Known-groups validity and responsiveness to change of the Patient Experience with Treatment and Self-management (PETS vs. 2.0): a patient-reported measure of treatment burden

David T. Eton[1,2] · Minji K. Lee[2] · Jennifer L. St. Sauver[1,2] · Roger T. Anderson[3]

Abstract

Purpose The purpose of this study was to test the known-groups validity and responsiveness to change of the Patient Experience with Treatment and Self-management (PETS, vs. 2.0), a measure of treatment burden.

Methods The PETS and other standard measures were mailed at baseline and 12-month follow-up to adults living with multiple chronic conditions in southeast Minnesota (USA). A sample of 365 people (mean age = 62.1 years) completed both surveys. Baseline, 12-month, and changes in PETS burden scores were examined. Clinical anchors used to test validity included number of diagnoses (2–4 vs. 5+), mental health diagnosis (yes/no), medication adherence and health literacy (suboptimal/optimal), and changes in self-efficacy, global physical, and global mental health (worsening/improving). Independent-samples t-tests were used to compare scores.

Results PETS scales showed good internal consistency (αs ≥ 0.80). There were few differences across number of diagnoses, but having a mental health diagnosis was associated with higher baseline PETS burden scores (Ps < .05). Suboptimal medication adherence and health literacy over time were associated with worse 12-month PETS burden scores (Ps < .05). Compared with improvements, declines over time in self-efficacy, global physical health, and global mental health were each associated with worsening change scores on PETS impact summary, medical expenses, and bother due to medication reliance and medication side effects (Ps < .05).

Conclusion Among multi-morbid adults, the PETS demonstrated evidence of known-groups validity and responsiveness to change across both objective (e.g., mental health diagnoses) and subjective anchors (e.g., changes in self-efficacy, global physical, and global mental health).

Keywords Questionnaire · Multi-morbidity · Patient-reported outcome · Patient-related experience

✉ David T. Eton
  eton.david@mayo.edu

1 Department of Health Sciences Research, Mayo Clinic, 200 First St SW, Harwick Building, Rochester, MN 55905, USA

2 Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA

3 Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA

Introduction

Treatment burden constitutes the workload of treatment and self-management (including its difficulty), challenges and stressors that can exacerbate it (e.g., financial challenges), and their impact on patient functioning and well-being [1, 2]. Treatment burden is particularly problematic to the growing population of patients who must cope with more than one chronic medical condition [3–6]. The burdens of treatment and self-management are associated clinically with poorer adherence to medical regimens [7–10], challenges in navigating the healthcare system [3, 5, 6], and poorer health-related quality of life (HRQoL) [7, 8, 11, 12]. Yet treatment burden is often invisible to health professionals, who may underestimate its impact or overestimate their patients' ability to integrate complex care into their lives [5, 13]. To

enhance clinical understanding of treatment burden, high-quality, comprehensive measures of it are needed.

Various research groups in the United States and Europe have attempted to operationalize subjective treatment burden in multi-morbid patients. The result has been the emergence of several patient-reported measures of treatment burden applicable to people with multi-morbidity, including the Treatment Burden Questionnaire (TBQ) [14], the Multimorbidity Treatment Burden Questionnaire (MTBQ) [12], the Healthcare Task Difficulty scale [11], the Medication-Related Burden Quality of Life tool [15], and our own Patient Experience with Treatment and Self-management (PETS) [7]. All have demonstrated validity in diverse patient samples, including patients with multi-morbidity [7, 8, 11, 12, 14–16].

The PETS version 2.0 is a comprehensive measure of treatment burden consisting of 60 items divided into 14 content domains (12 multi-item burden scales and 2 single-item indicators of medication bother). Previously, we validated a shorter 48-item version composed of 9 multi-item scales and the 2 medication bother items in a cross-sectional study of multi-morbid patients from two healthcare institutions. PETS scales were found to be reliable (i.e., internally consistent) and demonstrated good construct validity (i.e., factorial validity and convergence with other measures) [7]. Three domains were excluded from this analysis due to excessive missing data from the use of yes/no screening questions, including two designed to assess burdens associated with diet and exercise/physical therapy. Validation of these scales as well as analyses of responsiveness to change of the PETS has yet to be undertaken. Both are crucial to determining whether the PETS is suitable for use in prospective studies or clinical intervention trials. Therefore, the objectives of this study are to (a) test known-groups validity and responsiveness to change of PETS vs. 2.0 scales prospectively in patients with multi-morbidity, and (b) determine utility of two new PETS index scores.

Methods

Design and sample

We conducted a prospective study of community-dwelling adults living with multiple chronic conditions (MCCs) in southeast Minnesota (USA). Resources of the Rochester Epidemiology Project (REP) were used to identify a sample of adults living with MCCs in Olmsted County, Minnesota between July 1, 2015 and June 30, 2016 [17]. The REP electronically links medical records of local healthcare providers for almost the entire population of Olmsted County making it a valuable resource for population-level health studies [18].

A baseline survey was mailed to a random sample of 1496 adults enrolled in the REP, with the sample stratified by age (20–49, 50–64, 65+), race (white, non-white), number of medical-record confirmed diagnosed chronic conditions (2–3, 4+), and the presence/absence of an incident condition diagnosed within a year of the survey mailing. These four stratification factors defined 24 strata which were randomly filled with eligible persons from the REP database. Those eligible for the study had received an International Classification of Diseases (ICD) diagnostic code from a healthcare provider for one or more of 20 chronic conditions identified by the U.S. Department of Health and Human Services as important for studying MCCs [19, 20]. Furthermore, persons with ICD codes for anxiety, hearing problems, vision problems, irritable bowel/Crohn's disease, atopic dermatitis/psoriasis, back problems associated with osteopathic conditions, or headaches were also considered eligible as these conditions were identified by clinical co-investigators as having high treatment burden. Eligible persons could have any combination of these conditions. Those with severe cognitive impairments (e.g., dementia) or severe mental illness (e.g., psychotic disorder) were excluded. While the original sample included 1500 persons, four were excluded prior to the baseline mailing as two had died and two had a documented preference not to be contacted about research studies. A 40% response rate to the baseline survey was assumed ($N = 600$), allowing sufficient power for all planned psychometric analyses of the parent study.

