

A health-system-scale, episode-resolved multimodal atlas of cancer cachexia

Corresponding Author: Dr Ed Reznik

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

A version of this paper was originally rejected for publication by Nature Medicine, however that decision was reconsidered after appeal by the authors.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The data from MSK-IMPACT might be useful for analysis related to cachexia. A clear alignment with consensus-based conceptual framework and revised analysis plan clearly needs to be built from the ground up. It may be helpful to reach out to an expert on cancer cachexia diagnostic criteria.

To the question of Novelty and Significance.

The analysis of association of cachexia with tumor mutation is potentially novel, with the caveat that the classification of cachexia prevalence and severity needs to be completely redone. Therefore the association with tumor mutation must be redone.

The association of cachexia with routine blood work values is not novel. The researchers did not provide a clear justification for selecting standard of care blood work as the source data for cachexia biomarkers; the manuscript does not contain or refer to a literature review of the cachexia biomarkers space. For example, by 2008, classification systems for cachexia (PMID 18718696) were based on well-established association increased inflammatory markers (CRP, IL-6), Anemia (Hb <12 g/dL) and low serum albumin (<3.2 g/dL. According to Google Scholar this work has been cited 3196 times. In the meantime, there have been >50 publications attesting to the association of the Glasgow Prognostic Score, which includes abnormal albumin and CRP values, with cancer cachexia. Likewise, the association of neutrophil to lymphocyte ratio with cancer cachexia is covered in about 40 publications. As for KRAS mutation association with cancer cachexia, that also has been shown already, please see PMID 30932365. Also, molecular crosstalk mechanisms between skeletal muscle atrophy and KRAS-mutant pancreatic cancer, has been the subject of a literature review PMID: 39925900 and this is the reason that KRAS mutant pancreatic cancer cell lines are used in preclinical models (PMID: 28730707).

A. Does not meet minimum reporting requirements for a clinical study.

1. The data was collected with an unrelated purpose (mutational profiling), and this analysis was not planned a priori (i.e. not as part of the secondary analysis of the MSK – IMPACT project). Not a clinical trial in the conventional sense, participants had to have a cancer diagnosis, agree to mutational profiling and be seen at the Memorial Sloan Kettering cancer center. It will be necessary to provide a clear description of the cohort considered for the cachexia – related analysis and the criteria used to select them; a complete set of descriptive should be provided; only primary cancer is given. The only provided population characteristics is the primary disease. Looking to deduce some of this from the provided data set, immediately note that there are (for example) 138 lines of data noted with an age of 0 years (<1 y?), in total there are 1685 lines of data on patients under the age of 18 years at diagnosis. This raises questions of intent to study cachexia during a time normally characterized by longitudinal growth and the applicability of the criteria for defining cachexia in children (none exist). This suggests that the resource (>2.5 million lines of weight data) was not reviewed in relation to its suitability for a cachexia - related inquiry. (see next point)

2. Insufficient information is provided regarding the provenance of the data to be used for cachexia study and its suitability for a cachexia inquiry. Many questions are raised:

a) Did or does the MSK-IMPACT study have planned weight measurements (this is implied at one point), or were the weight

records pulled from one or more sources? MSK-IMPACT was started in January 2014, however in the data set we see cachexia episodes that predate this considerably, with the latest recorded "cachexia episode" being over 53 years prior to the date of diagnosis (so weight history must go back to 1972 or earlier).

b) From the present to 53 years ago, what data systems were accessed to obtain the weight information? By whom and when was it typically recorded (physical exam, chemotherapy dosing, nutrition/dietetic consultation, other?), For those years did MSK have an SOP for weighing patients? What weight scales and calibration SOPs were in place? This gets to a key point – what is the best estimate of measurement error which bears on the ability to detect a 5% difference. This kind of analysis predicated on a rather small percentage weight loss as the defining criteria used to make patient classification, would require a very clear representation of what constitutes a measurement error in the available data. There is some vague language about "smoothing" out the day-to-day variation, but this is neither clear nor sufficient to inform the reader as to the probability of detecting a relatively small change.

c) The extent coverage of patient's clinical trajectory with weight data needs to be assessed and some criteria set for adequacy of coverage. In the very long term follow up (see above), many patients would be seen only annually or even less frequently, so if weight changes by 5% between year 10 and year 12 of follow up, was this coded as a cachexia episode of 2 years duration? It is said that patients with just 2 weight values of record were included, however there needs to be a justification for including patients with 2 or few weight records, especially if their clinical record is missing weight information over long periods of time. It is necessary and essential to benchmark of the weight records relative to date of death (or censoring), not just to date of diagnosis. Absent this information, it is impossible to understand the definition of the "end" of a cachexia episode – was it defined merely by the last available weight record, death, or by stabilization or regain of lost body weight?

3. The analytic approach, as well as the diagnostic criteria for cachexia, are not based on the sound conceptual framework.

a) The ~57,000 patients included many individuals whose inclusion is questionable:

i) infants and children,

ii) persons reasonably considered cured of their cancer with long term survival > 10 years, and whose weight change, if any, reflected short term, reversible temporal bouts of weight loss likely to be driven by cancer therapy. Episodes of weight loss >10 years after primary cancer diagnosis are likely not even attributable to malignant disease. A clear part of the existing conceptual framework is that patients cured of cancer might sustain weight loss of short term associated with side effects of therapy, but unlike cachexia, those weight loss episodes are spontaneously reversible and have no bearing on the patients long-term outcome. In this analysis thousands of cachexia episodes that have been recorded at times > 10 years since diagnosis, and I would wonder why these would be included in the analysis since at 10 y or indeed, even 5 y out, some patients would be considered cured of their cancer. It makes no sense to record as cancer cachexia episodes of weight loss that would likely to be coming from non-malignant causes and at these far distant time points it is necessary to know whether they related to recurrence, a second malignancy, or a non-malignant acute or chronic condition. Consider separate evaluation of patients diagnosed at early stage of disease who did not show recurrence. Also, the MSK data set is very likely to include dates of systemic therapy, and it would be of great value to identify if therapies align with weight loss and the extent to which therapy-related weight loss is regained, especially in patients treated with curative intent.

iii) obese persons for whom 5% weight loss is documented to have minimal clinical impact. I address this point further in context of discussion of cachexia severity

b) The authors have not used either up-to-date or consensus criteria for the diagnosis, classification and severity of cachexia. They cite several key papers by the international consensus group (Fearon et al 2011 Lancet Oncology; Martin et al 2015 J Clin Oncol) but claiming the use of "simplification" revert to a crude historical definition of any weight loss >5%.

According to Fearon et al. 2011: the agreed diagnostic criteria:

- Weight loss >5% over past 6 months (in absence of simple starvation); or
- BMI <20 kg/m² and any degree of weight loss >2%; or
- skeletal muscle index consistent with sarcopenia and any degree of weight loss >2%;

The authors chose to "simplify" i.e. ignore 2 of these 3 criteria. As such, it is misleading to state that this work included cachexia criteria that "satisfy" international consensus criteria, since they were not used.

Furthermore, a key consensus statement was that severity can be classified according to the rate of ongoing loss of weight in combination with the concurrent degree of depletion of energy stores and body protein mass (which can be compounded by a low initial reserve). This tenet was formalized in the consensus group's 2015 analysis (Martin et al. 2015) with 12,000 patients to derive the grades of severity of cancer-associated weight loss. It does seem disconnected with current classification of cachexia, to sidestep the notion of severity. The weight loss grade schema makes clear the limitation of using a 5% across the board cutoff. For patients of BMI <20 kg/m² at the outset, a weight loss of 5% is Grade 3, whereas for a person of initial BMI 30, it is only Grade 1. As such, the current analysis is likely to contain large numbers of patients of high initial BMI who are miscoded as having had "cachexia episodes". The number of obese people in his cohort is not disclosed, but one can presume to be substantial after reviewing Supplemental Fig 2.

The present analysis reverts to pre-consensus criteria, which focused on an arbitrary minimum degree of weight loss (5%) and does not provide a meaningful integration of severity in the analysis. The analysis provided is the prevalence of >5% incident weight loss (primary outcome); this would be acceptable language to describe what was done.

4. The data set provided with the submission is entirely insufficient to recapitulate any of the presented analysis.

a) This data set appears to align with the results section in so far as it contains 59,493 unique individuals, and the total count of cachexia episodes is 37,545. However, the data set has no legend. It contains no relevant dates, no indication of the duration of follow-up for each patient, and indeed no weight data are included. The only information is minimal and is demographic (sex and age and primary tumour site (but not stage)); the start and end dates of "cachexia episodes" are the only available data and these unfortunately are not associated with any information as to the number of weight data per

patient, the duration of follow up and time from last weight record to death or censoring. The definition of “end date” is unclear – in some instances it might mean that weight was regained, in others either the patient died or the weight record was discontinued.

b) Certain data are very difficult to understand – some examples which make me wonder whether these data were carefully curated and examined.:

Patient P – 000-4032 a 7 year old male with anaplastic ganglioma had an episode of >5% weight loss 15,369 days after his diagnosis, some 42 years later. Patients with pediatric cancers do have very long-term surveillance but is this cachexia? Patient P0066131, a male infant with sarcoma 0 years of age, had an episode of >5% weight loss 3 days duration during his first year of life. is this cachexia?

P – 001-4433 a woman with gastrointestinal stromal tumour had an episode of 5% weight loss of 10 days duration 10.8 years after their diagnosis. A total of 2125 cachexia episodes occurred 3650 days (10 years) after diagnosis or later. Are these episodes related to malignant disease? Are these people being followed up annually at 10 years out and beyond, and was a cachexia episode be defined by 2 weight records a year apart?

Columns H and I defined the start day and the end day of each diagnosis of incident 5% weight loss. There is no explicit description in the method section of Day 0. However, time, since the date of diagnosis seems to be the metric in all the data presentations, so I've taken this to mean time from date of diagnosis. Then.... 907 lines of data have a start day of an episode on day zero. This is very difficult to understand. This could mean one of two things did the person get weighed twice on their date of diagnosis and one of those weights with 5% lower than the earlier one? Alternatively, are there weight records for individuals that precede with their date of diagnosis? Likewise, there are about 2000 low single digit start dates in column H some clarification is required as to where these came from. As mentioned above, the lack of explicit definition of "end date" of a cachexia episode detracts greatly from the clarity of this presentation.

(Remarks on code availability)

This code is unevaluable, as the data provided cannot be used to test it.

See comments to authors where I make the case that at least some of the data should have been excluded, that the classification criteria for cachexia and for episodes are unclear/incorrect.

Reviewer #2

(Remarks to the Author)

Castillon et al. present the manuscript “A health-system-scale, episode-resolved multimodal atlas of cancer cachexia.” Using longitudinal clinico-genomic data from 59,493 individuals with cancer, the authors integrate biomarker, genetic, and clinical time-series data to identify signals associated with cachexia across a broad range of malignancies, including 656 distinct cancer histologies. Overall, the study identifies serological and genetic correlates of cachexia, some of which appear conserved across multiple cancer types, suggesting shared biological mechanisms. The work addresses an important and unresolved question in cancer metabolism and leverages an unusually large real-world dataset. However, the terminology “cachectic episodes” is potentially misleading and requires clarification. The phrasing implies reversibility and cyclicity of cachexia, which stands in contrast with its current conceptualization as a progressive clinical syndrome characterized by muscle loss, systemic inflammation, metabolic dysregulation, and declining function. If the authors intend to propose that aspects of cachexia may be modifiable in specific clinical contexts, this point should be explicitly justified and discussed, as it challenges long-standing assumptions in the field.

Major comments

1. Cachexia definition and staging

The use of the Fearon criteria to define cachexia is appropriate, and the adjusted weight-loss threshold is reasonable. However, Fearon et al. differentiate between pre-cachexia, cachexia, and refractory cachexia. This distinction is well established and widely used. The lack of reference to staging may limit interpretability for readers, and this should be acknowledged—at minimum in the Limitations.

2. Impact of tumor resection

There is ongoing debate regarding whether complete tumor resection is required to halt or reverse cachexia. It is unclear whether surgical status or resection margins were accounted for in the analyses. If available, incorporating or commenting on these variables would strengthen the validity of conclusions.

3. Body composition data

The manuscript notes that fat-predominant weight loss may be reversible, whereas profound skeletal muscle loss is generally not. CT-based body composition assessment is mentioned. If such data exist for a subset of patients, their inclusion or descriptive reporting would meaningfully enhance the interpretation.

4. Clinical validity of “episodes”

A key unresolved question is whether the observed weight-loss patterns represent clinically defined cachexia or non-specific body-weight fluctuations unrelated to the severe metabolic and functional phenotype of cachexia. This distinction is essential and should be addressed directly.

In the README.me file the meta data include pathological stage (TAGE_CDM_DERIVED_GRANULAR) but clinical stage is not mentioned in the README.me file. In the manuscript, however, clinical stage is mentioned (line 364/365) but not pathological stage. Please clarify which stage was used.

Minor comments

1. Line 47: Please add a reference after the first sentence.

2. Line 57: Please add the prominent role of muscle loss related to cachexia.
3. Line 67: Please add that another major issue in cachexia research is inconsistent diagnostic definitions across studies.
4. Line 193: Use the term refractory cachexia, which is defined as cachexia at the end of life.
5. Given the rise in early-onset cancer often diagnosed at later stages, why did the authors not consider age at diagnosis? Particularly for colorectal cancer, which comprises a relatively large part of the study population and provided novel results.
6. BMI outliers (<10 and >100) are described in the ReadMe file. Please provide justification or reference for these thresholds.
7. The manuscript does not clearly describe the handling of missing data and other preparatory steps; given the scale of the dataset, this requires clarification. How many variables were complete if there was missing, was there imputation?
8. Ethical oversight remains unclear. Written informed consent is mentioned, but IRB or equivalent approval is not specified.
9. Two tables at the end of the manuscript PDF lack titles, legends, and formatting and require revision.
10. In the Editorial Policy checklist, the competing interests section appears incomplete.

(Remarks on code availability)

BMI outliers (<10 and >100) are described in the ReadMe file. It is unclear how these cut-offs were chosen. 13047 patients were excluded if only one BMI measurement was available and trajectory less than 180 days. Were complete data available remaining for the other patients. How were missing data handled for covariates, particularly in the survival analysis. In the README.me file the meta data include pathological stage (TAGE_CDM_DERIVED_GRANULAR) but clinical stage is not mentioned in the README.me file. In the manuscript, however, clinical stage is mentioned (line 364/365) but not pathological stage. Please clarify which stage was used.

Reviewer #3

(Remarks to the Author)

Summary of the key results

The paper builds a health-system-scale, episode-resolved atlas of cancer cachexia by detecting longitudinal BMI-defined weight-loss episodes across multiple solid tumors. It shows that these episodes are common, concentrate around periods of disease progression, and associate with worse overall survival using time-varying Cox models. The authors identify a broadly consistent serologic signature of cachexia (e.g., low albumin/hemoglobin, high ALK and neutrophils) and link specific somatic alterations (notably TP53, STK11, SMAD4) to increased cachexia risk. Finally, they develop cancer specific prediction models for cachexia risk using serologic and genomic features.

Originality and significance

The work is original in combining: (i) episode-resolved cachexia detection from routine EHR BMI data and (ii) multimodal linkage to serologic markers, genomics, and progression across many cancer types. They also predict cachexia risk in two cancer types. Prior studies have examined longitudinal weight loss and survival in specific cohorts (for example, a study using Non Small Cell Lung Cancer trials PMID: 32940014, and Gastrointestinal cancers PMID: 36165100), proposed algorithmic cachexia criteria on EHR data in two cancers (<https://doi.org/10.1002/rco2.107>), and there are emerging real-world cachexia prediction models using routine data (PMID: 39109281). This manuscript is more comprehensive and multimodal than those, so the contribution is both novel and timely.

Data & methodology: validity of approach, quality of data, quality of presentation

Using large-scale EHR data to algorithmically detect cachexia episodes is an appropriate and powerful approach. The EWMA-smoothed BMI trajectories, sliding-window episode detection, and edema-related QC are reasonable in spirit, but key implementation details (window step, merging rules, recovery definition) need clearer specification and some parameter choices (α , 180-day window, 15-day minimum) are insufficiently justified. The multimodal linkage (labs, genomics, treatments, progression) is a major strength, but the handling of missingness, progression annotation sources, and supportive care interventions is under-described. Overall data presentation (figures/tables) is clear, though some supplemental figure labeling and legends require minor corrections.

Appropriate use of statistics and treatment of uncertainties

Time-varying Cox models for survival, mixed-effects models for serologic signatures, and logistic / risk-prediction models are appropriate choices for the questions posed. However, the manuscript provides limited information about proportional hazards checks, calibration, or other model diagnostics, and treatment of uncertainty is mostly via confidence intervals rather than deeper sensitivity analysis. The suggested parameter and subgroup sensitivity analyses would substantially strengthen the statistical credibility.

Conclusions: robustness, validity, reliability

The core qualitative conclusions that BMI defined cachexia episodes cluster around progression, are associated with worse survival, and have consistent serologic, and some genomic correlations are well supported by the presented analyses and align with prior cachexia literature and guidelines. Quantitatively, the exact effect sizes and some cancer-specific genotype associations are more vulnerable to modeling choices (episode definition, end-of-life windows, treatment/tumor-burden adjustment). With the requested robustness checks and clearer definition of recovery and progression, the conclusions would be more reliable and easier to interpret and reproduce.

Suggested improvements (Major revisions)

1. Provide explicit pseudocode for episode detection and recovery, make sure to clarify sliding-window step size, merging rules, recovery (including criteria for sustained weight gain or stability).

2. Run lightweight sensitivity analyses varying EWMA α in addition to the current value of 0.2 (for example, 0.1 and 0.3), minimum episode duration in addition to the current 15 day minimum (for example 30 and 45 days), and edema QC thresholds (additional windows such as, 14 and 45 days and thresholds such as, 3% and 7%), and report how prevalence, episode counts (1, 2, and 3+), progression overlap, and serologic signatures change.
3. Please specify how progression is defined, strictly radiographic RECIST progression, clinical/biochemical/radiology-combined progression or whatever has been used. Whether the same criteria are used across all cancers, or disease-specific rules are applied, and in what proportions. The source of progression labels (radiology structured fields, tumor board/registry abstractors, manual chart abstraction, NLP from reports, or whatever has been used). Also specify any quality control/inter-rater metrics used for progression adjudication and the exact timestamp used for progression.
4. How are duplicate and overlapping progressions handled? Also specify the proportion of patients (by cancer type) with usable progression annotations and make missing progressions transparent.
5. Add end-of-life sensitivity (exclude last 30 and 60 days before death) and stage-stratified analyses for progression overlap and serologic signatures. Explain how pseudo progression (for example in patients receiving immune-checkpoint inhibitors) is being handled in this study.
6. For patients with 1, 2, 3+ episodes please report how many last episodes of cachexia extended to end of life or end of follow up. This would shed light on the proportion of episodes from which recovery is not achieved and help distinguish reversible vs terminal cachexia phases based on the simplified cachexia definition used in this study.
7. Further analysis of the number of cachexia episodes per patient excluding the last 30 and 60 days before death (for comparison with Figure 2C) will provide further insights since based on your analysis irrespective of cancer type (excluding GBM) more than 50% patients develop cachexia close to end of life.
8. You identify a largely disease agnostic serologic signature of cachexia. Please repeat key mixed-effects logistic models stratified by stage (I–II vs III–IV) to assess whether this signature is consistent in both early and late disease.
9. To support mechanistic plausibility, consider showing whether the cachexia serologic signature (low albumin/hemoglobin, high ALK, high neutrophils etc.) is stronger in TP53-mutant vs wild-type patients during cachexia windows
10. You focus on 13 cancer types for somatic genotype cachexia analyses. Please explain the criteria used to select these 13 cancers.
11. You build prediction models for cachexia risk in LUAD and COADREAD using somatic features and serologic markers. Please explain why only SMAD4 is used for cachexia risk prediction in COADREAD and not TP53 or KRAS (Figure 5 suggests KRAS also has a significant association), and the rationale for focusing on LUAD and COADREAD given the high cachexia prevalence in PAAD and STAD.
12. To strengthen the predictive model's feasibility in clinics it will be beneficial to explore similar cachexia risk models in PAAD and STAD (high prevalence) and possibly PRAD (low prevalence) to illustrate performance across different prevalence regimes.
13. For clinical utility, models need to generalize beyond the development cohort. Please validate your trained models for predicting cachexia risk if possible with an external dataset for any cancer type, evaluate whether your trained models retain discrimination and calibration there. Or at least use another internal dataset for validation. Explicitly report how C-indices/AUCs change under these validations.
14. Please clarify if patients received only standard of care therapy, or did some patients receive additional nutrition and exercise interventions (clinical nutrition consults, specialized programs) since this could influence episode detection or recovery.
15. Supplemental Figure 2, panel labels appear swapped (A and B). Please correct it.
16. Radiographic features are mentioned in the Introduction and Discussion, however, the results or analysis do not include radiographic features. It will be beneficial to clarify which radiographic features been used in this study.

References: appropriate credit to previous work?

The manuscript cites core consensus and guideline work on cachexia and its clinical impact.

It would benefit to place their work relative to recent longitudinal weight-loss and cachexia studies in real-world and trial cohorts and to emerging EHR-based or AI-based cachexia prediction approaches. Adding these references will help readers appreciate what is new (episode-resolved, multimodal, health-system-scale analysis across many cancers) versus what builds on prior work.

Clarity and context: abstract, introduction, conclusions

The abstract clearly states the motivation, high level methods, and key findings, but it tends to present the phenotype as "cachexia" rather than "algorithmic weight loss episodes consistent with cachexia criteria," which slightly overstates clinical specificity. The introduction is well contextualized in terms of cachexia burden and unmet needs but could better position the work relative to existing longitudinal and EHR based cachexia studies. The conclusions are generally clear and aligned with the results.

(Remarks on code availability)

The Github repo has good data format specifications, function-level documentation, and algorithm explanation. I was able to test their code using synthetic data that I generated myself.

Following changes to the repo will make it more user friendly.

1. add installation dependencies in a requirements.txt
2. edit main.py to change hardcoded paths or instruct the user to add relevant paths here.
3. instruct the user to run "python main.py"

nomogram.py requires additional data not listed in the README and so I did not test this.

The included CSV (59,493 patients) contains the processed output with anonymized cachexia episodes, and other output files have also been shared that were for the paper figures/tables. Raw BMI trajectories and mutation data are not included.

This code can be used by the community if they can format their data as required and detailed in the repo. Also, their code doesn't handle missing data, so all required patient information needs to be available in order to run their code. They have a check applied for having more than 200 patients per cancer type for genotype survival analysis only. This hardcoded threshold can be changed in the code depending on the amount of data the user is working with.

Overall, the repo looks good, and the code is functional.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Need for a complete data verification process and re-analysis of the data.

1. One of the first points made in responses to previous comments is that the approach to calling episodes has been revised. This includes addition of a 2% weight loss over six months as a cachexia criterion value, specifically for individuals with BMI < 20 kg/m². Absent a formal grading system for severity, this criterion was used in Fearon et al. 2011 to denote individuals of clearly depleted body stores. Reading this criterion, I went to the data set to observe the cachexia episodes meeting this criterion (low grade weight loss between -1.999 and -4.999% with BMI<20) and went on to discover multiple bizarre errors in the provided data file. The first instance I happened upon is patient 0020837. This person was coded as having a weight loss episode -4.09%, but a BMI above 20, so this person's cachexia episode was miscoded. Whilst continuing to peruse instances of low-grade weight loss, I next found patient 0076545, who had a 2.19% weight loss lowering BMI of 22 to 21.5, but column J this patient's BMI category is "obese", also incorrect. I checked the baseline BMI categorical value (column J baseline_bmi_cat) against the BMI at start of a cachexia episode. The baseline BMI categorical variable seems to have been scrambled and would not render proper results if it was used in any analyses. For example: n=129 patients in the "normal BMI" category had a BMI of less than 18.5 at the start of their cachexia episode, and n=1554 patients with a BMI between 25.0 and 57.4, have been included in the normal BMI category. In the "obese" category of baseline BMI, n=906 patients have a BMI <30 (and as low as 17.0 kg/m²). I could go on ... in the "overweight" category there are patients with BMI as low as 17.5, and as high as 75.4, and indeed the "underweight" category has patients with a baseline BMI of up to and including 46.7 kg/m².

In the data provided n=29,210 incident episodes of cachexia are found, as mentioned in the results section. I would parenthetically add that this data file is a lot more informative than the previous one, however the %WL column has 2794 episodes of incident weight loss with the value of 2% loss to 4.999% loss. These only qualify as cachexia episodes if BMI is <20 kg/m². Of 2794 patients with low-grade weight loss n=1856 of them had BMI >20 kg/m² (including 119 with class II obesity and 55 with class III obesity). It is with disbelief, I ran across P-0085852, who had a starting body mass index of 43.7 and had a "cachectic episode" comprising the loss of 2.006% of her body weight. Overall, 1856 episodes were improperly included.

Additional basic data quality concerns.

Whilst looking at the largest % weight loss values in the data set, I ran across P-0053346 with an implausible weight loss coded as -50.87% during an episode duration of 16 days (BMI 56.4 fell to BMI 27.7). There are other patients with these characteristics of implausibly large weight loss (up to -68.1%), and some of these did so in a very short period. Some of these huge losers are also people with implausibly large BMI; such values can result from typographical errors in either the weight or the height of the person. I am experienced in this type of data and know that there is a certain level of error inputting the data, such as a missed decimal place or missed digit. Because height is the denominator in BMI, errors in height can result in bizarre body mass index values. There are some extremely heavy people in this population, which bears discussion if their BMI and weight loss percentage can be verified. To achieve a BMI of 50.0 an average height man would have to weigh upwards of 350 pounds. To achieve some of the higher BMIs (max value is 97.8 kg/m²), weight would reach 675 pounds. I encourage therefore, the authors to access the height and the weight data and verify as much as possible that there are no errors in the input variables.

The columns hasWL10 and hasWL15 in the data set also appear to be incorrect and would not render proper results if these categorical values were used in any analyses. These are supposed to indicate incident weight loss exceeding -10%, and -15%, respectively. In the %weight loss column there are 6161 rows with a value ≥15% (range 15.0% to 68.2%), however only 3226 of the episodes of incident weight loss ≥15%, has a value of "1" in the categorical hasWL15 column. Likewise, the column hasWL10 has a value of "1" for 3489 episodes of WL ≥10%, however n=3072 patients with a weight loss ≥10% have a value of zero in the has WL10 column

The next discover is that the weight loss grades are not applied correctly. Only the weight loss incidents are classified, so the account of there being just 90 patients with weight loss grade of 0, in the population of over 52,000 patients is not correct. The grading system is applied to all patients, so in a population weight stable patients with a BMI >28 kg/m² (no weight loss detected) are graded 0. Weight stable patients with a BMI between 20 and 24.9 are Grade 1, and weight stable patients with a BMI<20 are Grade 3.

To this point in my commentary, my conclusion is that the manuscript is not evaluable until such time as all of these errors and miscoding are corrected in the data set.

2. Method of calling cachexia episodes. I continue to have several questions about this approach. This is described as a “sliding-window episode-detection framework” and illustrated in Suppl Fig 1C. This is a single depiction of a body weight trajectory, not claimed to be representative; it would be helpful to see other trajectories to understand the application of the cachexia episode calls. The illustrated trajectory shows gain, stepwise loss and then some limited regain. How many episodes show complete regain, partial regain or no regain? Did you consider a regain parameter? In advanced disease some of these trajectories just go straight down. How often do we see gain (from baseline) or regain following an episode and what is the range of magnitudes?

What I am trying to grasp is the “cachexia” in certain examples

P-0018242 had a loss-gain-loss (BMI 22.9 > 21.3 > 24.7 > 22.3); This has 2 episodes of loss but a 3.4 kg/m² gain. Is this cachexia if they started at 22.9 and ended at 22.3?

P-0023564 had a loss-gain-loss (BMI 36.2 > 34.2 > 35.7 > 32.2) Is this cachexia of concern given the high body weight?

P-0006195 has 2 episodes: 3% loss, regain, 4% loss Cachexia episodes minimal absolute value

I am not able to go through all of the data, these are just examples, with undulating BMI but very little variation- this strikes me as over-classification.

3. Severity of cancer cachexia- how to understand it?

This version is much clearer on the severity issue, but a “try everything” approach has been taken towards it. Several analyses include 5% weight loss alone, as well as %, 10% and 15% cutoffs for weight loss %. Then the 2% weight loss with BMI<20 is added, and then the Weight loss Grades 0-4 (Martin et al.) are added. The 5% criterion value is still taken as the “primary” definition. All of this could be collapsed into the Weight loss grades, as a single primary outcome for the following reasons: the lower threshold of 2.5% is included, data driven thresholds of 6%, 11% and 15% are included, and the BMI<20 with 2.5-4.99 % weight loss is included. Taking this approach would also accommodate the fact that your population trends significantly to being overweight and obese. A major concept in cachexia severity is that the impact of weight loss is not the same across the spectrum of BMI. I believe further consideration should be given to the impact of obesity, with n=9311, of 29,219 instances of weight loss, 31.9% occur in obese persons. N= 1357 of those meet criteria for Class III obesity and n=2363 meet criteria for Class II obesity; n=200 have a BMI> 50 kg/m² with the highest value being 97.8. Very heavy people do you have a reduced risk of mortality as shown by Martin and coworkers in 2015. A weight loss of 6% is Grade 2 if BMI is > 28, is Grade 3 if BMI is between 22 and 28 and is Grade 4 in persons with a starting BMI of 20 or below. Insofar as the current samples have a very high disproportion of extremely heavy people, it begs the question as to whether these authors should undertake to determine the relative risk of mortality according to weight loss across their 9311 obese patients.

Depending on the number of events, the sample may be sufficiently powered to do so. The population comes from a very high obesity risk country with a number of obesity-related cancers included, tending to skew the BMI distribution to the right. Other countries may have quite different demographics, so this sample may not be generalizable.

When stratifying patients according to Weight Loss Grades, I do suggest that the full scale be used, however if these are to be grouped for any reason, Grades 0,1 and 2 should be grouped as they largely encompass weight loss <6% and weight-stable patients, and this is quite distinct from Grades 3 and 4.

