Quality of life and symptom burden among myeloproliferative neoplasm patients: Do symptoms impact quality of life?

Blake T Langlais[a], **Holly Geyer**[b], **Robyn Scherber**[c,d], **Ruben A Mesa**[d], and **Amylou C Dueck**[a,*]

[a]Section of Biostatistics, Mayo Clinic, Scottsdale, Arizona, United States of America

[b]Department of Internal Medicine, Mayo Clinic, Phoenix, Arizona, United States of America

[c]Oregon Health and Science University, Portland, Oregon, United States of America

[d]University of Texas Health San Antonio Cancer Center – An NCI Designated Cancer Center, San Antonio, Texas, United States of America

Abstract

Patients with myeloproliferative neoplasms (MPNs) are faced with high disease-related symptom burden and quality of life (QoL) decrements. This analysis assesses the extent to which individual symptoms or summary measures across symptoms impact QoL overall and within MPN subgroups. Four sets of summary measures were constructed assessing symptom prevalence and severity within group-standardized and patient-individualized approaches. Among 1,416 international patients with MPNs, mean symptom severity and prevalence were highly correlated (P <.001). Individual symptoms most impacting QoL were inactivity ($R^2 = 0.29$), fatigue ($R^2 = 0.23$), and depression ($R^2 = 0.23$). Multiple symptom severity scores are needed to best predict QoL. Symptom severity at the patient-level is more predictive of QoL than severity at the group-level where a fewer number of symptoms are considered. Having at least one severe symptom and having multiple symptoms of moderate intensity are meaningfully predictive of QoL decrements. Results were highly consistent across disease subgroups.

Introduction

The myeloproliferative neoplasms (MPNs), including essential thrombocythemia (ET), polycythemia vera (PV), and myelofibrosis (MF), are characterized by the clonal proliferation of one or more hematopoietic cell lines, primarily in the bone marrow [1]. Patients with MPNs suffer from a range of debilitating disease manifestations (e.g. splenomegaly, erythrocytosis, thrombocytosis) resulting in severe symptom burden and quality of life (QoL) decrement. Characterizing symptom burden among these patients informs individual treatment plans and drives a path to abate diminishing QoL.

Equitable descriptions of symptom profiles and QoL were published in 2007 showing MPN patients reported increased symptom severity and decreased QoL compared to control

*Corresponding author: Amylou C. Dueck, Section of Biostatistics, Mayo Clinic, 13400 E. Shea Blvd., Scottsdale, AZ 85259, dueck.amylou@mayo.edu.

patients [2]. Further effort into constructing patient-reported outcome (PRO) tools capturing the MPN symptom experience lead to the MF-specific PRO tool, Myelofibrosis Symptom Assessment Form (MFSAF) in 2009 [3]. Expanding on this, Scherber and colleagues created and validated the MPN Symptom Assessment Form (MPN-SAF) functioning as a single PRO tool surveying a spectrum of MPN disease-related signs and symptoms as well as overall QoL [4]. The MPN-SAF measures 18 symptoms (abdominal discomfort, abdominal pain, bone pain, concentration problems, cough, depression, dizziness, early satiety, fatigue, fever, headache, inactivity, insomnia, itching, night sweats, numbness, problems with sexual desire/function, and weight loss) allowing MPN patients to fully profile their disease-related symptom experience.

A shorter version of the MPN-SAF was developed to reduce the survey load for patients called the MPN-SAF Total Symptom Score (TSS) [5]. The MPN-SAF TSS produces a single computed mean of the 10 most clinically important symptom scores of the MPN-SAF. Emanuel and colleagues previously showed that symptom burden as measured by the MPN-SAF TSS was strongly correlated with QoL (Pearson $|r|$=.51; P<.001) as measured by the Global Health Status/QoL (GHS/QoL) scale of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) [6].

To better understand the impact of specific and general symptom severity and prevalence on QoL, we conducted an analysis of symptom-derived predictors of QoL not previously undertaken. Here we assess which of the 18 individual symptoms from the MPN-SAF or summary measures across the symptoms best describe QoL as measured by the QLQ-C30 GHS/QoL scale.