Survey and procedure

Those responding to the baseline survey and verified to be alive were subsequently mailed follow-up surveys at 6, 12, and 24 months post baseline. We used responses from the baseline and 12-month surveys in this analysis to allow time for fluctuation of the constructs of interest. Participants were compensated $10 for each completed survey. The study was approved by both the Mayo Clinic and Olmsted Medical Center Institutional Review Boards (IRB #'s 14-008629 and 022-OMC-16).

The survey consisted of the PETS treatment burden measure (vs. 2.0, 60 items) and other measures for validation. Version 2.0 extends the originally validated PETS measure [7] by three domains: burdens associated with diet, exercise/physical therapy, and use of medical equipment. For this analysis, we studied six discrete PETS scales (diet, exercise/physical therapy, medical expenses, difficulty with healthcare services, medication reliance bother, medication side-effects bother) and two aggregate summary scores (Workload and Impact). The Workload and Impact summary scores have recently been supported by a confirmatory factor analysis of the PETS vs. 2.0 (data not shown, manuscript submitted for publication). Workload is an aggregate of the

PETS medical information, medications, medical appointments, and monitoring health scales, i.e., scales that assess the "work" associated with treatment and self-management. Impact is an aggregate of the PETS role/social activity limitations and physical/mental exhaustion scales, i.e., scales that assess the "impact" of treatment and self-management on well-being. Use of these two summary scores simplifies reporting of PETS data. Standard PETS scoring was used to derive scores [7]. Missing responses to PETS items may occur when the issue queried is not applicable to the respondent. To handle this, aggregated scale scores are prorated for missing data, provided that more than 50% of the total number of items in a scale are non-missing. The Workload and Impact summary scores are calculated as the mean score of the contributing scales provided that > 50% of those scales are non-missing. All PETS scores use the same 0 to 100 metric with a higher score indicating more treatment burden (see Appendix in Electronic Supplementary Material for PETS items and domains).

Other measures were included for validation. The following items were used to assess medication adherence and subjective health literacy, respectively: (1) "In a typical week, how close do you come to following your doctor's recommendations about medications? (Always take all of my medications, Usually take all of my medications [80% of the time], Sometimes take all of my medications [< 80% of the time])" [21], and (2) "How often do you have problems learning about your medical condition because of difficulty understanding written information? (All of the time, Most of the time, Some of the time, A little of the time, None of the time)" [22]. Both are valid measures of these concepts [21, 22]. The Patient-Reported Outcomes Measurement Information System-10 item general measure (PROMIS-10) was used to assess global physical health (GPH) and global mental health (GMH) [23]. Wallston's Perceived Medical Condition Self-Management scale (PMCSM) was used to assess self-efficacy for self-management (i.e., perceived confidence in the ability to self-manage one's health conditions) [24]. Higher scores on the PROMIS-10 and PMCSM indicate better health status and higher self-efficacy, respectively. Both have been shown to be valid in patients coping with chronic health conditions [23–25].

From the electronic medical record, we extracted the number and types of chronic condition diagnoses from among the conditions that determined study eligibility (see above).

Analyses

To confirm internal consistency reliability, Cronbach's alpha coefficients were calculated at baseline for the PETS multi-item domain scales used. Spearman rank-order correlation coefficients (ρ) were used to check the inter-correlations of the PETS scales constituting the two aggregated summary scores. Frequency distributions, medians, means, and standard deviations (SD) were used to describe the study sample and characterize the PETS scores and clinical anchors used for validation. PETS change scores were calculated as 12-month—baseline.

Independent-samples t-tests were used to compare PETS scores of distinct groups at baseline and at 12 months, separately. Baseline PETS scores were compared across diagnosed conditions extracted from the medical record: total number of chronic conditions (median split) and the presence or absence of a mental health condition involving depression or anxiety. Twelve-month PETS scores were compared across separate respondent groups defined by medication adherence and subjective health literacy status over time. Status was defined as either "optimal or improving" versus "suboptimal or worsening" based on the baseline and 12-month data for the anchor. For example, reports of medication adherence that were consistently good at both time points (i.e., "always take all my medications") or reports of improvement in medication adherence from baseline to 12 months were classified as "optimal or improving." Consistently poor medication adherence at both time points (i.e., "usually take all my medications" or "sometimes take all my medications") or reports of worsening medication adherence from baseline to 12 months were classified as "suboptimal or worsening." The same procedure was used to define separate optimal and suboptimal health literacy groups. Alpha for all t-tests was set at 0.05 (two-tailed). Given the number of outcomes and analyses, we also calculated effect sizes (Cohen's d) corresponding to group mean differences (mean difference divided by the pooled within-group SD) with $d = 0.2$ indicating a small effect, $d = 0.5$ indicating a medium effect, and $d = 0.8$ indicating a large effect [26].

Responsiveness analyses determine whether changes in PETS treatment burden scores over time (i.e., PETS change scores) correspond to status changes in perceived self-efficacy and health status. Independent groups for comparison were defined using the distributions of the baseline to 12-month change scores for the PMCSM, PROMIS GPH, and PROMIS GMH. The top and bottom quartiles of the change score distributions for each of these variables defined an "improving" group and a "worsening" group, respectively. PETS change scores (12-month—baseline) were compared between the improving and the worsening status groups using independent-samples t-tests ($\alpha = 0.05$, two-tailed). A change score was not estimated if either the baseline or 12-month PETS score was missing. Effect sizes (d) corresponding to mean differences of the PETS change scores for the two groups being compared were calculated.

