
Article:
Duncan, P. orcid.org/0000-0002-2244-3254, Scott, L.J. orcid.org/0000-0003-3129-5123, Dawson, S. orcid.org/0000-0002-6700-6445 et al. (6 more authors) (2024) Further development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ). BMJ Open, 14 (4). e080096. ISSN 2044-6055

https://doi.org/10.1136/bmjopen-2023-080096

BMJ Open

Further development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ)

Polly Duncan ,[1] Lauren J Scott ,[2,3] Shoba Dawson ,[1] Muzrif Munas,[1] Yvette Pyne,[1] Katherine Chaplin,[1] Daisy Gaunt,[3] Line Guenette ,[4] Chris Salisbury [1]

To cite: Duncan P, Scott LJ, Dawson S, *et al*. Further development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ). *BMJ Open* 2024;**14**:e080096. doi:10.1136/bmjopen-2023-080096

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-080096).

Received 21 September 2023
Accepted 24 January 2024

[1]Centre for Academic Primary Care, Bristol Medical School, University of Bristol, Bristol, UK
[2]National Institute for Health Research Applied Research Collaboration West, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
[3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[4]Faculty of Pharmacy and CHU de Québec Research Center, Université Laval, Quebec city, Quebec, Canada

Correspondence to
Dr Polly Duncan;
polly.duncan@bristol.ac.uk

ABSTRACT

Objectives To undertake further psychometric testing of the Multimorbidity Treatment Burden Questionnaire (MTBQ) and examine whether reversing the scale reduced floor effects.

Design Survey.

Setting UK primary care.

Participants Adults (≥18 years) with three or more long-term conditions randomly selected from four general practices and invited by post.

Measures Baseline survey: sociodemographics, MTBQ (original or version with scale reversed), Treatment Burden Questionnaire (TBQ), four questions (from QQ-10) on ease of completing the questionnaires. Follow-up survey (1–4 weeks after baseline): MTBQ, TBQ and QQ-10. Anonymous data collected from electronic GP records: consultations (preceding 12 months) and long-term conditions. The proportion of missing data and distribution of responses were examined for the original and reversed versions of the MTBQ and the TBQ. Intraclass correlation coefficient (ICC) and Spearman's rank correlation (R_s) assessed test–retest reliability and construct validity, respectively. Ease of completing the MTBQ and TBQ was compared. Interpretability was assessed by grouping global MTBQ scores into 0 and tertiles (>0).

Results 244 adults completed the baseline survey (consent rate 31%, mean age 70 years) and 225 completed the follow-up survey. Reversing the scale did not reduce floor effects or data skewness. The global MTBQ scores had good test–retest reliability (ICC for agreement at baseline and follow-up 0.765, 95% CI 0.702 to 0.816). Global MTBQ score was correlated with global TBQ score (R_s 0.77, p<0.001), weakly correlated with number of consultations (R_s 0.17, p=0.010), and number of different general practitioners consulted (R_s 0.23, p<0.001), but not correlated with number of long-term conditions (R_s −0.063, p=0.330). Most participants agreed that both the MTBQ and TBQ were easy to complete and included aspects they were concerned about.

Conclusion This study demonstrates test–retest reliability and ease of completion of the MTBQ and builds on a previous study demonstrating good content validity, construct validity and internal consistency reliability of the questionnaire.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Further psychometric testing was carried out on the Multimorbidity Treatment Burden Questionnaire, including test–retest reliability, construct validity and assessing ease of completing the questionnaire.
⇒ Study participants had three or more long-term conditions and an average age of 70 years.
⇒ Postal recruitment enabled people without a smartphone, computer, access to the internet or good information technology literacy skills to take part.
⇒ The study was designed to assess the primary outcome of test–retest reliability but was not necessarily large enough to detect multiple associations between treatment burden and some patient characteristics.

INTRODUCTION

Having a good measure of treatment burden for patients with multimorbidity is important given the ageing population and the associated increase in multimorbidity.[1] Interventions designed to reduce treatment burden require a measure of treatment burden to assess their effectiveness. Treatment burden is defined as the 'effort of looking after ones' health and the impact that this has on everyday life'.[2] This includes ordering and collecting medications, taking complex medication regimens, coordinating and attending healthcare appointments, monitoring one's health conditions and making lifestyle changes.

The Multimorbidity Treatment Burden Questionnaire (MTBQ) is a 10-question simply worded measure of treatment burden, developed and validated as part of the three-dimensional (3D) study, a multicentre cluster-randomised controlled trial in the UK that aimed to improve the management of patients with multimorbidity within primary care.[3 4] There are three additional optional questions, which had a high proportion of 'does not apply' responses in the original study but may be relevant to other populations. Study investigators can choose to use the 10-question or 13-question version of the

MTBQ. 1546 adult participants with multimorbidity (≥3 long-term conditions) and a mean age of 77 years, took part in the original study.[3] The MTBQ was developed using a framework of treatment burden derived from qualitative research in the USA[2] and demonstrated good content validity, construct validity and internal consistency reliability and preliminary evidence of responsiveness.[4] The MTBQ is widely used internationally and has been translated, culturally adapted and validated into several languages, including Danish,[5] German,[6] French-Canadian[7] and Chinese.[8]

We are aware of four other existing measures of treatment burden for patients with multimorbidity, but all have limitations.[9–13] The Treatment Burden Questionnaire (TBQ) is a 13-question measure originally developed in France[10] and subsequently translated, adapted and validated in English.[9] A limitation of the TBQ is that the English version was developed and validated in a relatively young (mean age 51 years) and highly educated (78% with a college education) population recruited from an online platform.[9] Some of the wording is quite complex requiring high literacy levels (eg, question 1: 'How would you rate the problems related to the taste, shape or size of your tablets and/or the annoyances caused by your injections (eg, pain, bleeding, bruising or scars?'). The Patient Experience with Treatment and Self-management (PETS) questionnaire is a comprehensive measure of treatment burden developed in the USA, including 48 questions.[11] For some study investigators, the length of the PETS questionnaire will be considered a strength, as it is able to capture a detailed picture of the different aspects of treatment burden. However, others may consider its length a limitation, being too time-consuming and burdensome for participants to complete. A shorter version of the PETS questionnaire, called the 'Brief PETS' questionnaire, has been developed.[14] The length of the questionnaire (32 questions) may still be considered too long for some study investigators, particularly clinical trialists where treatment burden is one of several secondary outcomes and included among a battery of other measures. The Multimorbidity Illness Perceptions Scale questionnaire was developed and validated in older people (mean age 70 years) in the UK and includes a six-question treatment burden subscale.[13] This excludes some important aspects of treatment burden, such as arranging appointments with healthcare professionals. The Healthcare Task Difficult questionnaire, developed in the USA, was only designed to measure one aspect of treatment burden (difficulty with health-related tasks, such as obtaining and taking medications) and was not designed to measure other aspects of treatment burden (eg, seeing different healthcare professionals).[12]

The MTBQ, and the original validation study,[4] have four important limitations. First, the data were positively skewed and there was a high floor effect, with 22% of participants scoring a global MTBQ score of 0 (no treatment burden). As it is not possible to improve from a score of 0, this can make it difficult to detect change. Similar floor effects have been shown from other existing treatment burden questionnaires for patients with multimorbidity.[9–12] Second, within the context of a trial, it was not possible to assess test–retest reliability. Third, in the original validation study, we were only able to test construct validity using indirect measures which we expected to correlate with high and low treatment burden scores, such as health-related quality of life score, rather than direct measures of treatment burden, such as number of healthcare appointments. Fourth, as the MTBQ was developed and validated as part of a trial, it may not be generalisable to non-trial populations. The ease of completing the MTBQ was assessed in the original study as part of the cognitive interviews (n=8) but has not been assessed in a larger sample of participants.

The purpose of this study was: (1) to examine whether reversing the scale of the questionnaire improved the floor effects and the skewness of the data; (2) to assess test–retest reliability; (3) to compare responses, construct validity and ease of completion of the MTBQ and a comparator questionnaire, the TBQ[9 10] and (4) to assess interpretability of the MTBQ in a non-trial population.

METHODS

Study population, eligibility criteria and recruitment

Participants were recruited from four General Practices serving a range of deprived, mid-deprived and affluent populations, from August 2018 to August 2019. Patients were eligible if they were aged ≥18 years and had ≥3 long-term conditions from 17 major long-term conditions included in the 2014 National Health Service Quality and Outcomes Framework (a UK programme which incentivises General Practices to deliver high quality healthcare).[15] Conditions were grouped into 12 types of condition with similar management considerations; for example, asthma and chronic obstructive pulmonary disease (COPD) within the same individual were counted as one condition. Patients who had taken part in the 3D study[3] or who were deemed unsuitable to take part by a clinician from the practice (eg, due to a recent bereavement, cognitive impairment or poor level of English to read and complete the questionnaire) were excluded.

Potentially eligible participants were identified via a standardised search of the electronic general practitioner (GP) records, which was used in the original validation study.[3 4 16] Similar conditions, such as asthma and COPD, within the same individual were counted as one condition (table 1). A random sample of potentially eligible participants was selected from each practice and was reviewed by a clinician in the practice to check whether it was appropriate to invite them. Eligible participants were sent an invitation letter, a participant information sheet, and a questionnaire booklet (with original MTBQ or version with the scale reversed; see below). Completion of the questionnaire implied consent, as stated in the participant information sheet. Those who responded were sent a follow-up questionnaire 1–4weeks after returning the