Association vs causation: a bundle of secondary analyses

The presentation of the data begins with what we are led to understand is the primary purpose: to detect the prevalence of incident weight loss (of varying degrees). As suggested above, this data presentation could be made more concisely.

There is an odd interjection of radiologically determined fat loss data, that doesn't add to the narrative in any substantial way. It's only a partial data set, and it is muscle rather than fat that sits at the centre of the cancer cachexia definition and shows association with cancer outcomes. I suggest that this be deleted. The authors cannot offer meaningful insight into the body composition piece.

- A large part of the paper concern association of incident weight loss with primary cancer site, cancer stage, progression on cancer therapy, and cancer driver mutations, “serological signatures” (standard clinical lab values), age, sex, performance status. The pieces of this are rather disjointed, and I would hope that these could be represented in a coherent model. A) Tumor activity drives the wasting syndrome in the first instance, with its stage, progression and cancer driver mutations being part of one bundle. B) the serological signatures do not represent any catabolic effector molecules i.e. there are no actors in the list capable of eliciting adipose tissue lipolysis, muscle proteolysis or suppress food intake. So, there is no possibility of causation, only association. It has not occurred to the authors that many of the blood changes are secondary i.e. myelosuppressive drugs such as Platinum agents: Carboplatin, Cisplatin, Oxaliplatin, Alkylating agents: Busulfan, Cyclophosphamide, Chlorambucil, Carmustine, Antimetabolites: Fluorouracil (5-FU), Capecitabine, Gemcitabine, Topoisomerase, Inhibitors: Topotecan, Irinotecan, Etoposide, Anthracyclines: Doxorubicin. Regimen (not included in this analysis) would be a more direct contributor to weight loss as systemic therapy drives wasting by generating GI symptoms such as nausea and vomiting, and also has direct cytotoxic effects on muscle and fat. PMID: 36942895 summarizes recent findings on different systemic therapy regimens, that induce double digit muscle loss in ~3 months and have specific cytotoxic effects on muscle. PMID: 37686641 is an interesting read, which shows that male sex, progressive disease on therapy are independent drivers of muscle loss in patients with pancreatic cancers, and that FOLFIRINOX is a specific driver of muscle loss while gemcitabine -nab paclitaxel is a fat loss driver. C) age, and sex have been included in some analyses, but any sex difference in cachexia prevalence or severity is not reported. I am not certain why ECOG PS is included as a factor, as it is more likely an outcome of cachexia (see Lancet Oncology 2011) than a driver of it.

(Remarks on code availability)

The data set is filled with errors. As such, I have not attempted to evaluate the code.

Reviewer #2

(Remarks to the Author)

The authors have undertaken a thorough and careful revision in response to all reviewer comments. Nevertheless, several important issues remain.

First, I agree with the previous reviewer that the inclusion of patients with follow-up extending beyond 5 years may be problematic in the context of cancer-related weight loss. A 10-year time window after diagnosis appears excessively broad, as many patients with such long-term follow-up are likely to be cured (unless there are secondary cancers or a late recurrence etc.). An analysis restricted to the first 5 years after diagnosis may therefore be more appropriate for capturing weight loss that is plausibly attributable to cancer and its treatment.

Second, the limitations should more explicitly address the challenges of interpreting weight loss in patients with obesity. In this group, clinically meaningful wasting may occur in the context of sarcopenic obesity, which is difficult to ascertain using routine clinical data. Weight loss alone cannot distinguish between loss of adipose tissue and loss of skeletal muscle, both of which may have different biological and clinical implications. This limitation should be acknowledged more clearly. However, the authors have addressed changes in VAT and SAT in the population which is highly appreciated.

Third, the application of the Fearon criteria represents an important improvement. However, although the authors note that insufficient data were available for complete classification, the current dichotomization into non-cachexia versus cachexia remains a limitation because it does not consider pre-cachexia as a clinically relevant intermediate stage. Please add to the limitations explicitly that assigning pre-cachectic individuals to the non-cachexia group may increase heterogeneity within the reference category and potentially obscure stage-specific associations.

Relatedly, Reviewer 3 raised the issue of cachexia episodes in the period preceding death. In this context, the authors should consider whether refractory cachexia can be evaluated. Based on the information provided, at least some of the required data appear to be available, particularly survival <3 months and performance status (ECOG), consistent with the framework proposed by Fearon et al. If a formal classification is not feasible, this should at minimum be discussed explicitly (see above)

Finally, I appreciate that the authors have now included age at diagnosis as a covariate in response to my earlier comment regarding early-onset cancer. Given the size and depth of this dataset, a stratified analysis by age at onset could be highly informative. If this lies beyond the scope of the current manuscript, it would still represent a valuable direction for a subsequent study and would likely be of considerable interest to investigators working in this area.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The authors have comprehensively addressed the majority of the reviewer comments. In particular, they have incorporated multiple sensitivity analyses across weight-loss thresholds, stage strata, and independent cohorts, with results presented throughout the manuscript and supplementary materials. The methodological clarity has substantially improved, especially regarding episode definition and progression annotation.

Notably, the revision reflects major improvements in several key areas: methodological rigor has been significantly strengthened, sensitivity analyses are now clearly and systematically presented, genomic findings have been reinforced through more robust and multivariable analyses, and external validation has been incorporated to support generalizability.

While some analyses remain primarily descriptive rather than fully statistical, the revisions sufficiently support the robustness of the conclusions. No further revisions are required.

(Remarks on code availability)

While the GitHub repository is well structured and documented, it would benefit from the inclusion of a requirements file (e.g., requirements.txt or environment.yml) to facilitate installation of dependencies and improve reproducibility for users.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

I would like to thank the authors for their thoughtful and detailed responses to my comments.

Regarding my first comment, my concern has been sufficiently addressed. I would encourage the authors to include the additional discussion on differences observed when restricting the analysis to patients within five years of diagnosis in the manuscript, as this would help readers better interpret the findings.

Regarding the second comment, the response was adequate, and I welcome the authors' decision to address this point in the discussion. The same applies to my third comment regarding the Fearon classification and to my fourth comment

regarding refractory cachexia. In both cases, the authors have addressed my earlier concerns in a constructive and scientifically appropriate manner.

I also appreciate the authors' willingness to add further nuance to the interpretation of the findings. Given the size and uniqueness of the dataset, some implausible observations, such as apparent drops or reversals in weight loss, may be easy to overlook. However, I agree with the concern that such patterns should ideally have been detected through routine data-quality checks. Even a brief histogram or distributional summary of percentage weight loss would have helped to identify and contextualize these observations.

Overall, the authors have addressed the main points raised in my earlier review. The additional clarification and willingness to revise the discussion improve the objectivity of the manuscript and will help readers interpret the findings more appropriately. My only remaining conceptual concern relates to the terminology of "cachexia episodes," which is not fully congruent with the traditional understanding of cachexia as a syndrome that is not readily reversible. However, I also recognize that this is an important issue for the field and that the manuscript may contribute meaningfully to an ongoing re-evaluation of current cachexia definitions.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I have reviewed the revised manuscript and the authors' point-by-point responses. My overall assessment is unchanged from the prior round, the study is methodologically sound, the sensitivity analyses are thorough, and the external validation supports the robustness of the central findings.

Having examined the data integrity concerns raised in this round in detail, I find the authors' responses persuasive. The majority of the flagged discrepancies arise from comparing quantities defined at different timepoints (anchor/diagnosis vs. episode start) or computed by distinct procedures. The worked trajectories, the anchor vs episode start comparison table, and the replication of analysis on an independent cohort adequately substantiate this. The column renaming, the added episode start and nadir fields, and the expanded data dictionary directly resolve the source of confusion, and I consider these concerns addressed.

A few of Reviewer 1's points are scientific rather than clerical, and the editor may wish to consider whether the authors address them more fully:

1. Treatment confounding of the serologic signature. Several of the reported shifts (anemia indices, albumin, transaminases, leukocyte populations) are also expected consequences of cytotoxic chemotherapy. Given that the serologic signature is a primary result, I recommend more than a Limitations acknowledgment, at minimum a clearer framing that these associations may partly reflect treatment, and if possible, a sensitivity check (e.g., whether the directionality persists outside peri-treatment windows or after adjusting for therapy exposure where available).
2. WLGS scope. As Reviewer 1 notes, the Martin et al. grading system is defined across the full population, whereas it was applied here only to detected episodes. The authors should state this scoping explicitly, so the grade distribution is not misread as a population-level classification.
3. Post-episode trajectory. Reviewer 1 raised a reasonable question about regain following episodes. The CT-based SAT/VAT analysis already addresses this in part, showing adipose regain in roughly 59–66% of episodes in the imaging subset, but a complementary BMI level quantification across the full episode set (complete/partial/no regain, with magnitude ranges) would extend this beyond the imaging cohort and more directly address the question raised.

On the body composition section, I would suggest retaining it, as it provides orthogonal evidence that episodes correspond to real adipose depletion, provided its claims remain appropriately bounded given the partial coverage.

Regarding the code, please ensure the new auditing columns and a requirements file are committed to the repository.

(Remarks on code availability)

I have reviewed the public analysis code (github.com/reznik-lab/cachexia). The episode detection logic confirms the post-hoc weight_loss field is computed start to end and is distinct from the threshold that triggers an episode call, which accounts for nearly all of the "miscoded" cases. In my assessment the data integrity objections are largely resolved by the code and documentation. I would hold the authors to their commitment to add a dependency/requirements file, which is still absent from the repository.

Version 3:

Reviewer comments:

Reviewer #3

(Remarks to the Author)

all my comments have been addressed by the authors.

(Remarks on code availability)
I had reviewed the code in last revision.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Thank you to each of the three Reviewers for the critiques and comments below. We have read all of the concerns raised by the Reviewers, and found them to be both reasonable and constructive. In the attached revision, we have now addressed each concern, and by consequence the rigor and insight delivered in this work has proportionately increased.

Before addressing each comment in the point-by-point response, we would like to briefly highlight **two key revisions** to the manuscript which are broadly relevant to each of the reviewers:

1. First, we have **revised our approach to calling episodes** to align with consensus weight loss criteria from the Fearon 2011 Lancet Oncology manuscript. Specifically, this means episodes must satisfy one of two criteria: 5% weight loss over a 6 month period, or 2% weight loss over a 6 month period for individuals with BMI < 20. Importantly, these weight loss criteria align with the exact eligibility criteria used for contemporary clinical trials including but not limited to the Phase 2 ponesimod trial¹ recently reported in the *New England Journal of Medicine*. Furthermore, to more granularly examine the effect of different degrees of weight loss, we have also repeated many of our analyses using the formalized weight loss grading system (WLGS), as well as *ad hoc* thresholds of at least 10% or 15% weight loss over a 6 month period. These complementary analyses deliver qualitatively identical findings, and emphasize the robustness of our observations.
2. In the 3 month period since receiving reviewer comments, we have secured an **external validation cohort** from Dana Farber Cancer Institute. We have used this cohort to corroborate (a) the prevalence of weight loss episodes across cancer types, (b) the discovery of specific tumor genotypes (especially *TP53* mutation across diseases) associated with weight loss episode risk, and (c) the prognostic risk model for impending weight loss in non-small-cell lung cancer.

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

The data from MSK-IMPACT might be useful for analysis related to cachexia. A clear alignment with consensus-based conceptual framework and revised analysis plan clearly needs to be built from the ground up. It may be helpful to reach out to an expert on cancer cachexia diagnostic criteria.

Response: Thank you for the above commentary and for the many points of feedback detailed below. We acknowledge and agree generally with nearly all of the critiques, and they are addressed in detail below.

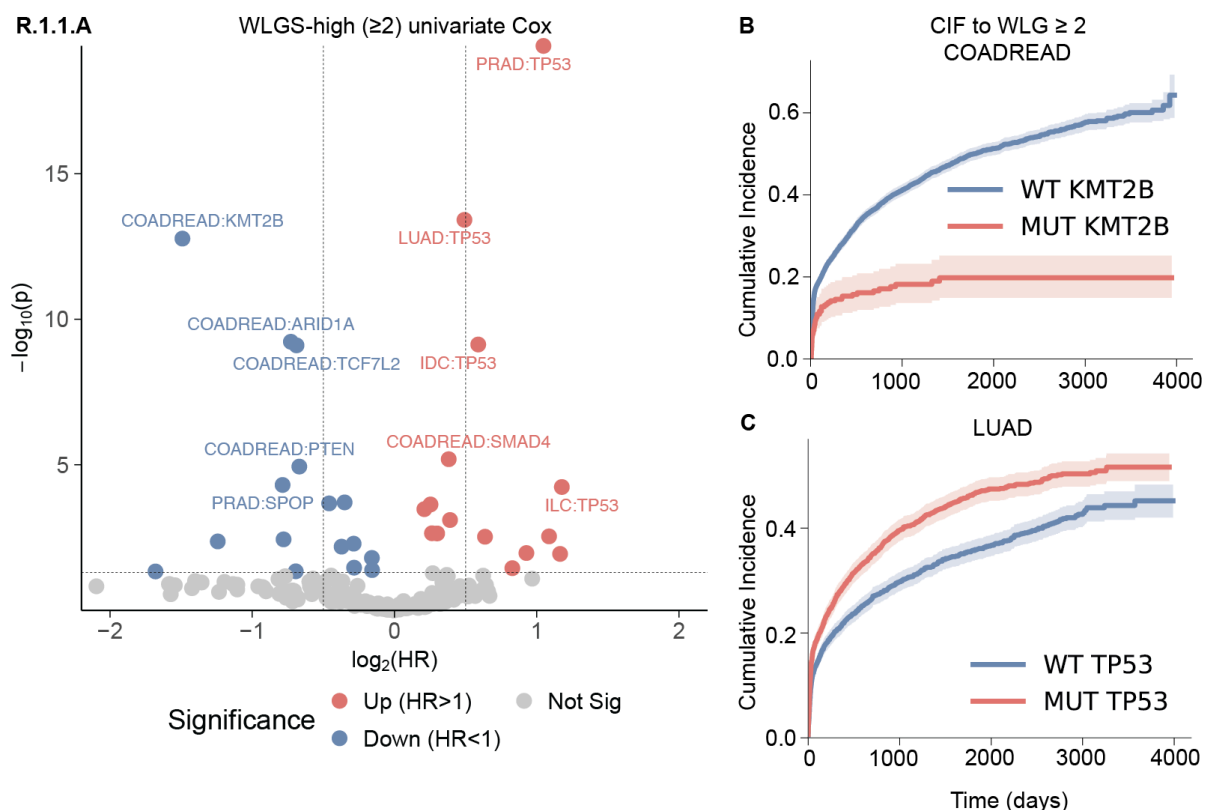
To the question of Novelty and Significance.

The analysis of association of cachexia with tumor mutation is potentially novel, with the caveat that the classification of cachexia prevalence and severity needs to be completely redone. Therefore the association with tumor mutation must be redone.

Response: Thank you for raising this point. The question of cachexia severity is raised in several comments, and so we have taken this point of feedback quite seriously. Accordingly, we rebuilt our cachexia classification pipeline to align more closely with the consensus definition, implementing both the six-month weight-loss criterion (>5% weight loss over 180 days) and the low-BMI adjustment (BMI <20 kg/m² with >2% weight loss over 180 days) (**Methods-Consensus-defined weight loss episode identification, Pages 20-21**). We were not able to implement the sarcopenia-based criterion requiring skeletal muscle index because standardized muscle measurements are not available across the full cohort, and we now state this explicitly as a limitation (**Page 17**).

Using the revised pipeline, we re-ran the full somatic mutation analysis and found that the key weight-loss associated alterations remained consistent. We identified 32 significant cancer-gene associations (FDR<0.1), including *TP53*, *STK11* in lung, *SMAD4*, *KMT2B* in colorectal, and *SPOP*, *TP53* in prostate adenocarcinoma. Among these associations, 24 remained significant when applying a 10% weight loss (WL) over 6 months threshold and 13 remained significant under a 15% WL over 6 months threshold (**Supplementary Figure 5A and 5B; Supplementary Table 13 and 14**).

To directly address severity, we additionally modeled mutation status against time to WLGS-high cachexia (WLGS ≥ 2), and the associations remained statistically significant, indicating these alterations are linked to risk of developing more clinically significant WL rather than being driven by marginal 5% WL events (**Fig. R.1.1.A–C, below**).



(A) Pan-cancer volcano plot of univariate Cox models, within each cancer type, the association between mutation status and time-to-first WLGS-high (≥ 2) cachexia event (B–C) Cumulative incidence curves (CIF) for WLGS-high (≥ 2) comparing mutant vs wild-type within representative cancer types (COADREAD *KMT2B* shown in B, LUAD *TP53* shown in C)

The association of cachexia with routine blood work values is not novel. The researchers did not provide a clear justification for selecting standard of care blood work as the source data for cachexia biomarkers; the manuscript does not contain or refer to a literature review of the cachexia biomarkers space. For example, by 2008, classification systems for cachexia (PMID 18718696) were based on well-established association increased inflammatory markers (CRP, IL-6), Anemia (Hb <12 g/dL) and low serum albumin (<3.2 g/dL. According to Google Scholar this work has been cited 3196 times. In the meantime, there have been >50 publications attesting to the association of the Glasgow Prognostic Score, which includes abnormal albumin and CRP values, with cancer cachexia. Likewise, the association of neutrophil to lymphocyte ratio with cancer cachexia is covered in about 40 publications.

Response: Thank you. The reviewer raises an important point, which is that we have not acknowledged the very long and significant history of findings associating standard-of-care blood work with cachexia. We have now amended this in several parts of the manuscript, including new references and text on **Pages 10-12**.

With regard to the question of the novelty of our association between blood work and cachexia, we firmly disagree. There are two key distinctions which render this specific analysis highly novel. *First*, the work is carried out systemically across cancer types, using patients as their own controls by separating cachectic and non-cachectic time periods and leveraging longitudinal serologic measurements. There is no such comprehensive assessment of serologic changes, granularly stratified by disease, available. *Second*, several specific findings in the blood biomarker analysis are indeed novel in their own right. The most apparent of these is the identification of generic elevation of liver enzymes during periods of significant weight loss. Our survey of the literature identifies some hints of this in the literature, but no systematic analysis. The nearest we could find is PMID 28193264² which constitutes a mixed cohort of cancer and non-cancer cachexia and has on the order of hundreds (as opposed to our tens of thousands) of patients.

As for KRAS mutation association with cancer cachexia, that also has been shown already, please see PMID 30932365. Also, molecular crosstalk mechanisms between skeletal muscle atrophy and KRAS-mutant pancreatic cancer, has been the subject of a literature review PMID: 39925900 and this is the reason that KRAS mutant pancreatic cancer cell lines are used in preclinical models (PMID: 28730707).

Response: Thank you. We acknowledge that prior literature associates certain somatic genotypes with cachexia risk (most prominently, *KRAS* and *STK11*), and this is now explicitly

stated on **Page 11**. However, there is simply no systematic analysis of the association between somatic genotypes and cachexia risk in the literature across somatic genotypes and cancer types. Even where this analysis does exist in specific genomic and disease contexts, the statistical methods employed (for example, chi-squared testing rather than time-to-event Cox proportional hazards models) are inappropriate. Our genomic discoveries reported here are therefore novel and now, through an external validation cohort from DFCI, have been validated (**Figure 5C**).

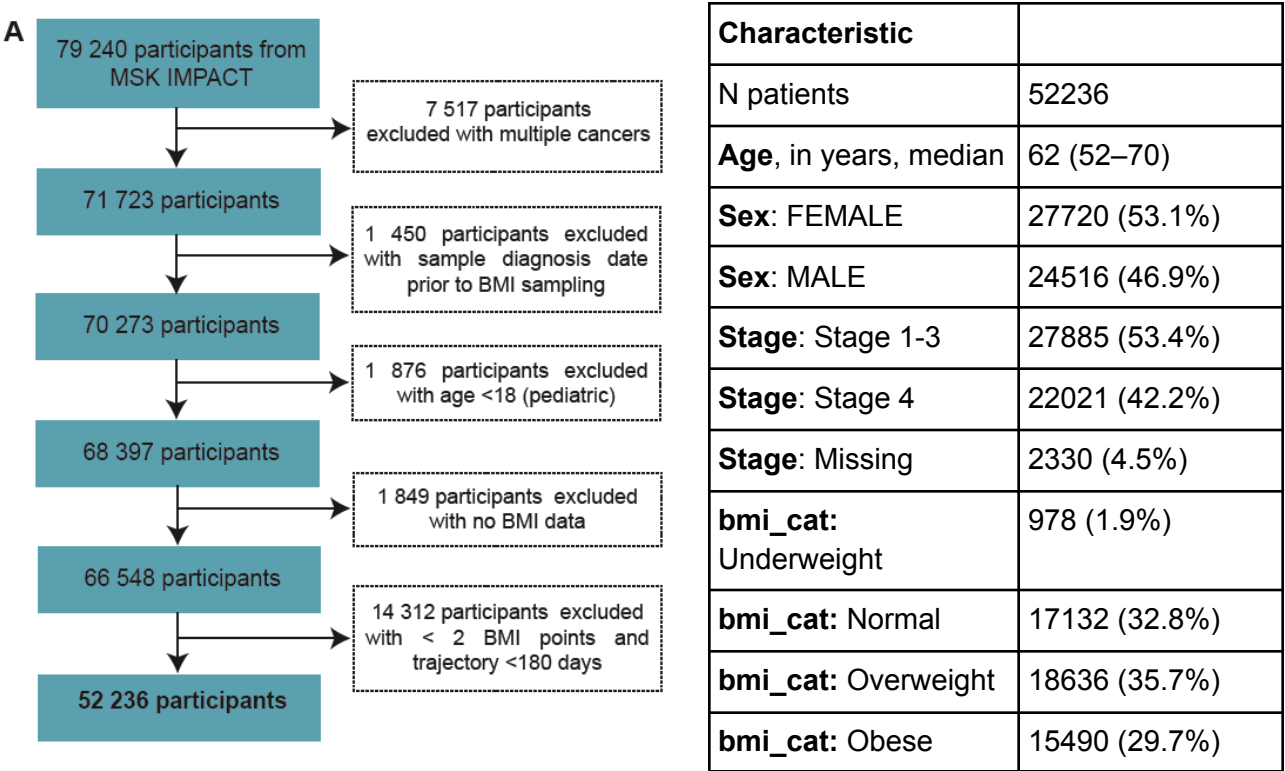
A. Does not meet minimum reporting requirements for a clinical study.

1. The data was collected with an unrelated purpose (mutational profiling), and this analysis was not planned a priori (i.e. not as part of the secondary analysis of the MSK – IMPACT project). Not a clinical trial in the conventional sense, participants had to have a cancer diagnosis, agree to mutational profiling and be seen at the Memorial Sloan Kettering cancer center. It will be necessary to provide a clear description of the cohort considered for the cachexia – related analysis and the criteria used to select them; a complete set of descriptive should be provided; only primary cancer is given. The only provided population characteristics is the primary disease. Looking to deduce some of this from the provided data set, immediately note that there are (for example) 138 lines of data noted with an age of 0 years (<1 y?), in total there are 1685 lines of data on patients under the age of 18 years at diagnosis. This raises questions of intent to study cachexia during a time normally characterized by longitudinal growth and the applicability of the criteria for defining cachexia in children (none exist). This suggest that the resource (>2.5 million lines of weight data) was not reviewed in relation to its suitability for a cachexia -related inquiry. (see next point)

Response: This is an absolutely critical point and we appreciate the reviewer raising it. We have revised the manuscript to provide an explicit, reproducible cohort definition for this retrospective observational study, leveraging longitudinal EHR-derived weight/BMI measurements linked to MSK-IMPACT–profiled patients.

We added a step-by-step cohort derivation schematic (**Supplementary Fig 1A**) and expanded the Methods (**Page 20-21**) to describe the data source and exclusion. Briefly, starting from 79,240 MSK-IMPACT–profiled patients, we (i) excluded patients with >1 distinct primary cancer (7,517 excluded; 71,723 retained) to avoid ambiguity in attributing longitudinal weight change to a single malignancy course; (ii) excluded patients with diagnosis/anchor dates prior to BMI sampling (1,450 excluded; 70,273 retained) to ensure adequate capture of longitudinal weight data; (iii) excluded pediatric patients (<18 at diagnosis) (1,876 excluded; 68,397 retained), as the reviewer noted that weight trajectories in children are shaped by normal growth and there are no established pediatric cachexia criteria in this setting. We therefore restricted all analyses to adults at diagnosis. (iv) excluded patients with no BMI data (1,849 excluded; 66,548 retained); and (v) required minimum longitudinal BMI coverage of at least 2 BMI measurements and ≥180 days between first and last measurement, excluding 14,312 patients. In the final cohort (n = 52,236), the median age at diagnosis date was 62 years (IQR 52–70), 53.1% were female and 46.9% male, and stage distribution was 53.4% Stage 1–3, 42.2% Stage 4, and 4.5% missing.

We now report this complete set of baseline characteristics for this cohort (**Table 1**), including age, sex, stage, baseline BMI, primary cancer type, follow-up duration, and survival status. All downstream analyses have been rerun in this revised cohort (**Pages 4, 18**).



2. Insufficient information is provided regarding the provenance of the data to be used for cachexia study and its suitability for a cachexia inquiry. Many questions are raised:
a) Did or does the MSK-IMPACT study have planned weight measurements (this is implied at one point), or were the weight records pulled from one or more sources? MSK-IMPACT was started in January 2014, however in the data set we see cachexia episodes that predate this considerably, with the latest recorded "cachexia episode" being over 53 years prior to the date of diagnosis (so weight history must go back to 1972 or earlier).

Response: MSK-IMPACT does not include protocolized or planned weight measurements. The longitudinal BMI/weight data used here were obtained separately from routine clinical records including those captured from the EHR during standard inpatient and outpatient visits. This is then linked to MSK-IMPACT profiled patients. In addition, the earliest BMI measurement available in the dataset used for this analysis dates to 2010. Although MSK-IMPACT was launched in 2014, patients were often diagnosed earlier and underwent sequencing later in their disease course.

Finally, episode timing is recorded as days since diagnosis (day 0); thus, a start_day value such as 19342 (~53 years) indicates an episode occurring ~53 years after the diagnosis (not before

diagnosis), which can occur for patients with very early diagnosis dates and BMI measurements captured decades later. We addressed this concern in our cohort construction, excluding patients whose diagnosis/anchor date predates the start of BMI availability in our extract (January 2010), avoiding analyses across intervals with no weight data capture. Furthermore, for patients with multiple diagnosis, we have kept the diagnosis date closest to the surgical path and sequencing date. This avoids historical tumor diagnosis dates that are years prior sequencing. We now state this explicitly in the Methods (**Page 18**).

b) From the present to 53 years ago, what data systems were accessed to obtain the weight information? By whom and when was it typically recorded (physical exam, chemotherapy dosing, nutrition/dietetic consultation, other?), For those years did MSK have an SOP for weighing patients? What weight scales and calibration SOPs were in place? This gets to a key point – what is the best estimate of measurement error which bears on the ability to detect a 5% difference. This kind of analysis predicated on a rather small percentage weight loss as the defining criteria used to make patient classification, would require a very clear representation of what constitutes a measurement error in the available data. There is some vague language about “smoothing” out the day-to-day variation, but this is neither clear nor sufficient to inform the reader as to the probability of detecting a relatively small change.

Response: Patient weights were obtained from routine clinical measurements recorded in the MSK electronic health record during standard care encounters (ex. clinic check-in vital signs and inpatient vital-sign documentation), typically entered by clinical staff as part of the vitals workflow. These data span multiple decades, and historical metadata on scale model, calibration, and institution-wide weighing SOPs are not uniformly available in the structured record; hence, we cannot provide a single validated, time-invariant estimate of instrument calibration error for the full time range and treat individual EHR weights as potentially noisy. We note this now on **Page 19** of the **Methods - Data Processing - Clinical Data**.

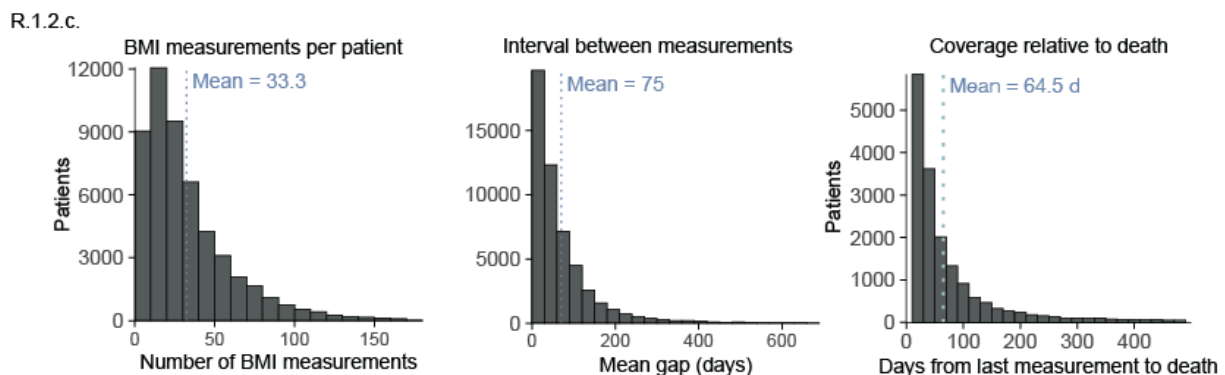
For this reason, to ensure that cachexia classification is not driven by single measurements, our episode calls are based on sustained trajectory-level change rather than isolated values. We first smooth each patient’s BMI trajectory using an exponentially weighted moving average (EWMA; $\alpha = 0.2$), which down-weights isolated outlier measurements and emphasizes persistent changes across sequential visits, and we then detect episodes using a sliding-window approach applied to this smoothed trajectory. We further reduce spurious calls by excluding rapid fluctuation patterns consistent with edema or recording artifacts and by requiring additional episode validity criteria (including a minimum episode duration of 15 days and a minimum net weight loss requirement of -2%). We revised the Methods to explicitly describe the smoothing, sliding-window logic, and these episode inclusion criteria (**Page 20-21; episode_identification_walkthrough.ipynb** (<https://github.com/reznik-lab/cachexia>)).

c) The extent coverage of patient’s clinical trajectory with weight data needs to be assessed and some criteria set for adequacy of coverage. In the very long term follow up (see above), many patients would be seen only annually or even less frequently, so if weight changes by 5% between year 10 and year 12 of follow up, was this coded as a cachexia episode of 2 years

duration? It is said that patients with just 2 weight values of record were included, however there needs to be a justification for including patients with 2 or few weight records, especially if their clinical record is missing weight information over long periods of time. It is necessary and essential to benchmark of the weight records relative to date of death (or censoring), not just to date of diagnosis. Absent this information, it is impossible to understand the definition of the "end" of a cachexia episode – was it defined merely by the last available weight record, death, or by stabilization or regain of lost body weight?