Methods

Patients and setting

This analysis included 1,416 adult patients with MPNs in our existing database having fully completed the symptom assessment packet (MPN-SAF and the EORTC QLQ-C30) at a single clinic visit at the time of the assessment. A detailed description of this cohort has been published elsewhere [5]. Briefly, an international group of collaborating medical centers prospectively administered the symptom assessment packet to patients with MPNs in their native language from November 2009 to January 2011. Approval for this research was given by each center's respective institutional review board.

Summary measures

To investigate the impact of high symptom severity on QoL, two sets of summary measures were constructed summarizing patient- and group-level severity. **(1)** Patient-level severity measures were intended to explore the impact of high severity regardless of the originating symptom item, whereby *a patient's highest scoring symptoms* may influence QoL. The first patient-level severity measure was defined as the highest symptom score within each patient, the second patient-level severity measure was defined as the average of the two highest symptom scores within each patient, the third patient-level severity measure was defined as the average of the three highest symptom scores within each patient, and so on up to the

eighteenth patient-level severity measure defined as the average of all 18 symptoms within each patient. For example, a patient whose three worst symptoms were fatigue (score=10), inactivity (score=9), and early satiety (score=5), the first patient-level severity measure resulted in a score of 10 (the score of the patient's 1 worst symptom), the second patient-level severity measure resulted in a score of 9.5 (average of the patient's 2 worst symptoms), the third patient-level severity measure resulted in a score of 8, and so on. **(2)** Group-level severity measures were intended to explore the impact of severity driven by the overall group, whereby *select symptoms* may influence QoL. The first group-level severity measure was defined as the study group's highest average symptom score, the second group-level severity measure was defined as the average of the study group's two highest scoring symptoms, the third group-level severity measure was defined as the average of the study group's three highest scoring symptoms, and so on up to the eighteenth group-level severity measure defined as the average of all 18 symptoms. Thus, the first group severity measure was the fatigue score (i.e., the most severe symptom later observed for the overall group) for each patient; the second group severity measure was the average of the fatigue and problems with sexual desire/function scores (i.e., the two most severe symptoms later observed for the overall group); the third group severity measure was the average of the fatigue, problems with sexual desire/function, and insomnia scores (i.e., the three most severe symptoms later observed for the overall group); and so on.

To investigate the impact of experiencing symptoms at or above each symptom score threshold, two sets of summary measures were constructed to summarize symptom prevalence and severity. **(1)** For prevalence, ten dichotomous summary variables were constructed as follows; the presence of at least one symptom with a score of 1 or higher, presence of at least one symptom with a score of 2 or higher, presence of at least one symptom with a score of 3 or higher, and so on up to a score of 10. **(2)** For severity, ten summary variables were created as follows; the total number of symptoms with a score of 1 or higher, total number of symptoms with a score of 2 or higher, total number of symptoms with a score of 3 or higher, and so on up to a score of 10.

Statistical analysis

Individual symptoms and constructed summary measures impacting QoL were evaluated using a linear regression model created for each single predictor (symptom or summary measure) with QLQ-C30 GHS/QoL score as the outcome. The resulting R^2 values, reflecting how well the model fit the data, were assessed for performance. Appropriately grouped sets of findings were compared using Wilcoxon rank-sum tests (patient- versus group-level measures, and prevalence threshold versus severity threshold measures).

We established multi-symptom model benchmarks using the R^2 values from a full MPN-SAF model (all 18 MPN-SAF items predicting QoL), reduced MPN-SAF model using forward selection regression, and MPN-SAF TSS model (MPN-SAF TSS as single predictor modeling QoL). Forward selection is a statistical technique aimed at reducing the number of independent variables (MPN-SAF items in our case) in a regression model. Variables are added sequentially starting with those which most improve the overall model fit. This process continues until the significance level of additional variables reaches a pre-specified

model-entry criterion. Due to the large sample size in this study, a strict model-entry criterion of $P<.0001$ for an F statistic was used. Analysis of variance (ANOVA) was used to compare characteristics among MPN disease subgroups. Pearson correlation coefficient (r) estimated the correlation between QoL score and individual MPN-SAF symptoms. Spearman correlation coefficient (rho) estimated the correlation between 18 symptom severities and prevalence. Mean and standard deviation (SD), or median and range are reported where appropriate. P values less than .05 were considered statistically significant. Analyses were executed using SAS statistical software (SAS Version 9.4, SAS Institute, Cary, NC).