Table 1 Participant characteristics

	Baseline survey (n=244)	Baseline original MTBQ (n=112)	Baseline reversed MTBQ (n=132)
Age (years; mean, SD)	69.9, 13.1	71.9, 11.6	68.1, 14.0
Age group (n, (%))			
18–50 years	18 (7.4)	4 (3.6)	14 (10.6)
51–60 years	37 (15.2)	15 (13.4)	22 (16.7)
61–70 years	55 (22.5)	25 (22.3)	30 (22.7)
71–80 years	82 (33.6)	42 (37.5)	40 (30.3)
81–90 years	47 (19.3)	22 (19.6)	25 (18.9)
91+ years	4 (1.6)	3 (2.7)	1 (0.8)
Missing	1 (0.4)	1 (0.9)	0 (0)
Gender (n, (%))			
Male	130 (53.3)	61 (54.5)	69 (52.3)
Female	113 (46.3)	50 (44.6)	63 (47.7)
Missing	1 (0.4)	1 (0.9)	0 (0)
Ethnicity (n, (%))			
White	229 (93.9)	106 (94.6)	123 (93.2)
Asian	5 (2.0)	2 (1.8)	3 (2.3)
Black/African/Caribbean	4 (1.6)	2 (1.8)	2 (1.5)
Mixed	4 (1.6)	1 (0.9)	3 (2.3)
Other	1 (0.4)	1 (0.9)	0 (0)
Missing	1 (0.4)	0 (0.0)	1 (0.8)
No of long-term conditions* (n, (%))			
3	136 (55.7)	58 (51.8)	78 (59.1)
4	73 (29.9)	35 (31.3)	38 (28.8)
≥5	35 (14.3)	19 (17.0)	16 (12.1)
Long-term conditions* (n, (%))			
Cardiovascular disease†	210 (86.1)	98 (87.5)	112 (84.8)
Stroke or transient ischaemic attack (TIA)	72 (29.5)	39 (34.8)	33 (25.0)
Diabetes	141 (57.8)	60 (53.6)	81 (61.4)
Chronic kidney disease	68 (27.9)	30 (26.8)	38 (28.8)
COPD or asthma	119 (48.8)	51 (45.5)	68 (51.5)
Epilepsy	20 (8.2)	12 (10.7)	8 (6.1)
Atrial fibrillation	79 (32.4)	41 (36.6)	38 (28.8)
Severe mental health problems	13 (5.3)	3 (2.7)	10 (7.6)
Depression	107 (43.9)	51 (45.5)	56 (42.4)
Learning disability	4 (1.6)	0 (0)	4 (3.0)
Rheumatoid arthritis	17 (7.0)	12 (10.7)	5 (3.8)
Heart failure	31 (12.7)	16 (14.3)	15 (11.4)
Age left full-time education (years) (n, (%))			
≤14	22 (9.0)	9 (8.0)	13 (9.8)
15 or 16	155 (63.5)	77 (68.8)	78 (59.1)
17 or 18	33 (13.5)	14 (12.5)	19 (14.4)
≥19	31 (12.7)	11 (9.8)	20 (15.2)
Missing	3 (1.3)	1 (0.9)	2 (1.5)

Continued

Table 1 Continued

	Baseline survey (n=244)	Baseline original MTBQ (n=112)	Baseline reversed MTBQ (n=132)
Employment status (n, (%))			
Fully retired from work	144 (59.0)	69 (61.6)	75 (56.8)
Employed	36 (14.8)	13 (11.6)	23 (17.4)
Other	64 (26.2)	30 (26.8)	34 (25.8)
Missing	0 (0)	0 (0)	0 (0)
Deprivation score‡ (n, (%))			
Quintile 1 (least deprived)	49 (20.1)	26 (23.2)	23 (17.4)
Quintile 2	49 (20.1)	22 (19.6)	27 (20.5)
Quintile 3	74 (30.3)	33 (29.5)	41 (31.1)
Quintile 4	47 (19.3)	21 (18.8)	26 (19.7)
Quintile 5 (most deprived)	25 (10.2)	10 (8.9)	15 (11.4)
Missing	0 (0)	0 (0)	0 (0)

*Long-term conditions from electronic GP records. Please note, similar long-term conditions are grouped together (eg, COPD/asthma, stroke/TIA).
†Cardiovascular disease includes coronary heart disease, hypertension and peripheral arterial disease.
‡Based on Townsend deprivation index scores.[18]
COPD, chronic obstructive pulmonary disease; GP, general practitioner; MTBQ, Multimorbidity Treatment Burden Questionnaire.

baseline questionnaire, including the same version of the MTBQ completed at baseline. Participants were sent a £5 Love2Shop voucher each time they returned a completed questionnaire.

Survey content

The questionnaire booklet included demographic information (age, gender, age left full-time education, employment status, ethnic group); the MTBQ (original or reversed version; see below)[4]; the TBQ comparator questionnaire[9 10] and four questions from the QQ-10 questionnaire to assess the ease of completion of the MTBQ and the TBQ.[17] Four different versions of the questionnaire booklet were created: original MTBQ followed by TBQ; reversed MTBQ followed by TBQ; TBQ followed by original MTBQ; TBQ followed by reversed MTBQ. Each booklet was colour coded and participants were sent the same version of the questionnaire at follow-up to assess test–retest reliability.

Participants could actively decline participation by ticking a box on the front page of the questionnaire booklet saying they did not wish to participate and returning the booklet in the FREEPOST envelope. For non-responders, a reminder letter was sent 10–14 days after the baseline questionnaire.

Data from electronic GP records

The following non-identifiable information was collected from the electronic GP records: Townsend Deprivation Index scores; long-term conditions; all consultations recorded in the preceding 12 months (including face-to-face, telephone, video and home visits), type of professional who performed the consultation (eg, GP, nurse) and a GP identifier if it was a GP consultation. The Townsend scores were used to calculate quintiles of deprivation based on the 2011 census data.[18]

Consultations were coded in the same manner as a previous study.[19 20] The number of consultations in the preceding 12 months was calculated by adding up all consultation entries where participants were seen by a GP, nurse or primary care paramedic (employed by the general practice as part of the clinical team). The number of different GPs seen in the preceding 12 months was calculated for each participant who had at least one GP appointment using the GP identifier codes assigned to the consultations listed above. For some GP appointments (19%), a GP identifier code was not assigned; these appointments were excluded from this analysis. We excluded participants from the analysis who had one or more GP consultations with no GP identifier assigned.

Patient involvement

Four members of the Patient Involvement in Primary Care Research group were involved in the study design. We worked closely with them to develop simply worded, concise and easy to read invitation letters, information sheets and questionnaire booklets, making the study more accessible to patients.

The original MTBQ

The original MTBQ comprises 10 questions including the following aspects of treatment burden: taking and collecting medications, monitoring health conditions, arranging and attending healthcare appointments with different healthcare professionals, making recommended lifestyle changes and having to rely on help from

family and friends.[4] There are three additional optional questions about paying for medicines and equipment, accessing healthcare in the evenings and weekends and getting help from community services (eg, physiotherapy, community nurses). In this study, the 13-question version of the MTBQ was used, including the three optional questions.

Participants score each of the questions on a five-point Likert scale ranging from 0 (not difficult), to 1 (a little difficult), to 2 (quite difficult), to 3 (very difficult) to 4 (extremely difficult). There is also an option of 'does not apply' (scores 0). A global MTBQ score can be computed by calculating the mean from the questions answered and multiplying this by 25 to give a score from 0 to 100.[4] A global score cannot be calculated if more than 50% of responses are missing. The global score based on the 10 core questions was used in most of our analyses; for some analyses, we have also calculated and presented the global score based on all 13 questions.

Reversing the scale

A new version of the MTBQ was developed where the order of responses was reversed, that is, the response option of 'extremely difficult' was listed first and 'not difficult' was listed last (online supplemental file 1). We hypothesised that this might frame difficulties as to be expected and reduce floor effects on the questionnaire. Participants were randomly sent either the original MTBQ or reversed MTBQ.

Data and statistical analysis

We used means and SDs, and medians and IQRs to summarise normally distributed and skewed data respectively. Categorical data were summarised using counts and percentages.

Objective 1: examine whether reversing the scale improved the floor effects and the skewness of the data

To assess the effect of reversing the scale, the count and percentage of each response to each question, as well as the floor effects for the global score (the proportion of participants with a global score of 0), were compared between the original and reversed versions of the MTBQ. The distribution of the MTBQ on each scale was presented as medians and IQRs. A χ^2 test was used to compare the floor effect between the original and reversed MTBQ. The analyses for objective 1 included data from the baseline questionnaire.

Objective 2: assess test–retest reliability

To assess test–retest reliability, we calculated the intraclass correlation coefficient (ICC) for agreement (and the 95% CI) between global MTBQ score at baseline and follow-up. An ICC>0.7 was considered acceptable.[21] A Bland-Altman plot was constructed, where the mean global MTBQ score from the two time points was plotted against the difference in global MTBQ score between the two time points.[22] All participants were included in these analyses, including those who were sent the original and reversed versions of the MTBQ scale. A sensitivity analysis was performed, including only those who were sent the original version.

Objective 3: compare responses, construct validity and ease of completion of the MTBQ and a comparator questionnaire, the TBQ

The TBQ was chosen as the comparator questionnaire for this study because it includes all aspects of treatment burden and is relatively short (13 questions),[9 10] and so was thought feasible for participants to self-complete. The proportion of missing data and not difficult/does not apply responses (floor effect) were examined for each question from the MTBQ alongside the comparable questions from the TBQ.

MTBQ construct validity was assessed using Spearman's rank correlation coefficients (R_s) with corresponding p value s for independence, for four prespecified hypotheses: the TBQ comparator (criterion validity) [9 10]; a positive association between treatment burden score and number of long-term conditions; a positive association between treatment burden score (global score) and number of primary care appointments in the prior 12-month period; and finally, a positive association between treatment burden score (global score) and the number of different GPs seen in the prior 12-month period.

Four statements from the QQ-10 questionnaire[17] were used to compare the ease of completion of the MTBQ with the TBQ: (1) the questionnaire was easy to complete; (2) the questionnaire included all aspects of my condition that I am concerned about; (3) the questionnaire was too long and (4) the questionnaire was too complicated. For each statement, participants could strongly disagree, mostly disagree, neither agree or disagree, mostly agree or strongly agree. The QQ-10 consists of 10 statements, however, only the 4 statements which appeared most relevant to assessing the ease of completion of the TBQs were selected to avoid overburdening participants. The proportions of each response to each of the four questions were examined for the MTBQ and the TBQ; responses were grouped as strongly agree/mostly agree versus neither agree or disagree/mostly disagree/ strongly disagree and formally compared using the McNemar test.[23]

Construct validity analyses included data from all baseline participants (original and reversed MTBQ scale). All other analyses for objective 3 included data from baseline participants who responded to the original MTBQ scale.

Objective 4: assess interpretability of the MTBQ

To assess interpretability of the MTBQ, we categorised the global MTBQ scores greater than 0 into tertiles to generate four categories: no burden (score 0), low burden (lowest tertile), medium burden (middle tertile) and high burden (upper tertile). The tertiles were based on the MTBQ baseline data of participants who completed the original and reversed MTBQ. We summarised the participant characteristics and key outcome variables, including number of long-term conditions, by the four categories.

Further, we dichotomised the burden categories into no/low burden versus medium/high burden and examined the effect of participant characteristics and key outcome variables on treatment burden using logistic regression. For these analyses, we collapsed some of the variables due to small numbers. We performed univariable analyses, in addition to adjusted analyses where each model was adjusted for age, gender, deprivation and number of comorbidities. Estimates are presented as ORs alongside 95% CIs and p values.

SAMPLE SIZE

Sample size calculations were performed so that the primary outcome, assessment of test–retest reliability, achieved an interval estimate with sufficient precision, rather than a specific power to test a hypothesis.[24] Using a 0.7 ICC with 95% CI having a width of 0.2 (ie, 0.6 to 0.8), 101 participants were required to complete the baseline and follow-up questionnaire. Based on the response rate of the 3D study[3] and the 'TBQ' validation study,[9] the anticipated response rate was 20%.

