A comparison of CT based measures of skeletal muscle mass and density from the Th4 and L3 levels in patients with advanced non-small-cell lung cancer

Bjørn H. Grønberg,1,2 Bjørg Sjøblom,3 Tore Wentzel-Larsen,4,5,6 Vickie E. Baracos,7 Marianne J. Hjermstad,8,9 Nina Aass,3,8 Roy M. Bremnes,10,11 Øystein Fløtten,12 Asta Bye,9,13 Marit Jordhøy8,14

1Department of Clinical and Molecular Medicine, Faculty of Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
2The Cancer Clinic, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway
3Department of Oncology, Oslo University Hospital, Oslo, Norway
4Norwegian Centre for Violence and Traumatic Stress Studies, Oslo, Norway
5Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, Oslo, Norway
6Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway
7Department of Oncology, Division of Palliative Care Medicine, University of Alberta, Edmonton, Canada
8Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
9European Palliative Care Research Centre, Department of Oncology, Oslo University Hospital, Oslo, Norway
10Department of Oncology, University Hospital North Norway, Tromsø, Norway
11Department of Clinical Medicine, Faculty of Medicine, University in Tromsø, Norway
12Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway
13Department of Nursing and Health Promotion, Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway
14Department of Internal Medicine, Innlandet Hospital Trust, Hamar, Norway

Keywords:

lean body muscle mass; skeletal muscle mass; prognostic factor; body composition
Abstract:

Background

Muscle mass and density assessed from CT-images at the L3 level are prognostic for survival and predict toxicity in cancer patients. However, L3 is not always included on routine CT-scans. We aimed to investigate whether images at the Th4 level may be used instead.

Methods

Patients from three chemotherapy trials in advanced NSCLC were eligible (n=1305). Skeletal muscle area (cm²), skeletal muscle index (SMI, cm²/m²) and skeletal muscle density (SMD) at Th4 and L3 levels were assessed from baseline CT-scans. SMI and SMD at the Th4 and L3 level were transformed into z-scores and the agreement between scores was investigated by Bland-Altman plots and estimated by intra-class correlation analyses. Linear regression was used to test if Th4 SMI and SMD z-scores predicted L3 SMI and SMD z-scores.

Results

CT-images from 401 patients were analyzable at both levels. There was a moderate agreement between Th4 and L3 SMI z-scores with an intra-class correlation of 0.71 (95% CI 0.64–0.77) for men and 0.53 (95% CI 0.41–0.63) for women. Regression models predicting L3 SMI z-scores from Th4 SMI z-scores showed coefficients of 0.71 (95% CI 0.62–0.80) among men and 0.53 (95% CI 0.40–0.66) among women. R-squares were 0.51 and 0.28 respectively, indicating moderate agreement. A similar, moderate agreement between Th4 and L3 SMD z-scores was observed.

Conclusion

There was only moderate agreement between muscle measures from Th4 and L3 levels, indicating that missing data from the L3 level cannot be replaced by analyzing images at the Th4 level.
Introduction

Changes in human body composition related to aging and disease is gaining increasing interest. A particular focus has been rendered to muscle wasting and thereby loss of lean body mass (LBM). In aging, muscular depletion is associated with frailty and several negative health outcomes, including mortality.\textsuperscript{1, 2} In cancer populations, an increasing body of evidence links this feature to cachexia,\textsuperscript{3} worse survival,\textsuperscript{4-7} and increased risk of toxicity from systemic cancer therapy.\textsuperscript{8-12} Associations with postoperative infections and delayed recovery after surgery for colorectal cancer have also been reported.\textsuperscript{13} Muscle wasting may occur in obese patients (sarcopenic obesity) as well as in those who are normal or underweight. It is, however, frequently undetected since both weight and body mass index (BMI) are poor indicators of LBM.\textsuperscript{14}