Results

Patient characteristics

Among 1,416 adult patients who provided complete symptom and QoL data, MPN disease type included ET (n=625, 44%), PV (n=456, 32%), and MF (n=335, 24%). The mean age was 59 years (SD=14, range 18, 91) and 53% of patients were female (n=739). Symptom assessment packets were provided in Chinese (n=517, 36.5%), French (n=361, 25.5%), Spanish (n=142, 10.0%), Italian (n=137, 9.7%), English (n=105, 7.4%), Swedish (n=92, 6.5%), and German (n=62, 4.4%). Study group characteristics are shown in Table 1

Individual symptoms

Fatigue, problems with sexual desire/function, and insomnia had the highest mean severity (4.0, 3.5, and 2.8, respectively) and were the most prevalent symptoms (85%, 64%, and 63%, respectively). Fever, abdominal pain, and weight loss had the lowest mean severity (0.4, 1.2, and 1.8, respectively) and were the least prevalent symptoms (19.0%, 40.9%, and 34.6%, respectively).Within disease subgroups, mean severity and prevalence for each symptom varied across disease subgroups with most symptoms becoming more severe and prevalent from ET to PV to MF groups. However, fatigue had the highest mean severity and was the most prevalent symptom within all three disease subgroups, and fever had the lowest mean severity and was the least prevalent symptom for all three disease subgroups. Mean QoL worsened across disease subgroups (ET 70.0, PV 66.7, and MF 59.2; $P<.001$; Supplemental Table 1).

Figure 1 depicts the mean symptom severities (height of each bar) in decreasing order along with each symptom's prevalence represented by shade in the overall group. Mean symptom severity and prevalence were highly correlated ($rho = 0.97$, $P<.001$). The relationship between symptom severity and prevalence was consistent within each disease subgroup as well (ET $rho = 0.97$, PV $rho = 0.99$, and MF $rho = 0.97$; all $P<.001$; Supplemental Figure 1).

The Pearson correlations between individual MPN-SAF symptom scores and QoL are also provided in Figure 1. Overall, the median of these correlations was −0.36 (range $r = -0.23$ to −0.54, all $P<.001$). The disease subgroups had median correlations of −0.31, −0.36, and −0.38 for ET, PV, and MF, respectively. The strongest correlations observed in the overall group were for inactivity ($r = -0.54$), fatigue ($r = -0.48$), depression ($r = -0.48$) and

dizziness ($r = -0.44$); the weakest were for fever ($r = -0.23$), night sweats ($r = -0.27$), cough ($r = -0.27$), and problems with sexual desire/function ($r = -0.30$). Consistent with the overall group, ET and PV patients had inactivity, depression, and fatigue as the top three most correlated symptoms with QoL. Within MF patients, inactivity, depression, and fatigue were also among the top most correlated symptoms with QoL.

Model measures

The multi-symptom model benchmarks performed relatively well as expected. The full MPN-SAF model with all 18 symptoms performed the best in modeling QoL ($R^2 = 0.40$). Followed by the forward selection model ($R^2 = 0.38$) then the MPN-SAF TSS model ($R^2 = 0.34$). The final forward selection model resulted in four symptoms being included (inactivity, fatigue, dizziness, and depression).