Results

Characteristics of the sample

Overall, 443 people returned a completed baseline survey (30% response). Four were subsequently excluded due to a recent diagnosis of an exclusionary condition. Our companion analyses showed a few statistically significant differences between survey responders and non-responders (data not shown, manuscript submitted). Briefly, responders were slightly older, of white race, and had more diagnosed conditions than non-responders. In this analysis, we focus on the 365 people who returned both baseline and 12-month surveys. Table 1 shows sample descriptive characteristics. Mean age was 62.1 years (SD = 15.9), and there were more women than men (64% vs. 36%). Most reported White/Caucasian race (78%), were married/partnered (61%) and college-educated (74%). Median number of diagnosed conditions was 5 (range: 2 to 13), with the most commonly observed diagnoses being hyperlipidemia (56%), hypertension (55%), low back disorder (52%), Type I or II diabetes (49%), and arthritis (49%). Mental health conditions such as depression and anxiety were also common (> 25%). Baseline descriptive data for the clinical anchors are also provided in Table 1.

Scale reliability, inter-scale correlations, and prospective PETS scores

As shown in Table 2, all ten of the multi-item PETS domain scales showed good internal consistency reliability. Cronbach's alpha coefficients for these scales at baseline all exceeded the 0.70 threshold for adequate reliability [27]. Internal consistency was not determined for either of the two single-item scales (medication reliance and side-effect bother) or the two summary indexes (Workload and Impact). However, inter-scale correlations of the PETS scales making up each summary score were generally high at both time points. Inter-correlations for the four scales making up the Workload summary ranged from $\rho = 0.49$ to 0.62 at baseline and $\rho = 0.57$ to 0.64 at 12 months. The correlations of the two scales making up the Impact summary were $\rho = 0.68$ at baseline and $\rho = 0.66$ at 12 months.

Unadjusted means of the PETS scores appear in Table 2. Sample sizes across and within the various PETS scores may vary due to missing data. Some "missingness" on items and domains of the PETS is to be expected given that a "not applicable" response to an item is treated as missing and items in the diet and exercise/physical therapy domains are preceded by a yes/no screening question that invokes a skip pattern for a "no" response. Change scores

for the sample as a whole appear in the last column of Table 2 with a positive score indicating worsening and a negative score indicating improvement. As shown in the table, the magnitude of changes was generally small for the overall sample.

Clinical anchor statistics

Descriptive statistics for the various clinical anchors are shown in Table 3. Sample sizes for the groups being compared in the cross-sectional analyses of known-groups are indicated in the top four rows of the table. For the responsiveness analyses, discrete groups were defined that showed evidence of change over time on three indicators (self-efficacy, GPH, and GMH) using baseline to 12-month change scores on the PMCSM and PROMIS-10. Independent worsening and improving groups were defined, respectively, as the lower and upper quartiles of the change score distributions for each anchor.

Known-groups comparisons of baseline PETS scores

Comparisons of the eight baseline PETS scores across clinical diagnoses appear in Fig. 1a, b. Having five or more diagnosed chronic conditions was associated with greater treatment burden Impact and being more bothered by medication reliance at baseline than having only 2–4 diagnosed chronic conditions (1a: Ps < 0.05). There were no significant differences across the other six PETS scores, and effect sizes of all differences were generally small (d's below 0.30). In contrast, having a diagnosis of a mental health condition (depression or anxiety) compared to not having one was associated with significantly greater treatment burden across all eight baseline PETS scores (1b: Ps < 0.05). Effect sizes (d) of these differences were generally moderate, ranging from 0.34 to 0.71.

Known-groups comparisons of 12-month PETS scores

Comparisons of 12-month PETS scores across medication adherence and health literacy status over time are shown in Fig. 2a, b. People who reported suboptimal medication adherence reported significantly greater treatment burden at 12 months across all of the PETS scores compared to those who reported having optimal medication adherence (2a: Ps < 0.05). Effect sizes of the group differences were generally moderate, ranging from 0.42 to 0.71. People who reported suboptimal health literacy reported significantly greater treatment burden at 12 months across 7 of the 8 PETS scores compared to those who reported optimal health literacy (2b: Ps < 0.05). Effect sizes of the significant group differences were moderate to high, ranging from 0.41

Table 1 Descriptive characteristics of sample at baseline assessment (*N* = 365)

Age	
Mean (SD)	62.1 (15.9) years
Range	20 to 98 years
Gender	
Female	232 (64%)
Male	133 (36%)
Race	
White/Caucasian	285 (78%)
Black/AA	27 (7%)
Asian	24 (7%)
Native American	11 (3%)
Mixed	3 (<1%)
Other	3 (<1%)
Unknown	12 (3%)
Ethnicity	
Hispanic	12 (3%)
Marital status	
Married/partnered	223 (61%)
Not married	130 (36%)
Missing	12 (3%)
Education status	
College educated	271 (74%)
No more than HS	83 (23%)
Missing	11 (3%)
Total number of diagnosed conditions	
Median	5.0 conditions
Range	2 to 13 conditions
2–3 conditions	79 (22%)
4–5 conditions	141 (39%)
6+ conditions	145 (40%)
Types of diagnosed conditions	
Hyperlipidemia	203 (56%)
Hypertension	200 (55%)
Low back disorder[a]	188 (52%)
Diabetes (type 1 or 2)	180 (49%)
Arthritis	179 (49%)
Depression	150 (41%)
Vision problems	141 (39%)
Cancer	100 (27%)
Anxiety	99 (27%)
Cardiac arrhythmia	93 (26%)
Coronary artery disease	66 (18%)
Hearing problems	60 (16%)
Chronic kidney disease	47 (13%)
COPD	38 (10%)
Substance abuse	30 (8%)
Headache	30 (8%)
Osteoporosis	30 (8%)
Congestive heart failure	25 (7%)
Psoriasis	17 (5%)
Crohn's disease	14 (4%)
Hepatitis	8 (2%)

Table 1 (continued)

HIV	2 (<1%)
Medication adherence ("In a typical week, how close do you come to following your doctors' recommendations about medications?")	
Always take all my meds	277 (76%)
Usually take all my meds. (80% of time)	55 (15%)
Sometimes take all my meds. (<80% of time)	8 (2%)
Missing	25 (7%)
Subjective health literacy ("How often do you have problems learning about your medical condition because of difficulty understanding written information?")	
None of the time	230 (63%)
A little of the time	83 (23%)
Some of the time	27 (7%)
Most of the time	8 (2%)
All of the time	8 (2%)
Missing	9 (3%)
Perceived Medical Condition Self-management scale (PMCSM)	
Mean (SD)	29.6 (6.5)
PROMIS Global Physical Health (GPH)	
Mean (SD)	45.1 (8.8)
PROMIS Global Mental Health (GMH)	
Mean (SD)	47.2 (9.8)

SD standard deviation, *AA* African-American, *HS* high school, *COPD* chronic obstructive pulmonary disease, *PROMIS* patient-reported outcomes measurement information system

[a]Includes osteopathic conditions such as disc displacement/degeneration, spondylosis, spinal stenosis, sciatica, and post-laminectomy syndromes

to 0.88. There was no significant difference in diet burden between the sub-optimal and optimal health literacy groups ($d = 0.18$).

