
<td></td>
</tr>
<tr>
<td><strong>Baseline scores of outcome measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean HCTD score§</td>
<td>1.14 (1.7, 143)</td>
<td>191 (12.4, 1458)</td>
</tr>
<tr>
<td>Mean self-reported disease burden score¶ (SD, N)</td>
<td>8 (3.2, 1543)</td>
<td></td>
</tr>
<tr>
<td>Mean number of self-reported conditions** (SD, N)</td>
<td>0.6 (0.3, 1542)</td>
<td></td>
</tr>
<tr>
<td>Mean quality of life score†† (SD, N)</td>
<td>2 (0.8, 1523)</td>
<td></td>
</tr>
<tr>
<td>Mean self-rated health score‡‡ (SD, N)</td>
<td>2.5 (1.0, 1232)</td>
<td></td>
</tr>
</tbody>
</table>

*For characteristics where there are no missing data, n is shown; for characteristics with missing data n/N is shown.
†Including schizophrenia and psychotic illness.
‡Individual Index of Multiple Deprivation score, 2010, for England, and Scottish Index of Multiple Deprivation score, 2012, for Scotland, based on participants’ postcodes. The lower quartile is the least deprived and the upper quartile is the most deprived.
§Calculation of global HCTD score: sum of scores where each question was scored 0 (no difficulty), 1 (some difficulty) or 2 (a lot of difficulty). Minimum score 0, maximum score 16. Missing data were scored 0 (not difficult), as suggested by the HCTD authors. Responses were excluded if participants ticked that they had a condition but did not score how much the condition limited their daily activity, or if they gave a score without ticking that they had the condition.
¶Sum of the weighted scores (each scored 1–5) from the Bayliss scale. Responses were excluded if participants ticked that they had a condition but did not score how much the condition limited their daily activity, or if they gave a score without ticking that they had the condition.
**Number of self-reported conditions from a list of 27 conditions itemised in the Bayliss scale.
††EuroQol five dimensions, five level questionnaire (EQ-5D-5L) score.
‡‡Single question: ‘In general, would you say your health is poor (1), fair (2), good (3), very good (4) or excellent (5)?’
§§Patient Assessment of Chronic Illness Care score.
responses. We set out to reduce the demands on investigators by providing clear instructions on how to calculate a global MTBQ score, including handling of missing data, and how to report and interpret these scores.

**Ethical approval and data sharing**

The study was registered under trial registration number ISRCTN06180958 (main trial results yet to be published). Data will be available from the University of Bristol Research Data Storage Facility after the main results of the 3D trial have been published in 2018.

**RESULTS**

**Participant characteristics**

One hundred and forty-three adults participated in the pilot study. From 1546 participants in the main 3D Study who completed the main baseline questionnaire, we were able to calculate an MTBQ score for 1524 (99%) individuals who completed at least half of the baseline MTBQ questions. At 9-month follow-up, 1356 returned the questionnaire and an MTBQ score could be calculated for 1299 (96%). The participants were mostly elderly (mean age 71 years for the main study), fully retired from work, had left school aged 16 years or younger, and 99% were white British (table 1). Around two-thirds of participants from England lived in areas of low deprivation (low or middle lower quartiles), whereas almost two-thirds of participants from Scotland lived in areas of high deprivation (middle upper or upper quartiles).

**Conceptual and measurement model**

**Conceptual framework**

The framework developed by Eton et al describes three major themes of treatment burden: the work required to look after one’s health (eg, self-monitoring, making lifestyle changes); tools and strategies patients use to reduce their treatment burden (eg, organising medication); and factors that increase burden (eg, poor continuity of care). We mapped the three existing treatment burden questionnaires against this framework, and discussed this with the PPI group who felt that all of the domains of treatment burden identified in the literature should be included in the PROM. We had initially considered excluding questions about costs since healthcare is mostly free under the National Health Service, but our PPI group argued that they still experienced additional costs from managing illness so this domain was retained in the first draft.

**Question properties**

The proportion of missing data for each question was between 1% and 3% (see table 2). Questions 3, 9 and 10 with a high proportion of ‘does not apply’ responses (table 2) were excluded from the main analysis. Since these questions might apply to other populations, we repeated Cronbach’s alpha including these questions in the various combinations (online supplementary appendix B). These extra questions may be considered as optional depending on the study population. Responses were positively skewed and a floor effect was found for some questions. However, the MTBQ had fewer floor effects than the comparator HCTD (online supplementary appendix C).