Response: We agree that longitudinal coverage should be made explicit, and we have revised the manuscript accordingly (**Page 4**). Importantly, sparse longitudinal records cannot generate long cachexia episodes in our framework: episodes are evaluated within fixed 180-day windows anchored to observed measurements, so widely separated weights (e.g., year 10 vs. year 12) do not, by themselves, yield multi-year cachexia episodes. We require a minimum of two BMI measurements so that each patient contributes at least one observable interval. In our cohort, patients contribute an average of 33.3 BMI measurements, with a mean interval of 75 days between measurements.

We also clarified how episode endpoints are defined (**Methods; Page 20-21**). Episodes are assessed within fixed 180-day windows anchored to observed BMI measurements; the episode end corresponds to the last recorded BMI measurement that still meets the weight-loss criterion within the evaluated windows, and overlapping windows are merged only when supported by intervening measurements. Thus, widely separated weights cannot generate episodes that span multiple years, and gaps in follow-up do not extend episode duration. To contextualize weight coverage near the end of follow-up, we report the time from the last available BMI measurement to death among deceased patients (mean 64.5 days; median 35 days [IQR 16–76]) and note that 79.6% (16,048/20,169) had a BMI measurement within 90 days of death.



R.1.2.c | BMI data coverage and cachexia episode endpoints. Left, distribution of the number of BMI measurements per patient (mean = 33.3 measurements per patient); Middle, distribution of the mean time gap (days) between consecutive BMI measurements within each patient (mean = 75 days). Right,(C) among deceased patients, distribution of the time from the last available BMI measurement to death

3. The analytic approach, as well as the diagnostic criteria for cachexia, are not based on the sound conceptual framework.

Response: Thank you for the comment. We recognize that aspects of the original cohort definition were overly permissive and required refinement. Below, we address these concerns.

a) The ~57,000 patients included many individuals whose inclusion is questionable:

i) infants and children:

Response: Thank you for pointing this out. We agree that pediatric patients should not be included, given growth-related weight change and the absence of validated pediatric cachexia criteria in this setting. We therefore now exclude all pediatric patients (<18 years at diagnosis; 1,876 unique patients) from the analytic cohort and re-ran all analyses in adults only (**Methods; Page 18, Supplemental Figure 1A**).

ii) persons reasonably considered cured of their cancer with long term survival > 10 years, and whose weight change, if any, reflected short term, reversible temporal bouts of weight loss likely to be driven by cancer therapy. Episodes of weight loss >10 years after primary cancer diagnosis are likely not even attributable to malignant disease. A clear part of the existing conceptual framework is that patients cured of cancer might sustain weight loss of short term associated with side effects of therapy, but unlike cachexia, those weight loss episodes are spontaneously reversible and have no bearing on the patients long-term outcome. In this analysis thousands of cachexia episodes that have been recorded at times > 10 years since diagnosis, and I would wonder why these would be included in the analysis since at 10 y or indeed, even 5 y out, some patients would be considered cured of their cancer. It makes no sense to record as cancer cachexia episodes of weight loss that would likely to be coming from non-malignant causes and at these far distant time points it is necessary to know whether they related to recurrence, a second malignancy, or a non-malignant acute or chronic condition. Consider separate evaluation of patients diagnosed at early stage of disease who did not show recurrence. Also, the MSK data set is very likely to include dates of systemic therapy, and it would be of great value to identify if therapies align with weight loss and the extent to which therapy-related weight loss is regained, especially in patients treated with curative intent.

Response: We agree that weight loss events occurring many years after diagnosis may reflect recurrence, a second malignancy, or non-malignant conditions rather than cancer-associated cachexia. Because recurrence status is not reliably available at scale in this dataset, we addressed this concern with two pragmatic sensitivity restrictions.

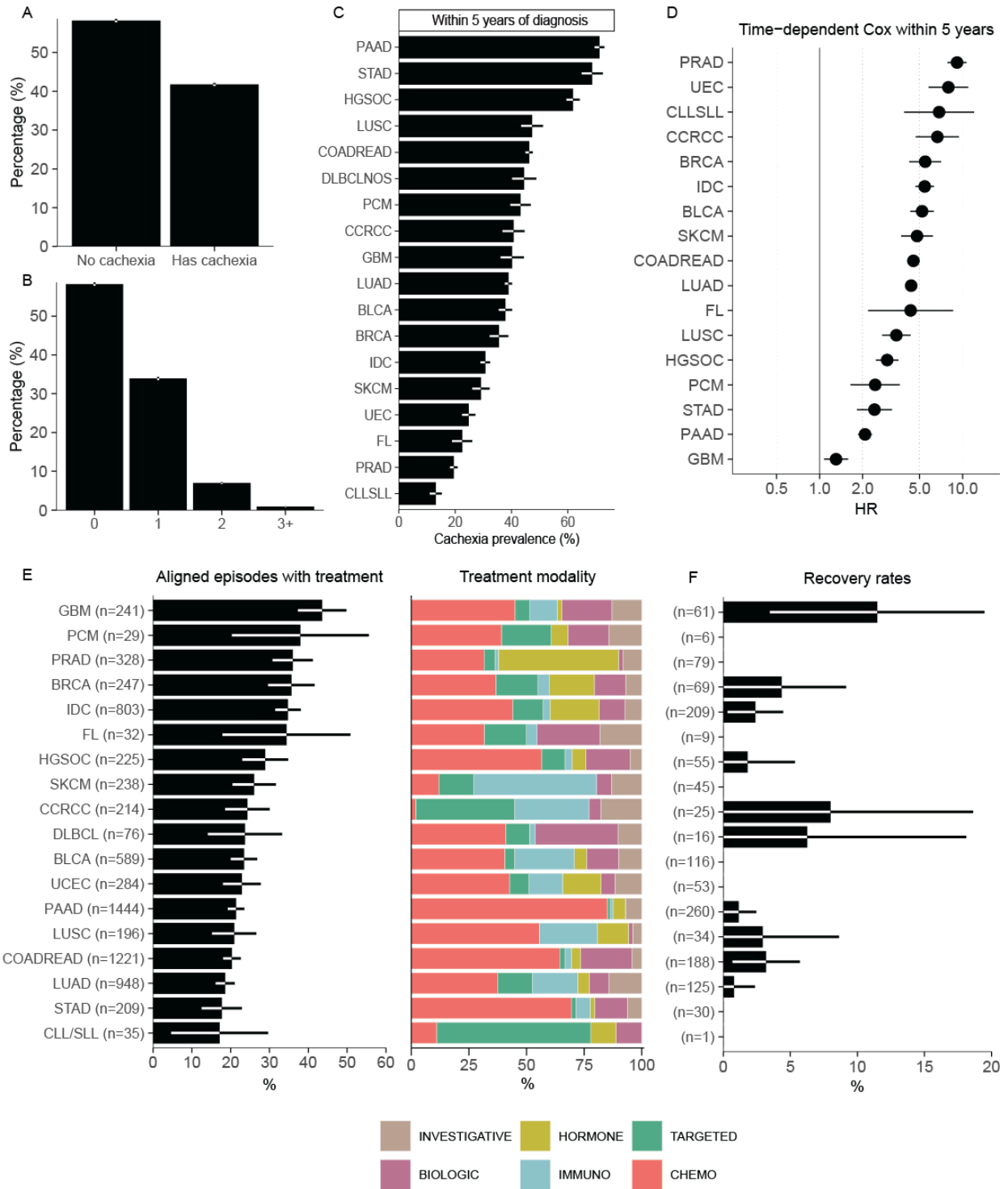
First, we repeated the primary analyses restricting cachexia episodes to those occurring within 5 years of diagnosis (**Fig. R3a.ii A–D, below**), thereby excluding long-latency events most susceptible to non-malignant attribution. In the revised cohort, late-onset events were uncommon: only 1,556 patients had their first episode >5 years after diagnosis, and 133 patients with >10 years after diagnosis. Restricting to this clinically plausible window did not materially change the overall conclusions. Specifically, the direction and relative magnitude of

the cancer-type survival associations and the serologic effect estimates were preserved under the 5-year restriction, and the progression-overlap enrichment signal remained consistent. This supports the primary signals we report reflect weight loss dynamics during the period clinically close to initial management rather than heterogeneous non-malignant causes that may accrue over long-term follow-up.

Second, to address the Reviewer's point that curative-intent therapy can cause transient, reversible weight loss, we performed a focused analysis in patients with Stage I–III disease. Using systemic therapy dates, we tested (i) whether cachexia episodes frequently occur during or near active treatment (**Fig. R3a.iiE, below**) and (ii) whether these episodes resolve promptly after treatment completion (**Fig. R3a.iiF, below**). We defined therapy alignment by selecting the closest systemic therapy that started within 180 days prior start and was ongoing at start. While a subset of early-stage episodes temporally overlap with therapy, recovery by our pre-specified criterion ($\geq 5\%$ BMI gain from the episode low) within 180 days after treatment end was infrequent. This pattern is not consistent with the majority of events representing short-lived, spontaneously reversible treatment-related weight loss, and supports that the episodes captured by our definition reflect more persistent weight loss dynamics even in patients treated with curative intent.

Figure R3a.ii. Sensitivity analyses for late-onset and therapy-associated weight loss. (A) Patient-level cachexia incidence within 5 years of diagnosis, showing the proportion of patients who ever meet cachexia criteria versus never cachectic. (B) Distribution of cachexia episode burden per patient (0, 1, 2, or ≥ 3 episodes). (C) Cancer type–stratified cachexia prevalence. (D) Cancer type–specific time-dependent Cox models (within 5 years of diagnosis) estimating the association between time-varying cachexia status and overall survival (E) Among Stage I–III cachexia episodes, proportion aligned with systemic therapy near episode onset (left) and treatment modality composition of the selected closest qualifying therapy among aligned episodes (right). (F) Recovery rates after treatment completion among Stage I–III therapy-aligned episodes where treatment ended before the cachexia episode ended, defined as reaching $\geq 5\%$ BMI gain from the episode low within 180 days after treatment end

R.3.a.ii



iii) obese persons for whom 5% weight loss is documented to have minimal clinical impact. I address this point further in context of discussion of cachexia severity

Response: Because a uniform WL5 cutoff could, in principle, overcall low-severity events in individuals with obesity, we assessed severity using the validated weight-loss grade system (WLGS; Martin et al., 2015)³, which explicitly incorporates baseline BMI (Supplemental Fig. 2C–D). Among episodes with baseline BMI >25, 56.0% fall into high-severity WLGS Grades 3–4 (10,883/19,448) and 33.3% are Grade 2, indicating consensus-defined events in higher-BMI patients are predominantly moderate-to-severe rather than weight loss attributed to fluctuations.

Additionally, we evaluated how often consensus-defined events in overweight and obese patients also met more stringent weight-loss thresholds (Fig. R3a.iii). Among consensus-positive patients with obesity, 47.2% also met a $\geq 10\%$ weight-loss definition and 18.8% met $\geq 15\%$; overweight patients showed similar overlap (42.5% and 15.3%, respectively). As expected, fewer episodes satisfy 10% or 15% WL by design, but the weight-loss distributions in both groups extend well beyond the 5% cutoff, indicating that a substantial fraction of consensus-defined events in higher-BMI patients reflect higher-magnitude, clinically meaningful loss rather than fluctuations (**Page 5-6**).

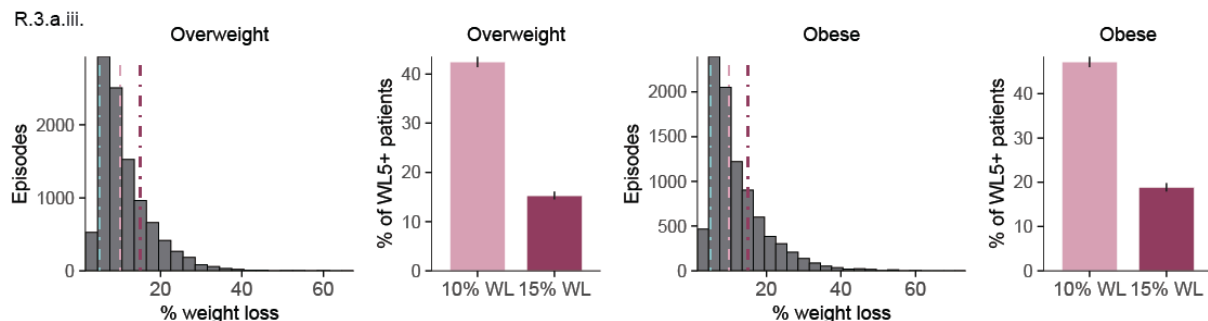


Figure R3a.iii. Severity of WL5-defined weight loss in overweight and obese patients. (Histograms) Net percent weight loss among WL5-defined episodes is shown for baseline overweight and obese groups, with reference thresholds at 5%, 10%, and 15%. (Bar plots) Among WL5-positive patients in each BMI group, the proportion also meeting $\geq 10\%$ and $\geq 15\%$ weight loss thresholds

b) The authors have not used either up-to-date or consensus criteria for the diagnosis, classification and severity of cachexia. They cite several key papers by the international consensus group (Fearon et al 2011 Lancet Oncology; Martin et al 2015 J Clin Oncol) but claiming the use of “simplification” revert to a crude historical definition of any weight loss >5%.

According to Fearon et al. 2011: the agreed diagnostic criteria:

- Weight loss >5% over past 6 months (in absence of simple starvation); or
- BMI <20 kg/m² and any degree of weight loss >2%; or
- skeletal muscle index consistent with sarcopenia and any degree of weight loss >2%;

The authors chose to “simplify” i.e. ignore 2 of these 3 criteria. As such, it is misleading to state that this work included cachexia criteria that “satisfy” international consensus criteria, since they were not used.

Response: Thank you for raising this concern. We agree with the Reviewer on the importance of addressing this and aligning this analysis with consensus weight loss criteria. We agree that our initial framing was overly simplified, and that we did not clearly state which elements of the international consensus definition were applied. We have now incorporated the low-BMI modifier, (ii) >2% weight loss when BMI <20 kg/m², in addition to the (i) >5% weight loss criterion over 6 months (**Methods, Page 20-21**). We cannot implement the sarcopenia-based criterion because skeletal muscle index is not available across the full cohort, and we now state this explicitly as a limitation (**Page 17**). As we state in the introduction to this rebuttal, these exact weight loss criteria have been used as eligibility criteria in the recent Phase 2 trial of ponesegromab, and thus have high real-world utility. **We have now re-run all downstream analyses using these updated consensus weight-loss criteria.** Finally, we removed language in the manuscript that could be interpreted as applying the full consensus definition.

Furthermore, a key consensus statement was that severity can be classified according to the rate of ongoing loss of weight in combination with the concurrent degree of depletion of energy stores and body protein mass (which can be compounded by a low initial reserve). This tenet was formalized in the consensus group's 2015 analysis (Martin et al. 2015) with 12,000 patients to derive the grades of severity of cancer-associated weight loss. It does seem disconnected with current classification of cachexia, to sidestep the notion of severity. The weight loss grade schema makes clear the limitation of using a 5% across the board cutoff. For patients of BMI <20 kg/m² at the outset, a weight loss of 5% is Grade 3, whereas for a person of initial BMI 30, it is only Grade 1. As such, the current analysis is likely to contain large numbers of patients of high initial BMI who are miscoded as having had "cachexia episodes". The number of obese people in his cohort is not disclosed, but one can presume to be substantial after reviewing Supplemental Fig 2.

Response: We agree with the reviewer that cachexia severity should be made explicit. Accordingly, we annotated 5% weight-loss episodes using the Martin et al. weight-loss grading system (WLGS) in our updated cohort, including the consensus low-BMI modifier (≥2% weight loss over 6 months if baseline BMI <20 kg/m²). We now report baseline BMI composition in the manuscript **Table 1**. The cohort includes a substantial higher-BMI fraction (35.7% overweight; 29.7% obese; 32.8% normal weight; 1.9% underweight), motivating a severity-aware view beyond a uniform 5% cutoff.

To directly address that uniform 5% cutoff could misclassify high-BMI patients as cachectic, we mapped each WL5 episode onto the Martin et al. WLGS matrix and summarized grades for episodes with baseline BMI of ≥28 kg/m². In this high-BMI subset, episodes were largely moderate-to-severe by WLGS (WLGS ≥2: 84.6% [11,211/13,258]; WLGS 0–1: 15.4% [2,047/13,258]), indicating that WL5 episodes in obese patients are not predominantly mild grade-1 events. As an additional severity-aware check, we repeated the pan-cancer mixed-effects serology analysis using WLGS 2–4 cachectic spans versus non-cachectic spans, and the resulting tissue-agnostic lab signature remained consistent. We also re-ran the somatic mutation analysis only on individuals with episodes of at least WLGS 2. Results likewise

remained consistent and are described in detail in our response to Reviewer 1 (**Comment 2 Novelty genomic analysis; Fig. R1.1.A**).

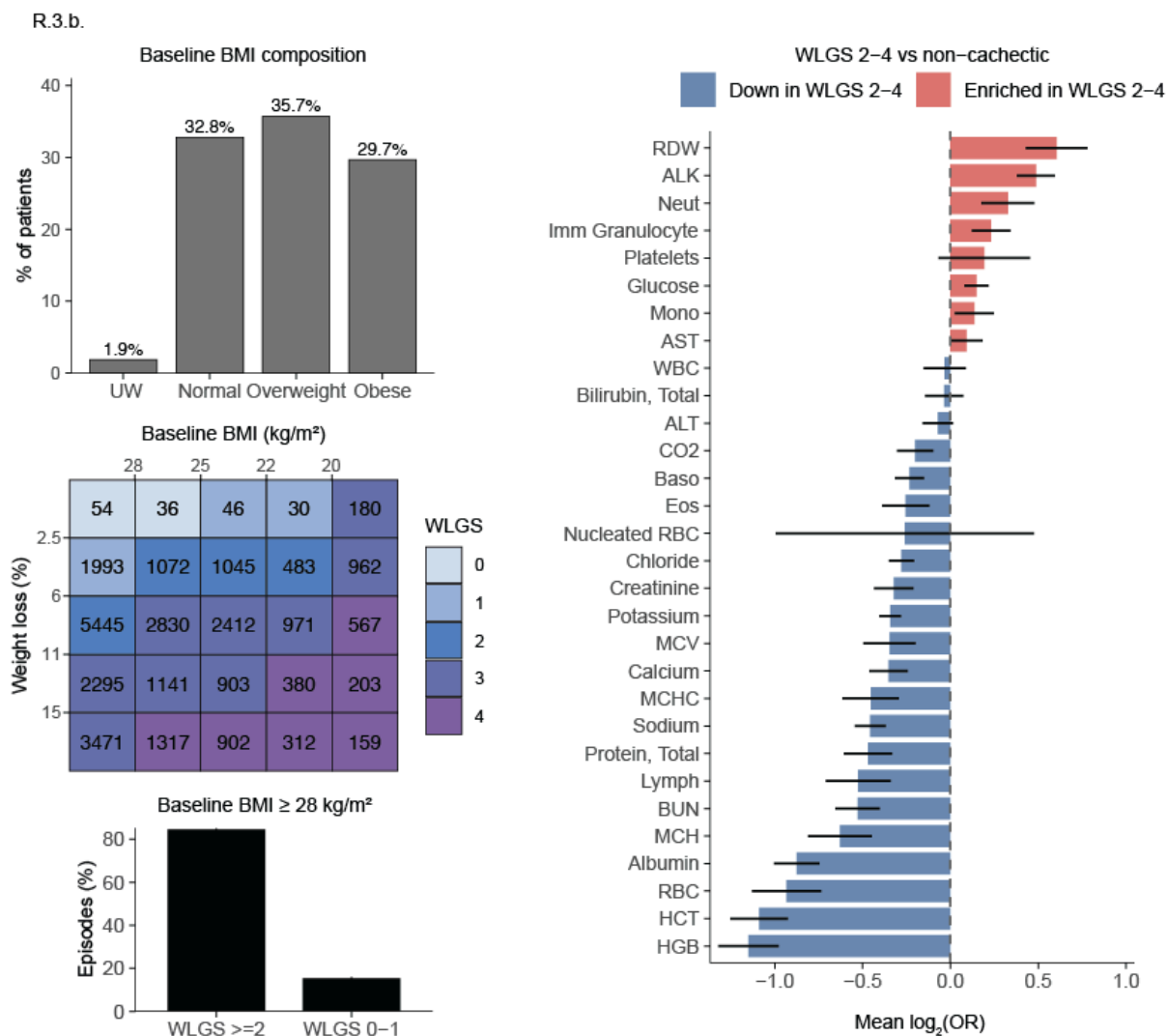


Figure R.3b. Baseline BMI composition of the revised cohort; WLGS matrix (baseline BMI $\times$ % weight loss) with episode counts; and, among episodes with baseline of BMI ≥ 28 kg/m², the proportion classified as WLGS ≥ 2 versus WLGS 0-1. *Right:* Tissue agnostic signature (WLGS 2-4 vs non-cachectic)

The present analysis reverts to pre-consensus criteria, which focused on an arbitrary minimum degree of weight loss (5%) and does not provide a meaningful integration of severity in the analysis. The analysis provided is the prevalence of $>5\%$ incident weight loss (primary outcome); this would be acceptable language to describe what was done.

Response: We agree that a uniform 5% threshold alone does not integrate severity. In the revision, we rebuilt the cachexia classification pipeline to align with the consensus definition by applying the 6-month weight-loss criterion ($>5\%$ over 180 days) together with the low-BMI

modifier (BMI <20 kg/m² with >2% loss over 180 days) (**Methods, Page 20-21**), and we now explicitly incorporate severity by annotating qualifying episodes using the Martin et al. weight-loss grade schema (WLGS) and reporting baseline BMI composition and WLGS distributions (**Page 5-6, Methods- Page 21; Supplemental Fig. 2F-H, 2I**). Where appropriate, we have also clarified the language to describe the primary outcome as incident ≥5% weight loss over 6 months (with the low-BMI adjustment), implementing weight-based components of the consensus.

4. The data set provided with the submission is entirely insufficient to recapitulate any of the presented analysis.

Response: Thank you for raising this. We have updated the shared materials and documentation to include the additional fields. We address each point below.

a) This data set appears to align with the results section in so far as it contains 59,493 unique individuals, and the total count of cachexia episodes is 37,545. However, the data set has no legend. It contains no relevant dates, no indication of the duration of follow-up for each patient, and indeed no weight data are included. The only information is minimal and is demographic (sex and age and primary tumour site (but not stage)); the start and end dates of “cachexia episodes” are the only available data and these unfortunately are not associated with any information as to the number of weight data per patient, the duration of follow up and time from last weight record to death or censoring. The definition of “end date” is unclear – in some instances it might mean that weight was regained, in others either the patient died or the weight record was discontinued.

Response: We agree and note that this overlaps with Comment 2.c regarding longitudinal coverage. To make the shared materials interpretable, we have added a data dictionary for **Supplementary Table 18 - De-identified episodes summary (data_dictionary.txt)** and **expanded the README** to define each file and column and to describe the cachexia episode-calling procedure. We also updated the shared episode-level table to include de-identified timing relative to diagnosis (day 0) and per-patient follow-up context, including BMI measurement count, follow-up duration, and time from last BMI measurement to death or censoring. Finally, we clarified that the episode end is defined as the last qualifying BMI measurement within the merged, window-defined episode (after merging overlapping 180-day windows). This endpoint is based only on observed BMI measurements and does not imply weight regain. If BMI measurements stop thereafter, including due to death, the episode does not extend beyond the last qualifying measurement (**Page 21**).

b) Certain data are very difficult to understand – some examples which make me wonder whether these data were carefully curated and examined.:

Patient P – 000-4032<tel:000-4032> a 7 year ols male with anaplastic ganglioma had an episode of >5% weight loss 15,369 days after his diagnosis, some 42 years later. Patients with pediatric cancers do have very long-term surveillance but is this cachexia?

Patient P0066131, a male infant with sarcoma 0 years of age, had an episode of >5% weight loss 3 days duration during his first year of life. is this cachexia?

P – 001-4433<tel:001-4433> a woman with gastrointestinal stromal tumour had an episode of 5% weight loss of 10 days duration 10.8 years after their diagnosis. A total of 2125 cachexia episodes occurred 3650 days (10 years) after diagnosis or later. Are these episodes related to malignant disease? Are these people being followed up annually at 10 years out and beyond, and was a cachexia episode be defined by 2 weight records a year apart?

Response: Thank you for highlighting these edge cases and we agree that they require careful cohort curation and clear documentation of inclusion criteria. As described in our response to *Comment A.1.*, we rebuilt the cohort with an explicit stepwise derivation (**Supp Fig. 1A**) and expanded Methods describing exclusions (Methods, Study design and cohort). Briefly, starting from 79,240 MSK-IMPACT–profiled patients, we excluded patients with >1 primary cancer, excluded anchors prior to the BMI sampling, excluded pediatric patients (<18 at diagnosis), excluded patients with no BMI data, and required minimum longitudinal coverage (≥ 2 BMI measurements and ≥ 180 days between first and last BMI). All analyses and shared files were rerun using this cleaned adult cohort.

Consistent with these criteria, two of the cited records are pediatric and are excluded in the revised cohort (P-0004032, age 7 at diagnosis; P-0066131, age 0 at diagnosis). We also reviewed the cited adult GIST example (P-0014433). In some cases, the EHR contains multiple tumor diagnosis-date entries, and using the earliest recorded diagnosis date as anchor can make a subset of episodes appear to occur very late in follow-up (≥ 10 years), even though the malignancy course represented by the MSK-IMPACT sample began later (**Methods, Page 18**). We have refined cohort anchoring by selecting the entry most consistent with the tumor resection/pathology and sequencing event for patients with multiple diagnosis entries. This reduces the misleading classification of episodes as occurring ≥ 10 years after diagnosis due to an earlier historical diagnosis-date entry, including the timing interpretation raised for P-0014433. In the revised cohort, late-onset events are uncommon. Only 133/23,372 cachectic patients (0.57%) have their first detected cachexia episode beginning ≥ 10 years after the diagnosis. Lastly, a cachexia episode cannot be defined by two annual weights a year apart in our framework because detection requires crossing the weight-loss threshold within a fixed 180-day evaluation window anchored to observed measurements.

Columns H and I defined the start day and the end day of each diagnosis of incident 5% weight loss. There is no explicit description in the method section of Day 0. However, time, since the date of diagnosis seems to be the metric in all the data presentations, so I've taken this to mean time from date of diagnosis. Then.... 907 lines of data have a start day of an episode on day zero. This is very difficult to understand. This could mean one of two things did the person get weighed twice on their date of diagnosis and one of those weights with 5% lower than the earlier one? Alternatively, are there weight records for individuals that precede with their date of diagnosis? Likewise, there are about 2000 low single digit start dates in column H some clarification is required as to where these came from. As mentioned above, the lack of explicit definition of "end date" of a cachexia episode detracts greatly from the clarity of this presentation.

Response: We have clarified the time origin and episode fields in the **Methods** and in the **data dictionary**. In the shared episode table, Day 0 corresponds to the diagnosis anchor, and all episode timing fields are recorded as days relative to this anchor. BMI measurements recorded before Day 0 are excluded from episode calling. For instance, *start_day* is the episode start day standardized relative to the anchored tumor diagnosis date. Importantly, *start_day* denotes the baseline day that anchors the 180-day evaluation window, not the day the $\geq 5\%$ threshold is first reached. Therefore, episodes can have *start_day* = 0 when a BMI measurement is available on Day 0 and the $\geq 5\%$ loss is observed within the subsequent 180 days. The threshold-crossing day is captured separately as onset day. Likewise, small single-digit *start_day* values reflect early baseline measurements shortly after the anchor date, not instantaneous weight loss at diagnosis. We also clarify that *end_day* is the last BMI measurement that still satisfies the weight-loss criterion within the merged, window-defined episode.

Reviewer #1 (Remarks on code availability):

This code is unevaluable, as the data provided cannot be used to test it. See comments to authors where I make the case that at least some of the data should have been excluded, that the classification criteria for cachexia and for episodes are unclear/incorrect.

Response: Thank you for the detailed feedback on cohort curation, episode definitions, and the shared materials. We have addressed these points throughout the revision by explicitly documenting cohort construction and exclusions (**Supp Fig. 1A; Methods Page 20-21**), clarifying the cachexia episode definitions, and regenerating all shared episode-level outputs using the curated cohort (**Supplemental Table 18, Data Dictionary**). To improve clarity and usability, we also added a data dictionary describing each column, and the key entries and added step by step methods to run episode identification (**episode_identification_walkthrough.ipynb, Sample mock data attached below**)

Reviewer #2 (Remarks to the Author):

Castillon et al. present the manuscript “A health-system-scale, episode-resolved multimodal atlas of cancer cachexia.” Using longitudinal clinico-genomic data from 59,493 individuals with cancer, the authors integrate biomarker, genetic, and clinical time-series data to identify signals associated with cachexia across a broad range of malignancies, including 656 distinct cancer histologies. Overall, the study identifies serological and genetic correlates of cachexia, some of which appear conserved across multiple cancer types, suggesting shared biological mechanisms. The work addresses an important and unresolved question in cancer metabolism and leverages an unusually large real-world dataset. However, the terminology “cachectic episodes” is potentially misleading and requires clarification. The phrasing implies reversibility and cyclicity of cachexia, which stands in contrast with its current conceptualization as a progressive clinical syndrome characterized by muscle loss, systemic inflammation, metabolic dysregulation, and declining function. If the authors intend to propose that aspects of cachexia may be modifiable in specific clinical contexts, this point should be explicitly justified and discussed, as it challenges long-standing assumptions in the field.