The best performing individual symptoms were inactivity ($R^2 = 0.29$), fatigue ($R^2 = 0.23$), and depression ($R^2 = 0.23$) (Figure 2A). However, no individual symptom fit as well as any of the model benchmarks including the MPN-SAF TSS. Fewer symptom scores were needed to reach the performance of the MPN-SAF TSS model for patient-level severity measures compared to group-level severity measures (5 vs 7) (Figure 2B). However, no difference was seen in overall performance between patient- and group-level severity measure sets ($P = .10$). The total number of symptoms at or above a severity score of 3 performed best among severity threshold measures ($R^2 = 0.32$) (Figure 2C). Although relatively low and entirely below the curve generated by the severity threshold measures, having at least one symptom at or above a score of 6 performed best among the prevalence thresholds ($R^2 = 0.16$). The summary measures for the total number of symptoms at/above each score performed better overall than those for presence of at least one symptom at/above each score ($P = .001$). When carrying out all modeling within disease subgroups separately, results remained highly consistent (Supplemental Figures 2-5).

Discussion

As a patient's ability to participate in daily activities pursuant to their prerogative recedes, QoL decrement logically follows [7]. Assessment of QoL is often a nebulously considered concept for patients amid highly symptomatic chronic disease. Adding to this complexity, the symptom burden experience among patients with MPNs is variable and undoubtedly bears heavily on QoL. The current literature outlining MPN disease-related symptom assessment methods are centered on monitoring disease progression and guiding therapeutic management. Though, should symptom burden and QoL assessments, two vital considerations to MPN patient trajectories, be utilized in concert? The MPN-SAF and QLQ-C30 GHS/QoL score are well published and rigorously validated PRO tools. When leveraged together, a better understanding of the impact symptom burden has on QoL can be achieved.

In this study of a large, international cohort of MPN patients we found that correlations between QoL decrement and individual MPN-SAF items are variable. Problems with inactivity, fatigue and depression were shown to be the most correlated with QoL decrement. However, these top correlations are only considered moderate in strength. Individual

symptom prevalence and severity also varied but were not unexpected for the MPN population. Interestingly, these data show that symptom prevalence and severity are highly correlated; however, neither high prevalence nor high severity was ubiquitously aligned with QoL decrement. For example, fatigue is known to be the foremost prevalent and severe symptom among MPN patients. Although this was also observed in our data, fatigue alone (mean severity score 4.0; $R^2 = 0.23$) modeled QoL just as well as depression alone (mean severity score 2.3; $R^2 = 0.23$) and worse than inactivity alone (mean severity score 2.2; $R^2 = 0.29$). Inactivity, fatigue, and depression can be considered as "global" symptoms indicative of a multifactorial process rather than as related to a single disease-related process like night sweats or fever, which may explain why they are more closely associated with QoL.

Individual symptom severity and prevalence varied across disease subgroups as expected based on our prior analysis of this cohort [5]. However, highly consistent correlations between individual symptoms and QoL across the disease subgroups suggest that the relationship between symptom burden and QoL is relatively stable independent of disease type. This was further supported by highly consistent results when modeling was carried out by disease subgroup.

Aside from highly prevalent MPN-related symptoms like fatigue, symptom burden profiles vary and patients experience different symptoms. Therefore there is value in assessing the most severe symptoms dynamically from patient-to-patient when considering QoL. Our study shows that patient-level summary measures better modeled QoL than the group-level summary measures. In Figure 2B we see that the curve incrementally incorporating the group's most severe symptoms improves model performance at a slower rate than the curve incrementally incorporating the patient's most severe symptoms. These two curves eventually reflect similar model performance where seven or more of the most severe symptoms are incorporated into the summary measures. Intuitively, this suggests that more can be gleaned of QoL using fewer symptoms by evaluating the most severe symptoms within individual MPN patients rather than the most severe symptoms among MPN patients overall.

The additional summary measures evaluating prevalence and severity at or above each severity threshold are consistent such that Figure 2C reveals that QoL is better modeled by the number of symptoms exceeding each threshold than whether or not a single symptom exceeds each threshold. This supports that QoL is driven by the symptom profile more than any individual symptom.