The global MTBQ scores were also skewed with 26% of pilot study participants and 22% of main study participants scoring 0 (online supplementary appendix D). Again, the HCTD had greater floor effects, with 54% of participants having a global score of 0.

**Dimensionality**

Both Kaiser’s ‘eigenvalue greater than one’ rule and Cattell’s scree plot criterion suggested a one-factor solution, and this explained 93% of the common variance. Loadings on this factor were uniformly greater than 0.4. The factor solution had high uniqueness for some items. This can sometimes indicate that the item is not strongly related to others, but because of the important content of these variables (eg, lifestyle changes, collecting medication), we chose to include them.

**Reliability**

Questions 1 and 2 have a high interitem correlation of 0.69, and questions 6 and 7 have an interitem correlation of 0.62 (online supplementary appendix E). Almost all of the other interitem correlations were in the ideal range of 0.2–0.4. A decision was made to include questions 1 and 2, and 6 and 7, despite the high interitem correlation coefficients because it was felt these questions were about different aspects of treatment burden. Cronbach’s alpha was 0.83, indicating a high level of internal reliability. Including the optional questions (questions 3, 9 and 10) in various combinations, Cronbach’s alpha ranged from 0.82 to 0.84, again demonstrating good internal consistency (see online supplementary appendix B).

**Validity**

**Face and content validity**

Participants from the PPI group commented that the wording was clear and easy to understand. One participant felt that accessing healthcare outside of usual general practitioner (GP) opening hours caused significant treatment burden for him. In response to this, we added a question about difficulty getting healthcare in the evenings and weekends (question 10). The remaining participants commented that the important areas of treatment burden were covered by the questionnaire.

**Construct validity**

As predicted, the global MTBQ score had a positive association with the comparator HCTD scale (r = 0.58, P < 0.0001), the Bayliss disease burden scale (r = 0.43, P < 0.0001) and the number of self-reported comorbidities (r = 0.32, P < 0.0001), and a negative association with the quality of life scale (r = −0.36, P < 0.0001) and self-rated health (r = −0.36, P < 0.0001) (table 3). This provides good evidence on the construct validity of the scale.
Regression analysis found that for every 1 SD (ie, 0.17) increase in EQ-5D-5L score between baseline and 9-month follow-up, the MTBQ score at follow-up was reduced by 1.7 (regression coefficient −0.14 multiplied by an SD change in MTBQ score of 11.9 (95% CI for regression coefficient −0.19 to −0.08), P<0.0001) (see table 4). This association was also seen after further adjusting the model for the specified covariates (regression coefficient −0.14 (95% CI −0.20 to −0.08), P<0.0001).

The equivalent model for PACIC score showed that for every 1 SD (ie, 0.86) increase in PACIC score between baseline and 9-month follow-up, MTBQ at follow-up was reduced by 1.9 (regression coefficient −0.16 multiplied by an SD change in MTBQ score of 11.9 (95% CI for regression coefficient −0.22 to −0.10), P<0.0001). A similar decrease was also seen after further adjusting the model for the specified covariates (regression coefficient −0.17 (95% CI −0.23 to −0.11), P<0.0001).

Interpretability of scores
Comparing participants across the four treatment burden groups (no burden, low burden, medium burden and high burden), female participants; younger participants; those with a greater number of long-term conditions; participants with depression, dementia and severe mental health problems listed on their GP records; and participants with worse EQ-5D-5L scores had higher disease burden scores and poor self-rated health were more likely to have a high treatment burden score, after adjusting for age, gender, number of comorbidities, age left full-time education and individual deprivation level (see table 5). There was no convincing association between deprivation score and treatment burden score.

Translation
Not applicable.
### Table 3  Association between global MTBQ score and global HCTD score, self-reported disease burden score, quality of life score, number of self-reported conditions and self-rated health at baseline

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Spearman’s rank correlations (r_s)</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global HCTD score*</td>
<td>141</td>
<td>0.58</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Self-reported disease burden score†</td>
<td>1443</td>
<td>0.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Number of self-reported conditions‡</td>
<td>1523</td>
<td>0.31</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Quality of life score§</td>
<td>1520</td>
<td>−0.36</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Self-rated health¶</td>
<td>1503</td>
<td>−0.36</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Calculation of global HCTD score: sum of scores where each question was scored 0 (no difficulty), 1 (some difficulty) or 2 (a lot of difficulty). Minimum score 0, maximum score 16. Missing data were scored 0 (not difficult), as suggested by the HCTD authors.†Sum of the weighted scores (each scored 1–5) from the Bayliss scale. Responses were excluded if participants ticked that they had a condition but did not score how much the condition limited their daily activity, or if they gave a score without ticking that they had the condition. ‡Number of self-reported conditions from the Bayliss scale. §EuroQol five dimensions, five level questionnaire (EQ-5D-5L) score. ¶Single question: ‘In general, would you say your health is poor (1), fair (2), good (3), very good (4) or excellent (5)’?