There are several options for body composition assessment, including bioelectrical impedance analyses (BIA), dual energy X-ray absorptiometry (DXA) and analyses of computed tomography (CT) images.\textsuperscript{15} The latter method is particularly convenient in oncology settings due to frequent, routine CT-imaging for diagnosis, staging, treatment evaluation and follow-up. In contrast to BIA and DXA, CT images provide specific details on muscle characteristics, adipose tissues and organs. Furthermore, skeletal muscle area quantified from a single CT slice at the third lumbar level (L3) is closely correlated to the estimated total lean body skeletal muscle mass (LBM).\textsuperscript{15, 16} Thus, utilizing CT images at the L3 level to assess body composition has become the gold standard in studies on cancer patients.\textsuperscript{3, 17}

CT based assessment makes it possible to measure skeletal muscle radiodensity (SMD) in addition to muscle mass. SMD is expressed as the mean Hounsfield Units (HU) of the measured cross sectional muscle area. Low values reflect increased fat deposits,\textsuperscript{18} are associated with older age,\textsuperscript{19, 20} and when measured at the lumbar level, they are also linked to worse survival in cancer patients.\textsuperscript{7, 21} In non-cancer populations, both SMD- and age-related differences between muscle groups have been found, indicating that the underlying etiological factors for muscle wasting may not affect all muscles similarly.\textsuperscript{19}

In non-small cell lung cancer (NSCLC), cachexia and muscle wasting are common and associated with worse prognosis and increased risk of treatment toxicity.\textsuperscript{7, 12, 22} However, diagnostic work-up of these patients is usually restricted to a CT-scan of the thorax and upper abdomen which often does not include the L3 level. Thus, CT-images at the fourth thoracic level (Th4) have been used to assess skeletal muscle mass and its relation to survival in lung cancer patients.\textsuperscript{23, 24} There is, however, limited knowledge about the agreement between muscle-measures at the L3 and at Th4 level,\textsuperscript{25} and none have compared muscular SMD at these levels in cancer patients. Based on data from three Norwegian randomized controlled trials (RCT) comparing first line chemotherapy regimens in advanced non-small cell lung cancer (NSCLC),\textsuperscript{26-28} we aimed at investigating whether L3 muscle mass and SMD might be reliably predicted from Th4 measures.
Methods

Study sample
The trials which this study is based upon were conducted from 2003 to 2009, and the main inclusion criteria were: Chemonaïve patients, age ≥18 years, stage IIIIB/IV NSCLC and performance status (PS) 0-2. In all trials, the diagnostic work-up included a CT scan of the thorax and upper abdomen obtained within four weeks before chemotherapy commenced. These CT scans were collected retrospectively for assessment of LBM. For the present study, we included patients if the baseline CT-scan included analysable images both at the Th4 and L3 levels.

Body composition assessments
The diagnostic CT scans were analysed using Slice-O-Matic software (v.4.3 Tomovision, Montreal Canada) by three similarly trained observers blinded for other patient data. The first image in the caudal direction where both vertebral transverse processes were visible was used to manually outline the skeletal muscle tissue at the Th4 and L3 level, respectively. Based on pre-established thresholds of Hounsfield Units (HU) in the range of –29 to + 150 HU, the cross-sectional areas (cm²) of the outlined muscle tissues at the Th4 and L3 levels were automatically calculated by the software, normalised for stature (height squared), and expressed as Th4 and L3 skeletal muscle index (Th4 SMI, cm²/m² and L3 SMI, cm²/m²). Optimally the whole circumference of the body should be included in the images at the L3 and Th4 levels to enable an exact quantification of the respective tissue areas. In some patients, parts of the muscular tissue were missing on the CT scans. If less than half of the circumference was missing, the total area was estimated by doubling the area of the opposite half of the body. If more than half of the circumference was missing, no quantification was possible and the patient was excluded from the analyses. SMD was assessed as the mean HU of the entire cross sectional muscle area at levels Th4 and L3.

The patients’ BMI (weight (kg)/height (m²)) were calculated based on baseline data from the RCTs. No systematic registration of weight loss at baseline was conducted, hence we used appetite loss registered on the European Organisation of Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) as a supplementary indicator of nutrition status.

