Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Organizing Benzodiazepine Withdrawal-Related Symptoms: A Provisional Ontology and Severity-Allocation Analysis

Valsa Madhava

vmadhava@brainbodymedical.com

Brain Body Medical https://orcid.org/0009-0007-1581-0582

Research Article

Keywords: Benzodiazepine withdrawal, Symptom ontology, Patient-reported outcomes, Symptom severity, Regulatory systems, Interoception, Phenomenology

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11607678/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: The author is the founder and clinical director of Brain Body Medical (The Benzo Taper Doctor), the practice from which this cohort was drawn.

Organizing Benzodiazepine Withdrawal-Related Symptoms: A Provisional Ontology and Severity-Allocation Analysis

Short title: Benzodiazepine Withdrawal Symptom Ontology

Author: Valsa S. Madhava, MD, MPH, MS

Brain Body Medical (The Benzo Taper Doctor), New York, NY, USA

Corresponding author: Valsa S. Madhava, Brain Body Medical, New York, NY, USA. Email: vmadhava@brainbodymedical.com

---

Abstract

Background. Benzodiazepine withdrawal can involve many symptoms across different body systems. Existing symptom lists and rating scales help identify and measure these symptoms, but they do not show how the full symptom vocabulary fits together. We organized a community-compiled withdrawal-symptom list into a provisional, phenomenology-first ontology and then examined where patient-reported severity concentrated within that framework.

Methods. We analyzed cross-sectional intake data from 39 adults presenting to a specialty benzodiazepine taper practice (37 presenting for taper care; 2 symptomatic after completing a taper). In Analysis A, 243 original symptom phrases from a publicly available benzodiazepine withdrawal-symptom list were decomposed into 278 analyzable rows and organized into 171 provisional symptom nodes. Each node was mapped by the author to one or more of 17 regulatory systems using a written codebook. In Analysis B, this frozen map was applied to patients' 0–10 ratings on a 233-item questionnaire using severity-conserving fractional attribution, with 95% bootstrap confidence intervals. Robustness was examined using alternative analytic choices and a separate blinded AI coding pass.

Results. The highest-rated individual symptoms were anxiety (mean 7.3/10), insomnia (6.5), hypersensitivity to stress (5.5), nervousness (5.4), loss of confidence (5.4), and depression (5.2). Under the author's coding, 157 of 171 symptom nodes (91.8%) mapped to two or more regulatory systems. The largest allocated severity shares mapped to interoceptive regulation (14.3%), state-level regulation (13.5%), cognitive control (10.2%), and threat-arousal regulation (10.0%). When severity was summarized as the mean rating of questionnaire items contributing to each system, cognitive control ranked highest (3.23/10), followed by affective/reward regulation (3.09/10) and threat-arousal regulation (2.79/10). Across robustness checks, interoceptive regulation, cognitive control, and threat-arousal regulation consistently ranked in the top five, and affective/reward regulation ranked in the top six. Two features were coding-dependent: under the blinded AI coding pass, 70 of 171 nodes (40.9%) were multi-system, and state-level regulation fell from second to twelfth, while affective/reward regulation rose to first.

Conclusions. This study organized a broad vocabulary of benzodiazepine withdrawal-related symptoms into 171 provisional symptom nodes mapped to 17 regulatory systems, providing a structured approach to understanding complex symptom experiences. Applying this framework

to patient-reported severity ratings identified interoceptive regulation, cognitive control, and threat-arousal regulation as consistently prominent across alternative analyses, with affective/reward regulation also prominent. Although some findings varied with coding decisions, the framework provides a transparent, auditable approach to organizing symptoms and examining their severity across overlapping regulatory systems. Further research may help refine the framework and assess its applicability to other clinical populations.

Introduction

Benzodiazepine withdrawal can be prolonged and disabling, with symptoms extending beyond anxiety and insomnia. Patients may report neurologic, autonomic, gastrointestinal, sensory, cognitive, affective, sleep, motor, immune-related, and perceptual symptoms, sometimes persisting well beyond the period of dose reduction. For some patients, withdrawal is mild and transient; for others, it is severe, multisystem, and difficult to interpret clinically. This variability has been described by patients and clinicians and documented in patient-advocacy resources and large patient surveys [1,2,3].

Several prior efforts have attempted to capture this broader clinical picture. The descriptive construct of benzodiazepine-induced neurological dysfunction (BIND) emphasizes that symptoms and functional consequences may emerge during benzodiazepine use or tapering and may continue after discontinuation [4]. Comprehensive symptom lists, including patient-advocacy resources and clinical guides such as the Ashton Manual, have also catalogued the range of reported withdrawal and post-withdrawal symptoms [3,5]. These resources are valuable for recognizing that benzodiazepine-related symptoms may be persistent and multisystem. However, symptom lists do not, by themselves, show how the reported symptom vocabulary is organized — that is, how symptoms relate to one another as experienced symptoms or how they may map onto broader regulatory systems.

Standardized withdrawal rating scales take a different approach. The Clinical Institute Withdrawal Assessment for Benzodiazepines (CIWA-B) comprises 20 items — 17 patient-rated and 3 observer-rated, each scored 0–4 — designed to quantify withdrawal severity [6]. The Benzodiazepine Withdrawal Symptom Questionnaire (BWSQ) is a brief self-report instrument scored 0–2 per item that emphasizes perceptual and somatic symptoms [7]. Such scales are useful for tracking severity over time, but they capture only part of the symptom range patients describe and do not show how the broader symptom vocabulary is organized.

The evidence base also remains limited in ways that make careful terminology important. A recent scoping review of long-term neurological consequences following benzodiazepine exposure identified 46 relevant studies. Of these, 28 studies (61%) contained findings suggestive of a protracted withdrawal phase, but these findings were often incidental rather than primary, prespecified outcomes, and the review could not determine the prevalence, trajectory, or causality of symptoms [8]. Clinical guidance reflects a related classification problem: the 2025 joint clinical practice guideline on benzodiazepine tapering notes that withdrawal symptoms can be difficult to distinguish from recurrence or rebound and states that its withdrawal-symptom table is not exhaustive [9]. These limitations point to a necessary first step: organizing reported symptoms descriptively before using them to infer biological mechanisms, clinical subtypes, or treatment implications.

Neurobiology provides a rationale for considering symptoms across broad regulatory systems, while also setting an important boundary. Benzodiazepines act primarily as positive allosteric modulators of GABAA receptors, and long-term exposure has been associated with neuroadaptive changes in inhibitory and excitatory signaling, including altered GABAA-receptor function, receptor trafficking and subunit composition, and compensatory glutamatergic adaptations [10,11,12,13]. Because inhibitory signaling participates broadly in neural systems involved in arousal, sensory processing, autonomic function, motor control, cognition, affect,

sleep–wake regulation, and state regulation, withdrawal-related neuroadaptation provides a plausible context for multisystem symptoms rather than a simple return to baseline physiology. Recent synaptic and circuit-level reviews similarly describe prolonged benzodiazepine exposure as involving changes in inhibitory and excitatory synaptic organization, while emphasizing that direct links between these changes and specific patient-level tolerance or withdrawal presentations remain limited [14]. Thus, this study uses regulatory systems as descriptive organizing labels for the symptom ontology; it does not test these mechanisms directly, and the proposed regulatory-system map should not be read as biologically validated by this literature.

Moving beyond symptom cataloguing can overclaim if not carefully bounded, especially when reported symptoms are used to imply biological subtypes without formal clustering or latent-class modeling, biomarkers, prospective outcomes, or independent validation. We therefore take a descriptive, phenomenology-first approach.

In this approach, the source symptom vocabulary is first organized according to shared reported experience rather than presumed mechanism. Here, the source vocabulary was the Benzodiazepine Information Coalition withdrawal-symptom list, treated as patient-reported symptom language rather than prevalence or epidemiologic data. The aim was to preserve clinically meaningful distinctions without multiplying labels unnecessarily. Only after completing this symptom-level organization were symptom nodes mapped to broader regulatory systems. Ordering the work this way — node formation first, regulatory-system assignment second — kept the description of the symptom vocabulary separate from its interpretation.

The resulting symptom-node and regulatory-system map was then applied to the 39 patients' original 0–10 questionnaire ratings to describe how patient-rated severity was distributed across regulatory systems.

Regulatory-system framework

In this study, "regulatory system" refers to one of 17 broad brain–body organizing labels used to describe where symptom experiences appeared to fall within the study's provisional regulatory map. These labels covered functions such as internal bodily sensing, autonomic regulation, arousal and threat, cognition, sensation, pain, movement, gastrointestinal regulation, affect/reward, sleep–wake regulation, immune-related sickness signaling, and self-perceptual integration. They are organizing labels, not a complete taxonomy of human physiology, discrete biological mechanisms, disease categories, patient subtypes, or validated anatomical circuits.

Because a single reported symptom can plausibly involve more than one regulatory system, the ontology permits one-to-many mapping: a symptom node can be assigned to more than one regulatory system. For example, dizziness may involve interoceptive regulation, autonomic regulation, sensory processing, and state-level regulation, while depersonalization may involve self-body/perceptual integration, state-level regulation, sensory processing, and threat-arousal regulation. The 17-system vocabulary was author-defined during ontology development to cover the range of symptom-node content observed after node formation, while keeping it small enough to define, apply consistently, and audit. The system labels were informed by broad clinical and neurobiological constructs, including interoceptive regulation, autonomic regulation, sensory processing, pain processing, affective/reward regulation, sleep-wake regulation, and state-level regulation. The vocabulary was fixed before regulatory-system assignment began and

was not changed during patient-level severity allocation; the item-to-node-to-system crosswalk, which records those assignments, was frozen once regulatory-system assignment was complete and before any patient-level severity rating was applied, and both were held fixed through the patient-level severity-allocation analysis. All system-level percentages reported in this study are conditional on this provisional 17-system framework and should not be treated as fixed biological properties of the symptoms. Full operational definitions, including assign/do-not-assign criteria, are given in Table 1 and S5 File.

Study aims

This study aimed to organize benzodiazepine-withdrawal symptom vocabulary descriptively and describe how patient-rated severity is distributed across broad regulatory systems, providing a provisional, hypothesis-generating foundation for future work.

Specifically:

- **Aim 1.** Construct and describe a provisional, phenomenology-first symptom ontology from the Benzodiazepine Information Coalition (BIC) withdrawal-symptom list, resolving reported phrases into provisional phenomenological symptom nodes and describing how those nodes map to a set of broad regulatory systems: the 171-node / 17-regulatory-system map.
- **Aim 2.** Apply the frozen 171-node / 17-system map to raw 0–10 symptom-severity ratings from 39 patients to estimate severity-weighted regulatory-system allocation and describe how that allocated severity is distributed across regulatory systems.

These aims are descriptive and hypothesis-generating. This study does not infer clusters, latent classes, or discrete clinical phenotypes; does not establish causal mechanisms; does not demonstrate biological validity of the regulatory map; and does not evaluate treatment response or claim treatment benefit. Its purpose is to provide an auditable, provisional description of how reported withdrawal-symptom vocabulary is organized and where patient-rated severity concentrates, so that these observations can be tested and evaluated in subsequent work.

Methods

Study design and cohort

This study had two linked components: symptom-ontology construction (Analysis A) and patient-level regulatory-system severity-allocation mapping (Analysis B). The study was observational and single-site. The analyses describe symptom vocabulary and severity distribution; they do not define patient subgroups or test any intervention (Figure 1). This study is reported in accordance with the STROBE guideline for observational studies; the completed checklist is provided as a supplementary reporting checklist.