HCTD, Healthcare Task Difficulty; MTBQ, Multimorbidity Treatment Burden Questionnaire.

### Table 4  Association between global MTBQ score and (1) quality of life (EQ-5D-5L) score and (2) PACIC score

<table>
<thead>
<tr>
<th>Outcome</th>
<th>N*</th>
<th>Linear regression coefficient of MTBQ standardised change score (95% CI)</th>
<th>P value</th>
<th>N</th>
<th>Adjusted linear regression coefficient of MTBQ standardised change score (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D-5L standardised change score</td>
<td>1270</td>
<td>−0.14 (−0.19 to −0.08)</td>
<td>&lt;0.0001</td>
<td>1239</td>
<td>−0.14 (−0.20 to −0.08)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>PACIC standardised change score</td>
<td>930</td>
<td>−0.16 (−0.22 to −0.10)</td>
<td>&lt;0.0001</td>
<td>914</td>
<td>−0.17 (−0.23 to −0.11)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Results were from linear regression model of standardised change.

*This analysis included participants who completed the outcome questionnaire (EQ-5D-5L or PACIC) and the MTBQ questionnaire at baseline and 9-month follow-up.

†Linear regression model further adjusted for age, gender, number of comorbidities, age left full-time education and individual deprivation score.

‡This analysis included participants who completed the outcome questionnaire (EuroQol five dimensions, five level questionnaire (EQ-5D-5L), PACIC or MTBQ) at baseline and 9-month follow-up.

MTBQ, Multimorbidity Treatment Burden Questionnaire; PACIC, Patient Assessment of Chronic Illness Care.

### Discussion

In this study, we have developed and validated a 10-item questionnaire, named the Multimorbidity Treatment Burden Questionnaire (MTBQ). The psychometric properties of the questionnaire meet the minimum standards for a PROM set out by ISOQOL, demonstrating good content validity, internal reliability consistency, construct validity and responsiveness. Three additional questions, including one question about the cost of treatment, had a high proportion of ‘does not apply’ responses in this study population and were omitted from the main analysis. However, these questions may be relevant to other populations (eg, countries where patients pay for prescriptions and healthcare), and the scale remained internally consistent and reliable when they were included, so they may be considered as optional.