The study design was assessed against the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist (online supplemental file 2).[25]

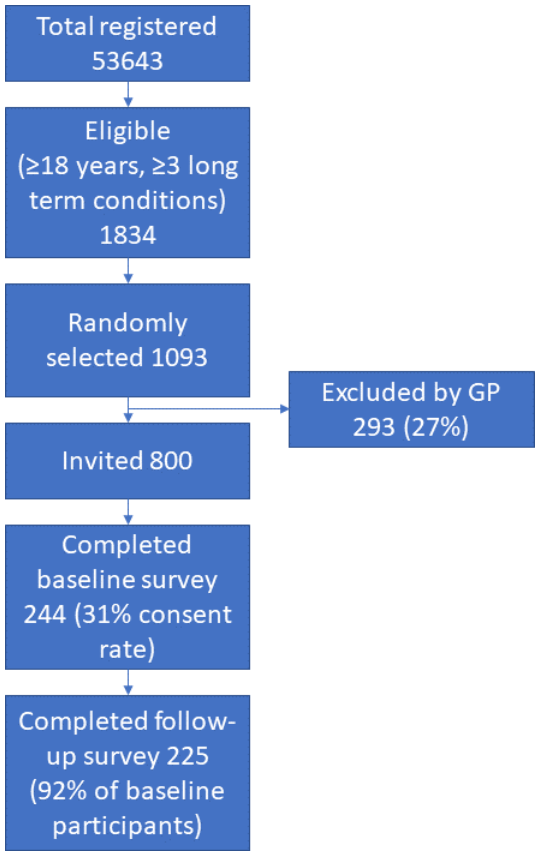


Figure 1 Participant flow diagram. GP, general practitioner.

RESULTS

Of the 800 adults invited, 244 completed the baseline survey (consent rate 31%, 112 and 132 completed the original and reversed scale versions, respectively) and 225 completed the follow-up survey (92% of participants who had completed the baseline survey, 105 and 120 completed the original and reversed scale version, respectively) (figure 1). The mean age of participants was 70 years (SD 13), 53% were male and 94% were of white British ethnicity (table 1). 56% had 3 long-term conditions, 30% had 4 and 14% had 5 or more. The most common long-term conditions were cardiovascular disease (86%), diabetes (58%), COPD or asthma (49%) and depression (44%). Seventy-three per cent left school aged 16 years or under, and 59% were fully retired from work. The sociodemographic characteristics and long-term conditions of those who completed the original and reversed scale versions of the MTBQ were similar (online supplemental file 3).

Objective 1: examine whether reversing the scale of the questionnaire improved the proportion of missing data, the floor effects and the skewness of the data

The proportion of missing data for each question was between 0% and 2% for the original version of the MTBQ, and between 0% and 3% for the reversed version (online supplemental file 4). The number of missing responses per participant was low for both versions of the questionnaire: 0 for 96% for the original version and 93% for the reversed version. The floor effect for the individual questions (the proportion of participants responding 'not difficult' or 'does not apply') was slightly higher for the reversed version compared with the original version, except for question 10. For both versions, the responses to individual questions were positively skewed, with a higher proportion of participants responding either 'a little difficult' or 'quite difficult', than 'very difficult' or 'extremely difficult'. The distribution of responses to individual questions was similar for the original and reversed versions (online supplemental file 4).

The median global MTBQ score was 17.1 (IQR 7.5–35.0) for the original MTBQ and 12.5 (IQR 5.0–27.5) for the reversed scale (online supplemental file 4). There were 11 (10%) participants with a global MTBQ score of 0 for the original version and 18 (14%) for the reversed version (p=0.35). The distribution of the TBQ global scores was also skewed and similar between participants who received the original version and reversed version of the MTBQ (online supplemental file 5).

Objective 2: assess test–retest reliability

The ICC for agreement between global MTBQ scores at baseline and follow-up was 0.768 (95% CI 0.705 to 0.818) and 0.765 (95% CI 0.702 to 0.816) for the 13-question version and 10-question version, respectively (includes all participants). Similarly, the agreement between baseline and follow-up for participants who completed the original MTBQ was 0.715 (95% CI 0.599 to 0.801) and 0.705 (95% CI 0.587 to 0.794) for the 13-question and

10-question scale, respectively. The Bland-Altman plot[22] suggests that there was no systematic bias between values at the two time points, with an average difference of only −0.5 (95% CI −24.8 to 23.8; online supplemental file 6).

Objective 3: compare responses, construct validity and ease of completion of the MTBQ and a comparator questionnaire, the TBQ

Of the participants who completed the original version of the MTBQ, the TBQ had similar floor effects for the global treatment burden score as the MTBQ, with 12% of participants scoring 0 for the TBQ compared with 10% of participants scoring 0 for the MTBQ (McNemar ratio 0.85, 95% CI 0.46 to 1.56, p=0.593; see also online supplemental files 5 and 7). The proportion of missing data for each question was between 0% and 2% for the MTBQ and 0%–5% for the TBQ. Ninety-two per cent of participants had no missing responses for the MTBQ, compared with 86% of participants for the TBQ (online supplemental file 7).

The floor effects for the individual questions (proportion of participants responding 'does not apply' or 'not difficult') ranged from 38% to 89% for the MTBQ, and from 36% to 71% for the TBQ (online supplemental file 7).

Regarding construct validity, the global MTBQ score had a strong positive correlation with the comparator TBQ scale (R_s 0.77, p<0.001 for 10-item MTBQ ; R_s 0.78, p<0.001 for 13-item MTBQ). A weak positive correlation was found between global MTBQ score, number of primary care appointments and number of different GPs consulted within the preceding 12-month period (table 2). Similar weak correlations were found between global TBQ score and the same variables. There was no correlation found between global MTBQ score or global TBQ score and number of long-term conditions.

Slightly more participants agreed that the MTBQ was easy to complete compared with the TBQ (86% vs 80%; p=0.013; table 3). For the MTBQ and TBQ, respectively, 66% and 67% agreed that the questionnaire included all aspects of their condition they were worried about (no significant difference between the two questionnaires, p=1.0; table 3). The proportion of participants who agreed or strongly agreed that the questionnaire was too long or too complicated was 12% and 13%, respectively, for the MTBQ and 13% and 12% for the TBQ (no significant difference between the questionnaires; table 2).

Objective 4: to assess interpretability of the MTBQ in a non-trial population

Grouping global MTBQ scores greater than 0 into tertiles, four categories were generated: no burden (score 0), low burden (score <11), medium burden (12–25), high burden (>25). Categorising treatment burden as medium to high (≥11) or low (0 to 10), younger participants (≥71 years vs 18–70 years; adjusted OR 0.24, 95% CI 0.13 to 0.44, p<0.001), and those with depression (adjusted OR 3.11, 95% CI 1.71 to 5.65, p<0.001), or rheumatoid arthritis (adjusted OR 4.34, 95% CI 1.14 to 16.48, p=0.031) were at greater risk of having high treatment burden (table 4). Treatment burden split by four categories is described in online supplemental file 8.

DISCUSSION

In this study, we examined test–retest reliability, construct validity and ease of completion of the MTBQ, and assessed whether reversing the scale (listing 'extremely difficult' first and 'not difficult' last) improved the floor effects and skewness of the data. There was good evidence for test–retest reliability and a strong positive correlation was found between global MTBQ score and global TBQ score (the comparator questionnaire). Global MTBQ score was weakly correlated with number of consultations and number of different GPs consulted but not with number of long-term conditions. Reversing the scale did not reduce the floor effects or skewness of the data. For both the MTBQ and TBQ, participants mostly agreed the questionnaires were easy to complete, included aspects of their condition they were worried about and were not too long

Table 2 Correlations between global MTBQ score and global TBQ score, number of long-term conditions, number of primary care appointments and number of different GPs consulted in the preceding 12 months

	Global MTBQ score (10 items)			Global MTBQ score (13 items)			Global TBQ score		
	N	R_s	P value	N	R_s	P value	N	R_s	P value
No of long-term conditions	243	−0.063	0.330	243	−0.050	0.441	242	−0.052	0.421
No of primary care appointments in the preceding 12 months	243	0.165	0.010	243	0.183	0.004	242	0.203	0.002
No of different GPs consulted in the preceding 12 months*	220	0.225	<0.001	220	0.235	<0.001	219	0.168	0.013

R_s is the abbreviation for Spearman's rank correlation coefficient.
Note, the analyses presented in table 2 included participants who responded to the original or reversed version of the MTBQ at baseline.
*These analyses only included participants who had at least one GP appointment. Further, 19% of GP appointments were excluded from the analysis as a GP ID was not recorded (see methods).
GP, general practitioner; MTBQ, Multimorbidity Treatment Burden Questionnaire; TBQ, Treatment Burden Questionnaire.

Table 3 Ease of completing the MTBQ and TBQ, measured using four questions from the QQ-10 questionnaire (n=112)

	MTBQ*	TBQ†	McNemar test for proportions‡	
			Relative risk (95% CI)	P value
Questionnaire was easy to complete (n, (%))				
Strongly disagree	1 (0.9)	2 (1.8)	1.10 (1.02 to 1.19)	0.013
Mostly disagree	2 (1.8)	5 (4.5)		
Neither agree or disagree	9 (8.0)	14 (12.5)		
Mostly agree	44 (39.3)	41 (36.6)		
Strongly agree	52 (46.4)	49 (43.8)		
Missing	4 (3.6)	1 (0.9)		
Questionnaire included all aspects of my condition I am worried about (n, (%))				
Strongly disagree	5 (4.5)	4 (3.6)	1.00 (0.88 to 1.13)	1.0
Mostly disagree	6 (5.4)	6 (5.4)		
Neither agree or disagree	22 (19.6)	22 (19.6)		
Mostly agree	50 (44.6)	50 (44.6)		
Strongly agree	24 (21.4)	25 (22.3)		
Missing	5 (4.5)	5 (4.5)		
Questionnaire was too long (n, (%))				
Strongly disagree	53 (47.3)	47 (42.0)	0.73 (0.45 to 1.19)	0.206
Mostly disagree	15 (13.4)	22 (19.6)		
Neither agree or disagree	25 (22.3)	23 (20.5)		
Mostly agree	8 (7.1)	12 (10.7)		
Strongly agree	5 (4.5)	3 (2.7)		
Missing	6 (5.4)	5 (4.5)		
Questionnaire was too complicated (n, (%))				
Strongly disagree	52 (46.4)	52 (46.4)	1.0 (0.62 to 1.61)	1.0
Mostly disagree	25 (22.3)	25 (22.3)		
Neither agree or disagree	16 (14.3)	18 (16.1)		
Mostly agree	6 (5.4)	10 (8.9)		
Strongly agree	8 (7.1)	3 (2.7)		
Missing	5 (4.5)	4 (3.6)		

*MTBQ is the 'Multimorbidity Treatment Burden Questionnaire'.[4]
†TBQ is a comparator questionnaire, the 'Treatment Burden Questionnaire'.[9 10]
‡For the McNemar tests, questionnaire responses were dichotomised into strongly agree/mostly agree versus neither agree or disagree/ mostly disagree/ strongly disagree.

or complicated. As the MTBQ and TBQ[9 10] performed similarly in this study, the choice of which questionnaire to use will likely come down to study preference.