Responsiveness to change of PETS scores

To determine responsiveness, PETS change scores (12-month—baseline) were compared between independent groups defined as either worsening or improving in (a) self-efficacy for self-management, (b) GPH, or (c) GMH. Results of these group comparisons are shown in Fig. 3a–c. As shown in Fig. 3a, most people whose self-efficacy scores worsened over time also reported worsening treatment burden, whereas those whose self-efficacy scores improved over time reported improving treatment burden. Mean treatment burden change scores were significantly different between the discrete self-efficacy groups on the following six PETS domains: Workload, Impact, exercise/physical therapy, medical expenses, medication reliance bother, and medication side-effects bother ($Ps < 0.01$). Effect sizes associated with these differences were moderate, ranging from 0.27 to 0.77.

Regarding discrete groups defined by PROMIS-10 scores, Fig. 3b shows that those who experienced worsening GPH over time also reported worsening treatment burden, whereas those who experienced improvement in their GPH over time reported improving treatment burden. Mean treatment burden change scores were significantly different between the groups on the following five PETS domains: Impact, exercise/physical therapy, medical expenses, medication reliance bother, and medication side-effects bother ($Ps < 0.05$). Effect sizes associated with these differences were moderate, ranging from 0.26 to 0.56. Similarly, those who experienced a drop in their GMH over time reported worsening treatment burden and those who experienced an improvement in their GMH over time reported improving treatment burden (see Fig. 3c). Mean treatment burden change scores were significantly different between the groups on the following five PETS domains: Workload, Impact, medical expenses, medication reliance bother, and medication side-effects bother ($Ps < 0.05$). Effect sizes associated with these differences were moderate, ranging from 0.25 to 0.43.

Discussion

This analysis extends findings from our initial cross-sectional validation [7] and supports the psychometric utility of version 2.0 of the PETS treatment burden measure. All PETS scales had good reliability and PETS scores were able to discriminate discrete groups across two time points. At baseline, we found that having a mental health diagnosis was associated with higher PETS treatment

Table 2 Mean PETS scores (range 0–100) and Cronbach's alphas for multi-item PETS scales

PETS score[a]	Baseline	12 month	Base. to 12-month change[e]
	N Mean (SD)	N Mean (SD)	N Mean (SD)
Workload summary index[b]	332	336	313
	23.6 (17.0)	23.4 (17.3)	− 0.2 (12.5)
Medical information ($\alpha = 0.92$)	333	341	318
	25.6 (19.8)	24.7 (19.1)	− 0.7 (16.2)
Medications ($\alpha = 0.90$)	341	348	331
	15.9 (16.9)	16.1 (17.3)	0.1 (14.6)
Medical appointments ($\alpha = 0.93$)	349	349	334
	21.6 (21.7)	23.0 (21.4)	1.0 (20.2)
Monitoring health ($\alpha = 0.89$)	313	322	288
	32.3 (26.2)	30.7 (25.2)	− 1.3 (22.5)
Impact summary index[c]	354	344	337
	24.4 (23.9)	25.1 (23.5)	0.7 (16.4)
Role activity limitations ($\alpha = 0.95$)	355	346	339
	20.0 (25.7)	20.1 (25.2)	0.0 (20.2)
Physical / mental exhaustion ($\alpha = 0.93$)	356	346	340
	28.7 (25.7)	30.1 (26.2)	1.2 (19.6)
Diet[d] ($\alpha = 0.80$)	161	142	112
	43.8 (23.4)	43.4 (23.2)	− 0.3 (23.8)
Exercise / physical therapy[d] ($\alpha = 0.83$)	210	194	138
	48.8 (24.6)	51.3 (23.6)	2.9 (19.1)
Medical expenses ($\alpha = 0.93$)	339	333	312
	41.0 (27.8)	38.9 (28.5)	− 2.9 (18.7)
Diff. with healthcare services ($\alpha = 0.86$)	315	312	282
	30.3 (20.2)	29.4 (19.6)	− 1.6 (17.7)
Medication reliance bother (α: NA)	342	348	332
	19.9 (26.1)	19.0 (26.5)	− 0.5 (25.6)
Medication side-effects bother (α: NA)	342	348	333
	18.2 (25.7)	16.7 (25.1)	− 1.4 (24.9)

NA not applicable

[a]Higher PETS score = more burden

[b]Aggregated mean of scores in the medical information, medications, medical appointments, and monitoring health scales. Calculated when > 50% of the four constituent scales are non-missing

[c]Aggregated mean of scores in the role/social activity limitations and physical/mental exhaustion scales. Calculated when > 50% of the two constituent scales are non-missing (i.e., when both scales are non-missing)

[d]Yes/no screener used for the diet and exercise/physical therapy domains

[e]Change score = 12-month PETS score – baseline PETS score. Sample sizes over time fluctuate due to missing data

burden scores. Robust associations of mental health problems and treatment burden have been observed in other studies [12, 28, 29] and may indicate that poor mental health is a risk factor for greater treatment burden among the multi-morbid. At 12 months, both suboptimal medication adherence and suboptimal health literacy status were each associated with greater PETS treatment burden scores when compared with an optimal status on these indicators.