The multi-symptom benchmark models exemplify how multiple symptoms better describe the impact of symptom burden on QoL. This can be seen across Figure 2 by the performance of the full model with all 18 MPN-SAF items as well as the shortened MPN-SAF TSS model. Although the full model performed best by measure of R^2, subsequent versions of the MPN-SAF were developed, in part, to reduce survey load for patients and provide more clinically informative 'snapshots' of symptom profiles for clinicians. Solely using the full 18-item MPN-SAF version in practice due to the statistical findings here may not be advantageous for the evolving needs of clinicians. That is, where the statistician would be pleased with more data, alas it is not necessarily ideal for the clinician or patient. For

example, the MPN-SAF TSS performed better than any individual symptom, and performed better than several of the summary measure models, further supporting the validity of MPN-SAF TSS. Although not originally intended to evaluate QoL, the MPN-SAF TSS appears to provide a profile of symptom burden meaningfully impacting QoL.

These findings merit discussion regarding current myelofibrosis disease response criteria in which a 50% reduction from baseline in MPN-SAF TSS is considered a symptomatic response [8]. Considerable data exist to support that a 50% reduction is clinically meaningful, although smaller reductions may also be meaningful. The current data, for example, suggest that even a single symptom with a score of 6 or higher can greatly impact QoL. Therefore, symptom response criteria could be personalized from patient-to-patient and focused on reducing the most problematic symptom(s), either in terms of percent reduction from baseline or in terms of reduction to below a given severity level. While our cross-sectional data cannot inform such modifications to the current symptom response criteria, these findings support further work in longitudinal studies to develop alternative symptom response criteria.

This study is not without limitation. Symptoms were included in the MPN-SAF TSS based on their *clinical importance* from the clinical investigators' perspective in describing patient-level disease status, not necessarily their *statistical importance* in modeling QoL. Therefore these clinical metrics are inherently not optimized to be robust statistical predictors of QoL and unmeasured constructs such as symptom interference, functioning domains, and perceived health status may also impact QoL.

Further study may include validating the best performing models on external data to develop statistically optimized summary measures of symptom burden predicting QoL. Longitudinal studies evaluating whether changes in symptoms impact QoL or whether the monitoring of QoL is clinically useful may extend the results from this cross-sectional study.

Concluding Remarks

The fundamental relationship between symptoms and QoL in MPN patients appears to be consistent across ET, PV, and MF disease subgroups. Multiple symptom severity scores are needed to best describe symptom impact on quality of life. Symptom severity at the patient-level is more predictive of quality of life than severity at the group-level where fewer symptoms are considered. Having at least one severe symptom and having multiple symptoms of moderate intensity are meaningfully predictive of quality of life decrements.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Passamonti F, Mora B, Maffioli M. New molecular genetics in the diagnosis and treatment of myeloproliferative neoplasms. Curr Opin Hematol. 2016 3;23(2):137–43. doi: 10.1097/MOH. 0000000000000218. PubMed PMID: 26825696. [PubMed: 26825696]

2. Mesa RA, Niblack J, Wadleigh M, et al. The burden of fatigue and quality of life in myeloproliferative disorders (MPDs): an international Internet-based survey of 1179 MPD patients. Cancer. 2007 1 1;109(1):68–76. doi: 10.1002/cncr.22365. PubMed PMID: 17123268. [PubMed: 17123268]

3. Mesa RA, Schwager S, Radia D, et al. The Myelofibrosis Symptom Assessment Form (MFSAF): an evidence-based brief inventory to measure quality of life and symptomatic response to treatment in myelofibrosis. Leuk Res 2009 9;33(9):1199–203. doi: 10.1016/j.leukres.2009.01.035. PubMed PMID: 19250674; PubMed Central PMCID: PMCPMC4419687. [PubMed: 19250674]

4. Scherber R, Dueck AC, Johansson P, et al. The Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF): international prospective validation and reliability trial in 402 patients. Blood. 2011 7 14;118(2):401–8. doi: 10.1182/blood-2011-01-328955. PubMed PMID: 21536863. [PubMed: 21536863]