A strength is that the study population comprised randomly selected older adults (mean age 70 years) with multimorbidity (≥3 long-term conditions)—the population for whom the MTBQ is intended. We recruited participants by post and so were able to include those who do not have a smartphone, computer, access to the internet or good information technology literacy skills. Further strengths are that we used questions from the validated QQ-10 questionnaire[17] to assess the ease of completing the MTBQ and the TBQ; and combined survey data with routinely collected data from the GP records.

The low baseline response rate of 31% is a weakness of the study since this may reduce the generalisability of the findings if those who participated differed from those who did not take part. Similar response rates have been reported by other study investigators validating measures of treatment burden internationally,[9–12] and in primary care survey studies in the UK.[26] The majority of study participants self-identified as white British ethnicity, which is also a limitation. We purposefully included two practices serving more ethnically diverse populations, but as our sample has a lower proportion of people from minority ethnic groups than the general UK population, this limits the generalisability of the findings and could potentially lead to selection bias. As the questionnaire was

Table 4 Participant characteristics by category of treatment burden (original version of 10-question MTBQ, n=243)

Characteristic	N	None or low (MTBQ score 0–10)	Medium or high (MTBQ score ≥11)	Univariable analysis		Adjusted analysis	
				OR* (95% CI)	P value	OR* (95% CI)	P value
Participants	243	99 (40.7%)	144 (59.3%)				
Age group (n, (%))							
18–70 years	110	27 (24.5%)	83 (75.5%)	1.00		1.00	
71+ years	132	72 (54.5%)	60 (45.5%)	0.27 (0.16 to 0.47)	<0.001	0.24 (0.13 to 0.44)	<0.001
Gender (n, (%))							
Male	130	56 (43.1%)	74 (56.9%)	1.00		1.00	
Female	112	43 (38.4%)	69 (61.6%)	1.21 (0.73 to 2.03)	0.460	1.26 (0.73 to 2.19)	0.403
Ethnicity (n, (%))							
White	229	94 (41.0%)	135 (59.0%)	1.00		1.00	
Non-white	13	4 (30.8%)	9 (69.2%)	1.57 (0.47 to 5.24)	0.466	1.30 (0.35 to 4.89)	0.694
Deprivation score (n, (%))‡							
Quintile 1–2 (least deprived)	98	47 (48.0%)	51 (52.0%)	1.00		1.00	
Quintile 3	74	24 (32.4%)	50 (67.6%)	1.92 (1.02 to 3.60)	0.042	1.56 (0.80 to 3.05)	0.197
Quintile 4–5 (most deprived)	71	28 (39.4%)	43 (60.6%)	1.42 (0.76 to 2.63)	0.272	0.87 (0.43 to 1.76)	0.696
No of long-term conditions (n, (%))†							
3	136	54 (39.7%)	82 (60.3%)	1.00		1.00	
4	73	29 (39.7%)	44 (60.3%)	1.00 (0.56 to 1.79)	0.998	1.38 (0.72 to 2.64)	0.335
≥5	34	16 (47.1%)	18 (52.9%)	0.74 (0.35 to 1.58)	0.437	1.27 (0.55 to 2.93)	0.577
Age left full-time education (years; n, (%))							
≤16	176	79 (44.9%)	97 (55.1%)	1.00		1.00	
17+	64	18 (28.1%)	46 (71.9%)	2.08 (1.12 to 3.87)	0.021	1.92 (0.99 to 3.73)	0.055
Employment status (n, (%))							
Retired	143	74 (51.7%)	69 (48.3%)	1.00		1.00	
Not retired	100	25 (25.0%)	75 (75.0%)	3.22 (1.84 to 5.63)	<0.001	1.86 (0.87 to 3.96)	0.108
Long-term conditions (n, (%))†							
Cardiovascular disease	209	91 (43.5%)	118 (56.5%)	0.40 (0.17 to 0.92)	0.032	0.48 (0.20 to 1.19)	0.113
Stroke or transient ischaemic attack (TIA)	71	31 (43.7%)	40 (56.3%)	0.84 (0.48 to 1.48)	0.552	0.99 (0.54 to 1.82)	0.978
Diabetes	141	59 (41.8%)	82 (58.2%)	0.90 (0.53 to 1.51)	0.681	0.83 (0.47 to 1.47)	0.522
Chronic kidney disease	68	35 (51.5%)	33 (48.5%)	0.54 (0.31 to 0.96)	0.035	0.58 (0.26 to 1.28)	0.177
COPD or asthma	118	49 (41.5%)	69 (58.5%)	0.94 (0.56 to 1.57)	0.809	0.70 (0.39 to 1.25)	0.229
Epilepsy	20	5 (25.0%)	15 (75.0%)	2.19 (0.77 to 6.22)	0.143	1.49 (0.49 to 4.52)	0.481
Atrial fibrillation	78	43 (55.1%)	35 (44.9%)	0.42 (0.24 to 0.73)	0.002	0.60 (0.31 to 1.15)	0.127
Severe mental health problems	13	5 (38.5%)	8 (61.5%)	1.11 (0.35 to 3.49)	0.864	1.11 (0.32 to 3.85)	0.866
Depression	106	25 (23.6%)	81 (76.4%)	3.81 (2.17 to 6.66)	<0.001	3.11 (1.71 to 5.65)	<0.001
Learning disability	4	1 (25.0%)	3 (75.0%)	2.09 (0.21 to 20.34)	0.527	0.92 (0.09 to 9.38)	0.947
Rheumatoid arthritis	17	3 (17.6%)	14 (82.4%)	3.45 (0.96 to 12.33)	0.057	4.34 (1.14 to 16.48)	0.031
Heart failure	31	17 (54.8%)	14 (45.2%)	0.52 (0.24 to 1.11)	0.091	0.52 (0.21 to 1.27)	0.151

*ORs>1 indicate a higher odds of medium/high treatment burden in patients with this characteristic compared with the reference characteristic (indicated by an OR of 1). For the multivariable analyses, each model was adjusted for age, gender, deprivation and number of comorbidities.
†Long-term conditions from electronic GP records. Please note, similar long-term conditions are grouped together (eg, COPD/asthma, stroke/TIA).
‡Based on Townsend Deprivation Index scores.[18]
COPD, chronic obstructive pulmonary disease; GP, general practitioner; MTBQ, Multimorbidity Treatment Burden Questionnaire.

self-administered, we were unable to include those with a poor level of English literacy to read the questionnaire. A further limitation is that reading similar questions from the MTBQ and TBQ in the same questionnaire pack could have influenced participants' responses, although we tried to mitigate the effects of this by randomising the order the MTBQ and TBQ questionnaires were presented in. One limitation is that, while the study was

designed to assess the primary outcome of test–retest reliability, it was not necessarily large enough to detect multiple associations between treatment burden and some patient characteristics. The sample size (n=243) was 'very good' according to the COSMIN criteria,[25] but for some subgroup analyses (eg, less common long-term conditions), sample sizes were low, precluding an adequate test of certain relationships. A final limitation is that we were unable to resolve the skewed distribution of responses. This merits further investigation, for example, through exploring other changes to response options; alternatively, it may indicate that experience of burden is inevitably skewed rather than being a problem of the measures.

Interestingly, we found that younger participants were more likely to report high treatment burden, a phenomenon found in the original MTBQ study,[4] Danish population survey[5] and in Tran's TBQ validation study.[10] This may be explained by younger people having more caring and work responsibilities, and reduced capacity to manage the workload of looking after their health.[27] In this and several other studies, including the original MTBQ validation study, people with depression were more likely to report high treatment burden.[4 12 13] In contrast to the original and much larger MTBQ study[4] but in keeping with studies in the USA,[14 28] we did not find an association between high treatment burden score and number of long-term conditions. We found modest associations between high treatment burden and number of consultations and poor continuity of care. We would not necessarily expect strong associations because the number of consultations and number of different healthcare professionals would only affect one aspect of treatment burden. Furthermore, the relationships between treatment burden, number of consultations and continuity of care are complex. A high number of consultations could lead to high treatment burden caused by having to arrange and attend multiple appointments, but this could also reflect good access to healthcare appointments, and subsequent reduced treatment burden. Similarly, seeing a healthcare professional whom you know and trust (good continuity of care) often comes at the expense of having to wait longer for an appointment, which could in turn increase treatment burden. The cut-off values for the four treatment burden groups were slightly higher in this study compared with the original study.[4] For studies using the MTBQ, we recommend using the original study cut-off values: no burden (score 0), low burden (score <10), medium burden (10–21) high burden (>21). Further research, such as anchor-based methods, is needed to determine the clinical significance of global MTBQ scores.

The MTBQ is a simply worded concise measure of treatment burden for patients with multimorbidity. This study provides further evidence of the scale's psychometric properties, including test–retest reliability, construct validity and ease of completion. These findings can be combined with the original validation study, where the MTBQ demonstrated good content validity, construct validity, internal consistency reliability and preliminary evidence of responsiveness. The MTBQ was developed and validated primarily as a research tool and has been widely used in interventional and observational studies. Further work is underway to develop and validate an adapted version of the MTBQ, known as the 'Short Treatment Burden Questionnaire', for use in clinical settings.[29]

X Polly Duncan @polly_duncan and Chris Salisbury @prof_tweet

Acknowledgements Appreciation is extended to the General Practices and their patients for taking part in the study, and to the PiPCare members (comprising people with two or more long-term conditions) for providing valuable patient and public insights into the study design. We also thank Dr Viet-Thi Tran for permission to use the Treatment Burden Questionnaire (TBQ).

Contributors PD led this project under the supervision of CS, and both were responsible for the concept and study design. CS provided methodological expertise in assessing the psychometric properties of the MTBQ. PD, SD, MM and KC were involved in inviting participants and transcribing data from the surveys into the database, extracting anonymous data from the electronic GP records. LJS led the analysis with input from DG, PD, YP, LG and CS. PD and LJS drafted the manuscript under supervision of CS. All authors critically reviewed the manuscript and approved the final version. PD is the guarantor.

Funding This study was part funded by the Avon Primary Care Research Collaborative (grant no: NA), an NIHR Doctoral Research Fellowship award to PD (NIHR301824) and an NIHR Senior Investigator award to CS (NIHR201314). It was also supported by the National Institute for Health Research (NIHR) Applied Research Collaboration West (ARC West) at University Hospitals Bristol and Weston NHS Foundation Trust (grant no: NA).

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care.