Response: Thank you for this feedback. In the manuscript, we use episode in a strictly operational sense to denote a predefined interval of clinically meaningful weight loss from longitudinal BMI data, periods that meet the weight-loss component of the consensus (Fearon) definition within measurement-anchored 180-day windows: (i) 5% WL over 180 days, and (ii) 2% WL if BMI < 20 kg/m². We have revised the text at first mention to make clear that this terminology reflects an episode-resolved application of the consensus weight-loss criterion, while the broader clinical syndrome of cachexia involves additional features such as skeletal muscle loss and functional decline that are not uniformly captured in this dataset (**Page 4**).

Major comments

1. Cachexia definition and staging

The use of the Fearon criteria to define cachexia is appropriate, and the adjusted weight-loss threshold is reasonable. However, Fearon et al. differentiate between pre-cachexia, cachexia, and refractory cachexia. This distinction is well established and widely used. The lack of reference to staging may limit interpretability for readers, and this should be acknowledged—at minimum in the Limitations.

Response: We agree that the distinction between pre-cachexia, cachexia, and refractory cachexia is clinically important. Pre-cachexia is characterized by early metabolic changes with limited weight loss, cachexia by ongoing weight loss accompanied by metabolic and functional impairment, and refractory cachexia by advanced disease with limited response to therapy and poor prognosis (*Fearon et al., Lancet Oncology, 2011; Martin et al., J Clin Oncol, 2015*). Staging was not performed in this work because key components required for classification, including longitudinal functional status, treatment resistance, and consistent muscle mass measurements, are not uniformly available at scale in this dataset. We now state this explicitly in the **Limitations (Page 17)**. As a complementary context, we report weight-loss severity using the

Martin et al. grading framework to stratify the magnitude of weight loss within BMI-defined cachexia intervals (**Page 5, 21**).

2. Impact of tumor resection

There is ongoing debate regarding whether complete tumor resection is required to halt or reverse cachexia. It is unclear whether surgical status or resection margins were accounted for in the analyses. If available, incorporating or commenting on these variables would strengthen the validity of conclusions.

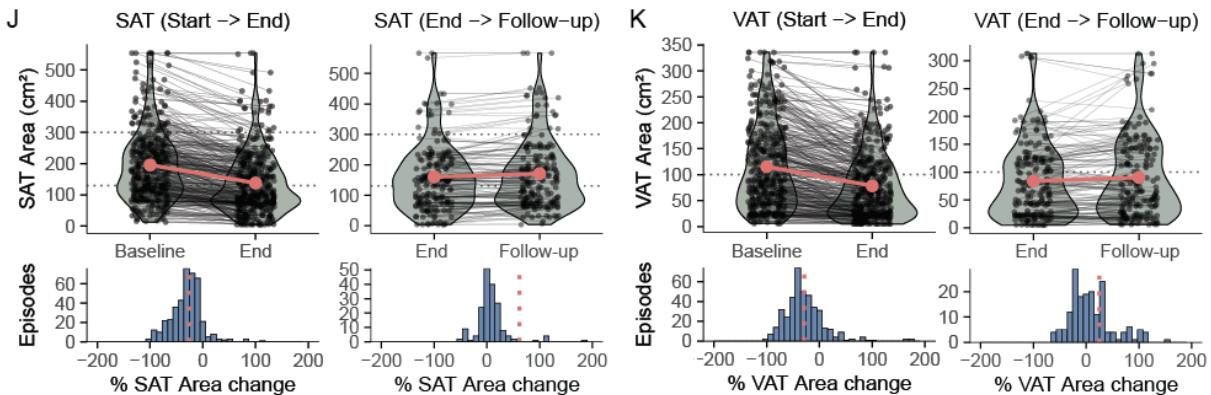
Response: We recognize that tumor resection, including complete resection, can influence weight trajectories and cachexia dynamics in some settings. However, surgical status and margin information are not available with sufficient completeness across cancer types in this dataset to support a consistent adjustment, and therefore were not included in the primary analyses. We now note this explicitly as a limitation (**Limitations; Page 17**).

3. Body composition data

The manuscript notes that fat-predominant weight loss may be reversible, whereas profound skeletal muscle loss is generally not. CT-based body composition assessment is mentioned. If such data exist for a subset of patients, their inclusion or descriptive reporting would meaningfully enhance the interpretation.

Response: For a subset of lung adenocarcinoma patients with available abdominal CT imaging, we incorporated CT-based body composition measurements to provide compartment-level context for the BMI-defined cachexia intervals. Using the identified BMI-defined weight-loss episodes, we selected up to three scan timepoints per episode: the scan closest to the episode start/baseline, the scan closest to the episode end, and the closest available follow-up scan occurring after the episode end/follow-up. Subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT) areas were quantified at the L3 level using the standardized segmentation workflow described in **Methods (Page 25)**. The segmentation workflow used for this subset quantified SAT/VAT and did not generate skeletal muscle measurements.

In paired “start–end” scans comparing the end of an episode to the beginning of an episode (loss cohort: 409 episodes, 228 patients), mean SAT decreased by 23.6% and mean VAT decreased by 27.5%, consistent with substantial adipose depletion over the BMI-defined weight-loss interval. In paired end–follow-up scans comparing followup to end of an episode (follow-up cohort: 188 episodes, 95 patients), mean SAT increased by 34.8% and mean VAT increased by 25.7%. Post-episode changes were heterogeneous across episodes; 66.1% showed SAT increase and 58.7% showed VAT increase at follow-up (**Page 6-7; Supplementary Fig. 2 J-K**). These measurements show marked adipose loss across compartments during episodes, with partial regain on follow-up, suggesting heterogeneous adipose dynamics during cachexia-associated weight loss.



Supplementary Figure 2J-K. Adipose compartment dynamics during and after BMI-defined weight-loss episodes in lung adenocarcinoma.

4. Clinical validity of “episodes”

A key unresolved question is whether the observed weight-loss patterns represent clinically defined cachexia or non-specific body-weight fluctuations unrelated to the severe metabolic and functional phenotype of cachexia. This distinction is essential and should be addressed directly.

Response: This is an important point. We identify sustained weight-loss intervals using the consensus weight-loss criterion applied to longitudinal BMI data, and we now state explicitly that the accompanying functional and metabolic features used to define cachexia cannot be uniformly assessed at this scale. To minimize the contribution of nonspecific body-weight fluctuations, these weightloss episodes are derived from smoothed BMI trajectories and evaluated within fixed 180-day windows anchored to observed measurements, with additional quality-control steps to exclude implausible patterns, including rapid fluctuations consistent with edema-related fluid shifts (**Methods, Page 20-21**). Finally, to address the concern that residual noise could drive our findings, we show that the same serologic and clinical associations are observed when (1) analyses are repeated under more stringent weight-loss thresholds (10% and 15%) and (2) when episodes are stratified by weight-loss grade (Martin et al.) (**Pages 5,6,8,10, 12-13; Supplementary Fig. 2B, 3C, 4D**).

In the README.me file the meta data include pathological stage (STAGE_CDM_DERIVED_GRANULAR) but clinical stage is not mentioned in the README.me file. In the manuscript, however, clinical stage is mentioned (line 364/365) but not pathological stage. Please clarify which stage was used.

Response: We used the AJCC clinical stage group at diagnosis (STAGE_CDM_DERIVED) for the stage-stratified analyses reported in the manuscript and have updated the README file.

Minor comments

Thank you for these helpful minor suggestions and have amended the manuscript accordingly, as detailed below.

1. Line 47: Please add a reference after the first sentence.

Response: We have amended the manuscript and added the following citations (Ref. 1 Rohm et al; Ref.2 Flint et al) (**Page 3**).

2. Line 57: Please add the prominent role of muscle loss related to cachexia.

Response: We have amended the text to explicitly emphasize skeletal muscle loss as a defining and clinically important component of cancer cachexia (**Page 3**).

3. Line 67: Please add that another major issue in cachexia research is inconsistent diagnostic definitions across studies.

Response: We agree and have added this point to the manuscript, with supporting citations. (**Page 3**).

4. Line 193: Use the term refractory cachexia, which is defined as cachexia at the end of life.

Response: We have revised the text to use the term *refractory cachexia* when referring to cachexia at the end of life (**Page 8**).

5. Given the rise in early-onset cancer often diagnosed at later stages, why did the authors not consider age at diagnosis? Particularly for colorectal cancer, which comprises a relatively large part of the study population and provided novel results.

Response: We agree that age at diagnosis is an important factor. Age at diagnosis was included as a covariate in the all key downstream analyses, and we have clarified this explicitly in the Methods (**Page 19, 22, 23**).

6. BMI outliers (<10 and >100) are described in the ReadMe file. Please provide justification or reference for these thresholds.

Response: We applied BMI <10 and >100 kg/m² as conservative filters to remove biologically implausible values most consistent with measurement or data-entry/unit errors. We have clarified this in the Methods (**Page 19**).

7. The manuscript does not clearly describe the handling of missing data and other preparatory steps; given the scale of the dataset, this requires clarification. How many variables were complete if there was missing, was there imputation?

Response: We have expanded the Methods to clarify preprocessing and missing-data handling. We did not perform imputation for clinical covariates; models were fit using available data for required variables, with complete-case analysis for the covariates included in each model (**Page 19**).

8. Ethical oversight remains unclear. Written informed consent is mentioned, but IRB or equivalent approval is not specified.

Response: We have added an explicit ethics oversight statement to the Methods indicating that the study was approved by the Memorial Sloan Kettering Cancer Center IRB under protocol 12-245 (Appendix C) (**Page 18**).

9. Two tables at the end of the manuscript PDF lack titles, legends, and formatting and require revision.

Response: We have revised these tables to include titles, legends, and consistent formatting (**Pages 38-45**).

10. In the Editorial Policy checklist, the competing interests section appears incomplete.

Response: We have updated the Editorial Policy checklist to provide a complete competing interests statement for all authors and added this in the Manuscript (**Editorial Policy checklist, Page 33**).

Reviewer #2 (Remarks on code availability):

BMI outliers (<10 and >100) are described in the ReadMe file. It is unclear how these cut-offs were chosen. 13047 patients were excluded if only one BMI measurement was available and trajectory less than 180 days. Were complete data available remaining for the other patients. How were missing data handled for covariates, particularly in the survival analysis. In the README.me file the meta data include pathological stage (TAGE_CDM_DERIVED_GRANULAR) but clinical stage is not mentioned in the README.me file. In the manuscript, however, clinical stage is mentioned (line 364/365) but not pathological stage. Please clarify which stage was used.

Response: The BMI plausibility filters (<10 and >100 kg/m²) were applied as conservative thresholds, as mentioned above, to remove biologically implausible values most consistent with measurement or data-entry/unit errors, and we now state this explicitly in the Methods (**Page 19**). Moreover, patients excluded for insufficient BMI coverage, with single BMI-measurement, were removed prior to all downstream analyses. We have also clarified missing-data handling for covariates, including in survival analyses where applicable (**Methods Page 19**). We did not impute covariates and instead used complete-case data for the covariates included in each model. Finally, we clarified the stage variable used for manuscript analyses STAGE_CDM_DERIVED was clinical stage from AJCC and updated the **README** to document.

Reviewer #3 (Remarks to the Author):

Summary of the key results

The paper builds a health-system–scale, episode-resolved atlas of cancer cachexia by detecting longitudinal BMI-defined weight-loss episodes across multiple solid tumors. It shows that these episodes are common, concentrate around periods of disease progression, and associate with worse overall survival using time-varying Cox models. The authors identify a broadly consistent serologic signature of cachexia (e.g., low albumin/hemoglobin, high ALK and neutrophils) and link specific somatic alterations (notably TP53, STK11, SMAD4) to increased cachexia risk. Finally, they develop cancer specific prediction models for cachexia risk using serologic and genomic features.

Originality and significance

The work is original in combining: (i) episode-resolved cachexia detection from routine EHR BMI data and (ii) multimodal linkage to serologic markers, genomics, and progression across many cancer types. They also predict cachexia risk in two cancer types. Prior studies have examined longitudinal weight loss and survival in specific cohorts (for example, a study using Non Small Cell Lung Cancer trials PMID: 32940014, and Gastrointestinal cancers PMID: 36165100), proposed algorithmic cachexia criteria on EHR data in two cancers (<https://doi.org/10.1002/rco2.107> <[https://urldefense.com/v3/__https://doi.org/10.1002/rco2.107_!!KVo1iE!QkqjQ3Ni8KL-7DcfKuQxLUayYxEuja12BLiZqcWmpUek8KjaVU8S5SDXpA0_5TOCrduC4JbqwR8_DRxLA\\$>](https://urldefense.com/v3/__https://doi.org/10.1002/rco2.107_!!KVo1iE!QkqjQ3Ni8KL-7DcfKuQxLUayYxEuja12BLiZqcWmpUek8KjaVU8S5SDXpA0_5TOCrduC4JbqwR8_DRxLA$>)), and there are emerging real-world cachexia prediction models using routine data (PMID: 39109281). This manuscript is more comprehensive and multimodal than those, so the contribution is both novel and timely.

Response: Thank you for placing our work in the context of prior longitudinal weight-loss and cachexia modeling work, and highlighting its novelty.

Data & methodology: validity of approach, quality of data, quality of presentation

Using large-scale EHR data to algorithmically detect cachexia episodes is an appropriate and powerful approach. The EWMA-smoothed BMI trajectories, sliding-window episode detection, and edema-related QC are reasonable in spirit, but key implementation details (window step, merging rules, recovery definition) need clearer specification and some parameter choices (α , 180-day window, 15-day minimum) are insufficiently justified. The multimodal linkage (labs, genomics, treatments, progression) is a major strength, but the handling of missingness, progression annotation sources, and supportive care interventions is under-described. Overall data presentation (figures/tables) is clear, though some supplemental figure labeling and legends require minor corrections.

Response: We have revised the **Methods (Page 20-21)** to more fully specify the episode-detection implementation, and we now justify key parameters and report sensitivity analyses demonstrating that the main findings are robust to reasonable parameter choices. We also expanded descriptions of multimodal linkage, including missingness and progression annotation sources, and corrected minor supplemental labeling/legend issues where needed. We address each item in full detail below under **Suggested Improvements**.

Appropriate use of statistics and treatment of uncertainties

Time-varying Cox models for survival, mixed-effects models for serologic signatures, and logistic / risk-prediction models are appropriate choices for the questions posed. However, the manuscript provides limited information about proportional hazards checks, calibration, or other model diagnostics, and treatment of uncertainty is mostly via confidence intervals rather than deeper sensitivity analysis. The suggested parameter and subgroup sensitivity analyses would substantially strengthen the statistical credibility.

Response: Thank you for this point. We have expanded the **Methods (Page 14, 22)** to report model diagnostics and validation procedures.

(A) We assessed proportional hazards for cachexia in each cancer-specific time-dependent Cox model using Schoenfeld residuals (*cox.zph*), and provide per-cancer test results in the table below with representative residual plots for COADREAD and PRAD. For cancer types showing evidence of non-proportionality, we additionally fit models with an explicit time-varying cachexia effect (*tt(cachexia)*) and report time-specific hazard ratios (at 6, 12, 24, and 60 months) to characterize how the association changes over follow-up. In these models, the cachexia association remained statistically significant and in the same direction, indicating that our survival conclusions are not driven by proportional-hazards violations (**Page 22, Supplementary Table 7**).

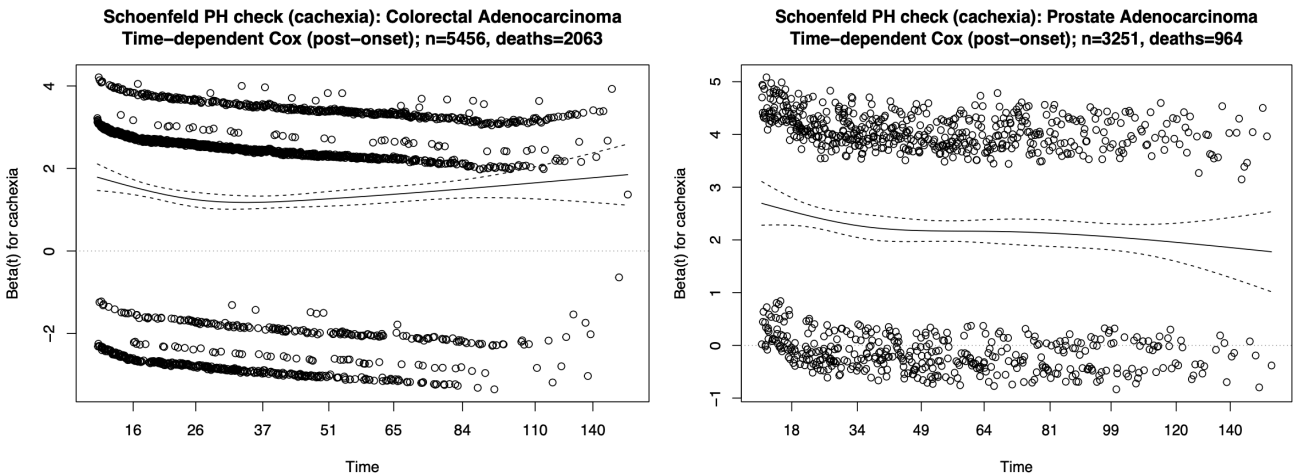


Figure 3.C. Representative Schoenfeld residual plots for the cachexia term in time-dependent Cox models.

Table 3.1. Proportional hazards diagnostics for cancer-specific time-dependent Cox models (Schoenfeld residual tests for the cachexia term and global test).

Cancer_Type	N_patients	N_deaths	PH_p_cachexia	PH_p_global	violated_cachexia
BLCA	1,625	532	0.199	0.001	No
IDC	2,818	898	0.185	9.1e-06	No

CLLSLL	946	61	0.031	0.583	Yes
COADREAD	5,456	2,063	0.011	2.9e-30	Yes
SKCM	840	284	2.2e-04	0.002	Yes
DLBCLNOS	508	127	0.983	0.154	No
FL	525	40	0.849	0.887	No
GBM	542	453	0.522	0.421	No
HGSOC	1,635	764	0.89	0.006	No
BRCA	791	316	0.186	0.023	No
LUAD	5,367	2,463	0.01	2.7e-12	Yes
LUSC	643	344	0.886	0.025	No
PAAD	2,608	1,789	0.076	5.9e-07	No
PCM	727	126	0.342	0.254	No
PRAD	3,251	964	0.002	3.5e-06	Yes
CCRCC	609	211	0.377	0.002	No
STAD	593	318	0.51	0.288	No
UEC	1,272	203	0.3	0.034	No

Table 3.2. Time-varying cachexia effects in cancer types with proportional hazards violations (HR estimates at 6, 12, 24, and 60 months from tt(cachexia) models) (**Supplementary Table 7**)

Cancer	p_ccx_base	p_ccx_tt	HR_at_6	HR_at_12	HR_at_24	HR_at_60
CLLSLL	1.03e-05	0.00232	45.02 (11.29–179.55)	22.87 (8.46–61.84)	11.57 (5.95–22.51)	4.69 (2.69–8.18)
COADREAD	8.55e-16	0.00203	5.86 (4.46–7.69)	4.97 (4.15–5.96)	4.22 (3.76–4.73)	3.39 (2.94–3.91)
SKCM	4.54e-09	0.000358	16.98 (8.30–34.74)	10.90 (6.65–17.86)	6.97 (5.13–9.48)	3.86 (2.90–5.13)
LUAD	1.59e-18	0.018	4.86 (3.96–5.97)	4.40 (3.84–5.04)	3.99 (3.64–4.36)	3.49 (3.08–3.96)
PRAD	4.08e-20	0.00101	17.89 (11.93–26.82)	14.35 (10.77–19.11)	11.49 (9.55–13.84)	8.56 (7.30–10.04)

(B) Mixed effects models (serologic labs): Models were fit at the level of individual lab draws using logistic mixed-effects regression with a random intercept for patient (MRN) to account for repeated measurements. We used complete-case analysis per lab, report OR with 95% CI, and control for multiple testing with Benjamini–Hochberg FDR. We repeated the models under 10% WL and 15% WL, as well as stratification analysis by grade, Stage 1-3 and Stage 4 (**Supplementary Fig. 4D**).

(C) Prediction models: We evaluated out-of-sample performance in an independent DFCI NSCLC cohort. Since ECOG-KPS was not available in the external dataset, we retrained the model in the MSK cohort, restricting to covariates shared across cohorts (clinical stage, baseline BMI, and TP53/STK11 mutation status), and then tested this model in the external cohort. In the external test set (n=1,159; 516 events), the model showed modest discrimination (C-index 0.6) with time-dependent AUCs of 0.58, 0.63, and 0.66 at 1, 2, and 3 years (**Supplementary Fig. 5K**), and predicted risk was reasonably aligned with observed incidence on calibration assessment (**Supplementary Fig. 5J**) (**Page 14**).

Conclusions: robustness, validity, reliability

The core qualitative conclusions that BMI defined cachexia episodes cluster around progression, are associated with worse survival, and have consistent serologic, and some genomic correlations are well supported by the presented analyses and align with prior cachexia literature and guidelines. Quantitatively, the exact effect sizes and some cancer-specific genotype associations are more vulnerable to modeling choices (episode definition, end-of-life windows, treatment/tumor-burden adjustment). With the requested robustness checks and clearer definition of recovery and progression, the conclusions would be more reliable and easier to interpret and reproduce.

Response: We performed the requested robustness checks (**Suggested improvements Major Revision 2, below**), including sensitivity analyses over episode-definition parameters, end-of-life window specifications, and additional adjustment for treatment/tumor-burden proxies where available. Across these analyses, the core qualitative conclusions remained consistent.

Suggested improvements (Major revisions)

1. Provide explicit pseudocode for episode detection and recovery, make sure to clarify sliding-window step size, merging rules, recovery (including criteria for sustained weight gain or stability).

Response: We have added explicit pseudocode for cachexia episode detection, episode merging, and recovery in **Methods (Page 20-21)**, and **Supplementary Text (episode_identification_walkthrough.ipynb)**.

Summary below:

1. **Prepare patient time series**

- For each patient, sort BMI observations by $t = \text{days_since_anchor}$ (or days since diagnosis/anchor).
- Compute $\text{smoothed_BMI}(t)$ using EWMA.

- Compute $\log_smoothed_BMI(t) = \log(smoothed_BMI(t))$.
- 2. Scan for candidate cachexia episodes (sliding window; step = 1 observation)**
- For each observation index i (each measured timepoint is a candidate start):
 - Set $t_0 = t[i]$ and baseline $b_0 = \log_smoothed_BMI[i]$.
 - Choose the weight-loss threshold:
 - Default: $thr = \log(1 - WL)$ ($WL \in \{0.05, 0.10, 0.15\}$)
 - If BMI<20 rule is enabled and $smoothed_BMI[i] < 20$: $thr = \log(1 - 0.02)$
 - Define the lookahead window: $(t_0, t_0 + 180]$ days.
 - Find all observations in that window where $\log_smoothed_BMI < b_0 + thr$.
 - If any exist:
 - $onset_day =$ earliest hit time
 - $end_day =$ latest hit time
 - Record a candidate episode with $(start_day=t_0, onset_day, end_day)$.
- 3. Merge overlapping candidate episodes**
- Sort candidate episodes by $start_day$.
 - Walk through them in order:
 - If the next episode starts on/before the current episode ends ($next.start_day \leq current.end_day$), merge:
 - $current.end_day = \max(current.end_day, next.end_day)$
 - $current.onset_day = \min(current.onset_day, next.onset_day)$
 - Otherwise, finalize the current episode and start a new one.
 - Result: merged episodes.
- 4. Assign recovery for each merged episode**
- For merged episode k , define the recovery search interval:
 - From end_day_k to $start_day_k + 1$ (if a next episode exists), otherwise to last follow-up.
 - Let end_idx be the first observation with $t \geq end_day_k$ (do not require exact equality).
 - Set $end_log = \log_smoothed_BMI[end_idx]$.
 - Recovery threshold: $end_log + \log(1.05)$.
 - Recovery day is the first time in the search interval where $\log_smoothed_BMI \geq end_log + \log(1.05)$.
 - (If you require sustained recovery, state the rule explicitly; otherwise this is the definition.)
- 5. Apply non-edema “valid episode” QC**
- For each merged episode:
 - Compute $duration_days = end_day - start_day$; discard if $< MIN_DUR$.
 - Compute $\%WL = (start_smoothed_BMI - end_smoothed_BMI) / start_smoothed_BMI \times 100$; discard if $< MIN_WL_PCT$.
 - Remaining episodes are valid episodes (post-QC).

6. Identify edema windows (per patient; log space)

- For each observation index i as an edema baseline:
 - $t0 = t[i]$, $b0 = \log_smoothed_BMI[i]$
 - Search window: $(t0, t0 + EDEMA_WINDOW]$ days
 - Find the first time within the window where $\log_smoothed_BMI \geq b0 + \log(1+UP_FRAC)$ (a rapid “up” swing).
 - After that peak, find the first time where $\log_smoothed_BMI \leq b0 + \log(1+RETURN_FRAC)$ (returns near baseline).
 - If both exist, record an edema window $[ed_start=t0, ed_end=return_time]$.
- Merge overlapping edema windows via interval union.

7. Apply edema QC

- For each valid episode, discard it if its interval $[start_day, end_day]$ overlaps any edema window for that patient
- Remaining episodes are valid episodes (post-QC + edema QC).

2. Run lightweight sensitivity analyses varying EWMA α in addition to the current value of 0.2 (for example, 0.1 and 0.3), minimum episode duration in addition to the current 15 day minimum (for example 30 and 45 days), and edema QC thresholds (additional windows such as, 14 and 45 days and thresholds such as, 3% and 7%), and report how prevalence, episode counts (1, 2, and 3+), progression overlap, and serologic signatures change.

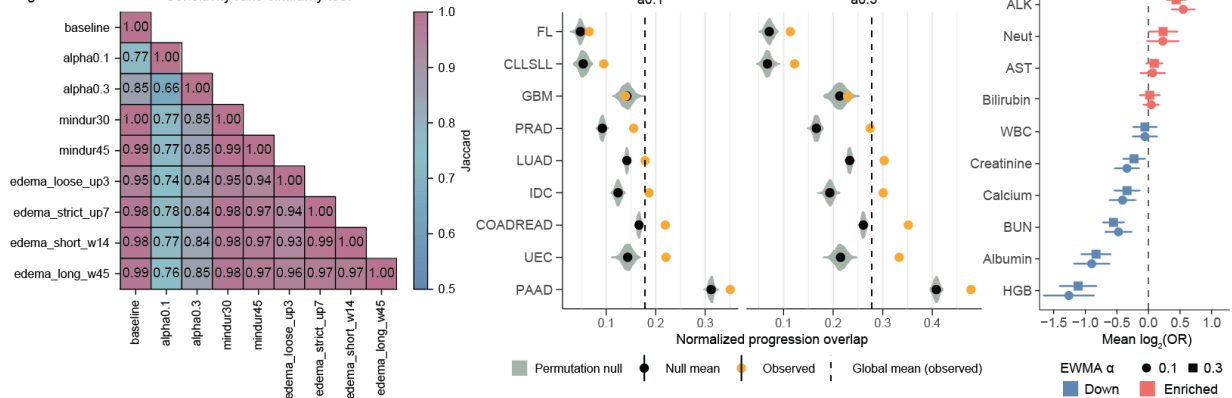
Response: We re-ran the requested lightweight sensitivity analyses varying EWMA smoothing ($\alpha = 0.1, 0.2, 0.3$), minimum episode duration (15, 30, 45 days), and edema QC thresholds (window length 14, 30, 45 days; reversible increase threshold 3%, 5%, 7%; return threshold held constant at 2%). Across all re-runs, the cohort denominator remained fixed ($N = 52,236$). Under the primary specification ($\alpha = 0.2$, minimum duration 15 days, edema window 30 days, 5% up / 2% return), we identified 30,303 candidate episodes in 24,036 patients prior to post-hoc QC (**Table 3.3, below**).

Across parameter settings, patient-level prevalence and episode-count distribution (1, 2, and 3+ episodes) were highly stable, with only modest absolute changes after varying smoothing factor for EWMA (**Table 3.3**). As expected, lower alpha parameter attenuates transient changes, whereas higher factor increases responsiveness which yields more candidate episode calls. To quantify concordance of episode calls across specifications, we computed pairwise Jaccard similarity of episode sets at patient-level episode intervals, observing consistently high agreement (mean Jaccard ≈ 0.8 – 0.9 across comparisons; **Fig.R.2.,below**). Importantly, downstream analyses were similarly robust, with estimates of progression overlap and the direction and relative ranking of serologic signatures were concordant across runs for varying smoothing factors, indicating that the main conclusions are not sensitive to reasonable choices of smoothing strength, minimum duration, or edema QC thresholds.

Table 3.3. Sensitivity analyses re-runs

Setting	EMW A	Min episode duration (days)	Edema QC threshold (%)	Edema QC threshold (days)	% patients with episodes post QC	Episode loss	% patients with 1,2,3+ episodes post QC
baseline	0.2	15	5	30	44.7% (23,372)	-664	35.3%,8.01%,1.46%
α low	0.1	15	5	30	35.7% (18,648)		30.76%,4.48%,0.45%
α high	0.3	15	5	30	50.4% (26,326)		37.55%, 10.2%, 2.65%
Min. dur. 30	0.2	30	5	30	44.5% (23,273)	-763	35.2%,7.92%,1.44%
Min. dur. 45	0.2	45	5	30	44.23% (23,102)	-934	35.0%,7.8%,1.42%
Edema QC looser	0.2	15	3	30	42.62% (22,261)	-1775	34.03%, 7.34%, 1.24%
Edema QC stricter	0.2	15	7	30	45.53% (23,783)	-253	35.68%, 8.30%, 1.55%
Edema QC short window	0.2	15	5	14	45.63% (23,837)	-199	35.71%,8.35%,1.56%
Edema QC long window	0.2	15	5	45	44.07% (23,022)	-1014	34.89%, 7.78%, 1.40%

Fig.R.2. Sensitivity runs similarity test



3. Please specify how progression is defined, strictly radiographic RECIST progression, clinical/biochemical/radiology-combined progression or whatever has been used. Whether the same criteria are used across all cancers, or disease-specific rules are applied, and in what proportions. The source of progression labels (radiology structured fields, tumor board/registry abstractors, manual chart abstraction, NLP from reports, or whatever has been used). Also specify any quality control/inter-rater metrics used for progression adjudication and the exact timestamp used for progression.