5. Emanuel RM, Dueck AC, Geyer HL, et al. Myeloproliferative neoplasm (MPN) symptom assessment form total symptom score: prospective international assessment of an abbreviated symptom burden scoring system among patients with MPNs. J Clin Oncol. 2012 11 20;30(33): 4098–103. doi: 10.1200/JCO.2012.42.3863. PubMed PMID: 23071245; PubMed Central PMCID: PMCPMC4872304. [PubMed: 23071245]

6. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993 3 3;85(5):365–76. PubMed PMID: 8433390. [PubMed: 8433390]

7. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995 1 4;273(1):59–65. PubMed PMID: 7996652. [PubMed: 7996652]

8. Tefferi A, Cervantes F, Mesa R, et al. Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. Blood. 2013 8 22;122(8):1395–8. doi: 10.1182/ blood-2013-03-488098. PubMed PMID: 23838352; PubMed Central PMCID: PMCPMC4828070. [PubMed: 23838352]
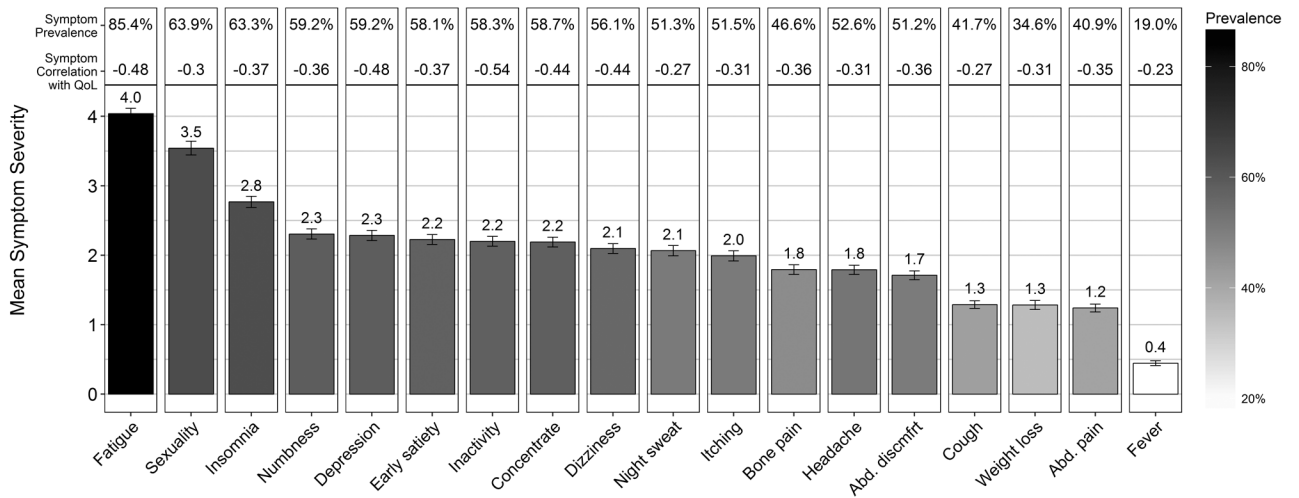
Figure 1.
Individual symptom severity, prevalence, and correlation with quality of life. Symptom severity was measured as 0 (absent) to 10 (worst imaginable) using the MPN Symptom Assessment Form. Prevalence was defined as a symptom score greater than or equal to 1. Pearson's correlations shown are between each symptom and quality of life as measured by the Global Health Status/Quality of Life scale of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 3.
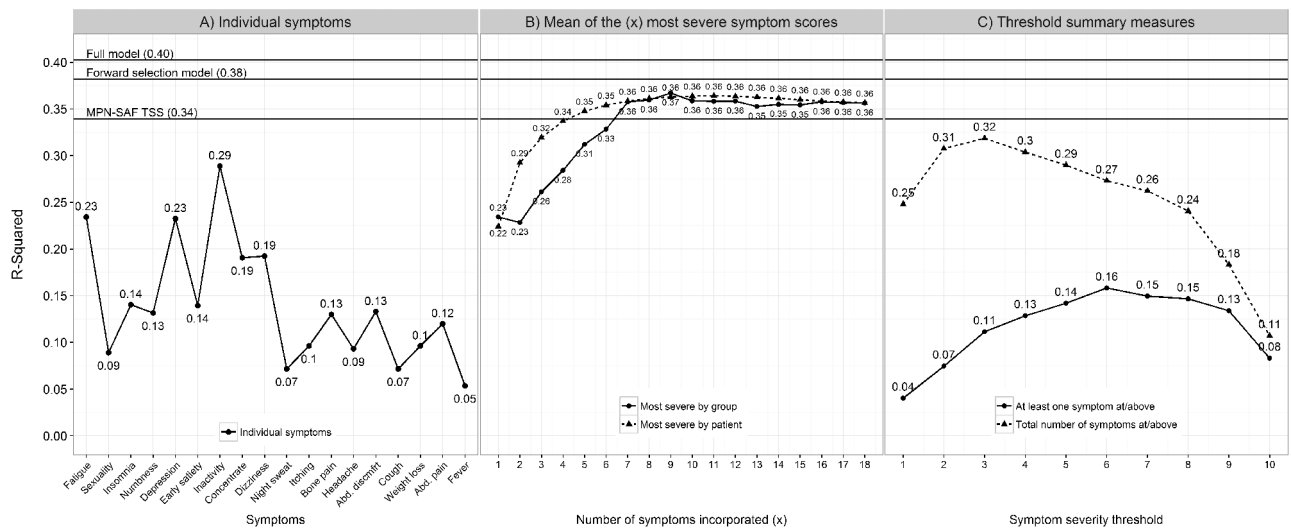
Figure 2.