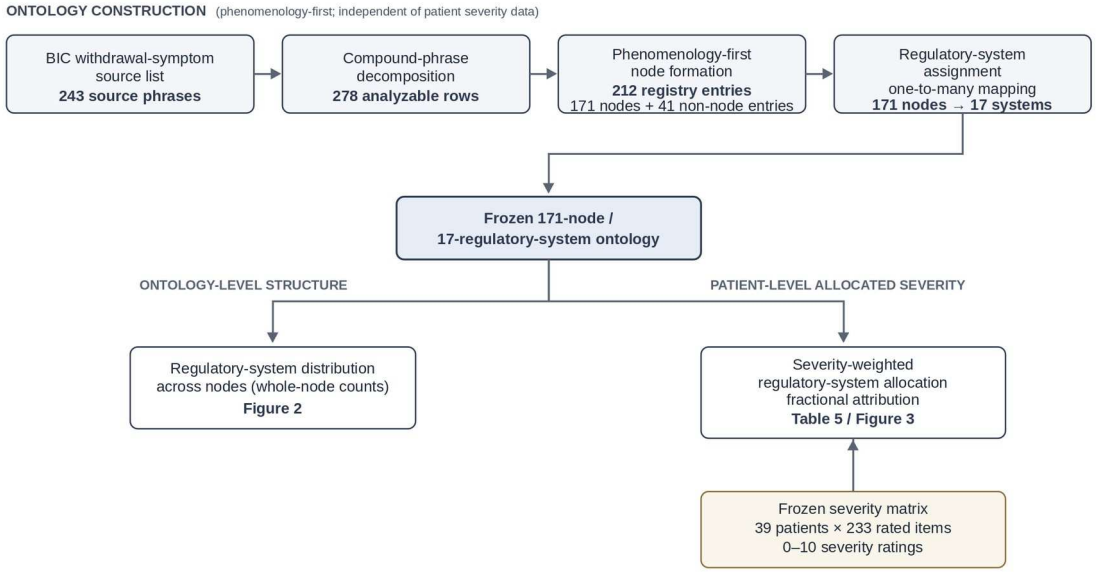


Figure 1. Study workflow.

Schematic of the two linked analyses. The Benzodiazepine Information Coalition withdrawal-symptom list was used to construct a phenomenology-first ontology: 243 source phrases were decomposed into 278 analyzable rows and organized into 212 registry entries, including 171 provisional symptom nodes and 41 non-node entries. The 171 nodes were then mapped to 17 regulatory systems. The frozen 171-node / 17-regulatory-system map was applied to the 39 patients' original 0–10 questionnaire ratings to estimate severity-weighted regulatory-system allocation. Ontology-level node counts and patient-level allocated severity are shown as separate outputs.

The patient sample comprised 39 consecutive adults evaluated in a specialty benzodiazepine taper program between approximately January 2022 and August 2025: 37 presented to begin a supervised taper, and two had completed tapering but remained symptomatic. Clinical records did not systematically distinguish whether patients were on stable doses, self-tapering, experiencing tolerance or interdose symptoms, or recovering from a recent dose reduction at the time of assessment. The ratings therefore describe symptoms reported at presentation for taper care rather than symptoms confirmed to arise from withdrawal. All patients completed a 233-item withdrawal-symptom questionnaire at intake, with each item rated from 0 to 10. These baseline severity ratings constituted the frozen patient input for Analysis B. Standard intake measures, including demographics, benzodiazepine and Z-drug exposure, and the PROMIS-29 v2.0 profile (Patient-Reported Outcomes Measurement Information System; a standardized,

population-normed measure of physical, mental, and social health domains) [16], are summarized as descriptive cohort context only. Continuous characteristics are reported as mean (SD) or median (IQR), and categorical characteristics as n (%) or %. PROMIS-29 raw domain scores were converted to T-scores using published conversion tables, with reverse-coding of Sleep item 18 and all Social Roles items; the number of patients contributing to each domain varies with item completeness.

This secondary, retrospective analysis used existing clinical intake data rather than a prospectively recruited study; consistent with the exempt-category determination governing this analysis, the investigator did not contact subjects, so no patient declined participation or was otherwise lost at the data-collection stage. All 39 patients who met the practice's intake criteria and completed the questionnaire during this period were included; no formal sample-size or power calculation was performed, consistent with the study's descriptive aims.

The 233-item questionnaire was completed at intake as part of routine clinical evaluation. The patient-facing form instructed patients to rate each symptom with a number between 0 ("not at all") and 10; no recall period or withdrawal-attribution instruction was specified. The questionnaire did not systematically distinguish symptoms occurring during active dose reduction, interdose periods, dose holds, or after taper completion. It also did not require patients to independently attribute each symptom to benzodiazepine withdrawal. Accordingly, patient-level severity allocation should be interpreted as describing symptoms reported by patients evaluated for benzodiazepine taper/withdrawal care, not as definitive attribution of every symptom to benzodiazepine withdrawal.

The BRANY IRB granted the study exempt status and a HIPAA authorization waiver for secondary research use of existing clinical data. De-identified analyses were performed on frozen data files.

This cohort has also been described in a separate preprint [15] that assigned each of the 233 questionnaire items to a single investigator-defined, mechanism-named axis and labeled each patient according to the axis with the greatest symptom severity. The present analysis differs in several respects: it starts from the source symptom vocabulary rather than from the questionnaire, separates symptom grouping from regulatory-system assignment, uses descriptive regulatory-system labels rather than mechanism-named axes, permits one-to-many mapping, documents its assignment rules in a codebook, and describes severity at the cohort level without classifying patients. The two analyses share the same cohort, questionnaire, cohort characteristics, and PROMIS-29 scores, but the present analysis did not use any results from the earlier categorization.

Source symptom corpus and questionnaire distinction

Two related symptom resources were used. The ontology source corpus for Analysis A was the BIC withdrawal-symptom list. From this list, 243 raw patient-reported symptom phrases were carried forward with original wording preserved. Source lines that contained more than one symptom were decomposed into component symptom phrases before analysis, yielding 278 analyzable rows. The BIC list was treated as an inventory of reported symptom language, not as prevalence, incidence, or epidemiologic data.

The patient questionnaire used for Analysis B was derived from the same BIC symptom vocabulary but was not identical to the 243-phrase source corpus. Differences arose from transcription into questionnaire format, seven rows that merged two distinct symptoms into a single rating, four symptoms omitted during transcription, and a duplicated Paraesthesia column; these differences are documented in S3 File. Excluding the duplicate Paraesthesia column left 232 distinct questionnaire items.

The 243-phrase source corpus used to build the ontology and the 233 rated questionnaire columns used to estimate allocated severity were connected through a frozen item-to-node-to-system crosswalk. The crosswalk was derived from the finalized ontology rather than constructed independently of it: for each of the 233 questionnaire columns, the author linked the corresponding source symptom phrase or phrases to their assigned provisional node or nodes and inherited each node's regulatory-system assignment or assignments from S4 Table. It is therefore a structured restatement of the frozen ontology at the questionnaire-column level, not a second, independent construction.

The full 233-item patient questionnaire is provided as S1 Table. The frozen item-to-node-to-system crosswalk is provided as S2 Table. Data provenance and matrix lineage are documented in S3 File. The full ontology registry, including the 171 provisional symptom nodes, regulatory-system assignments, confidence ratings, and retained non-node entries, is provided as S4 Table.

Analysis A: Ontology construction and regulatory-system mapping

The ontology was built using a phenomenology-first approach. Symptom phrases were grouped by shared reported experience before any biological or system-level interpretation was applied. A documented triage-and-grouping workflow organized symptom phrases into preliminary groups, applied a written sameness test, decomposed compound phrases into component symptoms, and retained non-node material under explicit labels.

The 278 analyzable rows resolved into 171 provisional phenomenological symptom nodes, which served as the primary ontology units for Analysis A. Material that did not meet criteria for a provisional node was retained separately rather than forced into the node set: 14 umbrella/general categories, 9 syndrome-level entries, 3 unresolved-singleton items, 1 deferred undecomposed compound, and 14 compound-parent traceability markers (41 non-node entries in total). Together, the 171 provisional nodes and 41 non-node entries make up the full 212-entry ontology registry (S4 Table); only the 171 nodes contribute to node-level counts and the primary patient-level severity-allocation analysis. The umbrella/general and syndrome-level entries represented broad symptom categories or syndromes rather than individual symptom concepts. Although excluded from the primary analysis, their names could still indicate one or more regulatory systems, allowing their inclusion in a separate sensitivity analysis (see Analysis B).

Regulatory-system assignment

With the 171 symptom nodes held fixed, each node was mapped to one or more of 17 controlled regulatory-system labels. These labels were used as descriptive organizing categories rather than as mechanisms, diagnoses, or validated anatomical circuits. The 17 systems were autonomic regulation; interoceptive regulation; sensory processing; motor regulation; threat-arousal regulation; sleep-wake regulation; cognitive control; affective/reward regulation;

self-body/perceptual integration; neuroimmune/inflammatory-sickness signaling; enteric/GI regulation; endocrine/metabolic/reproductive regulation; pain processing; barrier/tissue/structural integrity; sexual/reproductive function; oral/fluid regulation; and state-level regulation.

One-to-many mapping was permitted, so a single symptom node could be assigned to more than one regulatory system. Each regulatory system was specified by a written operational definition with explicit assign/do-not-assign criteria. Condensed definitions are provided in Table 1, and the full regulatory-system codebook, including assign/do-not-assign criteria and borderline-case guidance, is provided as S5 File. The regulatory-system labels draw on broad, established descriptions of interoceptive regulation [17], central autonomic network organization [18], and sensory/pain processing, including central sensitization [19], but the assignments in this study remain descriptive and provisional.

Table 1. The 17 regulatory systems and their operational definitions.

Systems are listed alphabetically. Full assign/do-not-assign criteria and borderline-case guidance appear in S5 File.

Regulatory system	Operational definition
Affective/reward regulation	Emotional-valence and motivational systems, including positive affect and motivation such as reward, pleasure, and drive, as well as negative affect such as distress, anger, shame, and dysphoria.
Autonomic regulation	Involuntary physiological control, including cardiovascular tone, thermoregulation, sweating, and pupillary and vascular responses.
Barrier/tissue/structural integrity	Tissue structure and surfaces, including skin, hair, nail, teeth, and gum findings; swelling; rash; delayed healing; and visible structural changes.
Cognitive control	Higher-order mental processes, including attention, memory, executive function, decision-making, and the flow and character of thought.
Endocrine/metabolic/reproductive regulation	Hormonal and metabolic regulation, including reproductive-hormonal state, appetite, weight, thyroid-related complaints, and menstrual or other hormonal complaints.
Enteric/GI regulation	The gastrointestinal tract and its nervous-system regulation, including motility, secretion, sphincter control, and visceral sensation.
Interoceptive regulation	Sensing and awareness of internal bodily states — the felt experience of what is happening inside the body, such as heartbeat awareness, internal pressure, or general "body sense."
Motor regulation	Control of voluntary and involuntary movement, muscle tone, and observable motor output, including tremor, weakness, spasm, and coordination.
Neuroimmune/inflammatory-sickness signaling	Immune-mediated "sickness behavior" and inflammatory processes producing systemic, flu-like, or malaise-type states.
Oral/fluid regulation	Fluid and secretion regulation broadly, including saliva, tears, mucosal secretions, and hydration or fluid-balance state.
Pain processing	Nociceptive or pain-specific processing — the sensation of pain, aching, or soreness regardless of location.
Self-body/perceptual integration	Integration of body ownership, body schema, and the sense of self or reality, including depersonalization and derealization.
Sensory processing	Processing of external sensory input, including vision, hearing, smell, taste, and touch, as well as vestibular/balance sensing, including hypersensitivity, distortion, or loss.
Sexual/reproductive function	The functional output of sexual response, including desire, arousal, and orgasm, distinct from the underlying hormonal system.
Sleep-wake regulation	The sleep-wake cycle, sleep architecture, and sleep-specific phenomena.
State-level regulation	Broad, whole-state configurations that cross systems rather than reducing to one system, including global altered-activation states and broad instability or fluctuation constructs.
Threat-arousal regulation	Fear, threat detection, vigilance, and fight-or-flight activation.