Response: Progression was defined as radiographic progression from radiology report impressions using the MSK-CHORD NLP pipeline (Jee et al., *Nature* 2024)⁴. Reports were segmented to the Impression section and classified with a transformer model fine-tuned against human-curated progression labels; “Progressing/Worsening/Enlarging” and “Mixed” were treated as progression-positive and all other categories as negative. A uniform NLP pipeline was applied, and is not disease-specific. This uniform, disease-agnostic pipeline yields dated, report-level progression calls that can generate multiple progression events per patient; the progression timestamp used in our analyses is the radiology report date corresponding to each NLP-derived progression-positive Impression. This has been added to the Methods (**Page 20**).

4. How are duplicate and overlapping progressions handled? Also specify the proportion of patients (by cancer type) with usable progression annotations and make missing progressions transparent.

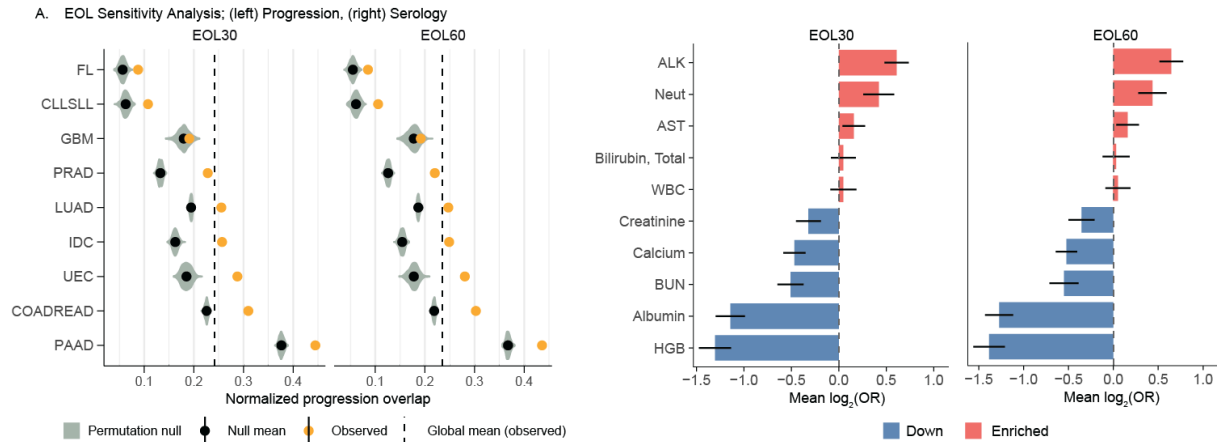
Response: Duplicate progression timestamps within a patient were retained once, with identical report dates collapsed. In the permutation test, we used the full set of de-duplicated timestamps and restricted to “Yes” progression annotations, corresponding to reports labeled as progressing, worsening, or enlarging as explained above. Progression annotations (any label with a valid report date) were available for 98.13% of patients overall (51,261/52,236). Among annotated patients, 72.07% had ≥ 1 progression-positive report (36,944/51,261). Among patients with ≥ 1 progression-positive report, the mean number of progression timestamps per patient was 5.40 (median 4, IQR 2–7) (**Page 20**).

5. Add end-of-life sensitivity (exclude last 30 and 60 days before death) and stage-stratified analyses for progression overlap and serologic signatures. Explain how pseudo progression (for example in patients receiving immune-checkpoint inhibitors) is being handled in this study.

Response: We performed end-of-life sensitivity analyses to assess whether terminal decline confounds the observed progression–cachexia overlap. Specifically, we excluded events occurring within 30 days and, separately, within 60 days prior to death among deceased patients. Under the 30-day restriction, 3,191 progression events were excluded; under the 60-day restriction, 7,115 progression events were excluded. After re-running the progression-overlap permutation analyses under both censoring windows, the magnitude and direction of the progression–cachexia temporal enrichment signal remained consistent with the primary analysis across cancer types, with GBM remaining non-enriched for progression–cachexia overlap relative to the permuted null (**Fig.R.5.,below**). We also re-ran

the serologic analyses on a subset of representative laboratory tests and observed consistent direction and comparable effect sizes after excluding end-of-life measurements (EOL30: 61519 measurements removed; EOL60: 118872 removed), supporting that these patterns are not driven by terminal decline near death.

Fig.R.5.



To further address potential confounding from terminal decline, we evaluated both the serology and progression-overlap analyses after stratifying the cohort by clinical stage (STAGE_CDM_DERIVED; Stage I–III vs Stage IV). For serology, concordance plots comparing per–cancer type lab $\log_2(OR)$ estimates from each stage-stratified model to the baseline unstratified analysis showed strong directional agreement overall, with only a small number of statistically significant discordant pairs:

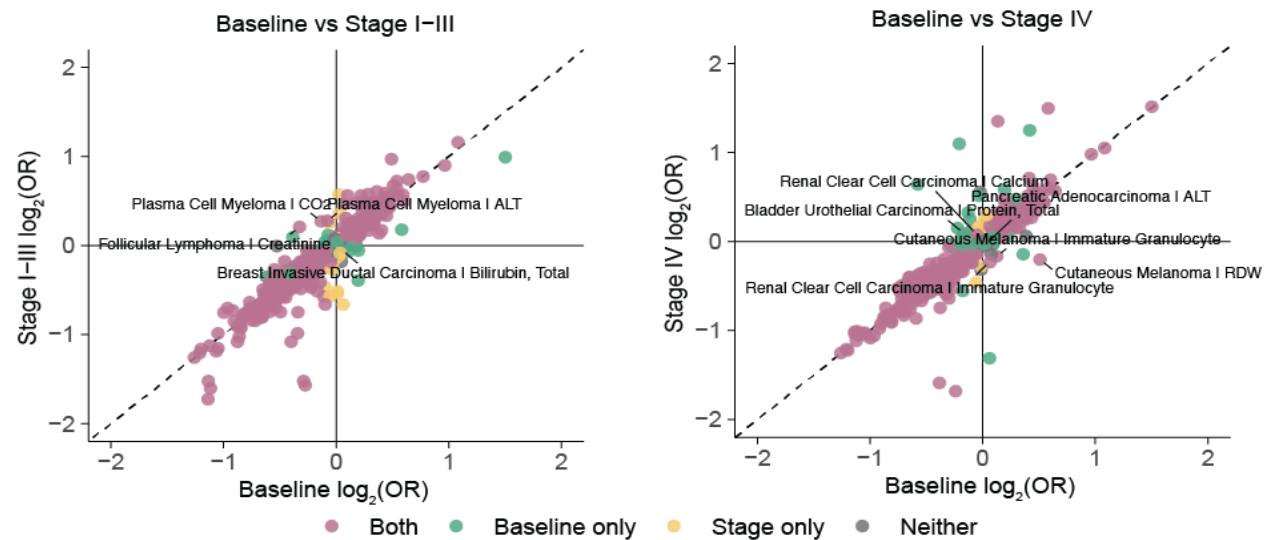
- Stage I–III: PCM- CO_2 , PCM-ALT, FL-Creatinine, IDC-Total Bilirubin;
- Stage IV: CCRCC-Immature Granulocyte, CCRCC-Calcium, SKCM -RDW, SKCM-Immature Granulocyte, BLCA-Total Protein, PAAD-ALT

For progression overlap, the stage-stratified permutation analyses produced results consistent with the primary analysis across cancer types, indicating that the progression–cachexia temporal overlap signal is not driven by differences in stage composition (**Fig.R.5.ii, below**).

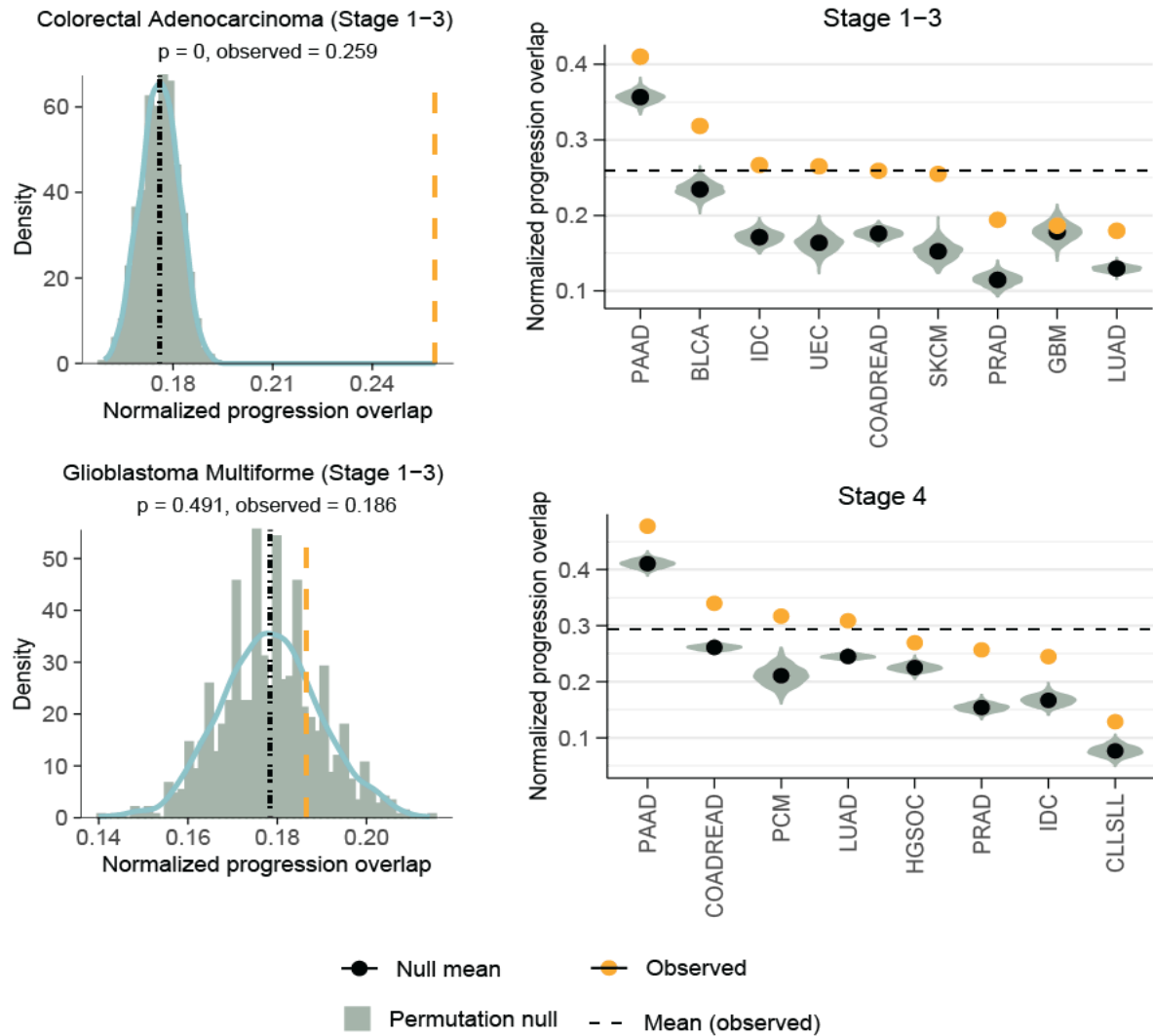
Lastly, we did not implement an immune-checkpoint–specific pseudo-progression adjudication step in this study. Progression events were defined using the radiology impression–derived progression annotations (Methods), and were analyzed as recorded.

Fig.R.5.ii.

A. Serology Analysis Stratified by Stage



B. Progression Analysis Stratified by Stage (n>500)

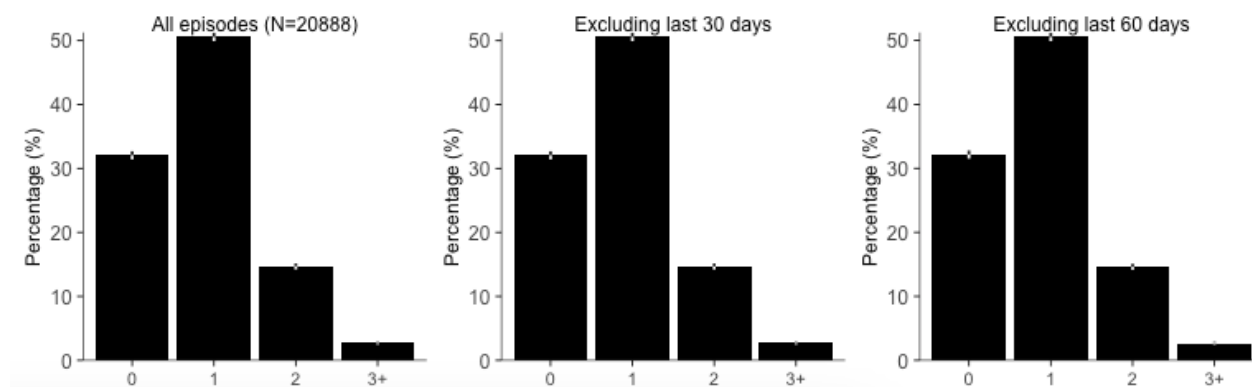


6. For patients with 1, 2, 3+ episodes please report how many last episodes of cachexia extended to end of life or end of follow up. This would shed light on the proportion of episodes from which recovery is not achieved and help distinguish reversible vs terminal cachexia phases based on the simplified cachexia definition used in this study.

Response: We assessed whether the last cachexia episode extended to end-of-life among deceased patients using a proxy definition in which the final cachexia episode ended within 30 days of the patient's last available BMI measurement before death and no recovery was observed. By this criterion, 31.96% (3,377/10,565) of deceased patients with 1 detected episode met the terminal-proxy definition, 38.29% (1,174/3,066) with 2 episodes and 36.90% (207/561) with ≥ 3 episodes. Weight-data coverage near death was similar across groups, with median last-BMI-to-death of 32, 30, and 30 days and IQRs of 15–68, 15–59, and 14–64 days, respectively, suggesting these differences are not driven primarily by differential end-of-life measurement density and are consistent with a larger subset of patients with recurrent episodes whose final cachexia interval does not show recovery prior to death.

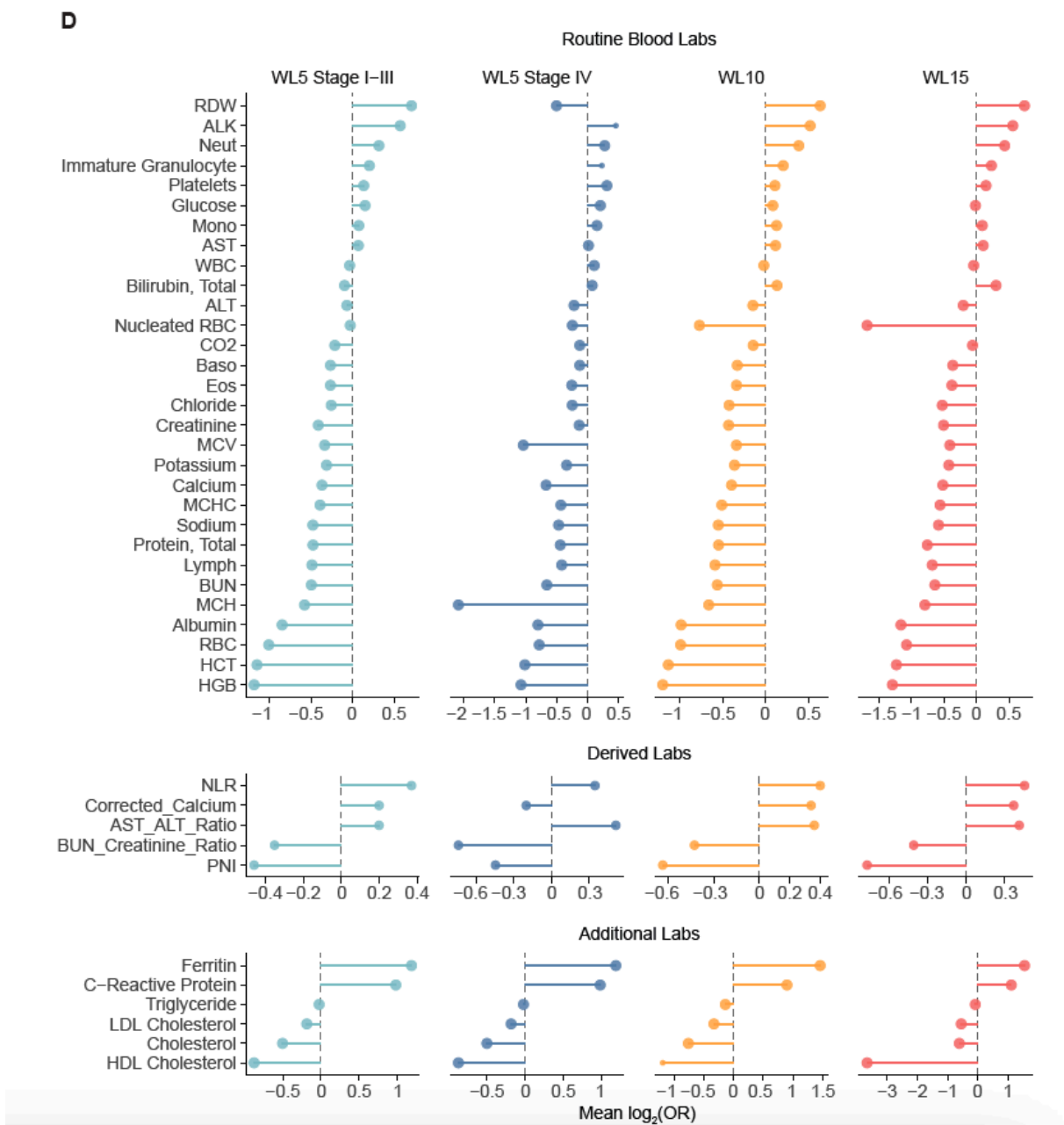
7. Further analysis of the number of cachexia episodes per patient excluding the last 30 and 60 days before death (for comparison with Figure 2C) will provide further insights since based on your analysis irrespective of cancer type (excluding GBM) more than 50% patients develop cachexia close to end of life.

Response: We repeated the cachexia episode-burden analysis among deceased patients in the cachexia cohort ($n = 20,888$), recomputing the number of detected episodes per patient after excluding episodes that started within the last 30 days or 60 days before death. The distribution was unchanged across these sensitivity windows: with no exclusion, 32.06% had 0 episodes, 50.58% had 1 episode, 14.68% had 2 episodes, and 2.69% had ≥ 3 episodes. After excluding the last 30 days, the corresponding proportions were 32.06%, 50.59%, 14.67%, and 2.69%, and after excluding the last 60 days they were 32.16%, 50.56%, 14.61%, and 2.67%. These results indicate that the overall episode-burden distribution is not driven primarily by episodes initiating in the immediate pre-death window.



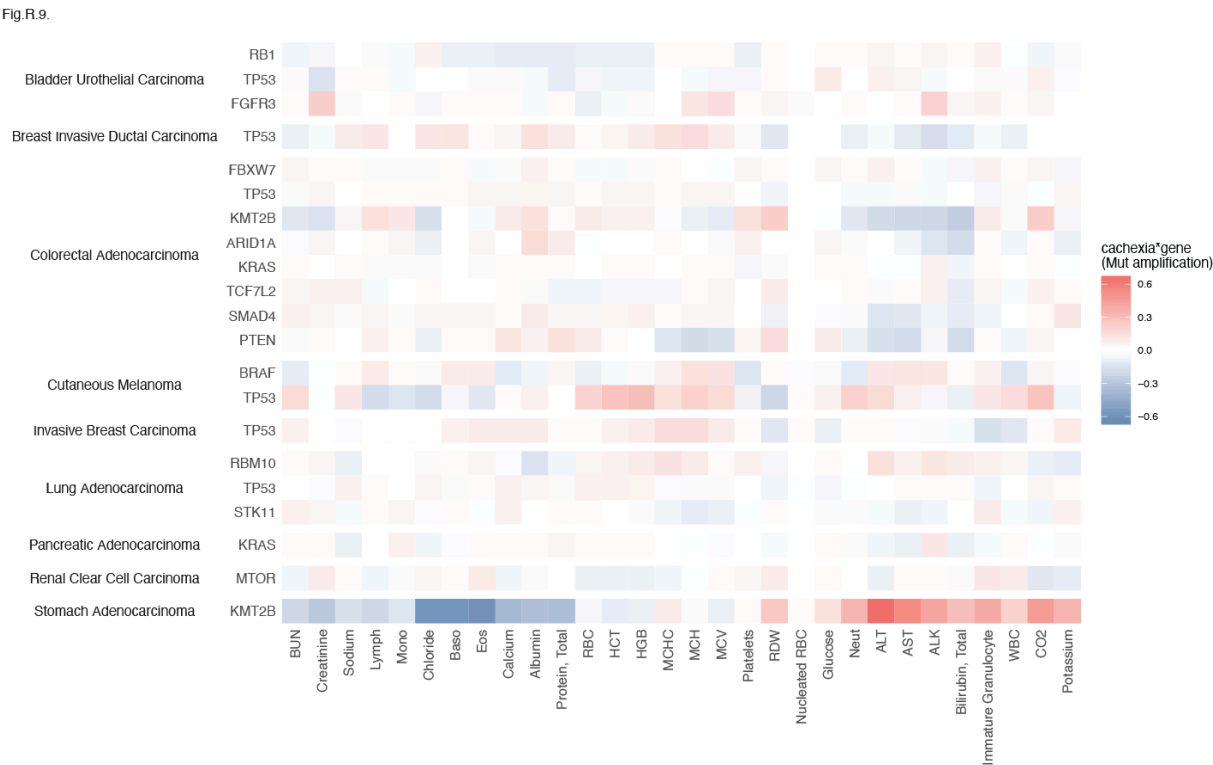
8. You identify a largely disease agnostic serologic signature of cachexia. Please repeat key mixed-effects logistic models stratified by stage (I–II vs III–IV) to assess whether this signature is consistent in both early and late disease.

Response: We repeated the mixed-effects logistic models for the serologic signature after stratifying by AJCC stage, comparing Stage I–III versus Stage IV. The same core markers showed concordant directionality across strata, indicating that the serologic signature is present in both earlier-stage and metastatic disease. Notable exceptions were RDW and corrected calcium, for which estimated mean $\log_2(\text{OR})$ changed directions between Stage I–III and Stage IV (Page 11, Supplementary Fig. 4D).



9. To support mechanistic plausibility, consider showing whether the cachexia serologic signature (low albumin/hemoglobin, high ALK, high neutrophils etc.) is stronger in TP53-mutant vs wild-type patients during cachexia windows

Response: We tested whether the cachexia-associated serologic signature was different in TP53-mutant vs TP53–wild-type patients during cachexia windows, using an interaction model evaluating effect modification of cachexia status by TP53 (lab ~ episode + TP53 + episode * TP53). Although a small number of individual gene–cancer–lab interactions reached significance, these effects were heterogeneous in direction and analyte and did not converge on a consistent mutant-amplified pattern. For example, in colorectal adenocarcinoma, TP53 showed a statistically significant interaction with neutrophils ($\beta = -0.0399$; 95% CI -0.0589 to -0.0209 ; $p = 3.84 \times 10^{-5}$), indicating a modest attenuation rather than amplification of the neutrophil shift during episodes in TP53-mutant patients. Likewise, significant interactions were observed for other gene–cancer pairs (e.g., COADREAD SMAD4 with ALK, $\beta = -0.0606$; $p = 3.29 \times 10^{-5}$; COADREAD KMT2B with platelets, $\beta = 0.1329$; $p = 2.09 \times 10^{-5}$), but these did not generalize across cancers or map onto a coherent strengthening of the episode-associated serologic signature in mutant tumors (**Fig.R.9., below**). We extended the same interaction framework across all significant genes and similarly did not identify a reproducible pattern of amplified episode-associated serologic shifts in mutant versus wild-type patients.



R.9. Gene-by-episode interaction effects on serologic biomarkers. Summary of significant gene × episode interaction terms from mixed-effects models across cancer types

10. You focus on 13 cancer types for somatic genotype cachexia analyses. Please explain the criteria used to select these 13 cancers.

Response: We restricted the somatic time-to-cachexia analyses to cancer types with ≥ 200 patients remaining after applying the analysis inclusion criteria and required covariate complete-case filtering; under these criteria, 13 cancer types met inclusion.

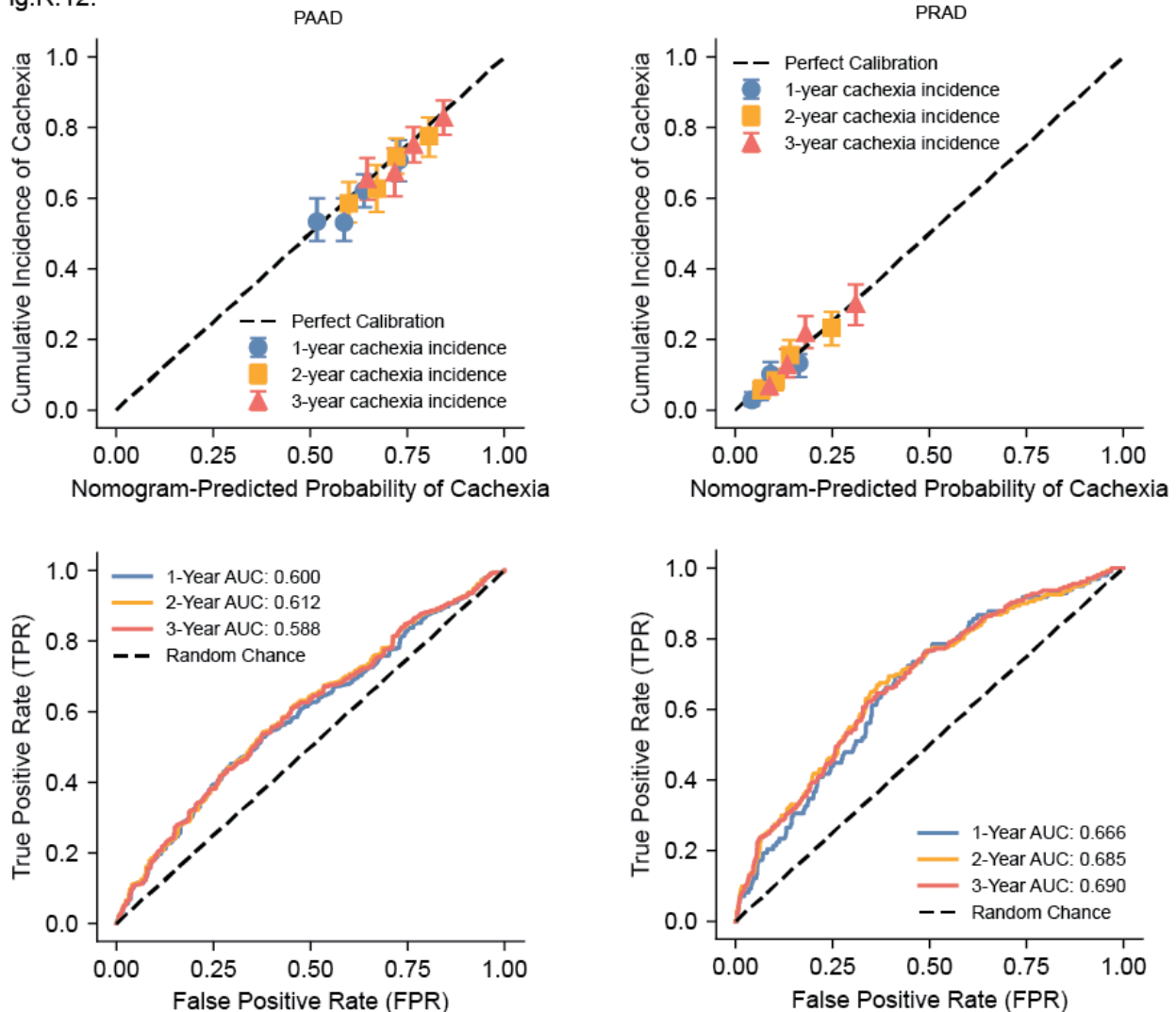
11. You build prediction models for cachexia risk in LUAD and COADREAD using somatic features and serologic markers. Please explain why only SMAD4 is used for cachexia risk prediction in COADREAD and not TP53 or KRAS (Figure 5 suggests KRAS also has a significant association), and the rationale for focusing on LUAD and COADREAD given the high cachexia prevalence in PAAD and STAD.

Response: Risk-model covariates were taken from the multivariate time-to-cachexia analysis, variables that remained significant after adjustment for the clinical covariates in the Cox model. In the original submission, SMAD4 was the only COADREAD somatic feature that remained significant after FDR correction; KRAS did not pass the $FDR < 0.10$, so it was not included. In the updated cohort using the revised cachexia definition, KRAS now remains significant after ($FDR = 0.047$) and we have added it to the COADREAD risk model (**Page 14**). We focused initially on LUAD and COADREAD because they are high-volume cancers. We now also include analogous models in PAAD and STAD to reflect high-prevalence cancers.

12. To strengthen the predictive model's feasibility in clinics it will be beneficial to explore similar cachexia risk models in PAAD and STAD (high prevalence) and possibly PRAD (low prevalence) to illustrate performance across different prevalence regimes.

Response: We fit the same cachexia risk model framework in PAAD and PRAD to illustrate performance across high- and low-prevalence settings. Model performance was evaluated using time-dependent discrimination (C-index and AUC at 1, 2, and 3 years) and calibration (predicted risk vs observed cumulative incidence at each timepoint). In PAAD, discrimination was modest (AUC = 0.600, 0.612, 0.588 at 1, 2, and 3 years, respectively) with C-index = 0.57. In PRAD, discrimination was higher (AUC = 0.666, 0.685, 0.690 at 1, 2, and 3 years, respectively) with C-index = 0.674, and calibration tracked observed incidence across timepoints despite the lower event rate (**Page 14**).

Fig.R.12.