We found that younger patients were more likely to report high treatment burden scores and, interestingly, Tran’s TBQ found the same phenomenon. There are several possible explanations for this. First, treatment burden may impact more on younger patients because they must juggle their appointments or complex medication regimens alongside having to work or look after dependants. Second, younger patients may have different expectations of how looking after one’s health might impact on their lives and, hence, suffer from a greater perceived treatment burden. As expected, we found that patients with mental health conditions including depression and dementia were more likely to have high treatment burden scores. Previous studies have reported similar findings. High treatment burden was also associated with having a greater number of long-term conditions. No individual physical condition was found to be associated with high treatment burden. This result differs from both the TBQ study, which found an association between treatment burden and diabetes, and the HCTD study, which found an association between treatment burden and stroke, congestive heart failure and falls. As expected,
Table 5  Characteristics by categories of treatment burden (main study baseline data)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>None (0)</th>
<th>Low (&lt;10)</th>
<th>Medium (10–22)</th>
<th>High (≥22)</th>
<th>Unadjusted OR*</th>
<th>Adjusted OR†</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td>1524</td>
<td>308</td>
<td>385</td>
<td>425</td>
<td>406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age (mean)</strong></td>
<td>1524</td>
<td>74</td>
<td>73</td>
<td>71</td>
<td>66</td>
<td><strong>0.96 (0.95 to 0.97)</strong></td>
<td><strong>0.96 (0.95 to 0.97)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Gender (n, %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>651</td>
<td>168 (22)</td>
<td>208 (28)</td>
<td>193 (26)</td>
<td>182 (24)</td>
<td><strong>0.74 (0.62 to 0.88)</strong></td>
<td><strong>0.73 (0.60 to 0.87)</strong></td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Number of long-term conditions (n, %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1217</td>
<td>246 (20)</td>
<td>323 (27)</td>
<td>335 (28)</td>
<td>313 (26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 or more</td>
<td>307</td>
<td>62 (20)</td>
<td>62 (20)</td>
<td>90 (29)</td>
<td>93 (30)</td>
<td><strong>1.21 (0.97 to 1.52)</strong></td>
<td><strong>1.38 (1.09 to 1.74)</strong></td>
<td>0.007</td>
</tr>
<tr>
<td><strong>Long-term conditions (n, %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>1423</td>
<td>294 (21)</td>
<td>367 (26)</td>
<td>389 (27)</td>
<td>373 (26)</td>
<td><strong>0.62 (0.44 to 0.91)</strong></td>
<td><strong>0.79 (0.54 to 1.14)</strong></td>
<td>0.208</td>
</tr>
<tr>
<td>Stroke/Transient ischaemic attack</td>
<td>517</td>
<td>127 (25)</td>
<td>140 (27)</td>
<td>135 (26)</td>
<td>115 (22)</td>
<td><strong>0.69 (0.57 to 0.83)</strong></td>
<td><strong>0.82 (0.67 to 1.01)</strong></td>
<td>0.059</td>
</tr>
<tr>
<td>Diabetes</td>
<td>800</td>
<td>158 (20)</td>
<td>200 (25)</td>
<td>211 (26)</td>
<td>231 (29)</td>
<td><strong>1.13 (0.94 to 1.35)</strong></td>
<td><strong>1.04 (0.87 to 1.26)</strong></td>
<td>0.633</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>454</td>
<td>101 (22)</td>
<td>121 (27)</td>
<td>115 (25)</td>
<td>117 (26)</td>
<td>0.86 (0.71 to 1.05)</td>
<td>1.10 (0.89 to 1.36)</td>
<td>0.356</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease or asthma</td>
<td>758</td>
<td>148 (20)</td>
<td>185 (24)</td>
<td>222 (29)</td>
<td>203 (27)</td>
<td><strong>1.08 (0.90 to 1.29)</strong></td>
<td><strong>0.91 (0.75 to 1.10)</strong></td>
<td>0.326</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>76</td>
<td>14 (18)</td>
<td>21 (28)</td>
<td>24 (32)</td>
<td>17 (22)</td>
<td>0.94 (0.63 to 1.41)</td>
<td>0.76 (0.50 to 1.17)</td>
<td>0.216</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>524</td>
<td>119 (23)</td>
<td>155 (30)</td>
<td>142 (27)</td>
<td>108 (21)</td>
<td>0.68 (0.56 to 0.82)</td>
<td>0.91 (0.74 to 1.12)</td>
<td>0.369</td>
</tr>
<tr>
<td>Severe mental health problems‡</td>
<td>66</td>
<td>7 (11)</td>