Structural analytics used whole-node counting: a node assigned to multiple systems contributed one count to each assigned system, so system counts were non-mutually exclusive, and percentages used the 171 nodes as the denominator. We also counted how many regulatory systems were assigned to each symptom node and how many nodes were shared by each pair of regulatory systems. Symptom-node formation preceded and was independent of regulatory-system assignment, so the system distribution is a description applied to an independently constructed node set rather than a structure built to confirm itself.

Procedural audit of ontology construction

To document procedural consistency, both ontology steps underwent a blinded AI coding pass using the written rules, with the original assignments withheld, followed by item-by-item adjudication by the author. The audit pass was performed using Claude Opus 4.8. Node formation was compared by whether the same symptom phrases were grouped into the same nodes; regulatory-system assignment was compared by whether each node was assigned to the same regulatory systems, using Jaccard similarity to measure the overlap between assignments regardless of their order. For node formation, first-pass agreement was classified as full when symptom-phrase groupings matched exactly, partial when they overlapped, and absent when they did not overlap. For regulatory-system assignment, agreement was classified as full when the assigned systems were identical, partial when they overlapped but were not identical, and absent when no regulatory system was shared. Disagreements in which both assignments satisfied the written rules were recorded as both-valid tie-breaks. We also compared the mean number of regulatory systems assigned per node under the author's coding and the blinded AI coding pass. The author (V.S.M.) performed all adjudication personally, applying the prewritten node-formation and regulatory-system-assignment rules to each disagreement; no other rater or adjudicator was involved. Because AI tools also supported development of the original taxonomy (see “Use of AI tools”) and the same author finalized and adjudicated the ontology, this audit is reported as a procedural consistency check rather than independent validation, human inter-rater reliability, or biological validation. Summary audit results are provided as S6 Table, and detailed audit materials are provided as S7 File. In addition, the blinded AI coding pass’s regulatory-system assignments were used as an alternative coding in a robustness analysis of severity allocation (see Analysis B).

Generative AI tools also assisted with ontology development, refinement of coding rules, and computational analysis. The author reviewed and adjudicated the resulting assignments and analytic outputs. The roles of these tools, human oversight, and associated limitations are detailed in the “Use of AI tools” section.

Analysis B: Patient-level severity-weighted regulatory-system allocation

Analysis B applied the frozen 171-node / 17-system ontology to the frozen 39-patient severity matrix to estimate how patient-reported severity was distributed across regulatory systems. Raw 0–10 symptom-severity ratings were the primary input; scores were not dichotomized for the primary analysis. Each rated symptom column was linked through the frozen crosswalk from raw symptom phrase to symptom number, provisional node or nodes, and regulatory system or systems. The duplicate Paraesthesia column (labeled “Paraesthesia.1” in the data file) was dropped. Severity that mapped only to non-node material, including umbrella/general, syndrome-level, unresolved-singleton, deferred-compound, and compound-parent entries, was retained as a separately reported excluded non-node bucket so that denominator reconciliation remained visible.

Each patient's total symptom severity was calculated by summing their ratings across all 233 questionnaire columns and summarized using the median, interquartile range (IQR), and range. To assess whether a small number of patients accounted for a large proportion of severity, we calculated the percentage of total mapped severity contributed by the five patients with the

highest mapped severity. Questionnaire items were ranked by their average 0–10 severity rating across patients, and the overall mean rating across the 232 distinct questionnaire items was also calculated.

The primary measure was the severity-weighted regulatory-system allocated severity share under severity-conserving fractional attribution. If a symptom column mapped to more than one node, its severity was split equally across those nodes. If a node mapped to more than one regulatory system, its node severity was split equally across those systems. Under this rule, the attributed severity summed exactly to the mapped-severity total, and the 17 system shares summed to 100%. For each system, allocated severity share was calculated as severity attributed to that system divided by total severity mapped to the 171 nodes. The primary allocated-severity share was calculated from pooled attributed severity across the cohort; patients with higher total mapped severity therefore contributed more severity points to the cohort-level distribution, so the measure is not an unweighted average of each patient's regulatory-system profile. Patient-level nonparametric bootstrap resampling provided 95% percentile confidence intervals for each system's allocated severity share, using 10,000 resamples, patient as the resampling unit, and seed 12345. These bootstrap intervals reflect uncertainty from patient-level resampling with the ontology, crosswalk, regulatory-system assignments, and equal-splitting rules held fixed; they do not incorporate uncertainty in node formation, regulatory-system assignment, or the attribution rule itself. Analyses were conducted in Python 3, using NumPy for numerical computation and openpyxl for reading source workbooks. Analytic code is provided as S9 File.

As a worked example, the questionnaire column "Earache and sinus problems" (total severity 79 across the cohort) mapped to two provisional nodes: Earache (N057) and Sinus problems (N148). Its severity was split equally between them, 39.5 points each. Earache, in turn, was assigned to two regulatory systems (pain processing; sensory processing), so its 39.5 points were split equally again, 19.75 points to each. Sinus problems was assigned to three systems (barrier/tissue/structural integrity; neuroimmune/inflammatory-sickness signaling; sensory processing), so its 39.5 points were split three ways, about 13.17 points to each. Sensory processing therefore received a contribution from both nodes (19.75 + 13.17 = 32.92 points) for this one questionnaire column, while the other systems each received a single, smaller share. Applied across the questionnaire data from all 39 patients, this item-level, severity-conserving splitting rule produces the system-level allocated severity shares reported in Table 5 and Figure 3.

During clinical data preparation, blank responses and 106 nonnumeric responses (for example, "?", "Unsure", or "N/A") were coded as zero in the original 39-patient severity matrix, and 61 other nonnumeric responses were converted to the number given. We reviewed the original patient questionnaires to distinguish unanswered responses from recorded numeric zero ratings. Unanswered items were defined as blank or absent responses, using 233 items as the denominator for every patient; nonnumeric responses were not counted as unanswered.

Missing-data sensitivity analyses tested three definitions of missing responses: (1) blank or absent responses only; (2) blank or absent responses plus nonnumeric responses coded as zero; and (3) blank or absent responses plus all nonnumeric responses. Each definition was evaluated using three approaches: leaving missing responses without estimated values, replacing them with the average rating for that questionnaire item across patients, or replacing them with the patient's

average rating across answered items. This produced nine scenarios. For each scenario, we recalculated regulatory-system allocated severity shares, rankings, and mean item ratings. The original zero-coded analysis remained the primary analysis. Imputed values were estimates rather than observed symptom ratings. Detailed methods and results are provided in S12 File.

To compare symptom-node representation with patient-rated severity, each of the 171 nodes was given equal weight, divided equally across its assigned regulatory systems. This produced a fractional node share that could be compared with each system's allocated severity share. The difference between allocated severity share and node share was calculated, with uncertainty estimated using the same patient-level bootstrap. Two sensitivity analyses were reported alongside the primary result but were not used as the primary basis for interpretation: full-credit attribution, which credited each node's full severity to every assigned system and was therefore not severity-conserving; and a binary severe-symptom analysis, which treated a symptom as present at severity ≥7 and a node as present if any mapped symptom was rated ≥7. The ≥7 threshold was used only in this sensitivity analysis. Pre-specified quality-control reconciliation confirmed the source matrix shape, 39 patients by 233 rated symptom columns, the frozen 171 nodes and 17 systems, and the severity denominators before interpretation.

Three additional post hoc sensitivity analyses, prompted by structural review of the attribution method and questionnaire, are reported in Results: allocated severity shares recomputed with the seven merged questionnaire rows excluded; allocated severity shares recomputed with state-level regulation excluded from the system set, with shares recalculated across the remaining 16 systems; and an item-normalized secondary descriptive measure, described next. Analytic code and full numeric output for these three analyses are provided in S10 File.

Because total allocated severity share is influenced by the number of questionnaire items mapped to each system, we also computed an item-normalized secondary descriptive measure. For each regulatory system, contributing items were defined as all distinct questionnaire columns for which at least one mapped node was assigned to that system. The item-normalized mean was calculated as the average raw 0–10 rating across all patients and all contributing items for that system. Items mapped to multiple systems contributed to each relevant system's item-normalized mean; therefore, this measure is not an independent biological system score, but a descriptive estimate of mean severity among questionnaire items touching each system.

Robustness analyses

Five further analyses, added after the primary analysis and reported as robustness checks, examined whether the leading regulatory systems depended on specific analytic choices. Because the pooled allocated severity share depends only on each questionnaire column's cohort total, these analyses could be computed from the crosswalk (S2 Table) and patient-level allocation outputs (S8 Table).

(1) Umbrella and syndrome items: 27 of the 29 questionnaire columns that mapped only to non-node entries (for example, brain fog, cognitive impairment, dizziness, and three sensory-hypersensitivity items; 1,892 severity points) were assigned directly to the regulatory system(s) their names denote under the codebook, and shares were recomputed; two columns ("Left/Right-sided symptoms" and "Ticking") had no defensible assignment and remained excluded.

(2) Duplicate items: because several questionnaire items map to the same node (for example, “Anxiety” and “Nervousness”; “Anhedonia,” “Joylessness,” and “Dysphoria”), each questionnaire column’s cohort total was first divided equally among its mapped nodes. For each node, these fractionally attributed contributions were averaged across the contributing columns, and that mean was then divided equally among the node’s assigned systems. This reduces the influence of multiple items expressing the same node.

(3) Patient weighting: each patient’s regulatory-system shares were computed separately and averaged, so that every patient contributed equally; 95% CIs were obtained from 10,000 patient-level bootstrap resamples (seed 12345).

(4) Patient exclusion: pooled shares were recomputed after excluding every possible combination of three patients (9,139 combinations), including the combination of the three patients with substantial missingness.

(5) Blinded AI coding pass: pooled shares were recomputed using the regulatory-system assignments produced in the separate blinded AI audit pass (S7 File) in place of the author's final assignments, with all other inputs unchanged. The mean number of regulatory systems assigned per node and the number of nodes assigned to multiple systems were also compared between the two coding sets. Code and full output are provided in S11 File.

Results

Cohort context

Baseline characteristics are summarized in Table 2. Demographic data were available for 38 of 39 patients. Mean age was 48.9 years (SD 14.5), 52.6% were female, 73.7% were White, and 71.1% were not Hispanic or Latino. Among 37 patients with exposure data, the mean diazepam-equivalent dose was 27.0 mg/day; among 36 patients with duration-of-use data, the mean reported duration of benzodiazepine use was 9.0 years.

PROMIS-29 scores indicated elevated Anxiety (68.8), Depression (62.3), Sleep Disturbance (61.0), and Fatigue (60.7), with reduced Physical Function (45.5) and Social Roles/Participation (40.7), relative to population norms (T-score mean 50, SD 10). Pain Interference (51.5) was near the population mean. These measures provide descriptive cohort context and were not used to define patient subgroups.

Per-patient total symptom severity across all 233 questionnaire items varied widely (median 406 points; IQR 163.5–608.5; range 72–1,504). The five patients with the highest mapped severity contributed 30.0% of pooled mapped severity (15,515 points). As previously reported, patients endorsed a median of 74 symptoms and rated a median of 24 as severe (≥7/10) [15].

Item-level missingness was low for most patients (median 0.4% of the 233 questionnaire items blank or absent; IQR 0.0%–1.7%), but three patients left more than 60% of items unanswered (61.8%, 63.9%, and 69.1%). These three patients accounted for approximately 82% of the 554 blank or absent responses across the cohort.

Table 2. Cohort characteristics, benzodiazepine/Z-drug exposure, and PROMIS-29 scores (N = 39).