This is consistent with findings from cross-sectional studies using two different treatment burden measures. Greater self-reported medication adherence and higher health literacy have been associated with lower PETS [7] and TBQ [8] treatment burden scores among multi-morbid patients. Interventions to improve adherence and shore-up deficits in health literacy could have added benefit by diminishing treatment burden.

Table 3 Clinical anchor descriptive statistics

Anchor	Group	N (%) or Mean change (Δ) from baseline to 12 months
Number of chronic condition diagnoses	2–4 conditions	N = 147 (40%)
	5 + conditions	N = 218 (60%)
Mental health diagnosis	Yes	N = 167 (46%)
	No	N = 198 (54%)
Medication adherence status[a]	Suboptimal group	N = 66 (18%)
	Optimal group	N = 262 (72%)[d]
Health literacy status[a]	Suboptimal group	N = 52 (14%)
	Optimal group	N = 291 (80%)[d]
Δ Self-efficacy (PMCSM)[b,c]	Worsening group	N = 97; Mean Δ = -6.6[e]
	Improving group	N = 109; Mean Δ = 5.8
Δ Global Physical Health (GPH)[c]	Worsening group	N = 82; Mean Δ = -6.6[e]
	Improving group	N = 80; Mean Δ = 6.0
Δ Global Mental Health (GMH)[c]	Worsening group	N = 84; Mean Δ = -7.6[e]
	Improving group	N = 98; Mean Δ = 6.6

[a]Suboptimal = consistently poor over time OR declining status from baseline to 12-month follow-up. Optimal = consistently good over time OR improving status from baseline to 12-month follow-up

[b]PMCSM: Perceived Medical Condition Self-Management scale

[c]Worsening group = lower quartile of change score distribution. Improving group = upper quartile of change score distribution

[d]Does not sum to 100% due to missing data on the anchor

[e]Negative change = worsened status on anchor over time; positive change = improved status on anchor over time

Interestingly, having more diagnosed chronic conditions overall was less consistently associated with PETS treatment burden scores. The association between number of chronic conditions and treatment burden has varied across studies with some showing a moderate positive relationship [12, 30] and others showing low or no relationship with burden [7, 31]. This might reflect differences in the types of conditions represented across study samples, the length of time that people have lived with their conditions, the presence/absence of available resources that may lessen burden [32], or some combination of these factors.

Several PETS scale and aggregate summary scores were found to be responsive to changes in self-efficacy and HRQoL. Compared with individuals who had improved in self-efficacy over time, those who had worsened in self-efficacy had significant worsening in 6 of 8 PETS treatment burden scores. Furthermore, compared with individuals who had improved in global physical or global mental health, respectively, those who had worsened in global physical and mental health had significant worsening in 5 of 8 PETS scores. Notably, the PETS difficulty with healthcare services scale failed to show any significant differences between groups. However, this may be due to the fact that the anchors tap change in psychosocial function and well-being, rather than changes in the perceived quality of care. Specifying changes in a quality of care anchor could be useful for future tests of the difficulty with health services scale.

Among existing measures of treatment burden, some assess the construct in full [7, 8, 14], while others address specific elements [11, 15], or embed treatment burden scales into a broader measure [29]. Many are short assessments of no more than 15 items [8, 11, 12, 14, 29]. However, in achieving brevity, they sacrifice comprehensiveness. Many of the scales produce a single "global" treatment burden score by aggregating conceptually disparate aspects of the construct. Unfortunately, this may gloss over important nuances in the diverse elements of treatment burden that patients have articulated [1, 2, 33, 34]. In our own work, it has become apparent that while there are certain cross-cutting elements that make up treatment burden (i.e., domains), there is also an idiosyncratic quality to treatment burden such that certain elements may be relevant to one person and less relevant to another. A multi-domain measure that employs profile scoring will more accurately represent the full landscape of treatment burden. Still, we recognize the diminished convenience of having to report a large number of scores for one measure. Hence, a critical result of this analysis was the demonstrated utility of the Workload and Impact summary index scores. Both of these aggregate index scores differentiated meaningful groups.