<td>10 (15)</td>
<td>17 (26)</td>
<td>32 (48)</td>
<td><strong>2.61 (1.64 to 4.15)</strong></td>
<td><strong>1.75 (1.08 to 2.82)</strong></td>
<td>0.022</td>
</tr>
<tr>
<td>Depression</td>
<td>553</td>
<td>85 (15)</td>
<td>105 (19)</td>
<td>169 (31)</td>
<td>194 (35)</td>
<td><strong>1.92 (1.59 to 2.32)</strong></td>
<td><strong>1.43 (1.16 to 1.77)</strong></td>
<td>0.001</td>
</tr>
<tr>
<td>Dementia</td>
<td>58</td>
<td>14 (24)</td>
<td>10 (17)</td>
<td>12 (21)</td>
<td>22 (38)</td>
<td>1.00 (0.99 to 1.01)</td>
<td>1.01 (0.99 to 1.02)</td>
<td>0.450</td>
</tr>
<tr>
<td>Learning disability</td>
<td>14</td>
<td>2 (14)</td>
<td>2 (14)</td>
<td>6 (43)</td>
<td>4 (29)</td>
<td><strong>1.47 (0.59 to 3.69)</strong></td>
<td><strong>1.07 (0.36 to 3.21)</strong></td>
<td>0.907</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>102</td>
<td>15 (15)</td>
<td>18 (18)</td>
<td>40 (39)</td>
<td>29 (28)</td>
<td><strong>1.41 (0.99 to 2.01)</strong></td>
<td><strong>1.28 (0.88 to 1.82)</strong></td>
<td>0.202</td>
</tr>
<tr>
<td>Heart failure</td>
<td>154</td>
<td>36 (23)</td>
<td>41 (27)</td>
<td>38 (25)</td>
<td>39 (25)</td>
<td>0.85 (0.63 to 1.14)</td>
<td>1.06 (0.77 to 1.44)</td>
<td>0.340</td>
</tr>
<tr>
<td>Age left full-time education (n, %)§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤16 years</td>
<td>681</td>
<td>164 (24)</td>
<td>172 (25)</td>
<td>177 (26)</td>
<td>168 (25)</td>
<td>1.00 (0.99 to 1.01)</td>
<td>1.01 (0.99 to 1.02)</td>
<td>0.450</td>
</tr>
<tr>
<td>England</td>
<td>1078</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td><strong>1.01 (1.00 to 1.01)</strong></td>
<td><strong>1.00 (0.99 to 1.01)</strong></td>
<td>0.904</td>
</tr>
<tr>
<td>Scotland</td>
<td>467</td>
<td>26</td>
<td>26</td>
<td>24</td>
<td>24</td>
<td><strong>1.00 (0.99 to 1.01)</strong></td>
<td><strong>0.99 (0.99 to 1.00)</strong></td>
<td>0.032</td>
</tr>
<tr>
<td>EuroQol five dimensions, five level questionnaire (EQ-5D-5L) 11 (mean)</td>
<td>1520</td>
<td>0.67</td>
<td>0.63</td>
<td>0.56</td>
<td>0.42</td>
<td><strong>0.11 (0.08 to 0.16)</strong></td>
<td><strong>0.09 (0.06 to 0.12)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Disease burden score 12 (mean)</td>
<td>1443</td>
<td>12.8</td>
<td>15.7</td>
<td>19.0</td>
<td>26.1</td>
<td><strong>1.06 (1.06 to 1.08)</strong></td>
<td><strong>1.07 (1.07 to 1.09)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Self-rated health (n, %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>315</td>
<td>36 (11)</td>
<td>42 (13)</td>
<td>75 (24)</td>
<td>162 (51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>674</td>
<td>112 (17)</td>
<td>168 (25)</td>
<td>216 (32)</td>
<td>178 (26)</td>
<td><strong>0.39 (0.30 to 0.50)</strong></td>
<td><strong>0.41 (0.31 to 0.53)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Good</td>
<td>422</td>
<td>111 (26)</td>
<td>138 (33)</td>
<td>116 (27)</td>
<td>57 (14)</td>
<td><strong>0.20 (0.15 to 0.26)</strong></td>
<td><strong>0.19 (0.14 to 0.26)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Very good</td>
<td>87</td>
<td>40 (46)</td>
<td>28 (32)</td>
<td>16 (18)</td>
<td>3 (3)</td>
<td><strong>0.08 (0.05 to 0.13)</strong></td>
<td><strong>0.08 (0.05 to 0.12)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Excellent</td>
<td>5</td>
<td>3 (60)</td>
<td>2 (40)</td>
<td>0</td>
<td>0</td>
<td><strong>0.04 (0.01 to 0.23)</strong></td>
<td><strong>0.03 (0.00 to 0.16)</strong></td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Continued
participants with low quality of life (EQ-5D-5L) score, high disease burden score and poor self-rated health were more likely to have high treatment burden. We also found that female participants were more likely to report high treatment burden compared with male participants. This has not been reported elsewhere. There was no association between deprivation level and treatment burden score. One might expect that people from more deprived areas might have fewer support networks and resources and so would experience higher treatment burden. Alternatively, one could argue that participants from more deprived areas might be more accepting of how looking after their health impacts on their day-to-day life and so report lower treatment burden.