Characteristic	Value
Demographics (n = 38)	
Age, years — mean (SD)	48.9 (14.5)
Age, years — range	21.9–76.3
Female sex — n (%)	20 (52.6)
Male sex — n (%)	18 (47.4)
White race — %	73.7
Not Hispanic or Latino ethnicity — %	71.1
Single — %	44.7
Married — %	39.5
Employed at intake — %	50.0
Benzodiazepine/Z-drug exposure (n = 37)	
Index agent: alprazolam — %	29.7
Index agent: clonazepam — %	27.0
Index agent: lorazepam — %	21.6
Index agent: diazepam — %	18.9
Z-drugs — %	2.7
Diazepam-equivalent dose, mg/day — mean (SD)	27.0 (24.6)
Diazepam-equivalent dose, mg/day — median (IQR)	18.0 (10.0–40.0)
Diazepam-equivalent dose, mg/day — range	0.4–100
Duration of use, years (n = 36) — mean (SD)	9.0 (8.2)
Duration of use, years (n = 36) — median (IQR)	7.0 (2.8–14.0)
PROMIS-29 v2.0 T-scores — mean (SD) (n by domain)	
Anxiety (n = 38)	68.8 (7.8)
Depression (n = 38)	62.3 (9.7)
Sleep Disturbance (n = 36)	61.0 (8.5)
Fatigue (n = 38)	60.7 (10.6)
Pain Interference (n = 38)	51.5 (12.4)
Physical Function (n = 37)	45.5 (10.7)
Social Roles/Participation (n = 37)	40.7 (7.5)

Note: Race and ethnicity were each missing for 15.8% of patients; marital and employment status were each missing for 10.5%. Other employment categories were retired (18.4%), unemployed (15.8%), and student (5.3%). PROMIS-29 v2.0 T-scores are standardized to the US general population (mean 50, SD 10); higher scores indicate greater symptom severity, whereas lower scores indicate greater impairment in Physical Function and Social Roles/Participation. PROMIS summaries, including standard deviations, are also reported in the companion preprint [15], Table 3. All values are descriptive cohort characteristics and were not used to define patient subgroups.

Highest-rated individual symptoms

Before any node or regulatory-system mapping, anxiety had the highest mean questionnaire rating across the 39 patients (7.3 of 10), followed by insomnia (6.5), hypersensitivity to stress (5.5), nervousness (5.4), loss of confidence (5.4), and depression (5.2). Racing thoughts, agitation, exhaustion, brain fog, and feelings of shame each averaged 4.7–4.9 (Table 3). The mean rating across all 232 analyzed items was 1.9.

In the earlier analysis of this cohort, anxiety, insomnia, and hypersensitivity to stress were also the symptoms most frequently rated severe (≥7/10 severity rating; 27, 24, and 21 of 39 patients, respectively) [15].

Several questionnaire items describe closely related symptoms (for example, anxiety and nervousness; joylessness, anhedonia, and dysphoria). Brain fog, the tenth-highest-rated item, was classified as an umbrella term and therefore excluded from the primary severity allocation. Both features are examined in the robustness analyses below.

Table 3. The 15 highest-rated questionnaire items.

Mean rating is the cohort mean of the raw 0–10 rating (N = 39). Node and regulatory-system assignments are the author's final assignments (S2 Table); "—" indicates an item that mapped only to a non-node (umbrella) entry and was excluded from the primary allocation.

#	Questionnaire item	Mean rating (0–10)	Provisional node	Regulatory systems (author coding)
1	Anxiety	7.33	Anxiety/nervousness	Threat-arousal regulation; Interoceptive regulation; Cognitive control
2	Insomnia	6.46	Insomnia	Sleep-wake regulation; State-level regulation; Threat-arousal regulation
3	Hypersensitivity to stress	5.54	Hypersensitivity to stress	Threat-arousal regulation; State-level regulation; Autonomic regulation
4	Nervousness	5.41	Anxiety/nervousness	Threat-arousal regulation; Interoceptive regulation; Cognitive control
5	Loss of confidence	5.36	Loss of confidence	Cognitive control; Affective/reward regulation
6	Depression	5.18	Depression (mood state)	Affective/reward regulation; State-level regulation; Cognitive control
7	Racing Thoughts	4.85	Racing thoughts	Cognitive control; Threat-arousal regulation; State-level regulation
8	Agitation	4.82	Agitation / psychomotor-affective activation	Threat-arousal regulation; Motor regulation; Affective/reward regulation
9	Exhaustion	4.77	Exhaustion (fatigue/low energy)	State-level regulation; Sleep-wake regulation; Interoceptive regulation; Endocrine/metabolic/reproductive regulation
10	Brain Fog	4.72	—	—
11	Feeling of shame	4.72	Feeling of shame	Affective/reward regulation; Cognitive control; Threat-arousal regulation
12	Joylessness	4.59	Anhedonia / dysphoria / joylessness	Affective/reward regulation; State-level regulation
13	Loss or changes in appetite	4.59	Appetite change (increase or decrease)	Endocrine/metabolic/reproductive regulation; Interoceptive regulation; Enteric/GI regulation
14	Appetite - either loss of, or voracious appetite with constant desire to eat	4.38	Appetite change (increase or decrease)	Endocrine/metabolic/reproductive regulation; Interoceptive regulation; Enteric/GI regulation
15	Apathy	4.36	Apathy / loss of interest	Affective/reward regulation; Cognitive control; State-level regulation

Note: Unanswered questionnaire items were scored as zero (see Methods). Item wording is reproduced from the questionnaire (S1 Table).

Analysis A: Ontology construction and regulatory-system mapping

The 243 raw BIC withdrawal-symptom phrases yielded 278 analyzable rows after compound decomposition and resolved into 171 provisional phenomenological symptom nodes, alongside 41 retained non-node entries. The non-node entries comprised 14 umbrella/general categories, 9 syndrome-level entries, 3 unresolved-singleton items, 1 deferred undecomposed compound, and 14 compound-parent traceability markers (Table 4; S4 Table). These non-node entries were retained in the full ontology registry (S4 Table) but excluded from node-level counts and the primary patient-level severity allocation.

Table 4. Corpus reduction and ontology structure.

Stage	Count
Corpus reduction	
Original BIC withdrawal-symptom phrases	243
Analyzable rows after compound decomposition	278
Provisional phenomenological symptom nodes	171
Non-node entries (41 total)	
Umbrella/general categories	14
Syndrome-level entries	9
Unresolved-singleton entries	3
Deferred-compound entries	1
Compound-parent traceability markers	14

Note: The 171 provisional nodes derive from the 243-phrase source vocabulary, not from the 233-item patient questionnaire; see S3 File for the full provenance of these counts. The 41 non-node entries were retained for completeness but did not contribute to patient-level allocated-severity estimates.

Across the 171 provisional symptom nodes, the most represented regulatory systems were interoceptive regulation (64 nodes, 37.4%), sensory processing (47, 27.5%), autonomic regulation and state-level regulation (45 each, 26.3%), motor regulation (36, 21.1%), and threat-arousal regulation (32, 18.7%). Because nodes could map to more than one regulatory system, these counts are non-mutually exclusive (Figure 2).

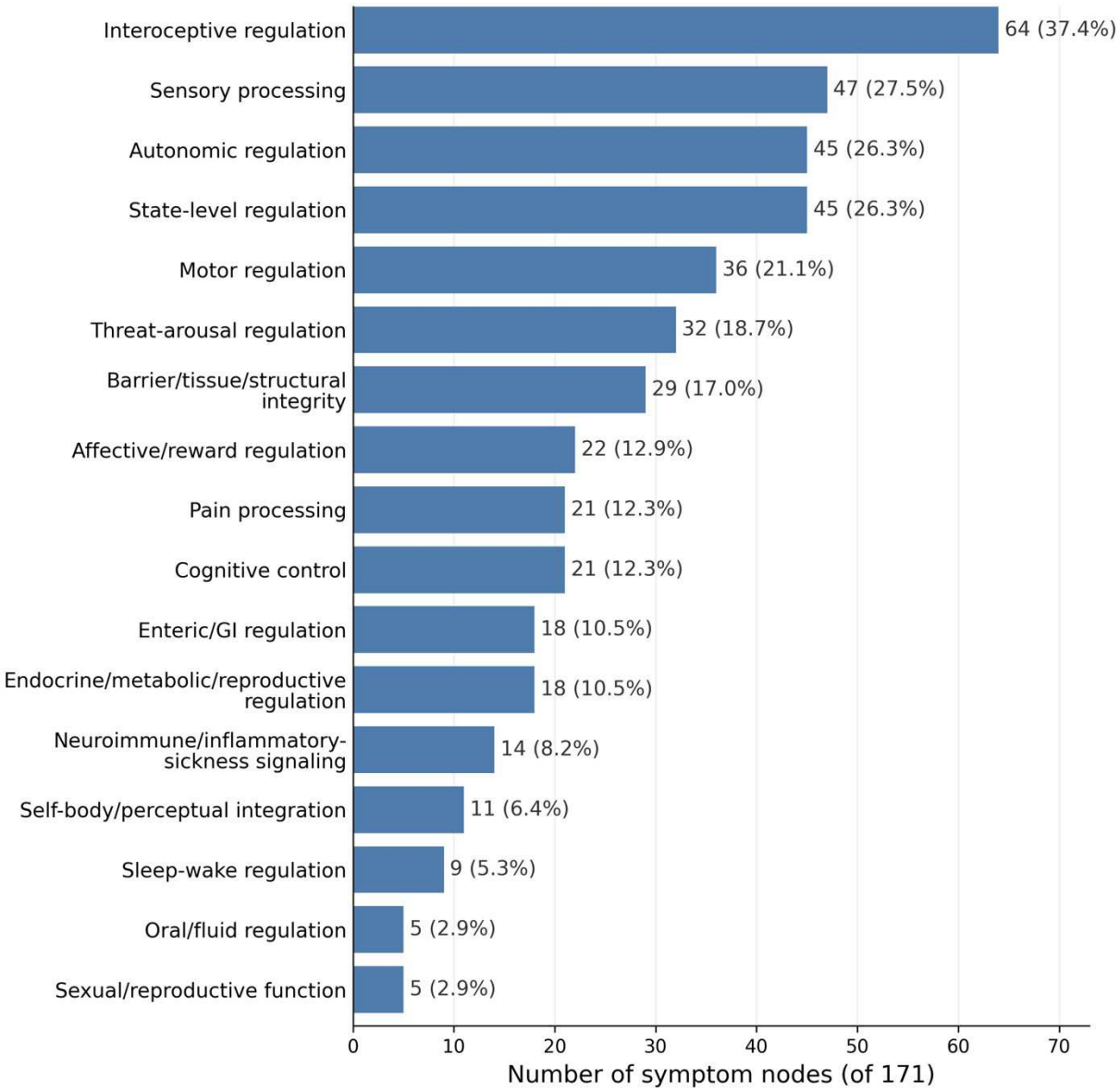


Figure 2. Ontology-level regulatory-system distribution across the 171 nodes.

Whole-node counts of the number of nodes mapped to each of the 17 regulatory systems. A node mapped to multiple regulatory systems contributes one count to each, so counts are non-mutually exclusive; percentages use the 171 provisional symptom nodes as the denominator. Values are ontology-level structural counts describing how the symptom vocabulary is organized, not patient prevalence or allocated severity.

Most symptom nodes mapped to more than one regulatory system. Of the 171 nodes, 157 (91.8%) mapped to two or more systems, including 52 nodes mapped to two systems (30.4%), 96 mapped to three systems (56.1%), and 9 mapped to four or more systems (5.3%); 14 nodes (8.2%) mapped to a single system. Three-system mapping was the most common pattern. The

sensitivity of multiple-system mapping to coding decisions is examined in the robustness analyses below.

The most frequent pairing of regulatory systems was autonomic regulation with interoceptive regulation (26 nodes). Other regulatory-system pairings were also observed, reflecting the multiple-system mapping of symptom concepts. These are ontology-level findings, not patient prevalence estimates or evidence that symptoms co-occurred within individual patients.