There are a few limitations to this study. First, the response rate to the baseline survey was 30%, leaving open the possibility of response bias. Low response rates in studies of treatment burden are not unheard of [8, 11, 29],
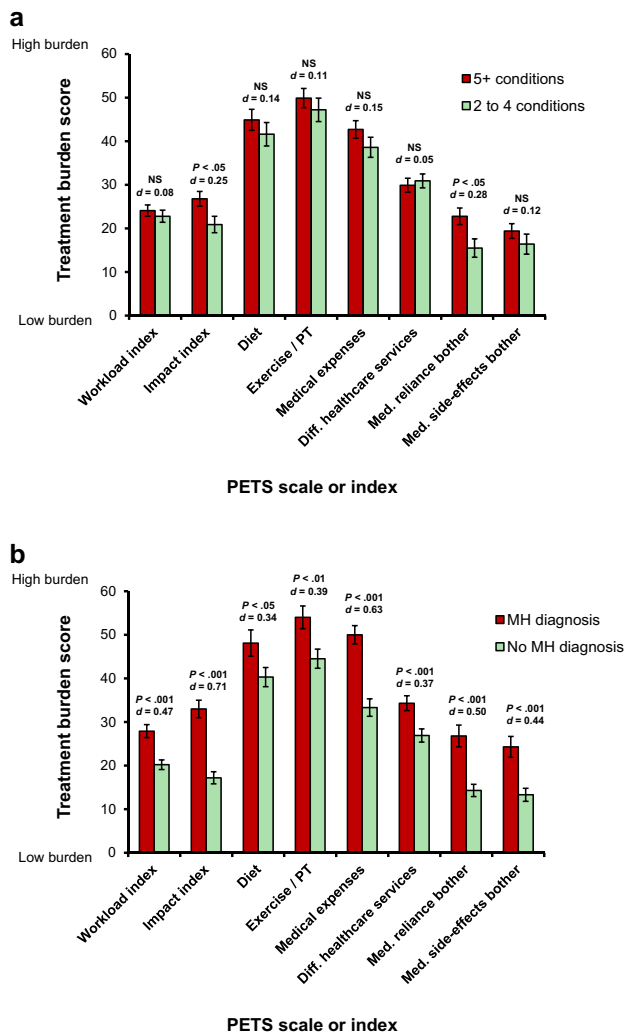
Fig. 1 a Mean PETS scores at baseline with standard error by number of chronic condition diagnoses, **b** Mean PETS scores at baseline with standard error by mental health diagnosis (MH diagnosis)
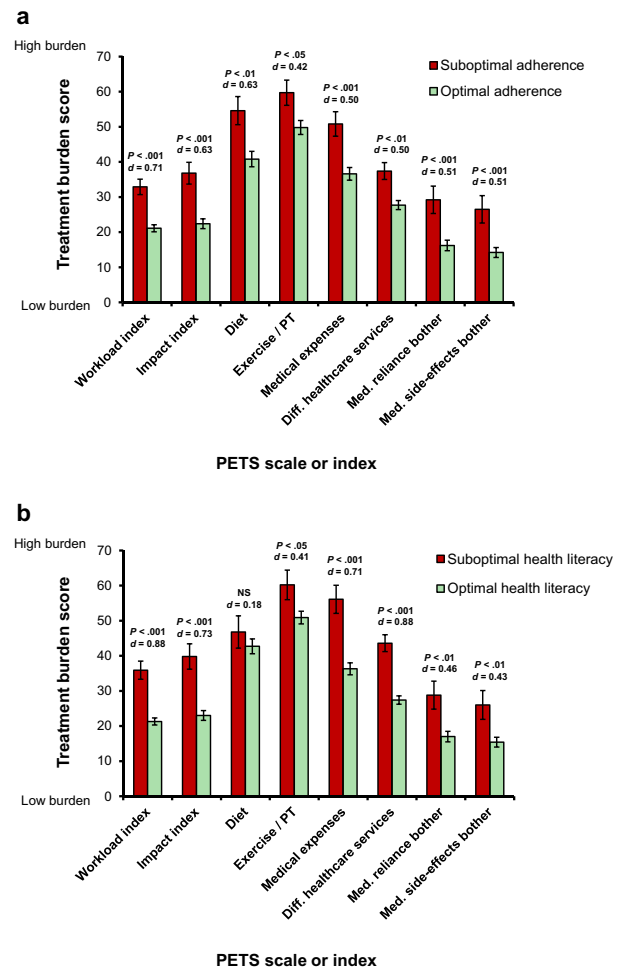
Fig. 2 a Mean PETS scores at 12 months with standard error by medication adherence status (suboptimal or optimal), **b** Mean PETS scores at 12 months with standard error by health literacy status (suboptimal or optimal)

perhaps due to the demands faced by people coping with a complex health situation. A responder analysis from our companion manuscript (submitted) indicates that our findings may be less representative of younger and non-white people and those with fewer diagnosed conditions. Second, sample sizes fluctuated across PETS domains over time. This is due in part to the relevance of PETS content at any given time. A response of "not applicable" to an item is treated as missing, based on the assumption that "burden" cannot be assessed unless the issue in question is personally salient. Furthermore, the yes/no screening questions for the diet and exercise/physical therapy domains reduce the overall sample for analysis of these domains. As both represent highly relevant lifestyle behaviors, we will remove these screening questions from future versions of the PETS. Third, given that many of the measures used were self-reports of

prior experience, we cannot rule out recall bias as a possible limitation. Finally, we used simple, single-item, self-reports to assess medication adherence and health literacy. Results may have been different if more sophisticated and objective means of assessing these had been used.

Conclusion

The PETS version 2.0 demonstrated evidence of known-groups validity and responsiveness to change across multiple objective and subjective anchors in community-dwelling adults with multi-morbidity. Findings support future use of the PETS as an outcome measure of treatment burden in intervention and other prospective studies. We are implementing means to make the measure easier to administer and report, including eliminating domain screening items and using aggregate summary index scores of related domains.
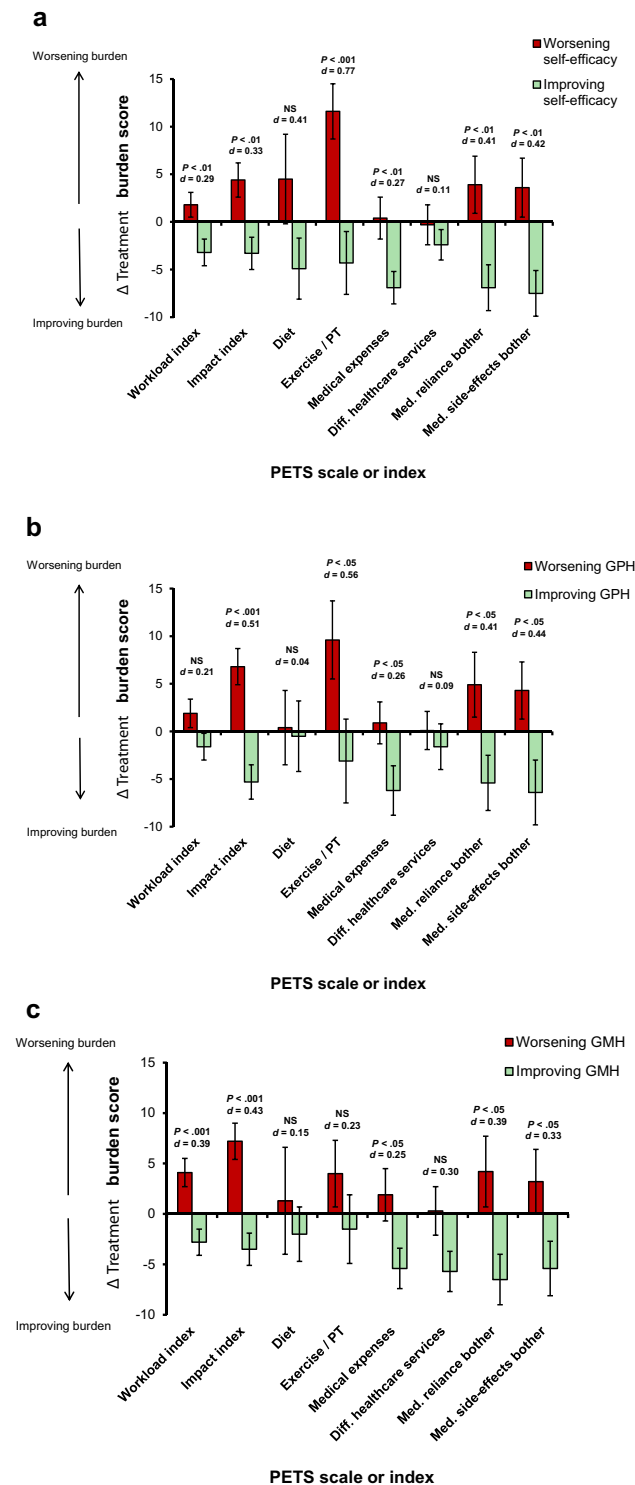
Our analyses support use of a summary "Workload" score that combines the medical information, medications, medical appointments, and monitoring health scales and a summary "Impact" score that combines the role/social activity limitations and physical/mental exhaustion scales. However, as each scale represents distinct content reflective of treatment burden, we feel that there is still value in maintaining scoring of the individual scales.