Individual symptoms and summary measure impact on quality of life. Full model: model with all 18 symptoms predicting quality of life; forward selection model: full model reduced by forward selection (included inactivity, fatigue, dizziness, and depression); MPN-SAF TSS: model with only the MPN-SAF TSS score. Abd discmfrt: abdominal discomfort; (A) individual symptoms along the x-axis are ordered from highest to lowest mean severity (same as Figure 1). (B) Most severe by group: the most severe symptoms were averaged together successively along the x-axis ordered highest to lowest mean severity (same as Figure 1), e.g. $x = 1$ represents the Fatigue score, $x = 2$ represents the average of the fatigue and sexuality scores, $x = 3$ represents the average of the fatigue, sexuality, insomnia scores, etc. up to the average of all 18 symptoms; most severe by patient: the most severe symptom scores within each patient were averaged successively. E.g. $x = 1$ represents the most severe symptom score within each patient, $x = 2$ represents the average of the two most severe symptom scores within each patient, $x = 3$ represents the average of the three most severe symptom scores within each patient, etc. up to the average of all 18 symptom scores. (C) Symptom severity is measured on a 0 (absent) to 10 (worst imaginable) scale.

Table 1.

Study Group Characteristics

Total study group:	n = 1416
Age, mean (SD)	58.5 (13.8)
Sex, n (%)	
Male	659 (47.1%)
Female	739 (52.9%)
MPN type, n (%)	
PV	625 (44.1%)
ET	456 (32.2%)
MF	335 (23.7%)
Duration of MPN diagnosis, median years (range)	5 (0-43)
Survey language, n (%)	
Chinese	517 (36.5%)
English	105 (7.4%)
French	361 (25.5%)
German	62 (4.4%)
Italian	137 (9.7%)
Spanish	142 (10.0%)
Swedish	92 (6.5%)
QLQ-C30 GHS/QoL, mean (SD)	66.4 (22.9)
MPN-SAF TSS, mean (SD)	2.0 (1.6)
Anemia [HGB <11], n (%)	
Missing	77 (5.4%)
Absent	1125 (79.4%)
Present	214 (15.1%)
Leukopenia [WBC <3.5], n (%)	
Missing	78 (5.5%)
Absent	1252 (88.4%)
Present	86 (6.1%)
Thrombocytopenia [PLT <150], n (%)	
Missing	81 (5.7%)
Absent	1188 (83.9%)
Present	147 (10.4%)

SD: standard deviation; HGB: hemoglobin; WBC: white blood count; PLT: platelet.