Procedural audit results

A blinded procedural audit was performed for both ontology steps using the written rules and withheld answer keys described in Methods. The audit pass used Claude Opus 4.8 and was followed by item-by-item adjudication by the author. For symptom-node formation, first-pass agreement with the original ontology was full for 66.7% of source phrases and partial for 33.3%, with no complete mismatches; after adjudication, the original assignment was retained for 95.9%, 4.1% were treated as both-valid tie-breaks, and no revisions were accepted into the ontology. For regulatory-system assignment, first-pass unordered-set agreement was full for 17.5% of nodes, partial for 80.7%, and absent for 1.8% (3 nodes with no shared regulatory system; mean Jaccard similarity 0.566); after adjudication, the original assignment was retained for 94.7%, 5.3% were treated as both-valid tie-breaks, and no revisions were accepted.

These findings describe procedural consistency and auditability, not human inter-rater reliability, independent validation, or biological validation, because AI tools were used both in the original taxonomy development and in this audit pass (see “Use of AI tools”). Aggregate counts and agreement statistics are provided in S6 Table, and item-level audit records in S7 File. Partial matches in regulatory-system assignment predominantly reflected the blinded pass assigning fewer systems per node (mean 1.5 versus 2.6 under the author’s coding); the consequences of this difference for severity allocation are reported under robustness analyses.

Analysis B: Patient-level severity-weighted regulatory-system allocation

Analysis B applied the frozen 171-node / 17-system ontology to the 39-patient severity matrix using raw 0–10 symptom-severity ratings and severity-conserving fractional attribution. Across all 233 questionnaire items, total patient-rated symptom severity was 17,491.0 points. Of this total, 15,515.0 points mapped to the 171 provisional symptom nodes and were carried forward into the primary regulatory-system severity-allocation analysis. Severity mapping only to retained non-node material accounted for 1,945.0 points; the dropped duplicate Paraesthesia.1 column accounted for 31.0 points, and no severity was unmatched after crosswalk reconciliation.

Severity-weighted allocation was distributed across all 17 regulatory systems rather than concentrated in a single system (Table 5, Figure 3; S8 Table). The largest allocated severity shares mapped to interoceptive regulation (14.3%; 95% CI, 13.5–15.1), state-level regulation (13.5%; 12.7–14.4), cognitive control (10.2%; 9.2–11.3), and threat-arousal regulation (10.0%; 9.3–10.9). Additional systems with substantial allocated severity included motor regulation (8.3%; 7.3–9.2), affective/reward regulation (8.2%; 7.4–9.3), sensory processing (8.1%; 6.9–9.2), and autonomic regulation (8.0%; 7.4–8.6). Each of the remaining systems received less than 5% of severity-weighted allocation.

Table 5 presents the primary allocated severity share alongside three alternative severity measures: full-credit share, binary severe-symptom share (≥7/10), and mean questionnaire-item rating (0–10). It also compares allocated severity with fractional node representation (the gap).

Table 5. Distribution of ontology nodes and patient-rated symptom severity across the 17 regulatory systems (Analysis B).

Regulatory systems are ordered by allocated severity share. Gap represents allocated severity share minus fractional node share, expressed in percentage points.

Regulatory System	Node fractional share %	Allocated severity share % (95% CI)	Gap, pp	Full-credit share %	Binary (≥7) share %	Mean item rating, 0–10
Interoceptive regulation	13.1	14.3 (13.5–15.1)	+1.2	15.8	12.0	1.93
State-level regulation	9.4	13.5 (12.7–14.4)	+4.1	13.9	14.2	2.51
Cognitive control	5.2	10.2 (9.2–11.3)	+5.0	8.9	11.2	3.23
Threat-arousal regulation	6.5	10.0 (9.3–10.9)	+3.6	10.9	11.2	2.79
Motor regulation	8.5	8.3 (7.3–9.2)	−0.3	7.6	6.7	1.66
Affective/reward regulation	4.8	8.2 (7.4–9.3)	+3.4	7.9	8.7	3.09
Sensory processing	13.0	8.1 (6.9–9.2)	−4.8	6.4	8.2	1.24
Autonomic regulation	9.3	8.0 (7.4–8.6)	−1.3	8.9	7.8	1.79
Pain processing	4.6	4.1 (3.6–4.5)	−0.6	4.1	3.6	1.68
Enteric/GI regulation	4.6	3.5 (2.9–4.1)	−1.1	3.4	2.8	1.63
Endocrine/metabolic/ reproductive regulation	4.0	3.2 (2.8–3.5)	−0.8	3.4	3.7	2.23
Barrier/tissue/structural integrity	8.0	2.6 (2.1–3.1)	−5.4	2.2	3.2	1.07
Neuroimmune/inflammatory-sickness signaling	3.2	2.0 (1.7–2.3)	−1.2	2.0	2.0	1.72
Sleep-wake regulation	1.7	1.6 (1.4–1.9)	0.0	2.0	2.1	2.59
Self-body/perceptual integration	2.2	1.2 (0.9–1.4)	−1.0	1.3	1.3	1.12
Oral/fluid regulation	1.0	0.6 (0.5–0.7)	−0.4	0.7	0.6	1.70
Sexual/reproductive function	0.9	0.5 (0.4–0.6)	−0.5	0.6	0.6	1.34

Note: Node fractional share describes the distribution of the 171 ontology nodes across regulatory systems, with each node contributing equally and its share divided among its assigned systems. Allocated severity share reflects the distribution of patient-rated symptom severity, using severity-conserving fractional attribution; each patient's mapped severity is counted only once. The gap compares these two measures: a positive value indicates that a system received a greater share of patient-rated severity than its fractional representation in the ontology, while a negative value indicates a smaller share. Gaps were computed from unrounded values, so they can differ by 0.1 from the difference between the rounded columns. The 95% confidence intervals were calculated from 10,000 patient-level bootstrap resamples (seed 12345), with the ontology, crosswalk, and allocation rules held fixed. Full-credit share credits each node's full severity to every regulatory system assigned to it, without conserving the total across systems. Binary (≥7) share uses whether mapped symptoms were rated severe (≥7/10) rather than their full severity ratings. Mean item rating (0–10) is a separate descriptive measure calculated by averaging raw questionnaire ratings across patients and distinct items mapped to each system; items mapped to multiple systems contribute to each relevant system's mean. Full patient-level and node-level allocation details are provided in S8 Table; the item-normalized calculation is provided in S10 File.

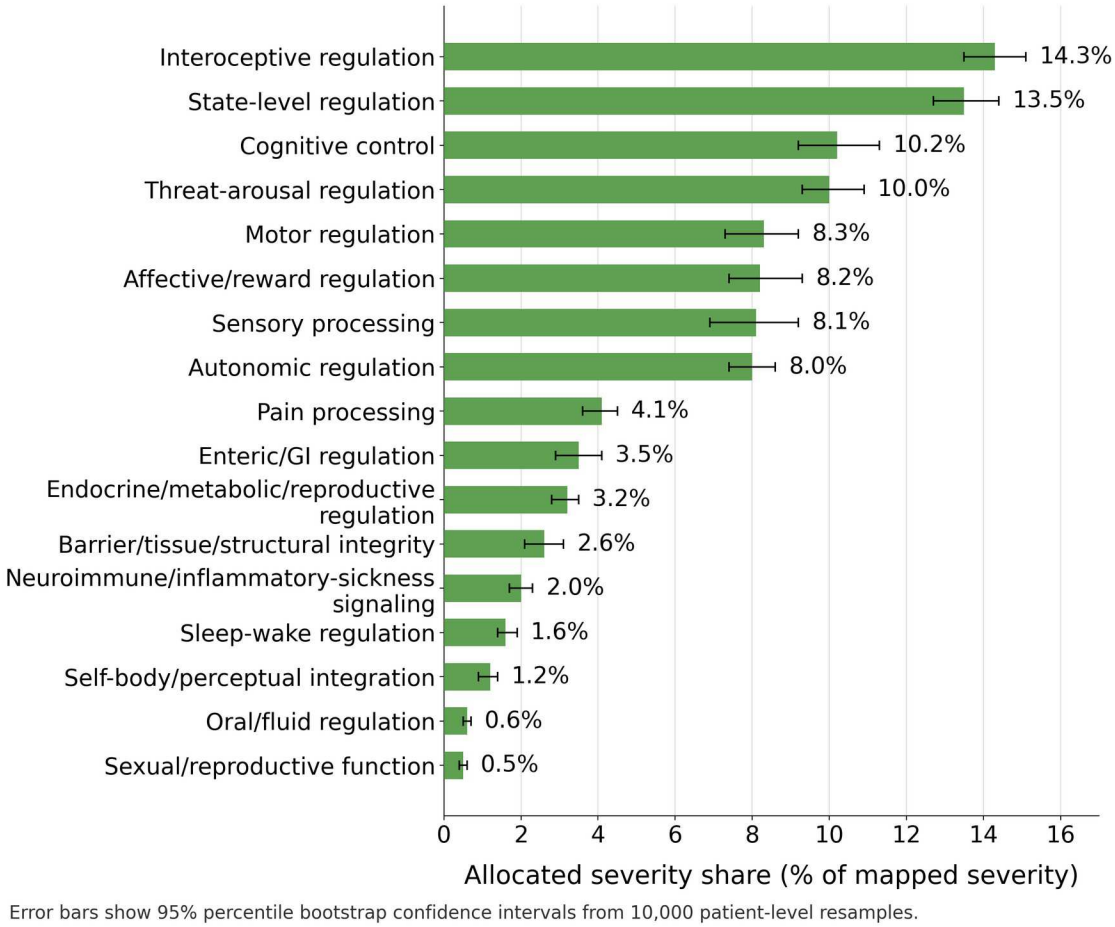


Figure 3. Patient-level severity-weighted regulatory-system allocation.

Allocated severity shares across the 17 regulatory systems, using severity-conserving fractional attribution; the denominator is total mapped severity (15,515.0 points across 171 nodes), not node counts. Error bars are 95% percentile bootstrap confidence intervals from 10,000 patient-level resamples (seed 12345). Values are patient-level and are distinct from the ontology-level node counts in Figure 2.

Allocated severity relative to ontology representation

The distribution of allocated severity did not simply mirror the distribution of symptom nodes in the ontology. Cognitive control received 10.2% of allocated severity despite representing only 5.2% of fractional node share, a difference of +5.0 percentage points. State-level regulation (+4.1 pp), threat-arousal regulation (+3.6 pp), and affective/reward regulation (+3.4 pp) also received greater shares of patient-rated severity than their fractional node representation. In contrast, barrier/tissue/structural integrity (−5.4 pp) and sensory processing (−4.8 pp) received substantially smaller severity shares relative to their representation in the ontology.

Motor regulation and sleep-wake regulation had the smallest gaps, with bootstrap intervals spanning zero (S8 Table). These comparisons show that the number and distribution of symptom concepts assigned to a regulatory system do not necessarily match the patient-rated severity attributed to it. The gaps are descriptive comparisons, not statistical tests or evidence that one regulatory system has greater biological importance than another.

Alternative severity measures

The two sensitivity analyses were broadly concordant with the primary severity-conserving analysis. Under full-credit attribution, the systems with the highest allocated severity were again interoceptive regulation (15.8%) and state-level regulation (13.9%), followed by threat-arousal regulation (10.9%), cognitive control (8.9%), autonomic regulation (8.9%), and affective/reward regulation (7.9%). Under the binary severe-symptom analysis, the leading systems were state-level regulation (14.2%), interoceptive regulation (12.0%), cognitive control (11.2%), threat-arousal regulation (11.2%), and affective/reward regulation (8.7%). Sexual/reproductive function, oral/fluid regulation, and self-body/perceptual integration remained among the lowest-allocation systems in both sensitivity analyses. Across the primary fractional, full-credit, and binary analyses, the same higher-allocation regulatory systems — interoceptive, state-level, cognitive-control, threat-arousal, and affective/reward regulation — recurred, suggesting that the main descriptive pattern was not driven solely by the attribution rule or by dichotomization. These sensitivity analyses are descriptive checks and do not validate the regulatory-system map or establish mechanism.