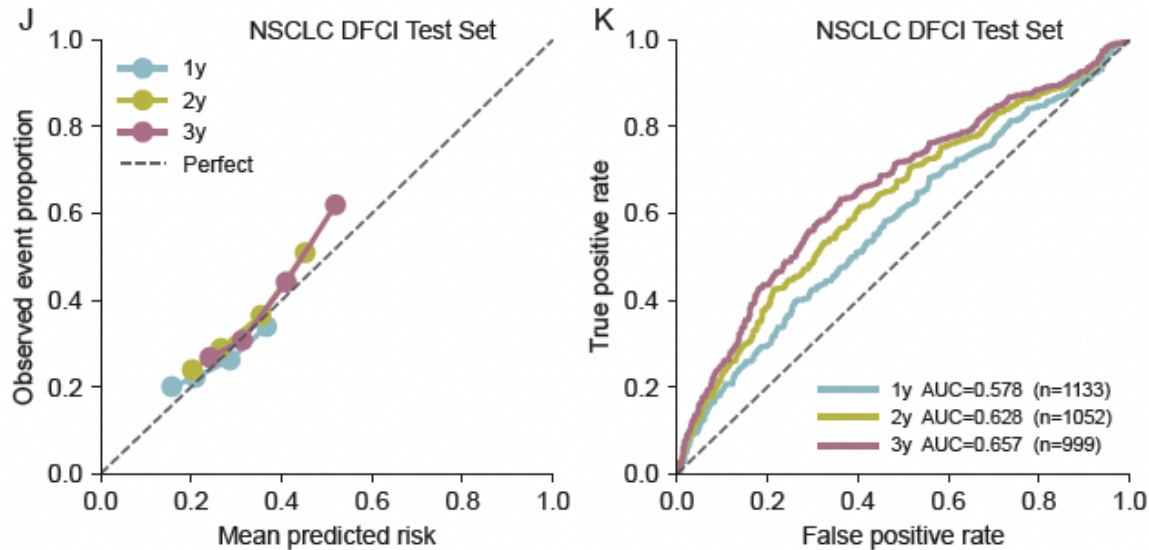


R.12. Cachexia risk model performance across high- and low-prevalence cancers. Calibration (top) and ROC curves (bottom) for PAAD (high prevalence) and PRAD (low prevalence).

13. For clinical utility, models need to generalize beyond the development cohort. Please validate your trained models for predicting cachexia risk if possible with an external dataset for any cancer type, evaluate whether your trained models retain discrimination and calibration there. Or at least use another internal dataset for validation. Explicitly report how C-indices/AUCs change under these validations.

Response: We evaluated out-of-sample performance in an independent DFCI NSCLC cohort. Since ECOG-KPS was not available in the external dataset, we retrained the model in the MSK cohort, restricting to covariates shared across cohorts (clinical stage, baseline BMI, and TP53/STK11 mutation status), and then tested this model in the external cohort. In the external test set (n=1,159; 516 events), the model showed modest discrimination (C-index=0.6) with

time-dependent AUCs of 0.58, 0.63, and 0.66 at 1, 2, and 3 years (**Supplementary Fig. 5K**), and predicted risk was reasonably aligned with observed incidence on calibration assessment (**Supplementary Fig. 5J**) (**Page 14-15**).



14. Please clarify if patients received only standard of care therapy, or did some patients receive additional nutrition and exercise interventions (clinical nutrition consults, specialized programs) since this could influence episode detection or recovery.

Response: These supportive-care interventions are not uniformly captured at scale across cancer types in the available structured data and therefore were not modeled as exposures in this work. This limitation has been added (**Page 17**).

15. Supplemental Figure 2, panel labels appear swapped (A and B). Please correct it.

Response: Thank you. We have updated the panel (**Supplemental Figure 2**).

16. Radiographic features are mentioned in the Introduction and Discussion, however, the results or analysis do not include radiographic features. It will be beneficial to clarify which radiographic features have been used in this study.

Response: Imaging information in our analyses includes (i) radiology report–derived progression timestamps (NLP applied to report impressions), used for the progression-overlap analyses, and (ii) CT image–derived body composition in a subset of lung adenocarcinoma patients, where we quantified L3 subcutaneous and visceral adipose tissue areas to provide compartment-level context for BMI-defined weight-loss episodes (**Supplementary Fig. 2J–K; Methods Page 25, Results Page 6-7**). We clarify that we did not analyze additional radiomic or lesion-level quantitative imaging features.

References: appropriate credit to previous work?

The manuscript cites core consensus and guideline work on cachexia and its clinical impact.

It would benefit to place their work relative to recent longitudinal weight-loss and cachexia studies in real-world and trial cohorts and to emerging EHR-based or AI-based cachexia prediction approaches. Adding these references will help readers appreciate what is new (episode-resolved, multimodal, health-system-scale analysis across many cancers) versus what builds on prior work.

Response: We updated the related-work discussion and citations to better position our study relative to recent longitudinal weight-loss/cachexia analyses in real-world and trial cohorts. These additions clarify what is distinct in our work, episode-resolved, multimodal, health-system-scale, pan-cancer versus what builds on prior approaches (**Discussion, Page 15**).

Clarity and context: abstract, introduction, conclusions

The abstract clearly states the motivation, high level methods, and key findings, but it tends to present the phenotype as “cachexia” rather than “algorithmic weight loss episodes consistent with cachexia criteria,” which slightly overstates clinical specificity. The introduction is well contextualized in terms of cachexia burden and unmet needs but could better position the work relative to existing longitudinal and EHR based cachexia studies. The conclusions are generally clear and aligned with the results.

Response: We revised the manuscript to consistently describe the phenotype as algorithmically detected weight-loss episodes consistent with consensus cachexia criteria. We also expanded the Introduction to more explicitly position our approach relative to prior longitudinal/EHR-based cachexia studies and to clarify the specific gap addressed by a pan-cancer, population-scale, longitudinal episode framework.

Reviewer #3 (Remarks on code availability):

The Github repo has good data format specifications, function-level documentation, and algorithm explanation. I was able to test their code using synthetic data that I generated myself. Following changes to the repo will make it more user friendly.

1. add installation dependencies in a requirements.txt
 2. edit main.py to change hardcoded paths or instruct the user to add relevant paths here.
 3. instruct the user to run "python main.py"
- nomogram.py requires additional data not listed in the README and so I did not test this.

The included CSV (59,493 patients) contains the processed output with anonymized cachexia episodes, and other output files have also been shared that were for the paper figures/tables. Raw BMI trajectories and mutation data are not included. This code can be used by the community if they can format their data as required and detailed in the repo. Also, their code doesn't handle missing data, so all required patient information needs to be available in order to run their code. They have a check applied for having more than 200 patients per cancer type for

genotype survival analysis only. This hardcoded threshold can be changed in the code depending on the amount of data the user is working with.

Overall, the repo looks good, and the code is functional.

Response: Thank you for testing the repository and for these suggestions. We updated main.py to remove hard-coded paths and clearly document required user-defined paths/inputs, and added explicit run instructions in the README (including python main.py) as well as *data_dictionary.txt* for explicit identification of data columns and files.

We also created a step-by-step walkthrough of the episode-detection pipeline to illustrate how consensus-defined weight-loss episodes are called from a longitudinal BMI trajectory. This is provided in the repository as **walkthrough/episode_identification_walkthrough.ipynb** (github.com/reznik-lab/cachexia) and uses a mock BMI trajectory included below as an example input (**example_patient_deid.csv**). The notebook visualizes the trajectory and demonstrates each processing step, from smoothing through sliding-window detection, episode merging, and quality-control filtering.

Example longitudinal data:

patient_id	days_since_anchor	bmi
EXAMPLE_PATIENT_001	0	23.3
EXAMPLE_PATIENT_001	15	23.6
EXAMPLE_PATIENT_001	35	23.4
EXAMPLE_PATIENT_001	55	23.7
EXAMPLE_PATIENT_001	75	23
EXAMPLE_PATIENT_001	86	25.4
EXAMPLE_PATIENT_001	94	25.4
EXAMPLE_PATIENT_001	97	25.6
EXAMPLE_PATIENT_001	104	22
EXAMPLE_PATIENT_001	105	23
EXAMPLE_PATIENT_001	114	22.9
EXAMPLE_PATIENT_001	127	22.7
EXAMPLE_PATIENT_001	130	23.1
EXAMPLE_PATIENT_001	142	22.8
EXAMPLE_PATIENT_001	153	23.2

EXAMPLE_PATIENT_001	170	23.6
EXAMPLE_PATIENT_001	190	22.9
EXAMPLE_PATIENT_001	230	23.1
EXAMPLE_PATIENT_001	300	23
EXAMPLE_PATIENT_001	388	22.7
EXAMPLE_PATIENT_001	460	23.2
EXAMPLE_PATIENT_001	520	23
EXAMPLE_PATIENT_001	565	23.1
EXAMPLE_PATIENT_001	580	23
EXAMPLE_PATIENT_001	604	23.2
EXAMPLE_PATIENT_001	608	23.5
EXAMPLE_PATIENT_001	612	23.7
EXAMPLE_PATIENT_001	625	24.3
EXAMPLE_PATIENT_001	645	23.5
EXAMPLE_PATIENT_001	700	23
EXAMPLE_PATIENT_001	780	23.2
EXAMPLE_PATIENT_001	860	22.9
EXAMPLE_PATIENT_001	940	22.6
EXAMPLE_PATIENT_001	1020	23
EXAMPLE_PATIENT_001	1120	22.7
EXAMPLE_PATIENT_001	1200	22.6
EXAMPLE_PATIENT_001	1260	22.9
EXAMPLE_PATIENT_001	1280	22.3
EXAMPLE_PATIENT_001	1300	22.4
EXAMPLE_PATIENT_001	1325	22
EXAMPLE_PATIENT_001	1350	21.6
EXAMPLE_PATIENT_001	1375	21.8
EXAMPLE_PATIENT_001	1414	21.2
EXAMPLE_PATIENT_001	1435	20

EXAMPLE_PATIENT_001	1460	19.4
EXAMPLE_PATIENT_001	1487	19.3
EXAMPLE_PATIENT_001	1500	19.6
EXAMPLE_PATIENT_001	1512	19.5
EXAMPLE_PATIENT_001	1538	19.8
EXAMPLE_PATIENT_001	1565	20.3
EXAMPLE_PATIENT_001	1607	20.8
EXAMPLE_PATIENT_001	1650	21.1
EXAMPLE_PATIENT_001	1750	21.6
EXAMPLE_PATIENT_001	1850	21.4
EXAMPLE_PATIENT_001	2020	21.6
EXAMPLE_PATIENT_001	2067	21.8
EXAMPLE_PATIENT_001	2120	22.3

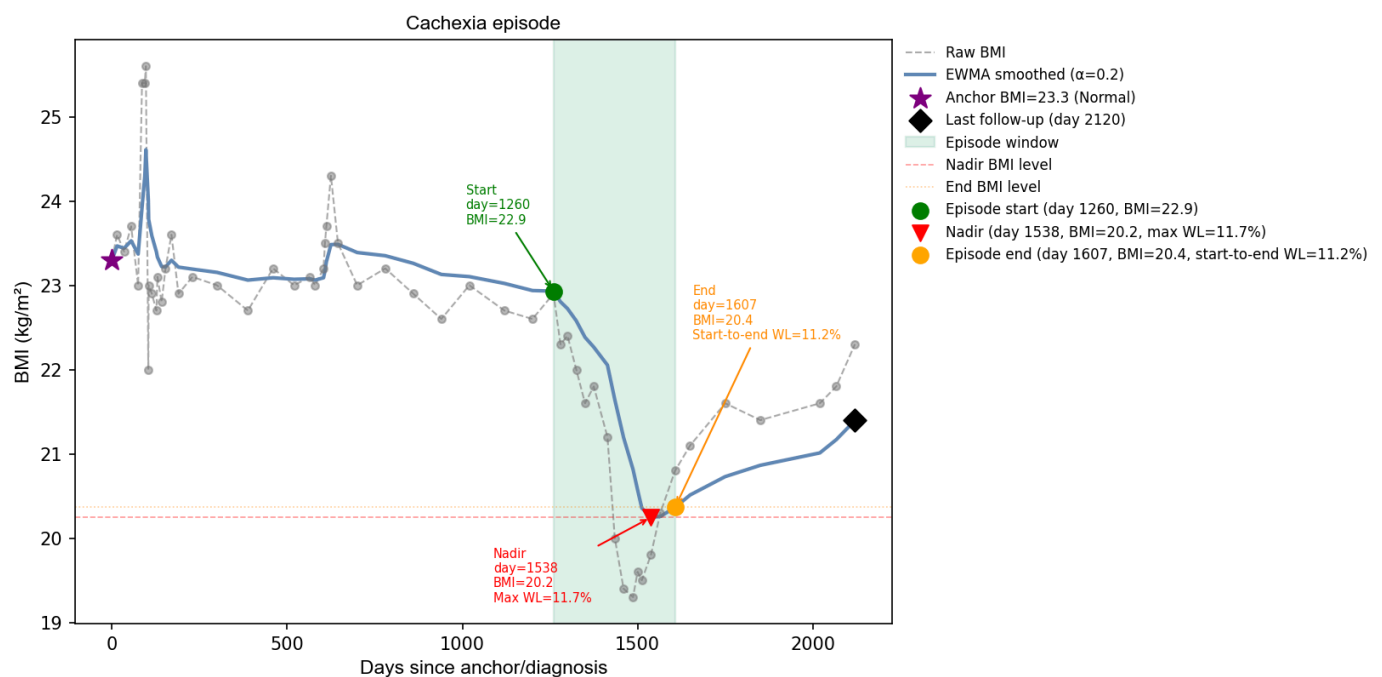
References:

1. Groarke, J. D. *et al.* Ponegromab for the Treatment of Cancer Cachexia. *N. Engl. J. Med.* **391**, 2291–2303 (2024).
2. Schwarz, S. *et al.* The clinical picture of cachexia: a mosaic of different parameters (experience of 503 patients). *BMC Cancer* **17**, 130 (2017).
3. Diagnostic Criteria for the Classification of Cancer-Associated Weight Loss | Journal of Clinical Oncology. <https://ascopubs.org/doi/10.1200/JCO.2014.56.1894>.
4. Automated real-world data integration improves cancer outcome prediction | Nature. <https://www.nature.com/articles/s41586-024-08167-5>.

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

We thank Reviewer 1 for the continued engagement with the manuscript. Several of the concerns raised relate to the interpretation of columns in the supplementary episode summary table (**Supplementary Table 18**), which are downstream outputs of a multi-step pipeline. We recognize that additional columns in the supplementary table may have aided interpretation and contributed to the concerns raised. We note however that the data dictionary attached to the original submission explicitly defines each column, and a step-by-step walkthrough notebook illustrating the full pipeline is provided in the GitHub repository (episode_identification_walkthrough.ipynb; github.com/reznik-lab/cachexia). The key timepoints and derived columns are defined and illustrated in the example trajectory below:



Need for a complete data verification process and re-analysis of the data.

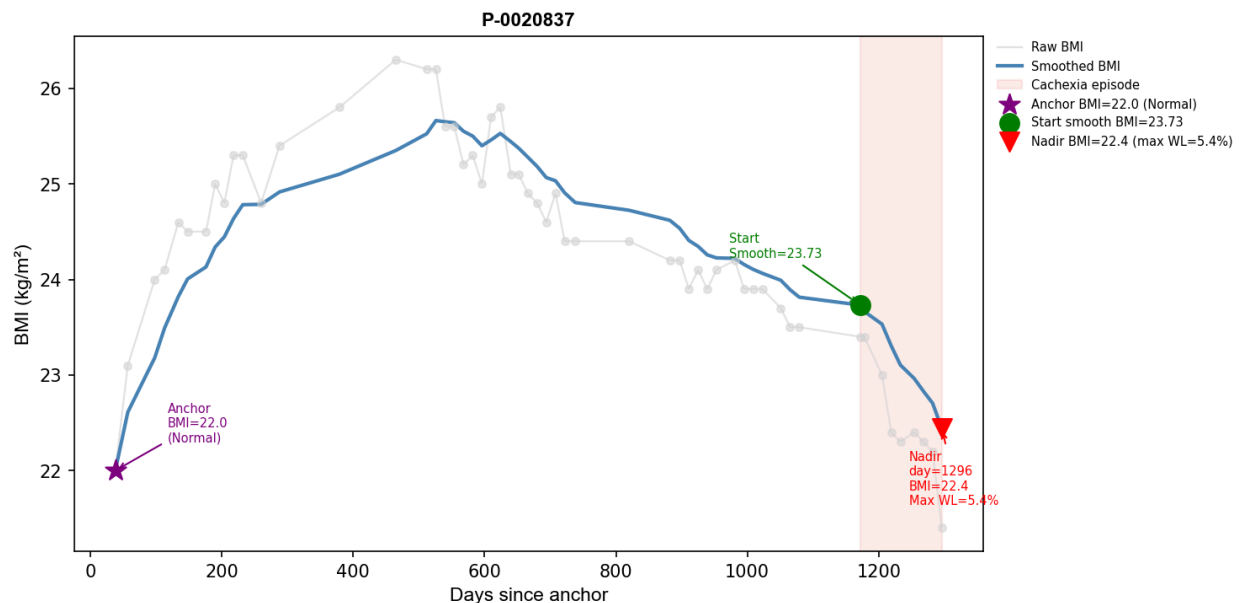
1. One of the first points made in responses to previous comments is that the approach to calling episodes has been revised. This includes addition of a 2% weight loss over six months as a cachexia criterion value, specifically for individuals with BMI < 20 kg/m². Absent a formal grading system for severity, this criterion was used in Fearon et al. 2011 to denote individuals of clearly depleted body stores. Reading this criterion, I went to the data set to observe the cachexia episodes meeting this criterion (low grade weight loss between -1.999 and -4.999% with BMI<20) and went on to discover multiple bizarre errors in the provided data file. The first instance I happened upon is patient 0020837. This person was coded as having a weight loss episode -4.09%, but a BMI above 20, so this person's cachexia episode was miscoded. Whilst continuing to peruse instances of low-grade weight loss, I next found patient 0076545, who had a 2.19% weight loss lowering BMI of 22 to 21.5, but column J this patient's BMI category is "obese", also incorrect. I checked the baseline BMI categorical value (column J baseline_bmi_cat) against the BMI at start of a cachexia episode. The baseline BMI categorical

variable seems to have been scrambled and would not render proper results if it was used in any analyses. For example: n=129 patients in the “normal BMI” category had a BMI of less than 18.5 at the start of their cachexia episode, and n=1554 patients with a BMI between 25.0 and 57.4, have been included in the normal BMI category. In the “obese” category of baseline BMI, n=906 patients have a BMI <30 (and as low as 17.0 kg/m²). I could go on ... in the “overweight” category there are patients with BMI as low as 17.5, and as high as 75.4, and indeed the “underweight” category has patients with a baseline BMI of up to and including 46.7 kg/m².

We thank the Reviewer for their careful examination of the supplementary episode summary table. In the data dictionary attached to the submission, **baseline_bmi_cat** is defined as ((STable18_cachexia_summary_episodes_datadictionary.txt): *“Baseline BMI category based on the first available BMI measurement on or after anchor (i.e., the earliest BMI record by days_since_anchor).”* This column (as explicitly stated in the Data Dictionary) reflects the patient's BMI category at their anchor/diagnosis date, not the BMI at the start of a cachexia episode. These are two distinct timepoints, and patients may gain or lose weight between their anchor date and the start of a weight loss episode. **baseline_bmi_cat** was included in the supplementary table as it was used as a covariate in survival and logistic regression analyses to adjust for baseline bmi category at presentation.

We also wish to clarify how weight loss is quantified in the supplementary table. The weight_loss column reflects percent weight loss from episode start to episode end, calculated as $(\text{start_bmi} - \text{end_bmi}) / \text{start_bmi} \times 100$. This is a post-hoc descriptive summary and may be less than the qualifying detection threshold if partial BMI recovery occurred before end_day. end_day is defined as the last BMI measurement still satisfying the weight loss criterion within the merged episode window and is **not** the lowest point in the trajectory (nadir).

We verified the two specific patients cited by the Reviewer:

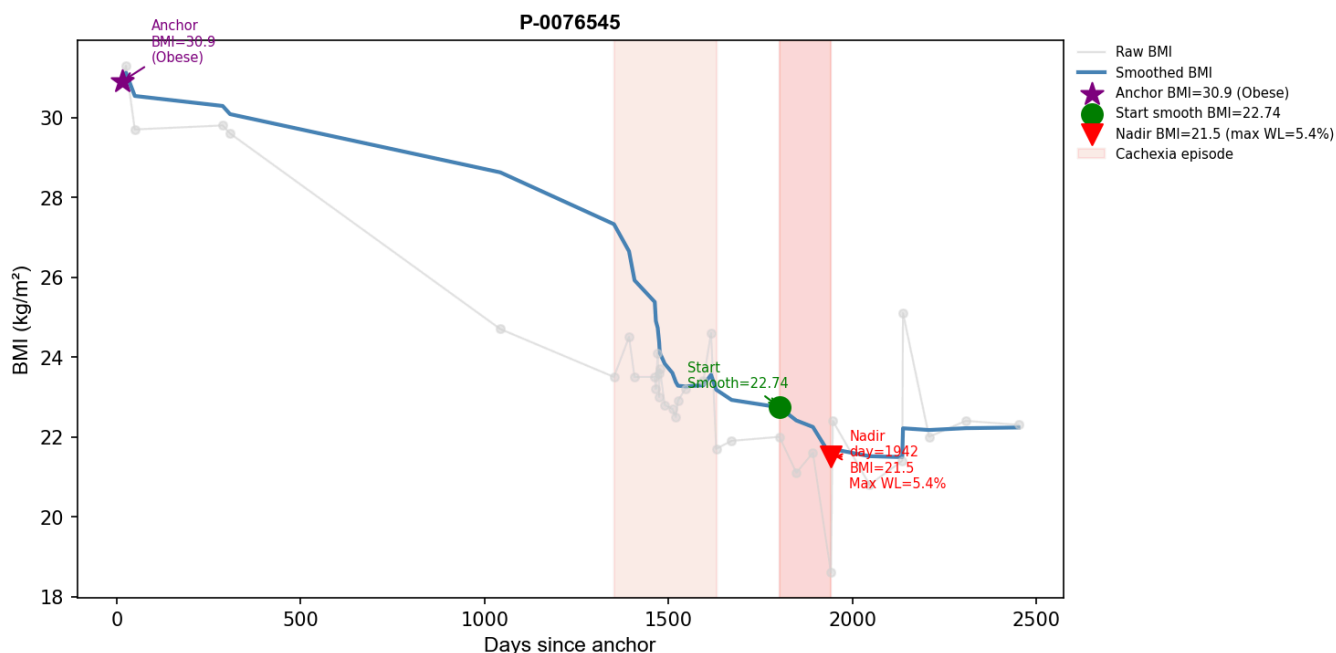


- **P-0020837:** We address two points. First, bmi_cat_at_anchor_date = Normal is correct. The patient's BMI at anchor (day 38) was 22.0 kg/m², within the Normal range. The patient gained weight over the following 3 years before the episode began at day 1172. Second, the episode was not miscoded. As described above, the weight_loss column reflects a post-hoc start-to-end summary and is not the value that triggered the episode call. The maximum loss from start to nadir (weight_loss_max) was 5.43%, confirming the episode correctly met the ≥5% threshold. The full trajectory (above) and 180-day window analysis (below) are provided. Each highlighted row (in red) indicates a timepoint in which a qualifying weight loss of 5% (or 2% if BMI<20) was detected within the subsequent 180 days.

Table 1. P-0020837 longitudinal BMI and exact timepoints the threshold was detected.

day	raw_bmi	smooth_bmi	max_wl_next_180d	first_day_5pct	days_to_cross	threshold_met
38	22	22	-2.78			
57	23.1	22.611	-2.52			
98	24	23.18	-1.34			
113	24.1	23.492	-1.4			
134	24.6	23.822	-0.77			
148	24.5	24.005	-0.52			
176	24.5	24.131	-0.87			
190	25	24.34	-0.44			
204	24.8	24.446	-0.78			
218	25.3	24.637	-0.59			
232	25.3	24.782	-0.02			
260	24.8	24.786	-0.52			
288	25.4	24.916	-0.74			
379	25.8	25.101	-0.99			
465	26.3	25.349	-0.2			
512	26.2	25.525	1.34			
526	26.2	25.663	2.33			
540	25.6	25.65	2.41			
554	25.6	25.64	2.86			
568	25.2	25.551	2.92			
582	25.3	25.5	2.73			
596	25	25.399	2.34			
610	25.7	25.46	2.57			
624	25.8	25.528	2.83			

640	25.1	25.442	2.82			
652	25.1	25.374	2.56			
666	24.9	25.279	2.2			
680	24.8	25.183	1.82			
694	24.6	25.066	1.37			
708	24.9	25.033	1.65			
722	24.4	24.906	1.49			
738	24.4	24.805	1.6			
820	24.4	24.724	2.29			
882	24.2	24.619	2.55			
897	24.2	24.535	2.62			
911	23.9	24.408	2.43			
925	24.1	24.346	2.19			
939	23.9	24.257	1.83			
953	24.1	24.226	1.7			
981	24.2	24.221	1.68			
995	23.9	24.156	1.76			
1009	23.9	24.105	1.83			
1023	23.9	24.064	1.66			
1051	23.7	23.991	2.86			
1065	23.5	23.893	3.3			
1079	23.5	23.814	3.57			
1172	23.4	23.732	5.43	1296	124	TRUE
1178	23.4	23.665	5.16	1296	118	TRUE
1205	23	23.532	4.63			
1219	22.4	23.306	3.7			
1233	22.3	23.105	2.86			
1254	22.4	22.964	2.26			
1268	22.3	22.831	1.7			
1282	22.2	22.705	1.15			
1296	21.4	22.444	0			



- P-0076545:** This patient's BMI at anchor (day 14) was 30.9 kg/m², correctly classified as Obese. The episode itself had a maximum weight loss of 5.37%, correctly meeting the ≥5% threshold. Each highlighted row (in red) indicates a timepoint in which a qualifying weight loss of 5% (or 2% if BMI<20) was detected within the subsequent 180 days.

Table 2. P-0076545 longitudinal BMI and exact timepoints the threshold was detected.

day	raw_bmi	smooth_bmi	max_wl_next_180d	first_day_5pct	days_to_cross	threshold_met
14	30.9	30.9	1.17			
25	31.3	31.122	1.87			
49	29.7	30.539	0			
288	29.8	30.289	0.68			
308	29.6	30.084	0			
1043	24.7	28.625	0			
1352	23.5	27.328	14.81	1408	56	TRUE
1393	24.5	26.648	12.7	1466	73	TRUE
1408	23.5	25.921	10.25	1475	67	TRUE
1463	23.5	25.378	8.66	1477	14	TRUE
1466	23.2	24.902	6.91	1511	45	TRUE
1471	24.1	24.73	6.26	1520	49	TRUE
1475	23	24.364	4.85			
1476	23.6	24.204	4.22			

1477	23.7	24.099	3.81			
1490	22.8	23.832	2.73			
1511	22.7	23.6	2.87			
1520	22.5	23.376	1.93			
1526	22.9	23.28	1.53			
1546	23.2	23.264	1.46			
1595	23.4	23.291	1.58			
1616	24.6	23.555	2.68			
1630	21.7	23.182	1.91			
1672	21.9	22.924	2.24			
1802	22	22.739	5.37	1942	140	TRUE
1848	21.1	22.41	3.99			
1893	21.6	22.247	3.29			
1942	18.6	21.517	0.01			
1947	22.4	21.694	0.82			
2047	20.8	21.515	0.11			
2136	21.4	21.492	-3.16			
2138	25.1	22.214	0.19			
2209	22	22.171	-0.21			
2309	22.4	22.217	-0.07			
2452	22.3	22.234	0			

The broader observation of the Reviewer that *baseline_bmi_cat* appears scrambled reflects a comparison between two distinct timepoints. Of the 129 patients classified as Normal at anchor date with BMI <18.5 at episode start, these patients lost weight between diagnosis and start of episode. The other 1,555 patients classified as Normal at anchor with BMI range of 25-58.4 at episode start reflect patients that gained weight between the anchored diagnosis date and the start of the episode. Lastly, 906 patients mentioned by the Reviewer classified as Obese at anchor with BMI<30 at episode start are patients that lost weight prior to episode onset, an example of this is P-0076545 as shown above. The table below shows the anchor-date BMI vs episode-start BMI for episodes classified as Normal at anchor date with BMI <18.5 at episode start (n=129 episodes) as mentioned by the Reviewer.