Competing interests PD and CS developed and validated the MTBQ. CS is an NIHR Senior Investigator.

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 Not required.

Ethics approval This study involves human participants and ethical approval was obtained by the Faculty of Health Sciences Research Ethics Committee (FREC), University of Bristol (18/LO/1051, IRAS 236536). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. The participants of this study did not give written consent for their individual anonymised data to be shared publicly, so the research supporting data are not available.

ORCID iDs
Polly Duncan http://orcid.org/0000-0002-2244-3254
Lauren J Scott http://orcid.org/0000-0003-3129-5123

Shoba Dawson http://orcid.org/0000-0002-6700-6445
Line Guenette http://orcid.org/0000-0001-9769-7550
Chris Salisbury http://orcid.org/0000-0002-4378-3960

REFERENCES

1 Salisbury C, Johnson L, Purdy S, *et al*. Epidemiology and impact of Multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011;61:e12–21.
2 Eton DT, Ramalho de Oliveira D, Egginton JS, *et al*. Building a measurement framework of burden of treatment in complex patients with chronic conditions: a qualitative study. *Patient Relat Outcome Meas* 2012;3:39–49.
3 Salisbury C, Man M-S, Bower P, *et al*. Management of Multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3d approach. *Lancet* 2018;392:41–50.
4 Duncan P, Murphy M, Man M-S, *et al*. Development and validation of the Multimorbidity treatment burden questionnaire (MTBQ). *BMJ Open* 2018;8:e019413.
5 Pedersen MH, Duncan P, Lasgaard M, *et al*. Danish validation of the Multimorbidity treatment burden questionnaire (MTBQ) and findings from a population health survey: a mixed-methods study. *BMJ Open* 2022;12:e055276.
6 Schulze J, Breckner A, Duncan P, *et al*. Adaptation and validation of a German version of the Multimorbidity treatment burden questionnaire. *Health Qual Life Outcomes* 2022;20:90.
7 Guénette L, Turcotte V, Bélanger L, *et al*. Multimorbidity treatment burden questionnaire (MTBQ): translation, cultural adaptation, and validation in French-Canadian. *Can J Aging* 2023;42:126–34.
8 Dou L, Huang J, Duncan P, *et al*. Translation, cultural adaptation and validation of the Chinese Multimorbidity treatment burden Questionnaire(C-MTBQ): a study of older hospital patients. *Health Qual Life Outcomes* 2020;18:194.
9 Tran V-T, Harrington M, Montori VM, *et al*. Adaptation and validation of the treatment burden questionnaire (TBQ) in English using an Internet platform. *BMC Med* 2014;12:109.
10 Tran V-T, Montori VM, Eton DT, *et al*. Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. *BMC Med* 2012;10:68.
11 Eton DT, Yost KJ, Lai J, *et al*. Development and validation of the patient experience with treatment and self-management (PETS): a patient-reported measure of treatment burden [Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2016]. *Qual Life Res* 2017;26:489–503.
12 Boyd CM, Wolff JL, Giovannetti E, *et al*. Healthcare task difficulty among older adults with Multimorbidity. *Med Care* 2014;52 Suppl 3(0 3):S118–25.
13 Gibbons CJ, Kenning C, Coventry PA, *et al*. Development of a Multimorbidity illness perceptions scale (multiples). *PLoS ONE* 2013;8:e81852.
14 Eton DT, Linzer M, Boehm DH, *et al*. Deriving and validating a brief measure of treatment burden to assess person-centered Healthcare quality in primary care: a multi-method study. *BMC Fam Pract* 2020;21:221.
15 Sutcliffe D, Lester H, Hutton J, *et al*. NICE and the quality and outcomes framework (QOF) 2009-2011. *Qual Prim Care* 2012;20:47–55.
16 Man M-S, Chaplin K, Mann C, *et al*. Improving the management of Multimorbidity in general practice: protocol of a cluster randomised controlled trial (the 3d study). *BMJ Open* 2016;6:e011261.
17 Moores KL, Jones GL, Radley SC. Development of an instrument to measure face validity, feasibility and utility of patient questionnaire use during health care: the QQ-10. *Int J Qual Health Care* 2012;24:517–24.
18 Yousaf S, Bonsall A. *UK Townsend Deprivation Scores from 2011 census data*. Colchester, UK: UK Data Service, 2017.
19 Murphy M, Scott LJ, Salisbury C, *et al*. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. *Br J Gen Pract* 2021;71:e166–77.
20 Murphy MT, Denholm A, Scott R, *et al*. Rapid COVID-19 intelligence to improve primary care response (RAPCI) final project report. University of Bristol Centre for Academic Primary Care (CAPC; 2020Sep.
21 Fayers PM, Machin D. Quality of life. In: *Quality of Life*. Assessment,Analysis & Interpretation: John Wiley & Sons Ltd, 18 April 2000.
22 Martin Bland J, Altman DouglasG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;327:307–10.
23 McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;12:153–7.
24 Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. *Stat Methods Med Res* 2004;13:251–71.
25 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
26 Warren FC, Abel G, Lyratzopoulos G, *et al*. Characteristics of service users and provider organisations associated with experience of out of hours general practitioner care in England: population based cross sectional postal questionnaire survey. *BMJ* 2015;350:h2040.
27 Shippee ND, Shah ND, May CR, *et al*. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. *J Clin Epidemiol* 2012;65:1041–51.
28 Eton DT, Anderson RT, St Sauver JL, *et al*. Longitudinal Trajectories of treatment burden: A prospective survey study of adults living with multiple chronic conditions in the midwestern United States. *J Multimorb Comorb* 2022;12:26335565221081291.
29 Johnson R, Kovalenko AG, Blakeman T, *et al*. Treatment burden in multiple long-term conditions: a mixed-methods study protocol. *BJGP Open* 2023;7:BJGPO.

Supplementary file 1: Sample of questions from the original and reversed versions of the MTBQ

Sample of questions from the original MTBQ

We are interested in finding out about the effort you have to make to look after your health and how this impacts on your day-to-day life.

Please tell us how much difficulty you have with the following:
(Please tick the box that most applies to you)

	Not Difficult	A little Difficult	Quite Difficult	Very Difficult	Extremely Difficult	Does not apply
1. Taking lots of medications	☐0	☐1	☐2	☐3	☐4	☐0
2. Remembering how and when to take medication	☐0	☐1	☐2	☐3	☐4	☐0
3. Paying for prescriptions, over the counter medication or equipment	☐0	☐1	☐2	☐3	☐4	☐0
4. Collecting prescription medication	☐0	☐1	☐2	☐3	☐4	☐0
5. Monitoring your medical conditions (e.g. checking your blood pressure or blood sugar, monitoring your symptoms etc.)	☐0	☐1	☐2	☐3	☐4	☐0
6. Arranging appointments with health professionals	☐0	☐1	☐2	☐3	☐4	☐0
7. Seeing lots of different health professionals	☐0	☐1	☐2	☐3	☐4	☐0

Permission must be sought to use the 'Multimorbidity Treatment Burden Questionnaire' (MTBQ)[4] via the website: https://www.bristol.ac.uk/primaryhealthcare/resources/mtbq/

Sample of questions from the reversed scale MTBQ

We are interested in finding out about the effort you have to make to look after your health and how this impacts on your day-to-day life.

Please tell us how much difficulty you have with the following:
(Please tick the box that most applies to you)

	Extremely Difficult	Very Difficult	Quite Difficult	A little Difficult	Not Difficult	Does not apply
1. Taking lots of medications	☐4	☐3	☐2	☐1	☐0	☐0
2. Remembering how and when to take medication	☐4	☐3	☐2	☐1	☐0	☐0
3. Paying for prescriptions, over the counter medication or equipment	☐4	☐3	☐2	☐1	☐0	☐0
4. Collecting prescription medication	☐4	☐3	☐2	☐1	☐0	☐0
5. Monitoring your medical conditions (e.g. checking your blood pressure or blood sugar, monitoring your symptoms etc.)	☐4	☐3	☐2	☐1	☐0	☐0
6. Arranging appointments with health professionals	☐4	☐3	☐2	☐1	☐0	☐0
7. Seeing lots of different health professionals	☐4	☐3	☐2	☐1	☐0	☐0

Permission must be sought to use the 'Multimorbidity Treatment Burden Questionnaire' (MTBQ)[4] via the website: https://www.bristol.ac.uk/primaryhealthcare/resources/mtbq/

Supplementary file 2: COSMIN study design checklist

General recommendations for the design of a study on measurement properties					
	Very good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Research aim					
1. Provide a clear research aim, including (1) the name and version of the PROM, (2) the target population, and (3) the measurement properties of interest.	Research aim clearly described				
PROM					
2. Provide a clear description of the construct to be measured.	Construct clearly described				
3. Provide a clear description of the development process of the PROM, including a description of the target population for which the PROM was developed.	Development process clearly described				
4. The origin of the construct should be clear: provide a theory, conceptual framework (i.e. reflective or formative model) or disease model used or clear rationale to define the construct to be measured.	Origin of the construct clear				
5. Provide a clear description of the structure of the PROM (i.e. the number of items and subscales included in the PROM, instructions given and response options) and its scoring algorithm.	Structure and scoring algorithm clearly described				
6. Provide a clear description of existing evidence on the quality of the PROM.	Existing evidence on the quality of the PROM clearly described				

7. Provide a clear description of the context of use*	Context of use clearly described
Target population	
8. Provide a clear description of in- and exclusion criteria to select patients, e.g. in terms of disease condition and characteristics like age, gender, language or country, and setting (e.g. general population, primary care or hospital/rehabilitation care)	In- and exclusion criteria for patients clearly described
9. Provide a clear description of the method used to select the patients for the study (e.g. convenience, consecutive, or random)	Method for patient selection clearly described
10. Describe whether the selected sample is representing the target population in which the PROM will be used in terms of age, gender, important disease characteristics (e.g. severity, status, duration)	Study sample representing the target population clearly described

* The context of use refers to the intended application of the PROM (e.g. for research or clinical practice), to a specific setting for which the PROM was developed (e.g. for use in a hospital or at home) or to a specific administration mode (e.g. paper or computer-administered). If the PROM was developed for use across multiple contexts, this should be described.

Measurement error and reliability					
	Very good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Design requirements					
1. Use at least two measurements.	At least two measurements				
2. Ensure that the administrations will be independent.	Independent measurements				
3. Ensure that the patients will be stable in the interim period on the construct to be measured.		Assumable that patients will be stable			
4. Use an appropriate time interval between the two measurements, which is long enough to prevent recall, and short enough to ensure that patients remain stable.	Time interval appropriate				
5. Ensure that the test conditions will be similar for the measurements (e.g. type of administration, environment, instructions).	Test conditions similar (evidence provided)				
6. Perform the analysis in a sample with an appropriate number of patients (taking into account expected number of missing values)	≥100 patients				
Statistical methods for measurement error					
7. For continuous scores: calculate an intraclass correlation coefficient (ICC)	ICC will be calculated, and model or formula of the ICC is clearly described*				
8. For dichotomous/nominal/ordinal scores: calculate kappa					Not applicable

9.	For ordinal scores: calculate a weighted kappa		Not applicable
10.	Provide a clear description of how missing items will be handled	The way missing items will be handled is clearly described.	