Availability of data and material De-identified data can be made available upon reasonable request of the principal investigator (D. Eton), pending approval by the Mayo Foundation for Medical Education and Research. All requests will be reviewed. The PETS measure is available by request to D. Eton.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Signed informed consent was not required as the study was approved as minimal risk with oral consent by both the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. Oral consent was provided in the form of a cover letter describing the survey procedures and consent to participate was implied based on the return of a completed survey.

Fig. 3 a Mean change in PETS scores with standard error by self-efficacy change (worsening or improving), **b** Mean change in PETS scores with standard error by global physical health (GPH) change (worsening or improving), **c** Mean change in PETS scores with standard error by global mental health (GMH) change (worsening or improving)

References

1. Eton, D. T., Ramalho de Oliveira, D., Egginton, J. S., Ridgeway, J. L., Odell, L., May, C. R., et al. (2012). Building a measurement framework of burden of treatment in complex patients with chronic conditions: A qualitative study. *Patient Related Outcome Measures, 3*, 39–49. https://doi.org/10.2147/PROM.S34681.

2. Eton, D. T., Ridgeway, J. L., Egginton, J. S., Tiedje, K., Linzer, M., Boehm, D. H., et al. (2015). Finalizing a measurement framework for the burden of treatment in complex patients with chronic conditions. *Patient Related Outcome Measures, 6*, 117–126. https://doi.org/10.2147/PROM.S78955.

3. May, C., Montori, V. M., & Mair, F. S. (2009). We need minimally disruptive medicine. *British Medical Journal, 339*, b2803. https://doi.org/10.1136/bmj.b2803.

4. May, C. R., Eton, D. T., Boehmer, K., Gallacher, K., Hunt, K., MacDonald, S., et al. (2014). Rethinking the patient: Using burden of treatment theory to understand the changing dynamics of illness. *BMC Health Services Research, 14*, 281. https://doi.org/10.1186/1472-6963-14-281.

5. Ortenblad, L., Meillier, L., & Jonsson, A. R. (2018). Multi-morbidity: A patient perspective on navigating the health care system and everyday life. *Chronic Illness, 14*(4), 271–282. https://doi.org/10.1177/1742395317731607.

6. Salisbury, C., Johnson, L., Purdy, S., Valderas, J. M., & Montgomery, A. A. (2011). Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. *British Journal of General Practice, 61*(582), e12–21. https://doi.org/10.3399/bjgp11X548929.

7. Eton, D. T., Yost, K. J., Lai, J. S., Ridgeway, J. L., Egginton, J. S., Rosedahl, J. K., et al. (2017). Development and validation of the patient experience with treatment and self-management (PETS): A patient-reported measure of treatment burden. *Quality of Life Research, 26*, 489–503. https://doi.org/10.1007/s11136-016-1397-0.

8. Tran, V. T., Harrington, M., Montori, V. M., Barnes, C., Wicks, P., & Ravaud, P. (2014). Adaptation and validation of the Treatment Burden Questionnaire (TBQ) in English using an internet platform. *BMC Medicine, 12*, 109. https://doi.org/10.1186/1741-7015-12-109.

9. Haynes, R. B., McDonald, H. P., & Garg, A. X. (2002). Helping patients follow prescribed treatment: clinical applications. *JAMA, 288*(22), 2880–2883. https://doi.org/10.1001/jama.288.22.2880.

10. Vijan, S., Hayward, R. A., Ronis, D. L., & Hofer, T. P. (2005). Brief report: the burden of diabetes therapy: implications for the design of effective patient-centered treatment regimens. *Journal of General Internal Medicine, 20*(5), 479–482. https://doi.org/10.1111/j.1525-1497.2005.0117.x.

11. Boyd, C. M., Wolff, J. L., Giovannetti, E., Reider, L., Weiss, C., Xue, Q. L., et al. (2014). Healthcare task difficulty among older adults with multimorbidity. *Medical Care, 52*(Suppl 3), S118–125. https://doi.org/10.1097/MLR.0b013e3182a977da.

12. Duncan, P., Murphy, M., Man, M. S., Chaplin, K., Gaunt, D., & Salisbury, C. (2018). Development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ). *British Medical Journal Open, 8*(4), e019413. https://doi.org/10.1136/bmjopen-2017-019413.

13. Dobler, C. C., Harb, N., Maguire, C. A., Armour, C. L., Coleman, C., & Murad, M. H. (2018). Treatment burden should be included in clinical practice guidelines. *British Medical Journal, 363*, k4065. https://doi.org/10.1136/bmj.k4065.

14. Tran, V. T., Montori, V. M., Eton, D. T., Baruch, D., Falissard, B., & Ravaud, P. (2012). Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. *BMC Medicine, 10*, 68. https://doi.org/10.1186/1741-7015-10-68.