A key strength of this study is that the MTBQ has been validated in a large sample of participants for whom it is intended—elderly multimorbid patients with a mean age of 71 years and three or more long-term conditions. In comparison, the English version of the Tran’s TBQ was validated in a younger computer-literate population with a mean age of 51 years. The MTBQ had good face validity, was found to be user-friendly and fits on a single page of A4 paper in size 14 font. All aspects of treatment burden identified in a comprehensive evidence-based framework are included in the questionnaire. In comparison, the most comprehensive existing questionnaire, the PETS questionnaire, includes 48 questions and is time-consuming to complete, and several of the other existing questionnaires focus on only some aspects of treatment burden. Preliminary assessment of responsiveness found that, as expected, a positive change in both quality of life (EQ-5D-5L) score and patient-centred care (PACIC) score between baseline and 9-month follow-up was associated with a reduction in treatment burden (MTBQ score). Of the other relevant PROMs, only the HCTD has been assessed for responsiveness, but the HCTD addresses fewer topics and has a narrower range of response options, possibly contributing to its greater problems with skewness and floor effects.

A limitation of this study is that the MTBQ was developed using a framework of treatment burden developed from qualitative study in the USA. However, apart from the issue of paying for care, we felt that other domains of treatment burden were likely to be generalisable, and we wanted to develop a measure that covered generic issues that would be relevant in a range of settings rather than specific to one healthcare system. Our measure was also informed by qualitative papers from different countries (including the UK) to ensure we included the important concepts. In cognitive interviews, participants with multimorbidity felt that the questionnaire captured the range of factors that contribute to treatment burden.

A further limitation is that the participants of this study were recruited into a trial, which creates potential for selection bias and may limit generalisability. However, the trial participants had similar characteristics to those invited but declining participation in respect of age, gender, number and type of long-term conditions (data will be reported with the 3D trial results). Almost all the participants of this study were white British and further work is planned to validate the questionnaire in other populations. We found high floor effects with 22% of participants scoring a global MTBQ score of 0. All of the other treatment burden measures also show similarly high floor effects. One explanation for this is a ‘response shift’, whereby patients adapt their everyday life so that looking after their health conditions becomes more acceptable to them over time and causes less perceived burden. The following are the implications of positively skewed treatment burden scores and high floor effects: first, this can make it difficult to detect change (ie, it is not possible to improve from a treatment burden score of 0); and second, mean treatment burden scores should be interpreted with caution. Preliminary analysis of responsiveness, however, has shown that changes in MTBQ score correlate as expected with changes in quality of life (EQ-5D-5L) score and patient-centred care (PACIC) over time. We recommend that, due to the skewness of global MTBQ scores, researchers should report the median and IQR rather than the mean and SD, and report the proportion of patients with high, medium, low or no treatment burden (MTBQ scores ≥22, 10–22, <10 and 0, respectively).

The MTBQ scale is a concise measure of treatment burden for patients with multimorbidity that has demonstrated good content validity, construct validity, internal consistency reliability and responsiveness. It is a useful research tool for assessing the impact of interventions on treatment burden for patients with multimorbidity. We anticipate the scale being used alongside other measures, such as disease burden, and that findings from the two measures will be related. The MTBQ could also be used in clinical practice to highlight problem areas for patients with multimorbidity, such as difficulties the patient may...
have with their medication or with making recommended lifestyle changes. Further work is needed to validate the MTBQ for use in a clinical setting.

**Acknowledgements** Appreciation is extended to members of the patient and public involvement group, known as the Patient Involvement in Primary Care Research (PiP-CaRe) group, who took part in the cognitive interviews. The PiP-CaRe group was formed for the purpose of the 3D Study and consists of people with two or more long-term conditions. We also thank all other members of the 3D research team and Professor Boyd for permission to use the Healthcare Task Difficulty questionnaire.

**Contributors** PD, M-SM and CS were responsible for study concept and design. PD, MM, DG and KC were involved in data extraction and analysis. PD drafted the manuscript. All authors critically reviewed the manuscript and approved the final version. All authors also had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. PD is the guarantor, PD led this project under the supervision of CS. She designed the study, undertook a literature review, developed the questionnaire, conducted and analysed the cognitive interviews, convened meetings with the patient and public involvement group, analysed the results, and drafted the paper. MM provided methodological expertise in assessing the psychometric properties of this new patient-reported outcome measure, including the approach to analysis and interpretation of the results. She critically appraised the paper and has approved the final version. M-SM provided methodological and practical expertise, and obtained ethical and governance approvals for this study. She critically appraised the paper and has approved the final version. KC acquired and cleaned the original data and produced the database used for analysis. She critically appraised the paper and has approved the final version. KC acquired and cleaned the original data and produced the database used for analysis. She critically appraised the paper and has approved the final version. KC acquired and cleaned the original data and produced the database used for analysis. She critically appraised the paper and has approved the final version. KC acquired and cleaned the original data and produced the database used for analysis. She critically appraised the paper and has approved the final version.

**Funding** This work was funded by the National Institute for Health Research Health Services and Delivery Research Programme (project number 12/130/15).

**Disclaimer** The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** The 3D Study was approved by South-West (Frenchay) NHS Research Ethics Committee (14/SW/0011).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Data will be available from the University of Bristol Research Data Storage Facility after the main results of the 3D trial have been published in 2018.

**Open Access** This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

**REFERENCES**