** The model (i.e. one-way random effect model or two-way random or mixed effect model), type (i.e. for single or multiple measurement) and definition (i.e. for consistency or absolute agreement) of the ICC that will be calculated is appropriately chosen and described (see 11); ** ICC formula does not correspond to the research question

Hypothesis testing for construct validity **A. Comparison with other outcome measurement instruments (convergent validity)**					
	Very good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Design requirements					
1. Formulate hypotheses about expected relationships between the PROM under study and other outcome measurement instrument(s).	Hypotheses formulated including the expected direction and magnitude of the correlations stated.				
2. Provide a clear description of the construct(s) measured by the comparator instrument(s).	Construct(s) measured by the comparator instrument(s) is/are clearly described.				
3. Use comparator instrument(s) with sufficient measurement properties.	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population.				
4. Perform the analysis in a sample with an appropriate number of patients (taking into account expected number of missing values)	≥100 patients				

5. Use an appropriate time schedule for assessments of the PROM of interest and comparison instruments.	PROM and comparison instrument(s) will be administered at the same time.
Statistical methods	
6. Use statistical methods that are appropriate for the hypotheses to be tested	Statistical methods will be appropriate.
7. Provide a clear description of how missing items will be handled	The way missing items will be handled is clearly described.

https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf

Hypothesis testing for construct validity **B. Comparison between subgroups (discriminative or known-groups validity)**					
	Very good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Design requirements					
1. Formulate hypotheses regarding mean differences between subgroups.	Hypotheses formulated including the expected directions and magnitude of the mean differences stated.				
2. Provide an adequate description of important characteristics of the subgroups, such as disease or demographic characteristics.	Adequate description of the important characteristics of the subgroups.				
3. Perform the analysis in a sample with an appropriate number of patients (taking into account expected number of missing values)	≥100 patients				
Statistical methods					
4. Use statistical methods that are appropriate for the hypotheses to be tested.	Statistical methods will be appropriate.				
5. Provide a clear description of how missing items will be handled	The way missing items will be handled is clearly described				

Supplementary file 3: Comparison of participant characteristics between baseline participants, participants who completed the baseline original version versus the reversed version of the MTBQ and participants who completed the follow-up survey.

	Baseline survey (n=244)		**Baseline original MTBQ (n=112)**		**Baseline reversed scale MTBQ (n=132)**		**Follow-up original MTBQ (n=105)**		**Follow-up reversed scale MTBQ (n=120)**	
	n	%	n	%	n	%	n	%	n	%
Age (years; mean, SD)	69.9	13.1	71.9	11.6	68.1	14.0	71.3	11.5	68.3	13.7
18-50 years	18	7.4	4	3.6	14	10.6	4	3.8	12	10.0
51-60 years	37	15.2	15	13.4	22	16.7	15	14.3	19	15.8
61-70 years	55	22.5	25	22.3	30	22.7	25	23.8	29	24.2
71-80 years	82	33.6	42	37.5	40	30.3	40	38.1	37	30.8
81-90 years	47	19.3	22	19.6	25	18.9	18	17.1	22	18.3
91+ years	4	1.6	3	2.7	1	0.8	2	2.9	1	0.8
Missing	1	0.4	1	0.9	0	0.0	0	0.0	0	0.0
Gender										
Male	130	53.3	61	54.5	69	52.3	56	53.3	62	51.7
Female	113	46.3	50	44.6	63	47.7	48	45.7	58	48.3
Missing	1	0.4	1	0.9	0	0.0	1	1.0	0	0.0
Ethnicity										
White	229	93.9	106	94.6	123	93.2	100	95.2	113	94.2
Asian	5	2.0	2	1.8	3	2.3	1	1.0	2	1.7
Black/African/Caribbean	4	1.6	2	1.8	2	1.5	2	1.9	1	0.8
Mixed	4	1.6	1	0.9	3	2.3	1	1.0	3	2.5
Other	1	0.4	1	0.9	0	0.0	1	1.0	0	0.0
Missing	1	0.4	0	0.0	1	0.8	0	0.0	1	0.8
Number of long-term conditions										
3	136	55.7	58	51.8	78	59.1	55	52.4	69	57.5
4	73	29.9	35	31.3	38	28.8	32	30.5	35	29.2
≥5	35	14.3	19	17.0	16	12.1	18	17.1	16	13.3
Long-term conditions*										
Cardiovascular disease	210	86.1	98	87.5	112	84.8	92	87.6	104	86.7
Stroke or transient ischaemic attack	72	29.5	39	34.8	33	25.0	37	35.2	30	25.0
Diabetes	141	57.8	60	53.6	81	61.4	56	53.3	76	63.3
Chronic kidney disease	68	27.9	30	26.8	38	28.8	29	27.6	35	29.2
Chronic obstructive pulmonary disease or asthma	119	48.8	51	45.5	68	51.5	48	45.7	63	52.5
Epilepsy	20	8.2	12	10.7	8	6.1	11	10.5	7	5.8
Atrial fibrillation	79	32.4	41	36.6	38	28.8	40	38.1	33	27.5
Severe mental health problems	13	5.3	3	2.7	10	7.6	2	1.9	10	8.3
Depression	107	43.9	51	45.5	56	42.4	49	46.7	51	42.5
Learning disability	4	1.6	0	0.0	4	3.0	0	0.0	3	2.5
Rheumatoid arthritis	17	7.0	12	10.7	5	3.8	9	8.6	5	4.2
Heart failure	31	12.7	16	14.3	15	11.4	14	13.3	12	10.0

*Long-term conditions from electronic GP records. Please note, similar long-term conditions are grouped together (e.g. COPD/asthma, stroke/TIA)

** Based on Townsend scores[18]

Supplementary file 3 (continued): Comparison of participant characteristics between participants who completed the baseline original version, baseline reversed scale version, follow-up original version and follow-up reversed scale version

	Baseline survey (n=244)		**Baseline original MTBQ (n=112)**		**Baseline reversed scale MTBQ (n=122)**		**Follow-up original MTBQ (n=105)**		**Follow-up reversed scale MTBQ (n=120)**	
	n	%	n	%	n	%	n	%	n	%
Age left full-time education (years)										
≤14	22	9.0	9	8.0	13	9.8	7	6.7	13	10.8
15 or 16	155	63.5	77	68.8	78	59.1	74	70.5	71	59.2
17 or 18	33	13.5	14	12.5	19	14.4	13	12.4	16	13.3
≥19	31	12.7	11	9.8	20	15.2	10	9.5	18	15.0
Missing	3	1.2	1	0.9	2	1.5	1	1.0	2	1.7
Employment status										
Fully retired from work	144	59.0	69	61.6	75	56.8	65	61.9	69	57.5
Employed	36	14.8	13	11.6	23	17.4	13	12.4	22	18.3
Other	64	26.2	30	26.8	34	25.8	27	25.7	29	24.2
Deprivation score**										
Quintile 1 (least deprived)	49	20.1	26	23.2	23	17.4	24	22.9	22	18.3
Quintile 2	49	20.1	22	19.6	27	20.5	19	18.1	23	19.2
Quintile 2	74	30.3	33	29.5	41	31.1	32	30.5	35	29.2
Quintile 4	47	19.3	21	18.8	26	19.7	20	19.0	25	20.8
Quintile 5 (most deprived)	25	10.2	10	8.9	15	11.4	10	9.5	15	12.5

Supplementary file 4: Responses to the original and reversed versions of the MTBQ (baseline questionnaire)

Please tell us how much difficulty you have with the following:	Original MTBQ (n = 112)		Reversed scale MTBQ (n = 132)	
	n	%	n	%
1. Taking lots of medications				
Does not apply	4	3.6	7	5.3
Not difficult	57	50.9	72	54.9
A little difficult	35	31.3	30	22.7
Quite difficult	9	8.0	12	9.1
Very difficult	5	4.5	8	6.1
Extremely difficult	2	1.8	2	1.5
Missing	*0*	*0.0*	1	0.8
2. Remembering how and when to take medications				
Does not apply	2	1.8	3	2.3
Not difficult	57	50.9	79	60.8
A little difficult	34	30.4	28	21.2
Quite difficult	9	8.0	12	9.1
Very difficult	6	5.4	5	3.8
Extremely difficult	3	2.7	3	2.3
Missing	1	0.9	2	1.5
3. Paying for prescriptions, over the counter medication or equipment				
Does not apply	85	75.9	106	80.9
Not difficult	15	13.4	18	13.7
A little difficult	4	3.6	1	0.8
Quite difficult	4	3.6	3	2.3
Very difficult	1	0.9	2	1.5
Extremely difficult	1	0.9	1	0.8
Missing	2	1.8	1	0.8
4. Collecting prescription medication				
Does not apply	21	18.8	16	12.5
Not difficult	36	32.1	56	43.8
A little difficult	21	18.8	32	24.2
Quite difficult	22	19.6	12	9.1
Very difficult	9	8.0	5	3.8
Extremely difficult	3	2.7	7	5.3
Missing	0	0.0	4	3.0
5. Monitoring your medical conditions (e.g. checking your blood pressure or blood sugar, monitoring your symptoms, etc)				
Does not apply	21	18.8	32	24.4
Not difficult	53	47.3	63	48.1
A little difficult	16	14.3	19	14.4
Quite difficult	14	12.5	13	9.8
Very difficult	3	2.7	1	0.8
Extremely difficult	5	4.5	3	2.3
Missing	0	0.0	1	0.8
6. Arranging appointments with health professionals				
Does not apply	6	5.4	5	3.9
Not difficult	38	33.9	51	39.2
A little difficult	32	28.6	41	31.1
Quite difficult	21	18.8	22	16.7
Very difficult	8	7.1	4	3.0
Extremely difficult	7	6.3	7	5.3
Missing	0	0.0	2	1.5

Supplementary file 4: Responses to the original and reversed versions of the MTBQ (baseline questionnaire)