Three further sensitivity analyses examined specific structural concerns about the attribution method and the questionnaire. First, excluding the seven questionnaire rows that merge two distinct symptoms into a single 0–10 rating (658.0 of 15,515.0 mapped severity points, 4.24%) left the two leading systems unchanged (interoceptive regulation, 14.8%; state-level regulation, 13.3%); threat-arousal regulation (9.9%) and cognitive control (9.6%) exchanged third and fourth rank, and autonomic regulation rose two ranks, from eighth to sixth (8.0% to 8.2%); affective/reward regulation and sensory processing each fell one rank, and no other system changed rank by more than one position.

Second, excluding state-level regulation entirely and recomputing shares among the remaining 16 systems left interoceptive regulation as the leading system (16.0%), followed by cognitive control (13.0%), threat-arousal regulation (12.2%), and affective/reward regulation (11.1%) — three of which (cognitive control, threat-arousal regulation, and affective/reward regulation) are among the systems highlighted in the allocation-relative-to-representation contrast above.

Third, because systems differ in how many questionnaire items contribute to them (from 5 items for sexual/reproductive function and oral/fluid regulation to 88 for interoceptive regulation), we examined the item-normalized mean rating (Methods). These item counts do not equal the number of symptom nodes assigned to each system, because multiple questionnaire items could map to one node, some items mapped to multiple nodes, and not every ontology symptom was represented separately in the questionnaire (S2 Table; S3 File). Under this measure, cognitive control (mean item rating 3.23), affective/reward regulation (3.09), threat-arousal regulation (2.79), and sleep-wake regulation (2.59) ranked highest, while interoceptive regulation ranked seventh (1.93) despite having the largest total allocated severity share.

This indicates that interoceptive regulation’s leading total allocated severity share was partly explained by its large number of contributing questionnaire items, whereas cognitive control’s high ranking under both measures was not attributable to item count alone. The item-normalized measure describes mean severity among questionnaire items associated with a system, rather than its share of total allocated severity. It is not a correction to the primary analysis, and neither measure is inherently more correct.

Together, these three analyses indicate that the main allocation pattern was not driven by the seven questionnaire rows combining distinct symptoms or by the inclusion of state-level regulation as a separate category. They also show that total allocated severity share and mean item ratings can provide different perspectives on the same regulatory system, most notably interoceptive regulation.

Missing-data sensitivity analyses

Because some questionnaire responses were unanswered, we examined whether different ways of handling these missing ratings would change our findings. We tested nine scenarios (three definitions of missingness, each handled three ways; Methods), as detailed in S12 File.

The same four regulatory systems remained the highest-ranked across all nine scenarios, although their order sometimes changed. Treating blank or absent responses as missing, rather than as zero, did not change the original severity shares; extending the missingness definition to all nonnumeric responses changed shares by approximately 0.1 percentage points. Replacing missing ratings with the average rating for that symptom across patients changed the shares by approximately 0.2 percentage points at most. Using each patient's average symptom rating instead produced changes of up to approximately 1.0 percentage point. The rankings of systems in fifth through eighth place varied somewhat depending on how missing responses were handled.

Mean item ratings (Table 5) increased by 0.05 to 0.48 points across the sensitivity scenarios, while the ranking of the eight highest-rated regulatory systems remained unchanged.

Robustness to analytic choices and to coder

The leading regulatory systems were generally stable across the alternative analytic approaches shown in Table 6. The primary analysis was compared with three alternatives: including previously excluded umbrella/general and syndrome-level entries assigned to the regulatory systems indicated by their names; counting each symptom concept once when multiple questionnaire items mapped to it; and calculating each patient's percentage of severity allocated to each regulatory system and then averaging those percentages across patients, so that every patient contributed equally regardless of total symptom severity. Under all three approaches, interoceptive regulation, state-level regulation, cognitive control, and threat-arousal regulation remained the four leading systems, although their order varied.

Under equal patient weighting, their allocated severity shares were 14.1% (95% CI, 13.0–15.2), 14.0% (13.0–15.0), 11.3% (9.9–12.7), and 11.0% (9.8–12.3), respectively. In a separate analysis, excluding every possible combination of three patients (9,139 combinations) never changed which four systems ranked highest and changed no system's allocated severity share by more than 0.9 percentage points. These findings indicate that the leading systems were not strongly dependent on differences in patients' total severity or on the inclusion of any three individual patients.

Results were more sensitive to differences in regulatory-system assignments than to the analytic choices examined above. The separate blinded AI coding pass assigned fewer regulatory systems per node than the author's coding (mean 1.5 versus 2.6). Consequently, 70 of 171 nodes (40.9%)

were assigned to multiple systems in the blinded AI coding pass, compared with 157 nodes (91.8%) under the author's coding. The largest difference involved state-level regulation, which was assigned to 7 nodes by the blinded AI coding pass compared with 45 under the author's coding.

These differences in assignments substantially changed some severity-allocation rankings. Under the blinded AI assignments, state-level regulation fell from second (13.5%) to twelfth (2.9%), affective/reward regulation rose from sixth (8.2%) to first (15.8%), and motor regulation rose from fifth (8.3%) to second (11.3%). However, interoceptive regulation (11.2%), cognitive control (10.5%), and threat-arousal regulation (9.6%) remained among the five leading systems.

Across the primary analysis and four alternative allocation approaches (umbrella items mapped, duplicates collapsed, equal patient weighting, and the blinded AI coding pass), interoceptive regulation, cognitive control, and threat-arousal regulation ranked in the top five without exception, and affective/reward regulation ranked in the top six. By contrast, the prominence of state-level regulation and the high proportion of multi-system nodes were not reproduced under the separate blinded AI coding assignments.

Table 6. Robustness of regulatory-system severity allocation to analytic choices and coding assignments.

Values are allocated severity share, expressed as percentages, with rank among the 17 regulatory systems in parentheses. Shares are normalized within each analysis using fractional node-to-system attribution. Confidence intervals for the equal-patient-weighting analysis and complete numerical results are provided in S11 File.

Regulatory system	Primary	Umbrella items mapped	Duplicates collapsed	Equal patient weighting	Blinded AI coding pass	Maximum share change after excluding three patients (pp)
Interoceptive regulation	14.3 (1)	13.2 (1)	12.5 (2)	14.1 (1)	11.2 (3)	0.4
State-level regulation	13.5 (2)	13.0 (2)	13.0 (1)	14.0 (2)	2.9 (12)	0.5
Cognitive control	10.2 (3)	11.0 (4)	9.8 (4)	11.3 (3)	10.5 (4)	0.6
Threat-arousal regulation	10.0 (4)	11.0 (3)	11.1 (3)	11.0 (4)	9.6 (5)	0.4
Motor regulation	8.3 (5)	8.0 (7)	7.0 (8)	7.6 (6)	11.3 (2)	0.6
Affective/reward regulation	8.2 (6)	8.0 (6)	8.9 (5)	9.4 (5)	15.8 (1)	0.5
Sensory processing	8.1 (7)	8.9 (5)	7.7 (7)	6.7 (8)	6.0 (8)	0.9
Autonomic regulation	8.0 (8)	7.6 (8)	8.6 (6)	7.5 (7)	7.1 (7)	0.4
Pain processing	4.1 (9)	4.3 (9)	3.4 (11)	3.9 (9)	7.5 (6)	0.3
Enteric/GI regulation	3.5 (10)	3.3 (11)	3.3 (12)	3.3 (11)	3.1 (11)	0.4
Endocrine/metabolic/ reproductive regulation	3.2 (11)	3.4 (10)	3.4 (10)	3.4 (10)	4.3 (9)	0.2
Barrier/tissue/structural integrity	2.6 (12)	2.3 (12)	3.7 (9)	2.0 (13)	3.2 (10)	0.4
Neuroimmune/ inflammatory-sickness signaling	2.0 (13)	2.3 (13)	2.5 (13)	1.8 (14)	1.4 (16)	0.2
Sleep-wake regulation	1.6 (14)	1.5 (14)	2.3 (14)	2.0 (12)	2.4 (13)	0.2
Self-body/perceptual integration	1.2 (15)	1.1 (15)	1.3 (15)	0.9 (15)	1.4 (14)	0.2
Oral/fluid regulation	0.6 (16)	0.6 (17)	0.9 (16)	0.5 (16)	1.4 (15)	0.1
Sexual/reproductive function	0.5 (17)	0.6 (16)	0.7 (17)	0.5 (17)	0.9 (17)	0.1

Note: Primary (author's coding) refers to the original severity-conserving fractional allocation. Umbrella items mapped include previously excluded umbrella/general and syndrome-level questionnaire items assigned to the regulatory systems indicated by their names. Duplicates collapsed calculates each node's severity as the mean of the fractionally attributed contributions from its contributing questionnaire items, so that a symptom concept represented by multiple items does not receive additional weight solely because it appears more than once. Equal patient weighting calculates each patient's allocated severity percentages separately and then averages them, giving every patient equal influence regardless of total symptom severity. Blinded AI coding pass uses the alternative regulatory-system assignments generated during the separate blinded AI procedural audit (S7 File). Maximum share change after excluding three patients is the largest absolute percentage-point change observed across all 9,139 possible combinations of three-patient exclusions.

Discussion

Principal findings

This study organized a large, community-compiled vocabulary of benzodiazepine withdrawal-related symptoms into 171 provisional symptom nodes mapped to 17 broad regulatory systems, and applied that frozen map to the 0–10 symptom ratings of 39 patients presenting to a specialty taper practice. Three findings stand out. First, the highest-rated individual items involved anxiety, insomnia, hypersensitivity to stress, loss of confidence, depression, racing thoughts, and exhaustion. Second, when severity was allocated across regulatory systems, it concentrated most consistently in interoceptive regulation, cognitive control, and threat-arousal regulation, with affective/reward regulation close behind, and this pattern was robust to patient exclusions, the handling of umbrella and duplicate items, patient weighting, alternative handling of missing questionnaire responses, and a separate blinded AI coding pass. Third, two findings changed substantially depending on how symptom nodes were assigned to regulatory systems: the proportion of nodes assigned to more than one system (91.8% under the author's coding versus 40.9% under blinded AI coding) and the ranking of state-level regulation (second under the author's coding versus twelfth under blinded AI coding) (Table 6).

Symptom-node counts and allocated severity told different stories. Sensory processing and barrier/tissue/structural integrity contributed many symptom nodes but proportionally little severity, whereas cognitive control and affective/reward regulation carried more severity than their share of nodes and also had the highest mean item ratings. Counting how many kinds of symptoms fall in a regulatory system therefore says little about where patients' reported severity lies.

Multi-system structure and the role of coding

Symptom nodes were formed before any regulatory-system assignment, so the regulatory system map describes a fixed set of symptom concepts rather than having shaped them. Even so, how many regulatory systems each node is assigned to is a coding decision. The author's assignments — 2.6 systems per node on average — reflect a deliberately inclusive reading of each symptom: dizziness, for instance, was treated as involving interoceptive, autonomic, sensory, and state-level regulation. A blinded AI coding pass applying the same written codebook assigned 1.5 systems per node on average. Both coding approaches followed the written codebook, but their differences show that the assign/do-not-assign criteria leave room for judgment, particularly for state-level regulation, which the codebook itself flags as the regulatory system most at risk of over-use. That withdrawal-related symptoms often span more than one regulatory system is supported under both coding approaches—between 41% and 92% of nodes were assigned to multiple systems—but the specific proportion should not be treated as an inherent property of the symptoms.

This multi-regulatory system character is clinically recognizable: air hunger, dizziness, tremor, sensory hypersensitivity, depersonalization, fatigue, and cognitive impairment are difficult to place in a single body-system category, and a framework that allows overlap avoids forcing them into mutually exclusive categories.