15. Mohammed, M. A., Moles, R. J., Hilmer, S. N., Kouladjian O'Donnel, L., & Chen, T. F. (2018). Development and validation of an instrument for measuring the burden of medicine on functioning and well-being: The Medication-Related Burden Quality of Life (MRB-QoL) tool. *British Medical Journal Open, 8*(1), e018880. https://doi.org/10.1136/bmjopen-2017-018880.

16. Rogers, E. A., Yost, K. J., Rosedahl, J. K., Linzer, M., Boehm, D. H., Thakur, A., et al. (2017). Validating the patient experience with treatment and self-management (PETS), a patient-reported measure of treatment burden, in people with diabetes. *Patient Related Outcome Measures, 8*, 143–156. https://doi.org/10.2147/PROM.S140851.

17. Rocca, W. A., Yawn, B. P., St Sauver, J. L., Grossardt, B. R., & Melton, L. J., 3rd. (2012). History of the Rochester Epidemiology Project: Half a century of medical records linkage in a US population. *Mayo Clinic Proceedings, 87*(12), 1202–1213. https://doi.org/10.1016/j.mayocp.2012.08.012.

18. St Sauver, J. L., Grossardt, B. R., Yawn, B. P., Melton, L. J., 3rd, Pankratz, J. J., Brue, S. M., et al. (2012). Data resource profile: The Rochester Epidemiology Project (REP) medical records-linkage system. *International Journal of Epidemiology, 41*(6), 1614–1624. https://doi.org/10.1093/ije/dys195.

19. Goodman, R. A., Posner, S. F., Huang, E. S., Parekh, A. K., & Koh, H. K. (2013). Defining and measuring chronic conditions: Imperatives for research, policy, program, and practice. *Preventing Chronic Disease, 10*, E66. https://doi.org/10.5888/pcd10.120239.

20. U.S. Department of Health and Human Services. (2010). *Multiple chronic conditions—A strategic framework: Optimum health and quality of life for individuals with multiple chronic conditions*. Washington, DC: U.S. Department of Health and Human Services.

21. Schoenthaler, A., Montague, E., Baier Manwell, L., Brown, R., Schwartz, M. D., & Linzer, M. (2014). Patient-physician racial/ethnic concordance and blood pressure control: the role of trust and medication adherence. *Ethnicity & Health, 19*(5), 565–578. https://doi.org/10.1080/13557858.2013.857764.

22. Chew, L. D., Griffin, J. M., Partin, M. R., Noorbaloochi, S., Grill, J. P., Snyder, A., et al. (2008). Validation of screening questions for limited health literacy in a large VA outpatient population. *Journal of General Internal Medicine, 23*(5), 561–566. https://doi.org/10.1007/s11606-008-0520-5.

23. Hays, R. D., Bjorner, J. B., Revicki, D. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of life Research, 18*(7), 873–880. https://doi.org/10.1007/s11136-009-9496-9.

24. Wallston, K. A., Osborn, C. Y., Wagner, L. J., & Hilker, K. A. (2011). The Perceived Medical Condition Self-Management Scale applied to persons with HIV/AIDS. *Journal of Health Psychology, 16*(1), 109–115. https://doi.org/10.1177/1359105310367832.

25. Wild, M. G., Wallston, K. A., Green, J. A., Beach, L. B., Umeukeje, E., Wright Nunes, J. A., et al. (2017). The Perceived Medical Condition Self-Management Scale can be applied to patients with chronic kidney disease. *Kidney International, 92*(4), 972–978. https://doi.org/10.1016/j.kint.2017.03.018.

26. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

27. Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., & Hays, R. D. (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health, 10*(Suppl 2), S94–S105. https://doi.org/10.1111/j.1524-4733.2007.00272.x.

28. Eton, D. T., Anderson, R. T., Cohn, W. F., Kennedy, E. M., St Sauver, J. L., Bucknell, B. J., et al. (2019). Risk factors for poor health-related quality of life in cancer survivors with multiple chronic conditions: Exploring the role of treatment burden as a mediator. *Patient Related Outcome Measures, 10*, 89–99. https://doi.org/10.2147/PROM.S191480.

29. Gibbons, C. J., Kenning, C., Coventry, P. A., Bee, P., Bundy, C., Fisher, L., et al. (2013). Development of a multimorbidity illness

perceptions scale (MULTIPleS). *PLoS ONE, 8*(12), e81852. https://doi.org/10.1371/journal.pone.0081852.

30. Schwartz, C. E., Zhang, J., Michael, W., Eton, D. T., & Rapkin, B. D. (2018). Reserve-building activities attenuate treatment burden in chronic illness: The mediating role of appraisal and social support. *Health Psychology Open, 5*(1), 2055102918773440. https://doi.org/10.1177/2055102918773440.

31. Song, M. K., Paul, S., Plantinga, L., Henry, C., & Turberville-Trujillo, L. (2019). Social networks of self-care and perceived treatment burden among patients on in-center hemodialysis. *Kidney Medicine, 1*(3), 97–103. https://doi.org/10.1016/j.xkme.2019.04.001.

32. Ridgeway, J. L., Egginton, J. S., Tiedje, K., Linzer, M., Boehm, D., Poplau, S., et al. (2014). Factors that lessen the burden of treatment in complex patients with chronic conditions: a qualitative study. *Patient Preference and Adherence, 8*, 339–351. https://doi.org/10.2147/PPA.S58014.

33. Gallacher, K., May, C. R., Montori, V. M., & Mair, F. S. (2011). Understanding patients' experiences of treatment burden in chronic heart failure using normalization process theory. *Annals of Family Medicine, 9*(3), 235–243.

34. Sav, A., Kendall, E., McMillan, S. S., Kelly, F., Whitty, J. A., King, M. A., et al. (2013). 'You say treatment, I say hard work': Treatment burden among people with chronic illness and their carers in Australia. *Health & Social Care in the Community, 21*(6), 665–674. https://doi.org/10.1111/hsc.12052.

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