	Original MTBQ (n = 112)		Reversed scale MTBQ (n = 132)	
Please tell us how much difficulty you have with the following:	n	%	n	%
7. Seeing lots of different health professionals				
Does not apply	13	11.6	16	12.2
Not difficult	37	33.0	62	47.3
A little difficult	26	23.2	24	18.2
Quite difficult	21	18.8	20	15.2
Very difficult	10	8.9	5	3.8
Extremely difficult	5	4.5	4	3.0
Missing	0	0.0	1	0.8
8. Attending appointments with health professionals (e.g. getting time off work, arranging transport etc)				
Does not apply	33	29.7	36	27.5
Not difficult	39	35.1	52	39.7
A little difficult	14	12.5	22	16.7
Quite difficult	13	11.6	16	12.1
Very difficult	6	5.4	4	3.0
Extremely difficult	6	5.4	1	0.8
Missing	1	0.9	1	0.8
9. *Getting healthcare in the* evenings and at weekends				
Does not apply	58	52.7	66	50.8
Not difficult	14	12.7	24	18.5
A little difficult	10	8.9	14	10.6
Quite difficult	11	9.8	12	9.1
Very difficult	8	7.1	8	6.1
Extremely difficult	9	8.0	6	4.5
Missing	2	1.8	2	1.5
10. Getting help from community services (e.g. physiotherapy, district nurses etc)				
Does not apply	61	55.5	69	53.1
Not difficult	21	19.1	26	20.0
A little difficult	11	9.8	13	9.8
Quite difficult	8	7.1	14	10.6
Very difficult	7	6.3	2	1.5
Extremely difficult	2	1.8	6	4.5
Missing	2	1.8	2	1.5
11. Obtaining clear and up-to-date information about your condition				
Does not apply	5	4.5	6	4.6
Not difficult	52	46.9	68	52.3
A little difficult	23	20.5	30	22.7
Quite difficult	19	17.0	23	17.4
Very difficult	8	7.1	2	1.5
Extremely difficult	4	3.6	1	0.8
Missing	1	0.9	2	1.5

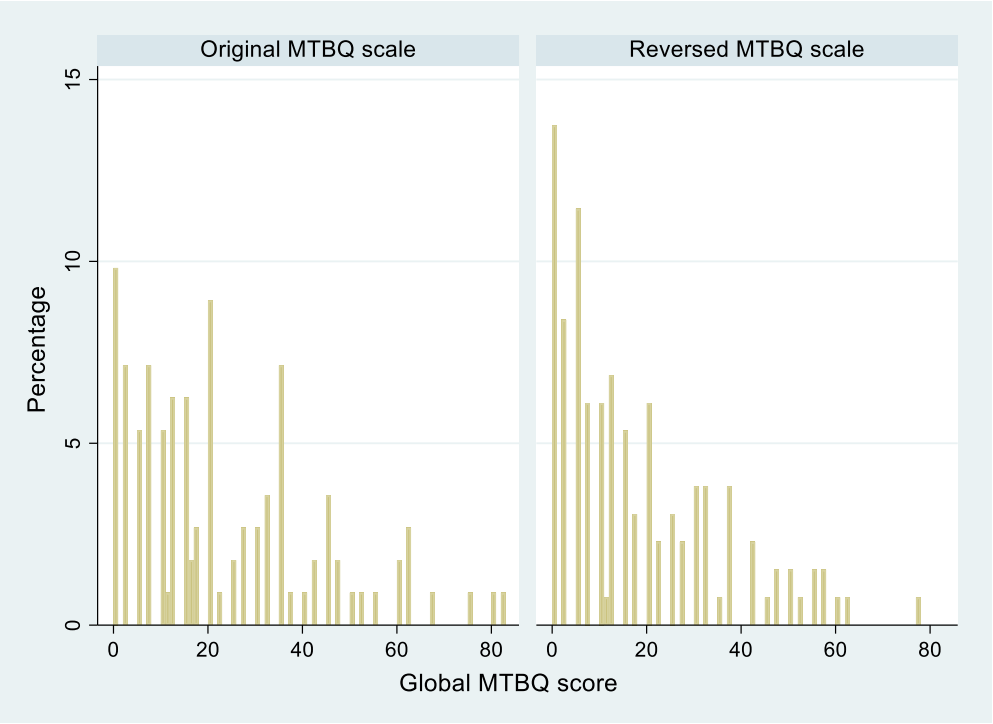
Supplementary file 4 (continued): Responses to the original and reversed versions of the MTBQ (baseline questionnaire)

	Original MTBQ (n = 112)		**Reversed MTBQ (n = 132)**	
Please tell us how much difficulty you have with the following:	n	%	n	%
12. Making recommended lifestyle changes (e.g. diet and exercise)				
Does not apply	9	8.0	12	9.2
Not difficult	34	30.4	48	36.9
A little difficult	26	23.2	32	24.2
Quite difficult	23	20.5	22	16.7
Very difficult	10	8.9	10	7.6
Extremely difficult	9	8.0	6	4.5
Missing	1	0.9	2	1.5
13. Having to rely on help from family and friends				
Does not apply	24	21.4	40	30.3
Not difficult	38	33.9	41	31.1
A little difficult	15	13.4	24	18.2
Quite difficult	20	17.9	11	8.3
Very difficult	8	7.1	8	6.1
Extremely difficult	7	6.3	8	6.1
Missing	0	0	0	0
Number of missing responses per participant				
0	103	92.0	122	92.4
1	8	7.1	6	4.5
≥2	1	0.9	4	3.1
Global MTBQ score (10-question version, excluding optional questions 2, 9 and 10)				
Median, IQR	17.1 (7.5, 35.0)		12.5 (5.0, 27.5)	
Score of 0 (n, (%)	11 (9.8)		18 (13.7)	
Chi-squared test for score of zero	p-value = 0.348			
Global MTBQ score (13-question version)				
Median (IQR)	15.4 (5.8, 32.7)		11.5 (3.8, 26.9)	
Score of 0 (n, (%))	11 (9.8)		17 (13.0)	
Chi-squared test for score of zero	p-value = 0.443			

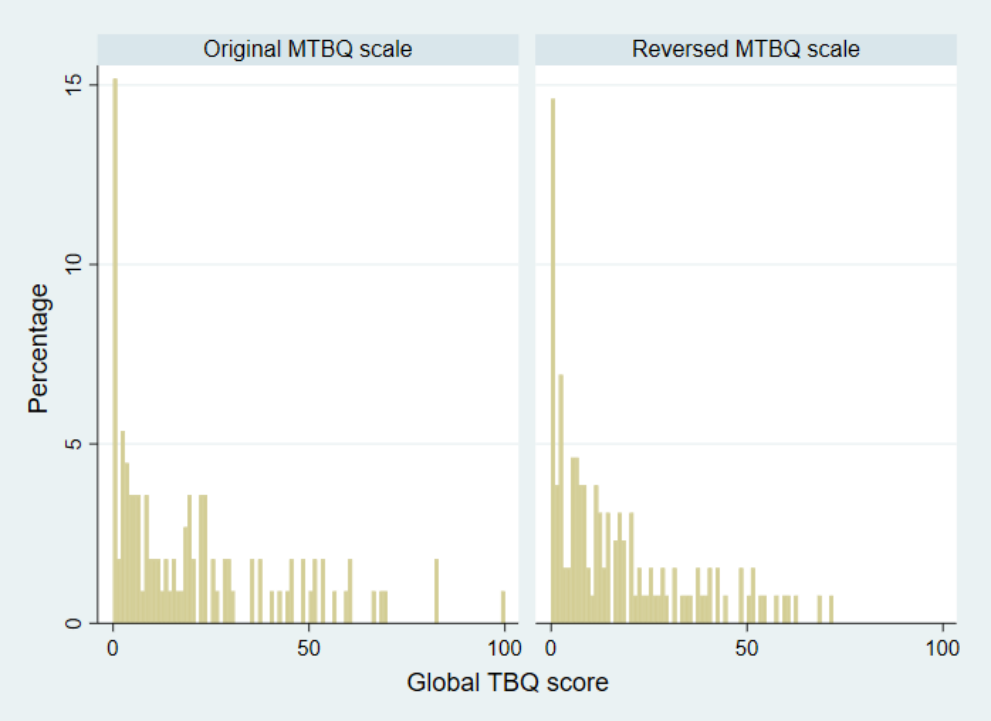
Note: Questions 2, 9 and 10 were excluded from the main analysis of the original MTBQ paper due to a high proportion of 'does not apply' responses. They are shown in italics. As they may be relevant to other populations, they can be considered as optional.

Supplementary file 5: Histogram of the global MTBQ score and global TBQ score, for the original and reversed versions of the MTBQ questionnaire

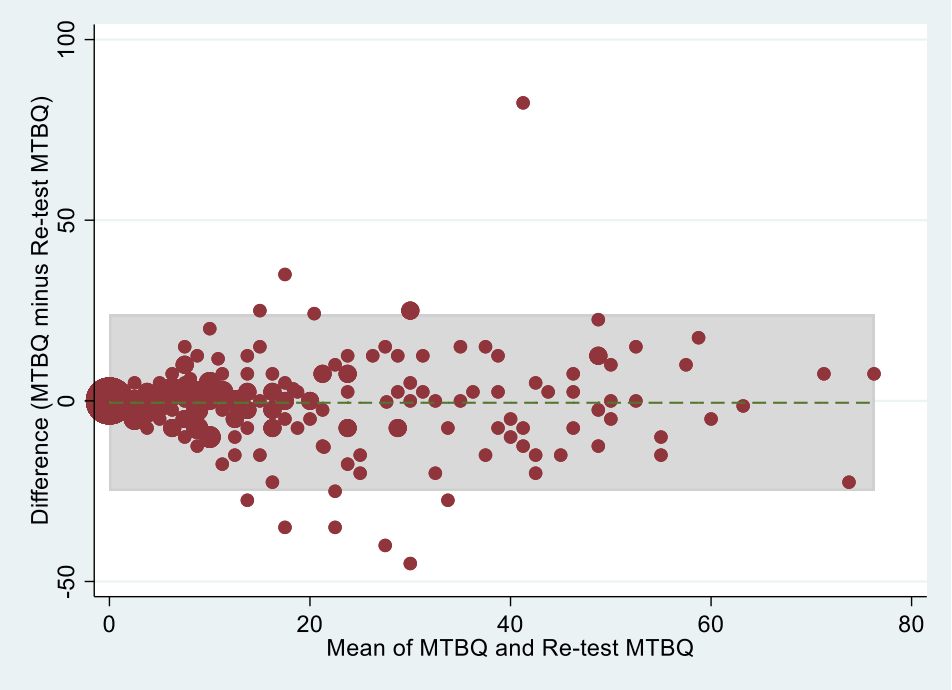
Distribution of global MTBQ scores for participants who received the original and reversed versions of the MTBQ



Distribution of global Treatment Burden Questionnaire (TBQ) scores for participants who received the original and reversed versions of the MTBQ



Supplementary file 6: Bland Altman plot showing the mean of the baseline and re-test global MTBQ score versus the difference between the two scores.