Interpretation of the leading regulatory systems

Different measures of symptom severity produce somewhat different rankings of the regulatory systems. Total allocated severity share is influenced by the number of questionnaire items contributing to each regulatory system: interoceptive regulation, fed by 88 of 232 items, ranked first on total allocation but seventh on mean item rating. Mean item rating, by contrast, highlights regulatory systems whose contributing questionnaire items have higher average severity ratings: cognitive control, affective/reward regulation, threat-arousal regulation, and sleep-wake regulation. Cognitive control and threat-arousal regulation ranked highly across the severity measures examined and under both coding approaches, and affective/reward regulation ranked first under blinded AI coding and second on mean item rating.

Under this framework, the leading severity allocations involved regulatory systems associated with internal bodily sensations, cognition, threat-arousal, and mood or motivation. Cognitive-control symptoms may be especially important clinically because they affect concentration, comprehension, memory, decision-making, and communication, although this study did not measure regulatory-system-specific functional impact.

Comparison with prior work

The prominence of anxiety, sleep disturbance, low energy, and cognitive complaints is consistent with a large patient survey. In a secondary analysis of an internet survey of 1,207 current and former benzodiazepine users, 88.1% reported anxiety, nervousness, or fear; 86.9% sleep disturbances; 86.2% low energy; and 85.3% difficulty focusing or distractedness [1,4]. The survey describes how frequently symptoms were reported; the present analysis examines their severity in a separate clinical cohort, where anxiety, sleep disturbance, low energy, and cognitive complaints were also prominent. Standardized scales such as the CIWA-B and BWSQ [6,7] sample a narrower range of symptoms, with the BWSQ emphasizing perceptual and somatic complaints. The prominence of cognitive, affective, and arousal-related symptoms in this cohort underscores the value of assessing symptoms beyond those captured by brief withdrawal scales.

An earlier analysis of the same 39-patient clinical cohort [15] summarized each patient by a single dominant, mechanism-named axis. The present framework is more conservative and more auditable: it starts from the source vocabulary, allows symptoms to belong to more than one regulatory system, uses descriptive rather than mechanistic labels, documents its rules, and makes no patient-level classification. The two analyses describe a cohort in which arousal, sleep, and cognitive-affective symptoms carry the greatest reported burden, but the present results should not be read as validating the earlier axis labels.

Clinical implications

For clinicians and patients, one central difficulty in benzodiazepine withdrawal is not simply the presence of symptoms, but their number, variety, severity, and apparent incoherence. Patients may report dozens of symptoms involving mood, cognition, autonomic function, sensation, digestion, movement, sleep, immune-related processes, and other bodily functions. Symptoms that initially seem unrelated can be considered within a shared framework of overlapping regulatory systems, rather than as entirely separate complaints.

In clinical use, the regulatory systems are best treated as organizing lenses rather than diagnoses or proven mechanisms. They offer a way to discuss which regulatory systems may be most relevant to a patient's reported symptoms, without assuming that those symptoms share a common biological cause. This gives patients and clinicians a more coherent language for describing and discussing complex symptom experiences.

The severity findings add a more specific point. The distribution of reported severity did not simply follow the number of symptom nodes assigned to each regulatory system. In this cohort, three separate regulatory systems—cognitive control, affective/reward regulation, and threat-arousal regulation—received larger shares of allocated severity than their fractional representation among symptom nodes. The highest-rated individual complaints included anxiety and insomnia. This suggests asking patients not only how many symptoms they have but which ones weigh most heavily.

These implications concern how withdrawal-related symptoms are understood, discussed, and tracked, not what should be prescribed. The framework has not been tested prospectively for prediction, stratification, or treatment selection.

Procedural audit and transparency

The blinded AI coding audit and severity-allocation reanalysis provide complementary assessments of agreement with the author's original coding decisions and the sensitivity of the findings to alternative coding assignments. The blinded AI coder agreed with the author on 66.7% of node-formation decisions, but assigned exactly the same set of regulatory systems for only 17.5% of nodes (mean Jaccard similarity 0.566). Repeating the severity-allocation calculations using the blinded AI assignments showed that some regulatory-system rankings were relatively stable, while others depended substantially on coding decisions. Several regulatory systems, including interoceptive regulation, cognitive control, and threat-arousal regulation, remained prominent under both coding approaches, although their rankings changed. In contrast, state-level regulation ranked much lower under blinded AI coding, and the proportion of symptom nodes assigned to multiple regulatory systems decreased from 91.8% to 40.9%. Because the same author built and adjudicated the ontology, and AI tools supported both development and audit, neither the agreement statistics nor the adjudication outcome constitutes independent validation. The audit’s main value is to identify areas of greater agreement and disagreement, helping guide future independent human validation.

Strengths

This study has several strengths. It begins with patient-reported symptom language rather than predefined diagnostic categories. Symptom-node formation and regulatory-system assignment were performed separately and frozen before patient severity data were applied. The analysis uses full 0–10 symptom ratings with a severity-conserving allocation rule and distinguishes the number of symptom concepts from their allocated severity. Finally, the documented coding rules, assignments, and adjudication decisions allow others to examine and independently recode the ontology. The blinded AI coding audit illustrates both the feasibility and the importance of such scrutiny.

Limitations

Several limitations qualify these findings.

First, the study is descriptive, cross-sectional, and based on a small cohort of 39 patients drawn from a single specialty benzodiazepine taper practice. Symptom-severity ratings were collected at a single point in time, so the analysis cannot address how regulatory-system allocated severity changes over the course of tapering or recovery. Because records did not distinguish whether patients were on stable doses, self-tapering, or experiencing interdose symptoms at intake, ratings cannot be attributed specifically to withdrawal rather than to tolerance, interdose effects, or co-occurring conditions. PROMIS-29 scores indicated a clinically impaired cohort at baseline (Table 2), consistent with referral to a specialty taper practice; the findings should not be generalized to all benzodiazepine users or all patients tapering benzodiazepines, who may present with a milder or different symptom profile. The allocated-severity estimates should therefore be read as characterizing this cohort rather than the benzodiazepine-withdrawal population at large. No causal or mechanistic inference is warranted.

Second, the source data were based on patient report. The symptom vocabulary, symptom-severity ratings, medical histories, medication histories, and substance-use histories were not independently adjudicated. The 233-item questionnaire was developed for clinical documentation and has not undergone formal psychometric validation. In addition, the Benzodiazepine Information Coalition source list is a patient-advocacy vocabulary rather than an epidemiologic or validated research instrument. These features are appropriate for a patient-language ontology but limit claims about prevalence, diagnostic specificity, and generalizability.

Third, the regulatory-system layer depends on an author-defined vocabulary of 17 regulatory systems. This set is a deliberate simplification of an inherently continuous and interacting physiology; a different but equally defensible regulatory system set — coarser, finer, or drawn along different boundaries — could redistribute both the symptom-node counts and the severity-weighted allocation. The specific percentages reported are therefore conditional on this regulatory system framework and should not be treated as fixed properties of the symptoms themselves.

Fourth, four symptoms from the original vocabulary — burping, catatonia, muscle stiffness, and muscle rigidity — were inadvertently omitted when the source list was transcribed into the patient questionnaire, so no patient-level severity data were collected for them in this cohort. Three of the four are motor-related symptoms; as a result, the patient-level severity-weighted allocation to the motor regulatory system may be affected, although the magnitude and direction of any change cannot be determined. The ontology itself, however, retains these symptoms as concepts, so the omission affects the patient-level allocated-severity estimates but not the ontology-level structural description. These are documented transcription artifacts rather than design decisions.

Fifth, 29 questionnaire items mapped only to non-node entries (16 umbrella/general, 9 syndrome-level, 3 unresolved-singleton, and 1 deferred-compound; for example, brain fog, cognitive impairment, dizziness, and sensory hypersensitivity) and were excluded from the primary allocation, removing 1,945 severity points (11.1% of total). Assigning 27 of these items to regulatory systems based on their names did not change the leading systems but increased the

allocated shares of threat-arousal regulation, cognitive control, and sensory processing by 0.8–1.0 percentage points.

In addition, unanswered items and some nonnumeric responses were coded as zero in the original severity matrix, and three patients left more than 60% of the questionnaire unanswered. The same four regulatory systems remained highest-ranked across nine alternative missing-data scenarios, with allocated severity shares changing by no more than approximately 1.0 percentage point. The leading four systems were also unchanged after every possible three-patient exclusion. These findings indicate that the leading severity-allocation pattern was relatively stable under the alternative assumptions examined. However, they do not establish that missing-as-zero coding was unbiased. Unanswered items may have differed systematically by symptom or patient, and the actual severity of missing responses cannot be determined from the available data.

Sixth, the ontology is provisional and under refinement. Confidence in assigning symptom nodes to regulatory systems varied: of the 171 symptom nodes, 87 had high-confidence assignments, 79 medium-confidence assignments, and 5 low-confidence assignments. These ratings identify assignments that warrant further review and may guide future refinement.

Seventh, a single author performed the regulatory-system assignment, and a separate blinded AI coding pass applying the same codebook produced substantially fewer multi-system assignments and a markedly lower share for state-level regulation. The proportion of multi-system nodes and the rank of state-level regulation should therefore be regarded as coder-dependent. The blinded AI coding pass used a single AI rater rather than a panel of clinicians, and AI tools also supported development of the ontology (see “Use of AI tools”), so neither the author’s assignments nor the blinded AI coding pass establishes inter-rater reliability; blinded human re-rating is needed.

Eighth, two structural features of the attribution method and questionnaire warrant caution despite the sensitivity analyses reported in Results. Allocated severity share is computed as a sum of item-level severity across all questionnaire items contributing to a regulatory system, so systems represented by more items can receive larger shares of total allocated severity. The item-normalized analysis illustrated this difference: interoceptive regulation ranked first in total allocated severity but seventh in mean item rating, whereas cognitive control ranked highly under both measures.

This distinction also helps explain why sleep-wake regulation received only 1.6% of total allocated severity, even though the cohort reported elevated sleep disturbance on the separate PROMIS-29 measure. Only 8 of the 232 questionnaire items contributed to sleep-wake regulation, and its mean item rating (2.59 out of 10) ranked fourth among the 17 systems, above interoceptive regulation (1.93). Thus, its low total allocated severity share may partly reflect limited questionnaire representation and the allocation rules rather than low severity among sleep-related items.

Seven questionnaire rows (4.24% of mapped severity) combined two distinct symptoms into a single 0–10 rating. Excluding these rows left the main findings materially unchanged, but their combined ratings limit symptom-level precision.

Separately, state-level regulation is explicitly flagged in the regulatory-system codebook as the regulatory system most at risk of over-use during coding, and it received the second-largest allocated severity share in the primary analysis but the twelfth-largest under blinded AI coding. Excluding it, or using the blinded AI assignments, did not change the identity of the other leading systems, but its allocated severity share should be interpreted cautiously.

Finally, this analysis did not include a control group, biomarkers, autonomic testing, neuroimaging, immunologic testing, or independent clinical outcomes. The symptom vocabulary, severity ratings, and resulting allocated-severity estimates derive from self-report, so any alignment among them reflects internal consistency rather than corroboration by an external or biological measure. The study also does not test whether regulatory-system severity allocation predicts taper tolerability, symptom trajectory, functional recovery, or response to clinical interventions. These questions require prospective validation in larger and more diverse cohorts.

Future directions

Several directions follow from this work. The most immediate is independent human validation: both symptom-node formation and regulatory-system assignments should be evaluated by independent clinical raters to estimate inter-rater reliability and identify where the codebook requires clarification, particularly for multi-system assignments. The ontology would also benefit from refinement of lower-confidence assignments and from restoring the omitted symptoms in a corrected questionnaire, so that the motor regulatory system in particular is more completely represented. A blinded second-rater protocol and coding packets for both ontology steps have already been prepared (S7 File) and can be used directly for this purpose.

Beyond the ontology itself, prospective and longitudinal studies in larger, more diverse cohorts are needed to determine whether regulatory-system severity allocation shifts across the course of tapering or recovery, and whether allocation profiles are associated with clinically meaningful outcomes such as taper tolerability, symptom trajectory, or functional recovery. Testing clinical predictions would require prospective outcome measurements, while investigating underlying mechanisms would also benefit from independent physiological, autonomic, immunologic, or neuroimaging measures to evaluate the proposed regulatory-system framework.