Table 3. 129 patients classified as Normal at anchor but presented underweight at episode start

DMP_ID	Anchor BMI	Anchor BMI Category	Episode Start Day	Episode Start BMI	Episode Start BMI Category
P-0005035	21.4	Normal	3233	18.4	Underweight
P-0010661	21.6	Normal	1045	18.3	Underweight

P-0001508	18.7	Normal	1156	17.8	Underweight
P-0069639	18.9	Normal	2994	16.4	Underweight
P-0005240	20.2	Normal	3485	17.8	Underweight
P-0002808	21.7	Normal	464	18.4	Underweight
P-0061778	21.2	Normal	1282	17.7	Underweight
P-0081518	19.4	Normal	1458	18.1	Underweight
P-0032816	21.6	Normal	2086	17.7	Underweight
P-0096101	19.3	Normal	572	17.1	Underweight
P-0036863	19.8	Normal	261	18.3	Underweight
P-0040166	24.9	Normal	3055	18.2	Underweight
P-0003691	19.3	Normal	4532	18.2	Underweight
P-0051949	20	Normal	1408	17.9	Underweight
P-0001483	24.9	Normal	1840	17.5	Underweight
P-0001483	24.9	Normal	2174	18.2	Underweight
P-0005985	19.4	Normal	2351	18.2	Underweight
P-0027364	19.9	Normal	558	17.9	Underweight
P-0027364	19.9	Normal	1894	18.3	Underweight
P-0027364	19.9	Normal	2437	17.5	Underweight
P-0027767	21.1	Normal	1852	17.8	Underweight
P-0001220	19.5	Normal	231	18.4	Underweight
P-0003843	23.9	Normal	1760	18.4	Underweight
P-0000699	22.8	Normal	2119	18.3	Underweight
P-0005510	19.8	Normal	2717	18	Underweight
P-0005510	19.8	Normal	2948	17.9	Underweight
P-0010927	20.5	Normal	2529	16.7	Underweight
P-0010927	20.5	Normal	3266	15.2	Underweight
P-0011163	19.9	Normal	301	18	Underweight
P-0011163	19.9	Normal	1135	18.4	Underweight
P-0001023	21.2	Normal	638	18.3	Underweight
P-0002228	22	Normal	2139	17.6	Underweight
P-0002228	22	Normal	3256	16.2	Underweight
P-0026193	23.6	Normal	1559	18.4	Underweight
P-0022127	22.9	Normal	2179	18.4	Underweight
P-0000923	20.4	Normal	1595	18.3	Underweight
P-0074740	22.3	Normal	3452	18.4	Underweight
P-0011578	24.9	Normal	1683	17.9	Underweight
P-0005840	20.3	Normal	572	16.3	Underweight
P-0008172	24.6	Normal	1354	17.7	Underweight
P-0008169	19.2	Normal	450	18.4	Underweight
P-0001475	19	Normal	1758	18.4	Underweight
P-0001475	19	Normal	2164	18.2	Underweight
P-0018030	19	Normal	818	18.4	Underweight
P-0035594	18.6	Normal	1501	17.3	Underweight
P-0062746	19	Normal	2117	17.7	Underweight
P-0016054	23.7	Normal	1885	18.4	Underweight

P-0004238	20.1	Normal	2664	16	Underweight
P-0004705	24.8	Normal	1202	16.7	Underweight
P-0067686	19	Normal	2203	18	Underweight
P-0010566	18.5	Normal	618	16.8	Underweight
P-0006467	19.5	Normal	761	18.4	Underweight
P-0008809	21	Normal	549	17.3	Underweight
P-0055169	24.8	Normal	1525	18	Underweight
P-0009356	24.4	Normal	2154	16.7	Underweight
P-0061128	19.8	Normal	985	16.4	Underweight
P-0009488	18.7	Normal	314	17.2	Underweight
P-0009948	19.3	Normal	2012	17.8	Underweight
P-0024120	21.6	Normal	293	18.3	Underweight
P-0024120	21.6	Normal	770	17.7	Underweight
P-0082346	22.8	Normal	2245	18.4	Underweight
P-0018701	20.3	Normal	1579	15.7	Underweight
P-0016092	19	Normal	1564	18.2	Underweight
P-0026707	22.4	Normal	593	18.2	Underweight
P-0023248	21.6	Normal	1971	18.1	Underweight
P-0033558	24.1	Normal	1961	17.6	Underweight
P-0041387	21.6	Normal	2837	17.8	Underweight
P-0017965	22.3	Normal	1756	18.3	Underweight
P-0017817	18.5	Normal	664	18.1	Underweight
P-0073120	18.7	Normal	1756	18	Underweight
P-0100867	19	Normal	1297	17.2	Underweight
P-0100867	19	Normal	1683	15.7	Underweight
P-0061565	22.6	Normal	2430	16.5	Underweight
P-0030376	18.5	Normal	1207	18.1	Underweight
P-0026533	22.5	Normal	558	17.8	Underweight
P-0040038	20.1	Normal	446	18.4	Underweight
P-0040038	20.1	Normal	720	18.3	Underweight
P-0035587	20.9	Normal	1510	17.9	Underweight
P-0030852	18.6	Normal	1580	18.4	Underweight
P-0036277	18.7	Normal	594	18	Underweight
P-0047284	19.9	Normal	481	17.2	Underweight
P-0046950	19.6	Normal	329	18.1	Underweight
P-0056318	20.3	Normal	428	18.4	Underweight
P-0053993	22.6	Normal	873	17.5	Underweight
P-0032053	18.8	Normal	338	17.7	Underweight
P-0034996	21.3	Normal	1978	16.3	Underweight
P-0057160	21.8	Normal	3483	18.3	Underweight
P-0037983	19.6	Normal	1254	17.5	Underweight
P-0056942	18.6	Normal	3672	16.3	Underweight
P-0037006	18.8	Normal	614	15.5	Underweight
P-0036604	18.5	Normal	574	18.3	Underweight
P-0038583	19.7	Normal	584	17.8	Underweight

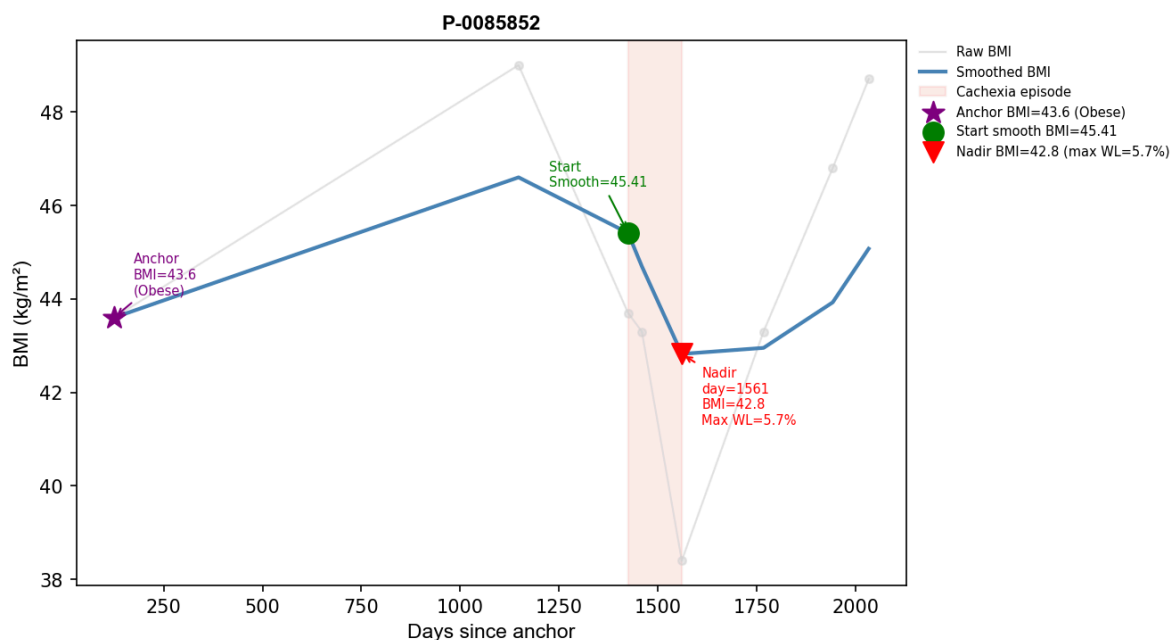
P-0044483	18.9	Normal	1988	18.3	Underweight
P-0038266	20.4	Normal	808	17	Underweight
P-0070765	19.1	Normal	1314	17.4	Underweight
P-0086434	20.5	Normal	1187	18.3	Underweight
P-0048457	19.3	Normal	1223	17.8	Underweight
P-0051055	22.9	Normal	930	17.9	Underweight
P-0054744	19	Normal	750	18.3	Underweight
P-0066239	23.3	Normal	1409	17.7	Underweight
P-0060190	18.5	Normal	1668	18.4	Underweight
P-0053869	21.2	Normal	812	17.5	Underweight
P-0054667	21.5	Normal	1333	17.2	Underweight
P-0057118	22	Normal	1022	16.6	Underweight
P-0063719	18.7	Normal	961	16.9	Underweight
P-0063719	18.7	Normal	1395	16.8	Underweight
P-0067569	21.6	Normal	1261	18.3	Underweight
P-0058577	18.5	Normal	687	18.3	Underweight
P-0060856	23	Normal	924	17.2	Underweight
P-0101048	21.9	Normal	708	18.3	Underweight
P-0062781	19.6	Normal	1341	18.2	Underweight
P-0064290	18.9	Normal	514	18.2	Underweight
P-0078263	20	Normal	683	17.8	Underweight
P-0067387	20.5	Normal	696	17.6	Underweight
P-0079564	21.1	Normal	352	18.2	Underweight
P-0066758	21.4	Normal	850	18.4	Underweight
P-0067897	22.8	Normal	1327	18.1	Underweight
P-0068612	23.8	Normal	480	16.9	Underweight
P-0077790	19.9	Normal	688	17.4	Underweight
P-0079635	18.6	Normal	318	17.4	Underweight
P-0078864	19.7	Normal	613	17.6	Underweight
P-0075531	19.9	Normal	1003	18.3	Underweight
P-0076882	20	Normal	658	18.3	Underweight
P-0082375	18.9	Normal	513	15.8	Underweight
P-0079671	18.8	Normal	764	18.4	Underweight
P-0084510	20.6	Normal	604	18.3	Underweight
P-0090706	19.7	Normal	314	18.1	Underweight
P-0093158	18.7	Normal	237	17.3	Underweight
P-0101938	20.4	Normal	2648	17.5	Underweight

We recognize that the naming of `baseline_bmi_cat` alongside episode-level columns in the supplementary table may have contributed to this misinterpretation, despite these columns being explicitly defined in the data dictionary attached to the original submission. In the updated table, we have renamed this column to `bmi_cat_at_anchor_date` to make the timepoint explicit, and have added `bmi_cat_at_episode_start`, `start_bmi_smooth`, `nadir_bmi` (lowest bmi point in the episode), `nadir_day` (associated relative day of `nadir_bmi`), and `weight_loss_max` to allow direct verification of episode calls within the same row. The data dictionary has been updated

accordingly to reflect all new and renamed columns. We believe these additions address the concerns raised and provide a fully auditable record of the detection framework for each episode.

In the data provided n=29,210 incident episodes of cachexia are found, as mentioned in the results section. I would parenthetically add that this data file is a lot more informative than the previous one, however the %WL column has 2794 episodes of incident weight loss with the value of 2% loss to 4.999% loss. These only qualify as cachexia episodes if BMI is <20 kg/m². Of 2794 patients with low-grade weight loss n=1856 of them had BMI >20 kg/m² (including 119 with class II obesity and 55 with class III obesity). It is with disbelief, I ran across P-0085852, who had a starting body mass index of 43.7 and had a “cachectic episode” comprising the loss of 2.006% of her body weight. Overall, 1856 episodes were improperly included.

Response: **P-0085852:** As described above, the weight_loss column in the supplementary table reflects a post-hoc start-to-end summary. The weight_loss_max for this patient was 5.70%, correctly meeting the ≥5%WL threshold. The full trajectory is shown below.



Additional basic data quality concerns.

Whilst looking at the largest % weight loss values in the data set, I ran across P-0053346 with an implausible weight loss coded as -50.87% during an episode duration of 16 days (BMI 56.4 fell to BMI 27.7). There are other patients with these characteristics of implausibly large weight loss (up to -68.1%), and some of these did so in a very short period. Some of these huge losers are also people with implausibly large BMI; such values can result from typographical errors in either the weight or the height of the person. I am experienced in this type of data and know that there is a certain level of error inputting the data, such as a missed decimal place or missed digit. Because height is the denominator in BMI, errors in height can result in bizarre body mass index values. There are some extremely heavy people in this population, which bears

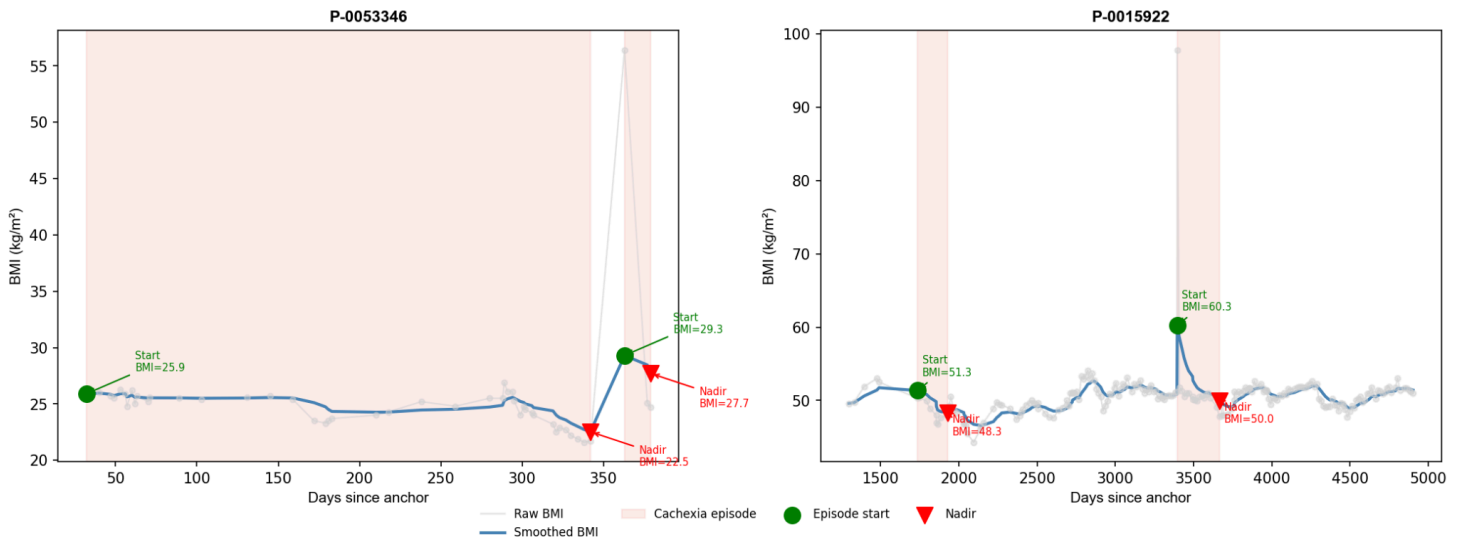
discussion if their BMI and weight loss percentage can be verified. To achieve a BMI of 50.0 an average height man would have to weigh upwards of 350 pounds. To achieve some of the higher BMIs (max value is 97.8 kg/ m²), weight would reach 675 pounds. I encourage therefore, the authors to access the height and the weight data and verify as much as possible that there are no errors in the input variables.

Response: We acknowledge that EHR data contains measurement and typographical errors in either weight and height, and that some extreme BMI values in the dataset likely reflect such errors. This is an inherent limitation of large-scale real-world data and not unique to this cohort. The minimum episode duration threshold of 15 days was a conscious design choice to exclude spurious short episodes that may arise from transient measurement errors, and the EWMA smoothing pipeline further mitigates the impact of anomalous single measurements by down-weighting them relative to the surrounding trajectory.

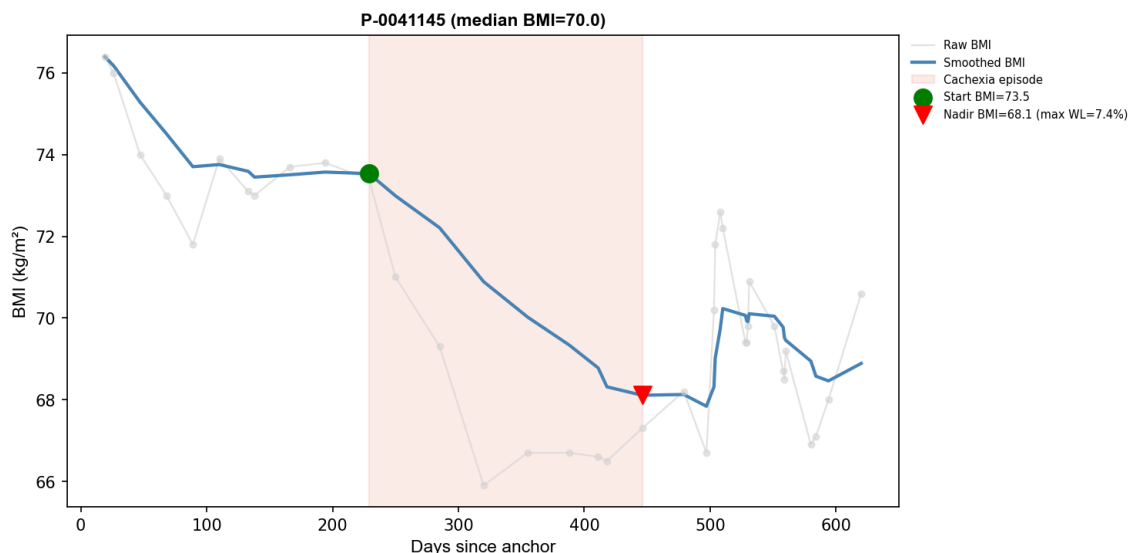
Table 4. Sensitivity analyses re-runs

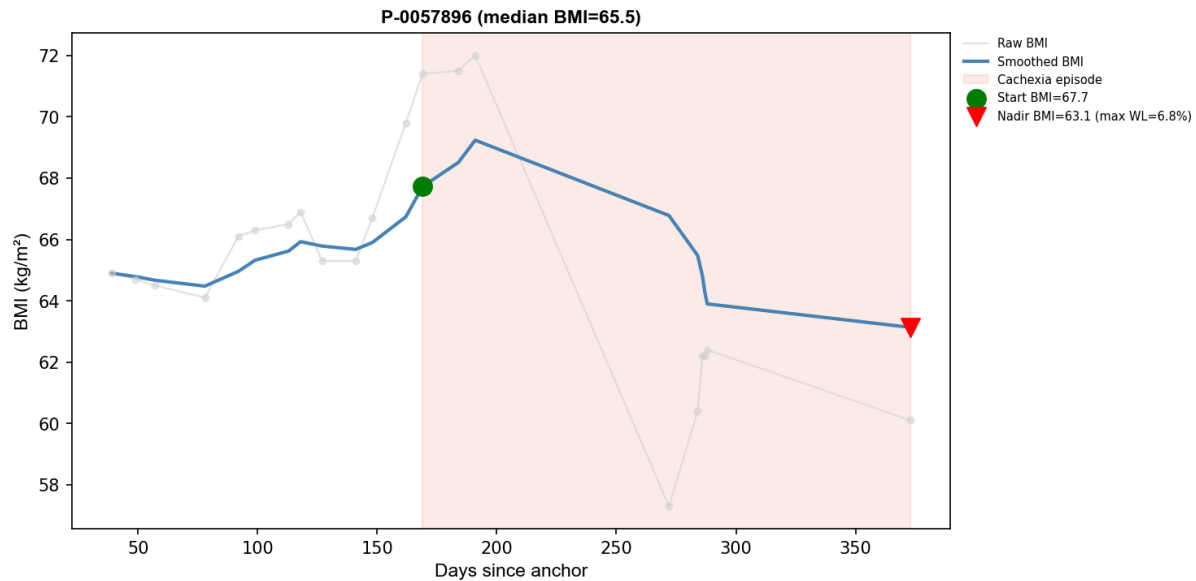
Setting	EMWA	Min episode duration (days)	Edema QC threshold (%)	Edema QC threshold (days)	% patients with episodes post QC	Episode loss	% patients with 1,2,3+ episodes post QC
baseline	0.2	15	5	30	44.7% (23,372)	-664	35.3%,8.01%,1.46%
α low	0.1	15	5	30	35.7% (18,648)		30.76%,4.48%,0.45%
α high	0.3	15	5	30	50.4% (26,326)		37.55%, 10.2%, 2.65%
Min. dur. 30	0.2	30	5	30	44.5% (23,273)	-763	35.2%,7.92%,1.44%
Min. dur. 45	0.2	45	5	30	44.23% (23,102)	-934	35.0%,7.8%,1.42%
Edema QC looser	0.2	15	3	30	42.62% (22,261)	-1775	34.03%, 7.34%, 1.24%
Edema QC stricter	0.2	15	7	30	45.53% (23,783)	-253	35.68%, 8.30%, 1.55%
Edema QC short window	0.2	15	5	14	45.63% (23,837)	-199	35.71%,8.35%,1.56%
Edema QC long window	0.2	15	5	45	44.07% (23,022)	-1014	34.89%, 7.78%, 1.40%

Critically, we note that in sensitivity analyses requested by Reviewer 3 and conducted by us in the prior submission, patients such as this would all be removed. These sensitivity analyses adjusted the quality control threshold for minimum episode durations of 30 and 45 days. Importantly, (1) changing these parameters increases specificity of cachexia calls but also induces loss of patients with *bona fide* weight loss meeting the cachexia criteria, and more importantly (2) in all cases, when adjusting these parameters, the core findings of the entire manuscript were preserved (Table 4). This establishes that our results are not driven by short-duration edge cases. In cases such as P-0053346 and P-0015922, the BMI does likely reflect an entry error as shown below:



We, however, note that to the Reviewer 1's point, *to achieve a BMI of 50.0 an average height man would have to weigh upwards of 350 pounds*, that these extremely high BMI values do exist in this population. In our dataset, there are 588 patients with a mean BMI greater than 45 kg/m² across their trajectory, reflecting a genuinely high-BMI subpopulation. These patients consistently have elevated BMI values across multiple measurements and are not artifacts of data entry errors. Two representative examples are provided below (P-0041145, P-0057896).





The columns `hasWL10` and `hasWL15` in the data set also appear to be incorrect and would not render proper results if these categorical values were used in any analyses. These are supposed to indicate incident weight loss exceeding -10%, and -15%, respectively. In the `%weight loss` column there are 6161 rows with a value $\geq 15\%$ (range 15.0% to 68.2%), however only 3226 of the episodes of incident weight loss $\geq 15\%$, has a value of “1” in the categorical `hasWL15` column. Likewise, the column `hasWL10` has a value of “1” for 3489 episodes of WL $\geq 10\%$, however $n=3072$ patients with a weight loss $\geq 10\%$ have a value of zero in the `hasWL10` column

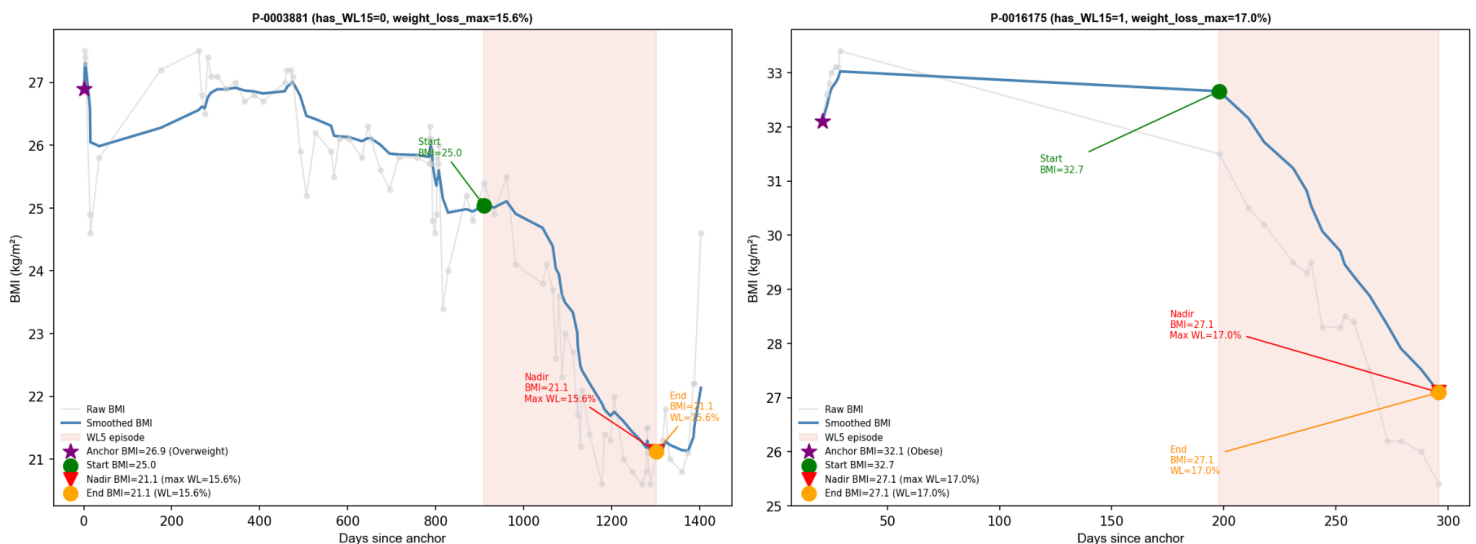
Response: The `has_WL10` and `has_WL15` columns are patient-level binary flags derived from independent reruns of the episode detection algorithm at 10% and 15% weight loss thresholds over 6 months respectively, as defined in the data dictionary attached to the original submission:

`has_WL10`: Patient-level; whether the patient had any validated cachexia episode under the independently re-run 10% weight-loss (WL10)

`has_WL15`: Patient-level; whether the patient had any validated cachexia episode under the independently re-run 15% weight-loss (WL15)

These flags are not derived by filtering the `weight_loss` column in this table. A patient may have `weight_loss` $\geq 15\%$ in the primary WL5 run but `has_WL15` = 0 if the independent 15% rerun did not detect a qualifying episode under its own 180-day windowing and merging rules. The two computations are entirely separate. We acknowledge that including these patient-level flags alongside episode-level columns in the same table may have caused confusion and are happy to provide the full episode-level output from the independent WL10 and WL15 reruns as separate supplementary tables if this would aid interpretation.

To illustrate this distinction, we provide two representative examples in the figures below. The first patient (P-0003881) has $\text{has_WL15} = 0$ despite a $\text{weight_loss_max} = 15.6\%$. This episode spans 392 days across merged windows and the gradual loss never exceeded 15% within any single 180-day window in the independent rerun. On the other hand, the second case patient (P-0016175) has $\text{has_WL15} = 1$ with $\text{weight_loss_max} = 17.0\%$. The nadir was reached rapidly within a short 98-day episode, triggering a qualifying episode in the independent 15% WL over 180 days (6 months) rerun.



The next discover is that the weight loss grades are not applied correctly. Only the weight loss incidents are classified, so the account of there being just 90 patients with weight loss grade of 0, in the population of over 52,000 patients is not correct. The grading system is applied to all patients, so in a population weight stable patients with a BMI >28 kg/m² (no weight loss detected) are graded 0. Weight stable patients with a BMI between 20 and 24.9 are Grade 1, and weight stable patients with a BMI <20 are Grade 3.

Response: We thank the Reviewer for this point. The Weight Loss Grading System was applied to detected episodes of weight loss in this manuscript, as the primary aim was to characterize the severity of clinically meaningful weight loss events. We acknowledge that the full Martin et al. framework can be applied also to patients with sub-threshold weight loss of 0-2.4%, and that applying WLGS more broadly across non-episode reference group would allow for a more granular characterization of patients into weight stable, weight gain, sub threshold weight loss. We recognize this as a valid point and will apply WLGS across the non episode group if invited to revise, but we emphasize that this analysis is trivial to execute and does not undermine any of the conclusions of the manuscript.

To this point in my commentary, my conclusion is that the manuscript is not evaluable until such time as all of these errors and miscoding are corrected in the data set.

2. Method of calling cachexia episodes. I continue to have several questions about this approach. This is described as a “sliding-window episode-detection framework” and illustrated in

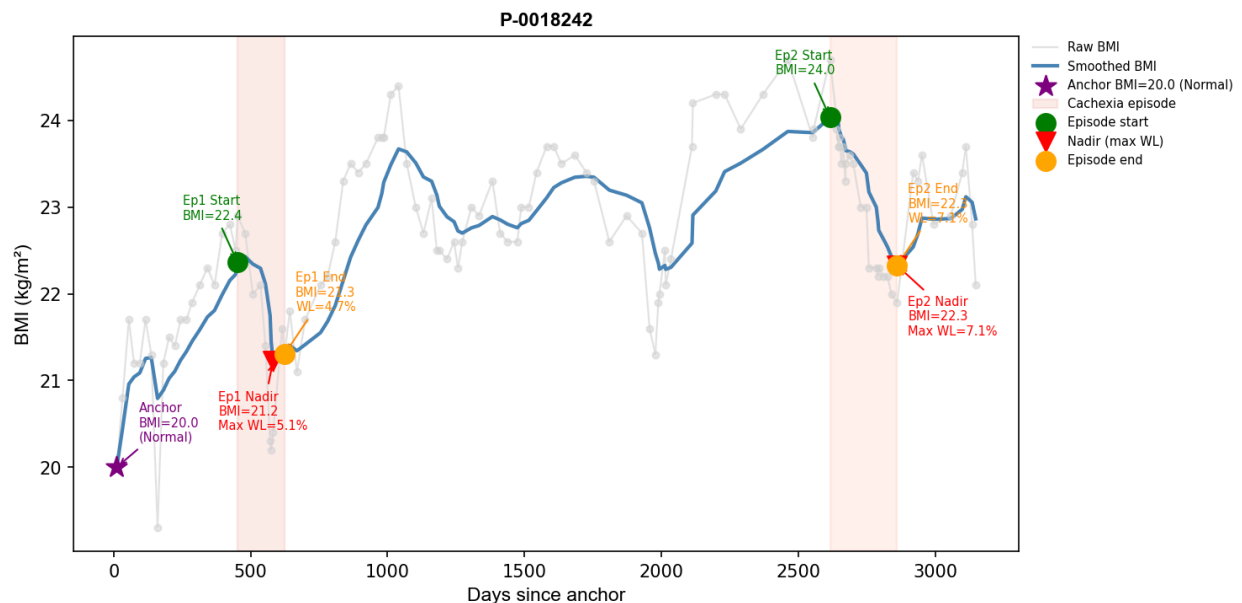
Suppl Fig 1C. This is a single depiction of a body weight trajectory, not claimed to be representative; it would be helpful to see other trajectories to understand the application of the cachexia episode calls. The illustrated trajectory shows gain, stepwise loss and then some limited regain. How many episodes show complete regain, partial regain or no regain? Did you consider a regain parameter? In advanced disease some of these trajectories just go straight down. How often do we see gain (from baseline) or regain following an episode and what is the range of magnitudes?

What I am trying to grasp is the “cachexia” in certain examples

P-0018242 had a loss-gain-loss (BMI 22.9 > 21.3 > 24.7 > 22.3); This has 2 episodes of loss but a 3.4 kg/m² gain. Is this cachexia if they started at 22.9 and ended at 22.3?