Supplementary file 7a: Comparison of floor effect and missing data between similar questions from the MTBQ and TBQ

MTBQ question (n = 112)		**Floor effect***	**Missing data**	**TBQ question with a similar latent construct (n = 112)**		**Floor effect**	**Missing data**
	Global MTBQ score	9.8%	8.0%		Global TBQ score	11.6%	14.3%
1.	Taking lots of medications	54.5%	0.0%	1b.	The number of times you should take your medication daily?	52.6%	5.4%
2.	Remembering how and when to take medication	52.7%	0.9%	1c.	The efforts you make not to forget to take your medications (for example: managing your treatment when you are away from home, preparing and using pillboxes…)	54.5%	3.6%
3.	*Paying for prescriptions, over the counter medication or equipment*	89.3%	1.8%	3b.	The financial burden associated with your healthcare (for example: out of pocket expenses or expenses not covered by insurance)?	64.3%	0.9%
5.	Monitoring your medical conditions (eg. checking your blood pressure or blood sugar, monitoring your symptoms etc)	66.1%	0.0%	2b.	Self-monitoring (for example, taking your blood pressure or checking your blood sugar): frequency, time spent and associated nuisances or inconveniences	70.5%	0.9%
6.	Arranging appointments with health professionals	39.3%	0.0%	2e.	Arranging medical appointments (doctors' visits, lab tests and other exams) and reorganizing your schedule around these appointments	50.9%	0.0%

Supplementary file 7a (continued): Comparison of floor effect and missing data between similar questions from the MTBQ and TBQ

MTBQ question (n = 112)		Floor effect	Missing data	TBQ question with a similar latent construct (n = 112)		Floor effect	Missing data
8.	Attending appointments with health professionals (e.g. getting time off work, arranging transport etc)	64.3%	0.9%	2e	Arranging medical appointments (doctors' visits, lab tests and other exams) and reorganizing your schedule around these appointments	50.9%	0.0%
12.	Making recommended lifestyle changes (eg. diet and exercise)	38.4%	0.9%	3c.	The burden related to dietary changes (for example: avoiding certain foods or alcohol, having to quit smoking...)?	50.9%	0.0%
				3d.	The burden related to doctors' recommendations to practice physical activity (for example: walking, jogging, swimming…)?	35.7%	0.9%
13.	Having to rely on help from family and friends	55.4%	0.0%	3e.	How does your healthcare impact your relationships with others (for example, needing assistance in everyday life, being ashamed to take your medication…)?	59.8%	0.0%

*The floor effect is the proportion of participants who responded 'does not apply' or 'not difficult'

We received permission to use the Treatment Burden Questionnaire (TBQ) for this study.[10 11] Please do not use the questions from the TBQ without permission. Contact information and permission to use the TBQ: Mapi Research Trust, Lyon, France – Internet: https://eprovide.mapi-trust.org. TBQ © Ravaud et al, 2012. All Rights Reserved.

Supplementary file 7b: Floor effect and missing data for MTBQ questions with no comparator question from the TBQ

		Floor effect	Missing data
4.	Collecting prescription medication	50.9%	0.0%
7.	Seeing lots of different health professionals	44.6%	0.0%
10.	*Getting help from community services (eg. physiotherapy, district nurses etc)*	73.2%	1.8%
11.	Obtaining clear and up-to-date information about your condition	50.9%	0.9%

*The floor effect is the proportion of participants who responded 'does not apply' or 'not difficult'

Supplementary file 7c: Floor effect and missing data for TBQ questions with no comparator question from the MTBQ (n=112)

		Floor effect	Missing data
1a.	The taste, shape or size of your tablets and/or the annoyances caused by your injections (for example, pain, bleeding, bruising or scars)?	54.5%	5.4%
1d.	The necessary precautions when taking your medication (for example: taking them at specific times of the day or meals, not being able to do certain things after taking medications such as driving or lying down…)	63.4%	2.7%
2a.	Lab tests and other exams (for example: blood tests or radiology): frequency, time spent and associated nuisances or inconveniences	59.8%	1.8%
2c.	Doctor visits and other appointments: frequency and time spent for these visits and difficulties finding healthcare providers	56.3%	0.9%
2d.	The difficulties you could have in your relationships with healthcare providers (for example: feeling not listened to enough or not taken seriously)	61.6%	0.0%
3a.	The administrative burden related to healthcare (for example: all you have to do for hospitalizations, reimbursements and/or obtaining social services)?	73.2%	1.8%
4.	'The need for medical healthcare on a regular basis reminds me of my health problems'	42.9%	0.9%

*The floor effect is the proportion of participants who responded 'does not apply' or 'not difficult'

We received permission to use the Treatment Burden Questionnaire (TBQ) for this study.[10 11] Please do not use the questions from the TBQ without permission. Contact information and permission to use the TBQ: Mapi Research Trust, Lyon, France – Internet: https://eprovide.mapi-trust.org.

Supplementary file 7d: Comparison of missing responses per participant between the MTBQ and TBQ (n=112)

Multimorbidity Treatment Burden Questionnaire (MTBQ)		**Treatment Burden Questionnaire (TBQ)**	
Number of missing responses per participant (n, (%))		**Number of missing responses per participant (n, (%))**	
0	103 (92.0)	0	96 (85.7%)
1	8 (7.1)	1	10 (8.9%)
≥2	1 (0.9)	≥2	6 (5.4%)

Supplementary file 8: Characteristics by category of treatment burden (original version of 10-question MTBQ, n=243)

Characteristic	N	None (MTBQ = 0)	Low (MTBQ=1-10)	Medium (MTBQ=11-25)	High (MTBQ>25)
Participants	243	29	70	69	75
Age group (n, (%))					
18-50 years	18	0 (0.0%)	3 (16.7%)	7 (38.9%)	8 (44.4%)
51-60 years	37	3 (8.1%)	8 (21.6%)	6 (16.2%)	20 (54.1%)
61-70 years	55	3 (5.5%)	10 (18.2%)	20 (36.4%)	22 (40.0%)
71-80 years	82	14 (17.1%)	31 (37.8%)	22 (26.8%)	15 (18.3%)
81-90 years	46	8 (17.4%)	18 (39.1%)	11 (23.9%)	9 (19.6%)
>90 years	4	1 (25.0%)	0 (0.0%)	2 (50.0%)	1 (25.0%)
Missing	1	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)
Gender (n, (%))					
Male	130	14 (10.8%)	42 (32.3%)	33 (25.4%)	41 (31.5%)
Female	112	15 (13.4%)	28 (25.0%)	35 (31.3%)	34 (30.4%)
Missing	1	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)
Ethnicity (n, (%))					
White	229	27 (11.8%)	67 (29.3%)	63 (27.5%)	72 (31.4%)
Asian	4	0 (0.0%)	1 (25.0%)	2 (50.0%)	1 (25.0%)
Black	4	1 (25.0%)	1 (25.0%)	0 (0.0%)	2 (50.0%)
Mixed	4	0 (0.0%)	1 (25.0%)	3 (75.0%)	0 (0.0%)
Other	1	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)
Missing	1	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Number of long-term conditions (n, (%))					
3	136	19 (14.0%)	35 (25.7%)	39 (28.7%)	43 (31.6%)
4	73	3 (4.1%)	26 (35.6%)	21 (28.8%)	23 (31.5%)
≥5	34	7 (20.6%)	9 (26.5%)	9 (26.5%)	9 (26.5%)

Supplementary file 8 (continued): Characteristics by category of treatment burden (original version of 10-question MTBQ, n=243)

Characteristic	N	None (MTBQ=0)	Low (MTBQ=1-10)	Medium (MTBQ=11-25)	High (MTBQ>25)
Long-term conditions*(n, (%))					
Cardiovascular disease	209	26 (12.4%)	65 (31.1%)	59 (28.2%)	59 (28.2%)
Stroke or transient ischaemic attack	71	6 (8.5%)	25 (35.2%)	20 (28.2%)	20 (28.2%)
Diabetes	141	20 (14.2%)	39 (27.7%)	35 (24.8%)	47 (33.3%)
Chronic kidney disease	68	9 (13.2%)	26 (38.2%)	17 (25.0%)	16 (23.5%)
Chronic obstructive pulmonary disease or asthma	118	15 (12.7%)	34 (28.8%)	30 (25.4%)	39 (33.1%)
Epilepsy	20	1 (5.0%)	4 (20.0%)	4 (20.0%)	11 (55.0%)
Atrial fibrillation	78	15 (19.2%)	28 (35.9%)	19 (24.4%)	16 (20.5%)
Severe mental health problems	13	3 (23.1%)	2 (15.4%)	4 (30.8%)	4 (30.8%)
Depression	106	3 (2.8%)	22 (20.8%)	37 (34.9%)	44 (41.5%)
Learning disability	4	1 (25.0%)	0 (0.0%)	3 (75.0%)	0 (0.0%)
Rheumatoid arthritis	17	1 (5.9%)	2 (11.8%)	10 (58.8%)	4 (23.5%)
Heart failure	31	5 (16.1%)	12 (38.7%)	7 (22.6%)	7 (22.6%)
Age left full-time education (years; n, (%))					
≤14	21	2 (9.5%)	9 (42.9%)	6 (28.6%)	4 (19.0%)
15 or 16	155	22 (14.2%)	46 (29.7%)	41 (26.5%)	46 (29.7%)
17 or 18	33	3 (9.1%)	4 (12.1%)	13 (39.4%)	13 (39.4%)
≥19	31	2 (6.5%)	9 (29.0%)	8 (25.8%)	12 (38.7%)
Missing	3	0 (0.0%)	2 (66.7%)	1 (33.3%)	0 (0.0%)

*Long-term conditions from electronic GP records. Please note, similar long-term conditions are grouped together (e.g. COPD/asthma, stroke/TIA) ** Based on Townsend deprivation index scores[18]

Supplementary file 8 (continued): Characteristics by category of treatment burden (original version of 10-question MTBQ, n=243)

Characteristic	N	None (MTBQ=0)	Low (MTBQ=1-10)	Medium (MTBQ=11-25)	High (MTBQ>25)
Employment status (n, (%))					
Fully retired from work	143	22 (15.4%)	52 (36.4%)	40 (28.0%)	29 (20.3%)
Employed	36	1 (2.8%)	8 (22.2%)	13 (36.1%)	14 (38.9%)
Other	64	6 (9.4%)	10 (15.6%)	16 (25.0%)	32 (50.0%)
Missing	0	0	0	0	0
Deprivation score (n, (%))					
Quintile 1 (least deprived)	49	9 (18.4%)	11 (22.4%)	18 (36.7%)	11 (22.4%)
Quintile 2	49	6 (12.2%)	21 (42.9%)	12 (24.5%)	10 (20.4%)
Quintile 3	74	9 (12.2%)	15 (20.3%)	22 (29.7%)	28 (37.8%)
Quintile 4	46	2 (4.3%)	16 (34.8%)	11 (23.9%)	17 (37.0%)
Quintile 5 (most deprived)	25	3 (12.0%)	7 (28.0%)	6 (24.0%)	9 (36.0%)