Finally, because the item-to-node-to-system crosswalk is explicit, the same approach could be applied to symptom data from other conditions in which patients report complex, multi-system experiences, allowing the framework to be tested, revised, or contested outside benzodiazepine withdrawal.

Conclusion

A large and heterogeneous vocabulary of benzodiazepine withdrawal-related symptoms was organized into 171 provisional symptom nodes mapped to 17 broad regulatory systems. Applying this framework to patients presenting to a specialty taper practice, we allocated their symptom-severity ratings across those regulatory systems in a transparent, auditable way.

The highest-rated individual symptoms were anxiety, insomnia, hypersensitivity to stress, nervousness, loss of confidence, and depression. Allocated severity concentrated most

consistently in interoceptive regulation, cognitive control, and threat-arousal regulation, which remained among the five leading systems across the tested allocation analyses and the author and blinded AI assignments; affective/reward regulation also remained prominent. The proportion of symptom nodes assigned to multiple regulatory systems and the prominence of state-level regulation, by contrast, varied depending on how symptom nodes were assigned to regulatory systems.

These findings are descriptive and hypothesis-generating: they do not identify patient subtypes, establish mechanism, or demonstrate biological validity. Their contribution is a transparent, auditable framework for organizing complex symptom experiences in ways that may help patients and clinicians describe and discuss them. The framework also provides a foundation for independent human validation and replication.

Acknowledgements

The author thanks the patients whose de-identified questionnaire responses formed the basis of the patient-level analysis, and Khristine Anne Garcia for assistance with patient care and data organization. The author thanks the Benzodiazepine Information Coalition (Benzoinfo.com) for making available the withdrawal-symptom list from which the symptom corpus was derived.

Funding

The author received no specific funding for this work.

Competing interests

The author is the founder and clinical director of Brain Body Medical (The Benzo Taper Doctor), the practice from which this cohort was drawn.

Author contributions

V.S.M. is the sole author and is responsible for all aspects of this work: conceptualization, study design, data acquisition, formal analysis, interpretation, drafting, critical revision, and final approval of the manuscript. AI tools assisted with manuscript drafting and editing, development and refinement of coding rules, code drafting, and consistency checks, as described in “Use of AI tools”; the author reviewed these contributions and made the final scientific and interpretive decisions. The author independently reviewed the source data, verified the reported numeric results against analytic outputs, adjudicated all ontology and regulatory-system decisions, and takes responsibility for the accuracy and integrity of the work.

Use of AI tools

The author (V.S.M.) designed the study, established the symptom-node and regulatory-system framework, and made the final decisions on node formation, regulatory-system assignment, analytic methods, interpretation, and conclusions. ChatGPT and Claude were used as assistants for: (1) drafting and editing manuscript text, which the author reviewed and revised; (2) drafting and refining written coding rules, which were reviewed and finalized by the author; (3) drafting analysis code, which the author reviewed and checked against outputs; and (4) consistency checks and preparation of tables, figures, and working materials.

The author reviewed the source data and analytic outputs, adjudicated all ontology and assignment decisions, and takes responsibility for the accuracy and integrity of the work. The author reviewed and approved the scientific interpretations and conclusions included in the final manuscript.

A separate blinded coding pass using Claude Opus 4.8 was performed from written rules and withheld answer keys to document procedural consistency and points of ambiguity. Because AI tools were used during development as well as audit, this pass is reported as a transparency and procedural-consistency check, not as independent validation, human inter-rater reliability, or biological validation. The robustness and missing-data sensitivity analyses, including the

reanalysis using the blinded AI coding pass, were coded with AI assistance; the author reviewed the code and outputs.

Ethics statement

This study was determined to be exempt from IRB review by the Institutional Review Board of the Biomedical Research Alliance of New York (BRANY) on 3 October 2025, under 45 CFR 46.104(d) category (4)(ii) (BRANY IRB File #25-12-541-2328; determination #262681). A full waiver of HIPAA authorization was granted under 45 CFR 164.512(i)(2)(ii). All analyses were performed on de-identified data files. Individual informed consent was not required for this exempt secondary analysis of existing clinical data.

Data availability

De-identified patient data underlying this study are available from the corresponding author on reasonable request. Supporting materials, including the ontology, crosswalk, codebook, analytic outputs, and supplementary analyses, are publicly available on Zenodo (DOI: 10.5281/zenodo.23510456).

References

1. Finlayson A, Macoubrie J, Huff C, Foster D, Martin P. Experiences with benzodiazepine use, tapering, and discontinuation: an internet survey. Ther Adv Psychopharmacol. 2022;12:20451253221082386.
2. Huff C, Finlayson A, Foster D, Martin P. Enduring neurological sequelae of benzodiazepine use: an internet survey. Ther Adv Psychopharmacol. 2023;13:20451253221145561.
3. Benzodiazepine Information Coalition. Withdrawal and post-withdrawal symptoms list [Internet]. Benzoinfo.com [accessed 2026 Aug 1].
4. Ritvo AD, Foster DE, Huff C, Finlayson AJR, Silvernail B, Martin PR. Long-term consequences of benzodiazepine-induced neurological dysfunction: a survey. PLoS One. 2023;18(6):e0285584.
5. Ashton H. Benzodiazepines: how they work and how to withdraw. Newcastle upon Tyne, UK: Newcastle University; 2002.
6. Busto UE, Sykora K, Sellers EM. A clinical scale to assess benzodiazepine withdrawal. J Clin Psychopharmacol. 1989;9(6):412–416.
7. Tyrer P, Murphy S, Riley P. The Benzodiazepine Withdrawal Symptom Questionnaire. J Affect Disord. 1990;19(1):53–61.
8. Shade KN, Ritvo AD, Silvernail B, et al. Long-term neurological consequences following benzodiazepine exposure: A scoping review. PLoS One. 2025;20(8):e0330277.
9. Brunner E, Chen C, Klein T, Maust D, Mazer-Amirshahi M, Mecca M, et al. Joint Clinical Practice Guideline on Benzodiazepine Tapering: Considerations When Risks Outweigh Benefits. J Gen Intern Med. 2025;40(12):2814–2859.
10. Tan KR, Rudolph U, Lüscher C. Hooked on benzodiazepines: GABAA receptor subtypes and addiction. Trends Neurosci. 2011;34(4):188–197.
11. Vinkers CH, Olivier B. Mechanisms underlying tolerance after long-term benzodiazepine use: a future for subtype-selective GABAA receptor modulators? Adv Pharmacol Sci. 2012;2012:416864.
12. Lorenz-Guertin JM, Jacob TC. GABA type A receptor trafficking and the architecture of synaptic inhibition. Dev Neurobiol. 2018;78(3):238–270.
13. Navarrete F, Marín-Mayor M, Martínez-Hostyn L, Rubio G, Manzanares J. Benzodiazepine Dependence: Clinical and Molecular Aspects, Preventive Strategies and Therapeutic Approaches. Int J Mol Sci. 2026;27(3):1430. doi:10.3390/ijms27031430.
14. Piot A, Jovanovic JN. Synaptic correlates of benzodiazepine tolerance. Mol Psychiatry. 2026;31:3775–3785. doi:10.1038/s41380-026-03512-8.
15. Madhava VS. Multisystem Symptom Burden and Functional Impairment in a Specialty Benzodiazepine Taper Cohort. medRxiv. 2025. doi:10.1101/2025.10.07.25336923.
16. HealthMeasures. PROMIS Adult Profile Instruments: Scoring Manual [Internet]. HealthMeasures; 2020 Feb 13.
17. Craig AD. How do you feel? Interoception: the sense of the physiological condition of the body. Nat Rev Neurosci. 2002;3(8):655–666.
18. Benarroch EE. The central autonomic network: functional organization, dysfunction, and perspective. Mayo Clin Proc. 1993;68(10):988–1001.
19. Woolf CJ. Central sensitization: implications for the diagnosis and treatment of pain. Pain. 2011;152(3 Suppl):S2–S15.

Figure legends

Figure 1. Study workflow. Schematic of the two linked analyses. The Benzodiazepine Information Coalition withdrawal-symptom list was used to construct a phenomenology-first ontology: 243 source phrases were decomposed into 278 analyzable rows and organized into 212 registry entries, including 171 provisional symptom nodes and 41 non-node entries. The 171 nodes were then mapped to 17 regulatory systems. The frozen 171-node / 17-regulatory-system map was applied to the 39 patients' original 0–10 questionnaire ratings to estimate severity-weighted regulatory-system allocation. Ontology-level node counts and patient-level allocated severity are shown as separate outputs.

Figure 2. Ontology-level regulatory-system distribution across the 171 nodes. Whole-node counts of the number of nodes mapped to each of the 17 regulatory systems. A node mapped to multiple regulatory systems contributes one count to each, so counts are non-mutually exclusive; percentages use the 171 provisional symptom nodes as the denominator. Values are ontology-level structural counts describing how the symptom vocabulary is organized, not patient prevalence or allocated severity.

Figure 3. Patient-level severity-weighted regulatory-system allocation. Allocated severity shares across the 17 regulatory systems, using severity-conserving fractional attribution; the denominator is total mapped severity (15,515.0 points across 171 nodes), not node counts. Error bars are 95% percentile bootstrap confidence intervals from 10,000 patient-level resamples (seed 12345). Values are patient-level and are distinct from the ontology-level node counts in Figure 2.

Supporting information

The complete supplementary package (S1–S12), including the ontology crosswalk, procedural audit materials, analytic code, and analytic outputs, is available through Zenodo (DOI: 10.5281/zenodo.23510456). The Zenodo package includes a README describing the file inventory and how to run the scripts.

S1 Table. The 233-item withdrawal-symptom questionnaire patients rated from 0 to 10 at intake.

S2 Table. Frozen item-to-node-to-system crosswalk linking each of the 233 questionnaire columns to its provisional ontology node(s) and regulatory system(s).

S3 File. Provenance note documenting the 243 → 233 → 171 corpus-reduction and questionnaire-transcription relationships, including the 7 merged rows, 4 omitted symptoms, and duplicate Paraesthesia column.

S4 Table. Full ontology registry: the 171 provisional symptom nodes with regulatory-system assignments and confidence ratings, plus the 41 retained non-node entries.

S5 File. Full operational codebook for the 17 regulatory systems, including assign/do-not-assign criteria and borderline-case guidance.

S6 Table. Summary results of the procedural audit for both ontology steps (node formation and regulatory-system assignment).

S7 File. Detailed audit materials, including rater packets, scored outputs, and adjudication logs underlying S6 Table.

S8 Table. Detailed patient-level allocated-severity results for Analysis B, including the primary and sensitivity analyses, node share, gap, and 95% confidence intervals.

S9 File. Analytic code for the patient-level severity-allocation analysis (Analysis B).

S10 File. Analytic code and full numeric output for the three post hoc sensitivity analyses (merged-rows-excluded, state-level-regulation-excluded, and item-normalized severity).

S11 File. Analytic code and full numeric output for the robustness analyses (umbrella-item mapping, duplicate-item collapse, equal patient weighting, leave-any-three-out, and blinded AI coding reanalysis), together with the umbrella-mapping assignment table and the missingness-provenance note documenting these analyses.

S12 File. Analytic code, methods note, item-level counts of unanswered responses, and full numeric output for the missing-data sensitivity analyses (blank responses treated as missing, item-mean imputation, and patient-mean imputation). Patient-mean imputation and per-patient missingness summaries require restricted patient-level data; their results are in the output file.

STROBE checklist. Completed STROBE reporting checklist for observational studies.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S1TableQuestionnaire233Items.xlsx
- S4TableOntologyRegistry171nodes41nonnode.xlsx
- STROBEChecklist.docx
- S5FileRegulatorySystemCodebook.docx