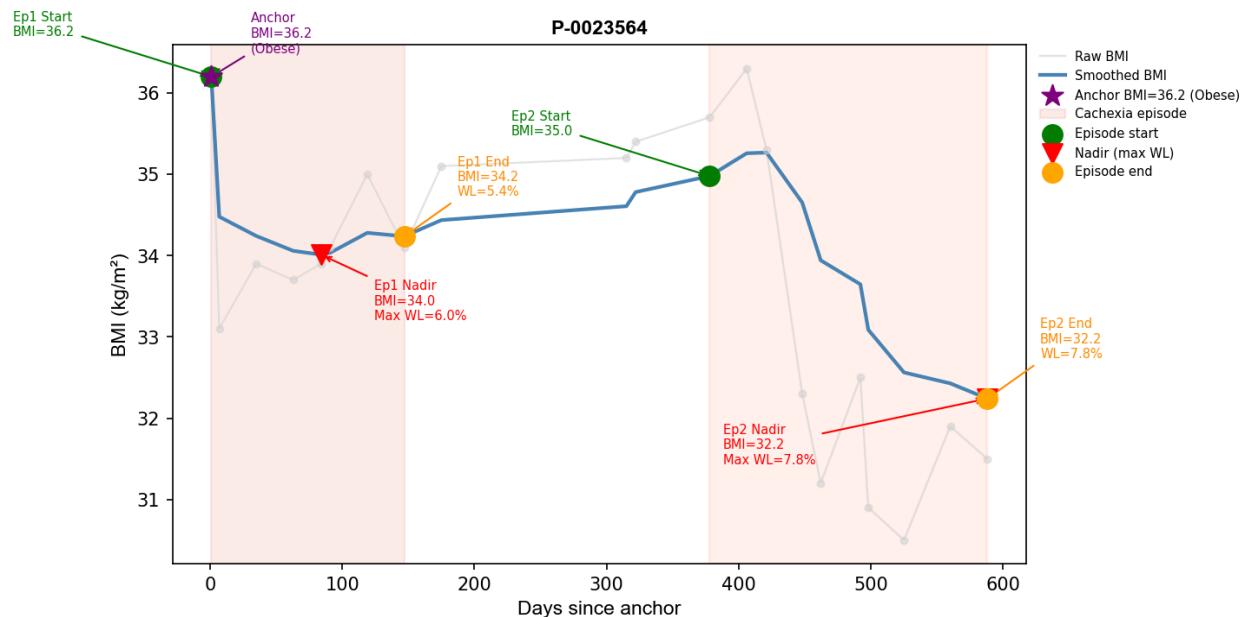
Response: Before addressing each patient in turn, we would emphasize that not only is the full code for episode calling available, but we also provide example data to illustrate the mechanics of calling. The Reviewer has decided not to examine this code (although other Reviewers have, and have commented on its clarity).

P-0018242: This patient had two discrete episodes each independently meeting the $\geq 5\%$ threshold within their respective 180-day detection windows (Episode 1: weight_loss_max=5.05%; Episode 2: weight_loss_max=7.10%). Our framework detects discrete episodes of clinically meaningful weight loss as they occur throughout the disease course and does not require that weight loss be sustained across the full observation period. The inter-episode weight gain reflects a period of recovery between two independent loss events. Each episode is evaluated within its own 180-day window consistent with the Fearon 2011 consensus definition.



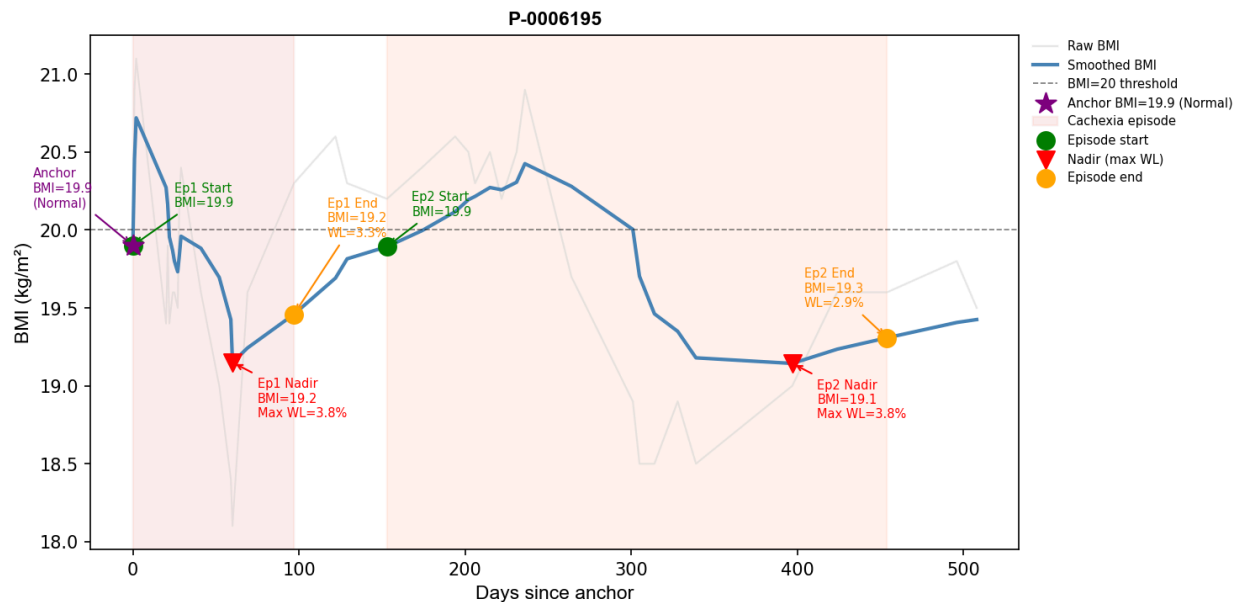
P-0023564 had a loss-gain-loss (BMI 36.2 > 34.2 > 35.7 > 32.2) Is this cachexia of concern given the high body weight?

Response: This patient had two discrete episodes each independently meeting the $\geq 5\%$ threshold within their respective 180-day detection windows (Episode 1: weight_loss_max= 6.05% ; Episode 2: weight_loss_max=7.83%), separated by a period of weight stability. Clinically meaningful weight loss occurs across the full BMI spectrum and is not limited to patients with low or normal BMI.



P-0006195 has 2 episodes: 3% loss, regain, 4% loss Cachexia episodes minimal absolute value I am not able to go through all of the data, these are just examples, with undulating BMI but very little variation- this strikes me as over-classification.

Response: This patient's episodes qualify under the low-BMI modifier of 2% WL if BMI<20. Both episodes have start_bmi_smoothed <20, triggering the 2% weight loss threshold which explicitly recognizes that patients with depleted body stores are at clinically meaningful risk even at lower absolute weight loss values (Episode 1: weight_loss_max=3.77%; Episode 2: weight_loss_max=3.77%). Importantly, these episodes are called precisely because of the insistence of Reviewer 1 in applying the Fearon 2011 criteria of 2% weight loss in BMI<20 patients, which (in our initial submission) we indicated would potentially lead to over-classification. It appears the Reviewer has come around to this idea. Importantly, we note that because we apply several different weight-loss calling criteria, an external analyst such as the Reviewer could easily identify and remove this patient applying their criteria of choice. This emphasizes the strength of data transparency which is unique to our cachexia dataset. The full trajectory for this patient is provided in the figure below.



3. Severity of cancer cachexia- how to understand it?

This version is much clearer on the severity issue, but a “try everything” approach has been taken towards it. Several analyses include 5% weight loss alone, as well as %, 10% and 15% cutoffs for weight loss %. Then the 2% weight loss with BMI<20 is added, and then the Weight loss Grades 0-4 (Martin et al.) are added. The 5% criterion value is still taken as the “primary” definition. All of this could be collapsed into the Weight loss grades, as a single primary outcome for the following reasons: the lower threshold of 2.5% is included, data driven thresholds of 6%, 11% and 15% are included, and the BMI<20 with 2.5-4.99 % weight loss is included. Taking this approach would also accommodate the fact that your population trends significantly to being overweight and obese. A major concept in cachexia severity is that the impact of weight loss is not the same across the spectrum of BMI. I believe further consideration should be given to the impact of obesity, with n=9311, of 29,219 instances of weight loss, 31.9% occur in obese persons. N= 1357 of those meet criteria for Class III obesity and n=2363 meet criteria for Class II obesity; n=200 have a BMI> 50 kg/m² with the highest value being 97.8. Very heavy people do you have a reduced risk of mortality as shown by Martin and coworkers in 2015. A weight loss of 6% is Grade 2 if BMI is > 28, is Grade 3 if BMI is between 22 and 28 and is Grade 4 in persons with a starting BMI of 20 or below. Insofar as the current samples have a very high disproportion of extremely heavy people, it begs the question as to whether these authors should undertake to determine the relative risk of mortality according to weight loss across their 9311 obese patients. Depending on the number of events, the sample may be sufficiently powered to do so. The population comes from a very high obesity risk country with a number of obesity-related cancers included, tending to skew the BMI distribution to the right. Other countries may have quite different demographics, so this sample may not be generalizable. When stratifying patients according to Weight Loss Grades, I do suggest that the full scale be used, however if these are to be grouped for any reason, Grades 0,1 and 2 should be grouped

as they largely encompass weight loss <6% and weight-stable patients, and this is quite distinct from Grades 3 and 4.

Response: We appreciate these comments on severity classification. Again, we emphasize that Reviewer 1 specifically requested the implementation of these criteria in the first review, but now takes issue with the fact that we have implemented them.

The multiple weight loss thresholds, 10% WL and 15% WL over 6 months were included to facilitate comparability with existing literature that uses varying definitions, with the (i) 5% WL criterion and (ii) 2% WL if BMI<20 retained as the primary definition per Fearon 2011 consensus. WLGS-stratified analyses were already conducted for both genomic associations and survival analyses in the prior revision. Furthermore, regarding the obesity point in this patient population, 31.9% of episodes in this cohort occur in obese patients, and we recognize that more granular dichotomization of severity using the WLGS scale for sub-threshold weight loss and weight stable patient would capture more granular weight loss distribution across the BMI spectrum. Lastly, we note that the high obesity prevalence in this US-based cohort may limit generalizability to populations with leaner BMI distributions, and future work in non-US cohorts would be a valuable direction. The core findings including prevalence and genomic analyses were validated in an independent external cohort from Dana Farber Cancer Institute with fundamentally different patient characteristics, supporting the robustness of the conclusions.

Association vs causation: a bundle of secondary analyses

The presentation of the data begins with what we are led to understand is the primary purpose: to detect the prevalence of incident weight loss (of varying degrees). As suggested above, this data presentation could be made more concisely.

Response: We thank the Reviewer for these comments. We will address each point in turn.

There is an odd interjection of radiologically determined fat loss data, that doesn't add to the narrative in any substantial way. It's only a partial data set, and it is muscle rather than fat that sits at the centre of the cancer cachexia definition and shows association with cancer outcomes. I suggest that this be deleted. The authors cannot offer meaningful insight into the body composition piece.

Response: We acknowledge this point. We note that Reviewer 2 specifically requested this analysis and found it valuable (see Comment below). We defer to the editor on whether to retain or remove this section.

- A large part of the paper concern association of incident weight loss with primary cancer site, cancer stage, progression on cancer therapy, and cancer driver mutations, "serological signatures" (standard clinical lab values), age, sex, performance status. The pieces of this are rather disjointed, and I would hope that these could be represented in a coherent model. A) Tumor activity drives the wasting syndrome in the first instance, with its stage, progression and cancer driver mutations being part of one bundle. B) the serological signatures do not represent any catabolic effector molecules i.e. there are no actors in the list capable of eliciting adipose

tissue lipolysis, muscle proteolysis or suppress food intake. So, there is no possibility of causation, only association. It has not occurred to the authors that many of the blood changes are secondary i.e. myelosuppressive drugs such as Platinum agents: Carboplatin, Cisplatin, Oxaliplatin, Alkylating agents: Busulfan, Cyclophosphamide, Chlorambucil, Carmustine, Antimetabolites: Fluorouracil (5-FU), Capecitabine, Gemcitabine, Topoisomerase Inhibitors: Topotecan, Irinotecan, Etoposide, Anthracyclines: Doxorubicin. Regimen (not included in this analysis) would be a more direct contributor to weight loss as systemic therapy drives wasting by generating GI symptoms such as nausea and vomiting, and also has direct cytotoxic effects on muscle and fat. PMID: 36942895 summarizes recent findings on different systemic therapy regimens, that induce double digit muscle loss in ~3 months and have specific cytotoxic effects on muscle. PMID: 37686641 is an interesting read, which shows that male sex, progressive disease on therapy are independent drivers of muscle loss in patients with pancreatic cancers, and that FOLFIRINOX is specific driver of muscle loss while gemcitabine + nab paclitaxel is a fat loss driver. C) age, and sex have been included in some analyses, but any sex difference in cachexia prevalence or severity is not reported. I am not certain why ECOG PS is included as a factor, as it is more likely an outcome of cachexia (see Lancet Oncology 2011) than a driver of it.

Response:

We agree entirely that causation cannot be established from these analyses, and have stated this explicitly in the Limitations section of the manuscript: "*as with all observational analyses, our findings establish associations rather than causation.*" We also acknowledge that treatment regimens are important contributors to the serologic changes observed during episodes and that many blood changes may reflect chemotherapy-induced effects. As noted in the Limitations, sophisticated statistical models will be necessary to adequately capture the convolved effects of therapy on cachexia, and this represents an important direction for future work.

Regarding demographics, 45.0% of female patients (12,475/27,720) and 44.4% of male patients (10,897/24,516) had at least one detected episode, and in adjusted pan-cancer analyses, we reported that male sex was associated with slightly lower odds of meeting the episode criterion ($\log_2\text{OR} = -0.19$, Supplementary Table 2). Similarly, cachexia episode prevalence was comparable across age groups, with 44.6% of patients under 60 (10,149/22,775) and 44.9% of patients aged 60 and older (13,223/29,461) having at least one detected episode. Sex-stratified and age-stratified analyses of severity may represent a valuable direction for future work. We, lastly, acknowledge that ECOG may partly reflect cachexia severity rather than drive it, and note that the predictive model was validated in an independent external cohort without ECOG with all performance of the model preserved.

Reviewer #1 (Remarks on code availability): The data set is filled with errors. As such, I have not attempted to evaluate the code.

The step-by-step walkthrough notebook linked below illustrates the full detection pipeline using an example patient trajectory and directly addresses how the framework is operationalized on longitudinal data.

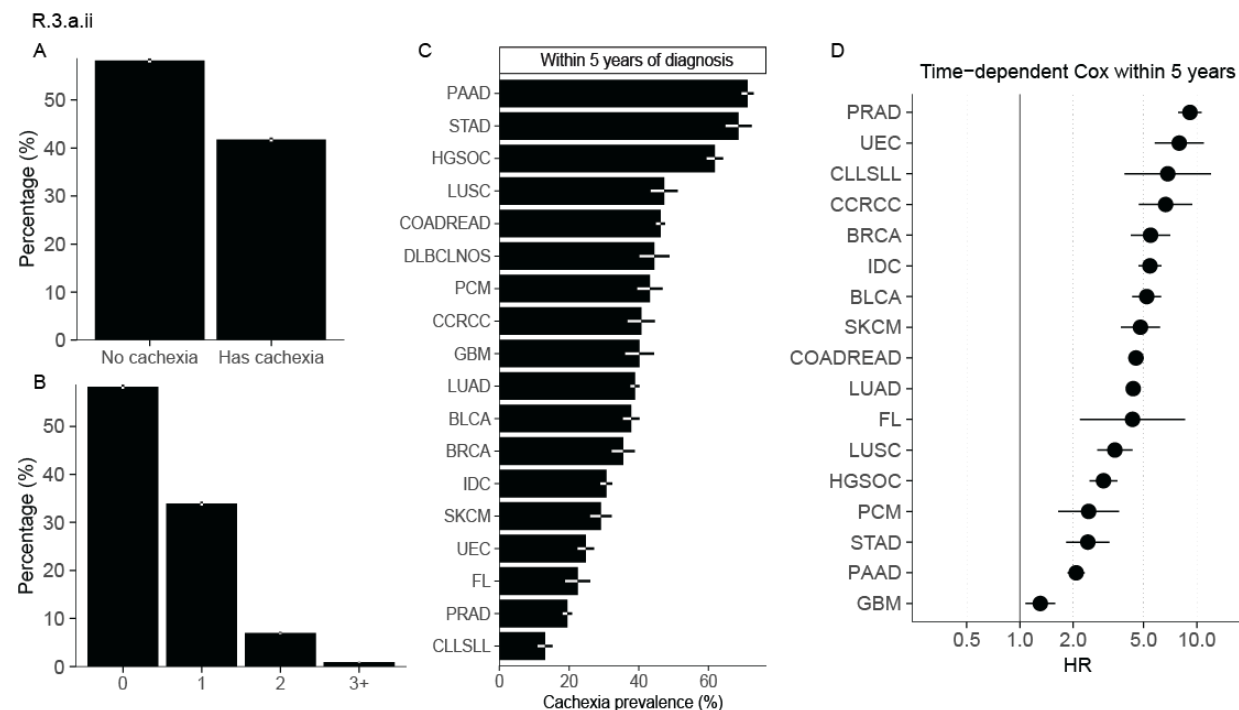
https://github.com/reznik-lab/cachexia/blob/main/walkthrough/episode_identification_walkthrough.ipynb

Reviewer #2 (Remarks to the Author):

The authors have undertaken a thorough and careful revision in response to all reviewer comments. Nevertheless, several important issues remain.

First, I agree with the previous reviewer that the inclusion of patients with follow-up extending beyond 5 years may be problematic in the context of cancer-related weight loss. A 10-year time window after diagnosis appears excessively broad, as many patients with such long-term follow-up are likely to be cured (unless there are secondary cancers or a late recurrence etc.). An analysis restricted to the first 5 years after diagnosis may therefore be more appropriate for capturing weight loss that is plausibly attributable to cancer and its treatment.

Response: We agree this is a reasonable concern. In the revision we had already conducted a sensitivity analysis restricting episodes to within 5 years of diagnosis, and the core findings were preserved as shown below. We are happy to include this if invited to revise our manuscript. We will address this further.



Second, the limitations should more explicitly address the challenges of interpreting weight loss in patients with obesity. In this group, clinically meaningful wasting may occur in the context of sarcopenic obesity, which is difficult to ascertain using routine clinical data. Weight loss alone cannot distinguish between loss of adipose tissue and loss of skeletal muscle, both of which may have different biological and clinical implications. This limitation should be acknowledged more clearly. However, the authors have addressed changes in VAT and SAT in the population which is highly appreciated.

Response: We thank the Reviewer for this comment and for acknowledging the VAT and SAT analysis. We agree that the challenges of interpreting weight loss in patients with obesity, particularly in the context of sarcopenic obesity, warrant more explicit discussion in the Limitations section.

Third, the application of the Fearon criteria represents an important improvement. However, although the authors note that insufficient data were available for complete classification, the current dichotomization into non-cachexia versus cachexia remains a limitation because it does not consider pre-cachexia as a clinically relevant intermediate stage. Please add to the limitations explicitly that assigning pre-cachectic individuals to the non-cachexia group may increase heterogeneity within the reference category and potentially obscure stage-specific associations.

Response: We acknowledge this limitation and agree that assigning pre-cachectic individuals to the non-episode reference group may introduce some heterogeneity within that category. We will add this explicitly to the Limitations section in revision.

Relatedly, Reviewer 3 raised the issue of cachexia episodes in the period preceding death. In this context, the authors should consider whether refractory cachexia can be evaluated. Based on the information provided, at least some of the required data appear to be available, particularly survival <3 months and performance status (ECOG), consistent with the framework proposed by Fearon et al. If a formal classification is not feasible, this should at minimum be discussed explicitly (see above)

Response: In our prior revision we addressed the question of end-of-life cachexia episodes raised by Reviewer 3, demonstrating that 31-38% of deceased patients had their final episode extend to end of life with no observed recovery. We agree that a formal proxy classification of refractory cachexia using available data including survival less than 3 months and ECOG performance status, consistent with the Fearon et al. framework, is feasible and represents a meaningful extension of this analysis. We will add explicit discussion of this to the Limitations section and are prepared to conduct this analysis.

Finally, I appreciate that the authors have now included age at diagnosis as a covariate in response to my earlier comment regarding early-onset cancer. Given the size and depth of this dataset, a stratified analysis by age at onset could be highly informative. If this lies beyond the scope of the current manuscript, it would still represent a valuable direction for a subsequent study and would likely be of considerable interest to investigators working in this area.

Response: We thank the Reviewer for this suggestion and will note age-stratified analysis explicitly as a valuable direction for subsequent work.

Reviewer #3 (Remarks to the Author):

The authors have comprehensively addressed the majority of the reviewer comments. In particular, they have incorporated multiple sensitivity analyses across weight-loss thresholds, stage strata, and independent cohorts, with results presented throughout the manuscript and supplementary materials. The methodological clarity has substantially improved, especially regarding episode definition and progression annotation.

Notably, the revision reflects major improvements in several key areas: methodological rigor has been significantly strengthened, sensitivity analyses are now clearly and systematically presented, genomic findings have been reinforced through more robust and multivariable analyses, and external validation has been incorporated to support generalizability.

While some analyses remain primarily descriptive rather than fully statistical, the revisions sufficiently support the robustness of the conclusions. No further revisions are required.

Reviewer #3 (Remarks on code availability):

While the GitHub repository is well structured and documented, it would benefit from the inclusion of a requirements file (e.g., `requirements.txt` or `environment.yml`) to facilitate installation of dependencies and improve reproducibility for users.

Response: We thank Reviewer 3 for their thorough re-evaluation of the manuscript. We are pleased that the revision comprehensively addressed the majority of concerns, and that the methodological rigor, sensitivity analyses, genomic findings, and external validation were found to be sufficient to support the robustness of the conclusions. We will add the `requirements.txt`.

Reviewers' Comments:

Reviewer #2 (Remarks to the Author):

I would like to thank the authors for their thoughtful and detailed responses to my comments.

Regarding my first comment, my concern has been sufficiently addressed. I would encourage the authors to include the additional discussion on differences observed when restricting the analysis to patients within five years of diagnosis in the manuscript, as this would help readers better interpret the findings.

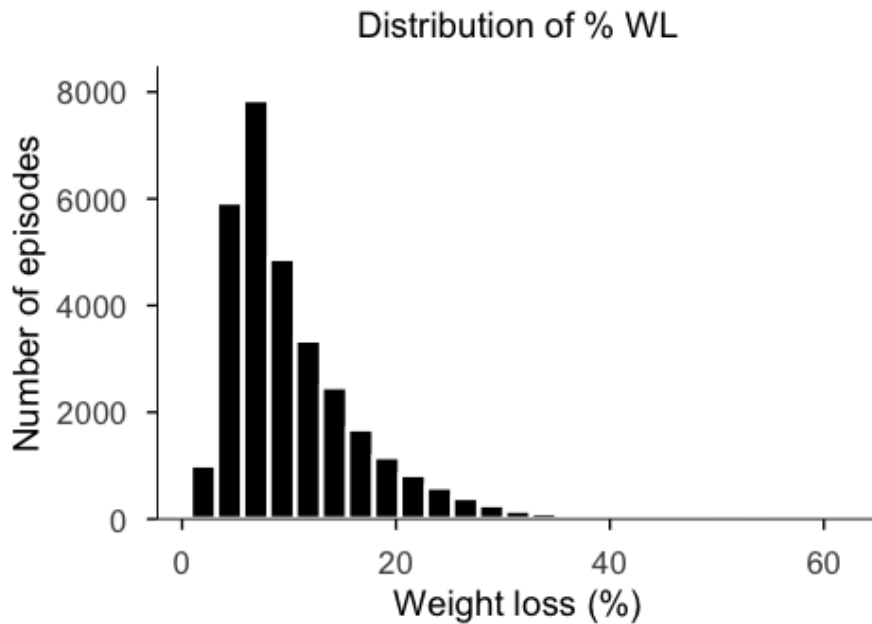
Response: Thank you, we have now included discussion on this point on **Page 6**.

Regarding the second comment, the response was adequate, and I welcome the authors' decision to address this point in the discussion. The same applies to my third comment regarding the Fearon classification and to my fourth comment regarding refractory cachexia. In both cases, the authors have addressed my earlier concerns in a constructive and scientifically appropriate manner.

Response: Thank you, we have now included additional text on these points in the **Limitations** on **Page 15**.

I also appreciate the authors' willingness to add further nuance to the interpretation of the findings. Given the size and uniqueness of the dataset, some implausible observations, such as apparent drops or reversals in weight loss, may be easy to overlook. However, I agree with the concern that such patterns should ideally have been detected through routine data-quality checks. Even a brief histogram or distributional summary of percentage weight loss would have helped to identify and contextualize these observations.

Response: Thank you for the comment. We have now included a histogram of percentage weight loss across cachexia episodes in **Supplementary Figure 2A** and **Page 4**. The distribution of weight loss (%) episodes has a mean of 10.4%, consistent with most episodes reflecting moderate clinically meaningful weight loss.



Overall, the authors have addressed the main points raised in my earlier review. The additional clarification and willingness to revise the discussion improve the objectivity of the manuscript and will help readers interpret the findings more appropriately. My only remaining conceptual concern relates to the terminology of “cachexia episodes,” which is not fully congruent with the traditional understanding of cachexia as a syndrome that is not readily reversible. However, I also recognize that this is an important issue for the field and that the manuscript may contribute meaningfully to an ongoing re-evaluation of current cachexia definitions.

Response: Thank you for the feedback. We acknowledge that the nomenclature of cachexia episodes is unfamiliar to the field at large. However, it is not incongruent with cachexia as not readily reversible, but instead (1) emphasizes that weight loss occurs over specific time periods, and (2) anchors analysis around weight loss time periods. Furthermore, as requested by Reviewer 3, we now tabulate and demonstrate that a significant fraction of patients recover weight subsequent to cachexia. We agree that these points are the fruit of exciting ongoing and future conversation in the cachexia field, and mention this in the **Discussion** on **Page 13**.

Reviewer #3 (Remarks to the Author):

I have reviewed the revised manuscript and the authors' point-by-point responses. My overall assessment is unchanged from the prior round, the study is methodologically sound, the sensitivity analyses are thorough, and the external validation supports the robustness of the central findings.

Response: We thank Reviewer 3 for their careful and thorough re-evaluation of the manuscript.

Having examined the data integrity concerns raised in this round in detail, I find the authors' responses persuasive. The majority of the flagged discrepancies arise from comparing quantities defined at different timepoints (anchor/diagnosis vs. episode start) or computed by distinct procedures. The worked trajectories, the anchor vs episode start comparison table, and the replication of analysis on an independent cohort adequately substantiate this. The column renaming, the added episode start and nadir fields, and the expanded data dictionary directly resolve the source of confusion, and I consider these concerns addressed.

Response: Thank you.

A few of Reviewer 1's points are scientific rather than clerical, and the editor may wish to consider whether the authors address them more fully:

1. Treatment confounding of the serologic signature. Several of the reported shifts (anemia indices, albumin, transaminases, leukocyte populations) are also expected consequences of cytotoxic chemotherapy. Given that the serologic signature is a primary result, I recommend more than a Limitations acknowledgment, at minimum a clearer framing that these associations may partly reflect treatment, and if possible, a sensitivity check (e.g., whether the directionality persists outside peri-treatment windows or after adjusting for therapy exposure where available).

Response: Thank you for the comment. We agree that several components of the serologic signature, particularly anemia indices, transaminases, and leukocyte populations, are plausible consequences of cytotoxic chemotherapy in addition to cachexia-intrinsic biology. We have added text to the Results (**Page 9,10**) explicitly acknowledging this. As a partial sensitivity check, we note that the signature's direction was preserved when restricting to early-stage patients relative to Stage 4 patients (Supplementary Fig. 4D), a population with a different treatment intensity and timing profile, which is consistent with (though does not fully isolate from) a

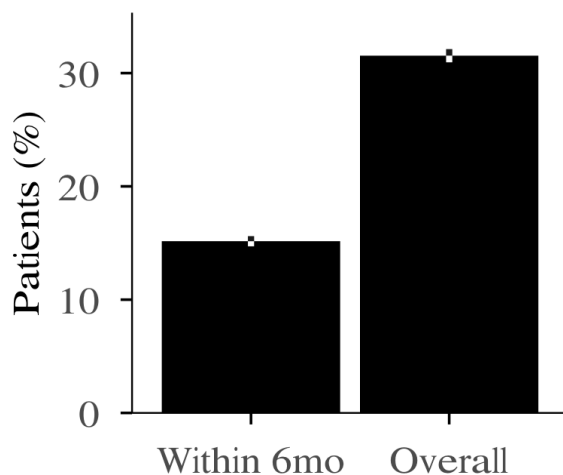
treatment-independent component of the signature. We agree that a direct adjustment for treatment-exposure timing would more definitively resolve this question and identify this as a valuable direction for future work using treatment-timing data linked to this cohort.

2. WLGS scope. As Reviewer 1 notes, the Martin et al. grading system is defined across the full population, whereas it was applied here only to detected episodes. The authors should state this scoping explicitly, so the grade distribution is not misread as a population-level classification.

Response: Thank you, this is now reflected in the text on **Page 5**, and made clearer in the Methods (**Page 19**).

3. Post-episode trajectory. Reviewer 1 raised a reasonable question about regain following episodes. The CT-based SAT/VAT analysis already addresses this in part, showing adipose regain in roughly 59–66% of episodes in the imaging subset, but a complementary BMI level quantification across the full episode set (complete/partial/no regain, with magnitude ranges) would extend this beyond the imaging cohort and more directly address the question raised.

Response: Thank you. We have now included a complementary BMI-level analysis of weight regain (**Methods, Page 20**) following cachexia episodes across the full cohort on **Page 6**. Of the 30,928 cachexia episodes identified, 26.7% showed significant weight regain ($\geq 5\%$ BMI gain from episode end) at any point following episode end, and 12.5% within 6 months, consistent with partial post-episode weight regain in the LUAD imaging subset.



On the body composition section, I would suggest retaining it, as it provides orthogonal evidence that episodes correspond to real adipose depletion, provided its claims remain appropriately bounded given the partial coverage.

Response: We agree, thank you.

Regarding the code, please ensure the new auditing columns and a requirements file are committed to the repository.

Response: We have now confirmed that the new auditing columns (with data dictionary) and requirements files are committed to the repository (github.com/reznik-lab/cachexia), including separate dependency files for R (requirements_R.txt) and Python (requirements_py.txt).

Reviewer #3 (Remarks on code availability):

I have reviewed the public analysis code (github.com/reznik-lab/cachexia). The episode detection logic confirms the post-hoc weight_loss field is computed start to end and is distinct from the threshold that triggers an episode call, which accounts for nearly all of the "miscoded" cases. In my assessment the data integrity objections are largely resolved by the code and documentation. I would hold the authors to their commitment to add a dependency/requirements file, which is still absent from the repository.

Response: Thank you so much, we have now added a dependency/requirements file.