Statistics
Data from all RCTs were analysed jointly. Body composition measures were compared between men and women by independent sample t-tests, and all analyses investigating agreement between measures at the Th4 and L3 level were done for each gender separately.
First, we investigated the agreement between the L3 skeletal muscle area, SMI and SMD and the corresponding measures at the Th4 level using scatterplots. Then, the SMI and SMD from both levels were transformed into z-scores, separately for men and women. The agreement between Th4 SMI z-scores and L3 SMI z-scores were investigated by Bland-Altman diagrams with locally fitted smooth (loess) curves, and by intraclass correlation. Whether Th4 SMI and SMD z-scores could predict L3 SMI and SMD z-scores were tested using linear regression. Finally, we tested the precision with which individual missing L3 SMI and SMD values could be estimated by using the patients’ z-scores from the corresponding Th4 SMI and SMD values. L3 SMI was recomputed using the mean L3 SMI for the cohort + SD x Th4 SMI z-score. The L3 SMD was recomputed similarly. The agreement between actual and recomputed L3 SMI and SMD were then examined by scatter plots.

All p-values were two-sided and p-values < 0.05 were used to define statistical significance. The statistical analyses were performed using IBM SPSS version 18 (IBM Corporation, Armonk, NY, USA).

Ethics

The study was performed according to the Helsinki declaration and approved by the Regional Committee for Medical and Health Research Ethics in South-East Norway.

Results

Overall, we were able to retrieve CT scans from 1119 of the 1305 study participants (85.7%). Among these, 688 scans did not include images at the levels of interest or enough of the circumference, or the quality was too poor for the analyses (Figure 1). Furthermore, 30 patients were excluded due to missing data on SMD either at the L3 or Th4 level (24 patients) or on relevant baseline characteristics (e.g. height and weight) (6 patients). Thus, 401 patients (30.7%) were included in the present study (Figure 1). The main baseline characteristics of these patients are presented in Table 1. 220 were men (54.9%); mean age was 66 years; 100 (25%) were younger than 60 years, 79 (19.7%) were 75 years or older; 316 patients (78.8%) had stage IV disease; and 89 (22.2%) had PS 2.

Body composition

The mean cross-sectional muscle area (cm²) and the SMI (cm²/m²) of the overall study sample were larger at the Th4 level than at the L3 level: 176.4 cm² versus 130.6 cm², and 60.0 cm²/m² versus 44.5 cm²/m². Th4 SMD was also higher than the L3 SMD in the overall sample (41.5 HU vs 36.9 HU) both among men (42.0 HU vs. 37.2) and women (40.8 vs 36.5) (Table 2). Comparing men to women, muscle area and SMI were
significantly larger in men, whereas no significant difference between genders was found for SMD. The muscle measures were close to normally distributed.

**Agreement between thoracic and lumbar muscle measures**

Scatterplots of the Th4 and L3 muscle area (cm²), and Th4 and L3 SMI (cm²/m²) showed a substantial spread around the lines of complete agreement, indicating only moderate agreement (Figure 2).

A Bland Altman plot (Figure 3A) investigating the agreement between Th4 and L3 SMI, transformed into corresponding z-scores, showed no substantial systematic deviation between the two levels and no substantial difference by gender. There was, however, a considerable spread in the difference between Th4 and L3 z-scores, and the intraclass correlation (single measures) was 0.71 (95% CI 0.64 – 0.77) for men and 0.53 (95% CI 0.41 – 0.63) for women, i.e. consistent with a medium agreement. Regression models predicting L3 SMI z-scores from Th4 SMI z-scores showed coefficients of 0.71 (95% CI 0.62 - 0.80) in the male population and 0.53 (95% CI 0.40 - 0.66) among females. The R squares for these models were 0.50 and 0.28 respectively, indicating that the Th4 SMI z-scores were only moderately related to the L3 SMI z-scores.

Regarding the agreement between z-scores transformed from Th4 and L3 SMD, the Bland Altman plot (Figure 3B) showed results fairly consistent with those for the SMI, except that the spread of differences was considerably larger. The intraclass correlation (single measures) between Th4 SMD and L3 SMD z-scores was 0.71 (95% CI 0.64 – 0.77) for men, and 0.76 (95% CI 0.70 – 0.82) for women. The regression models predicting L3 SMD z-scores from Th4 SMD z-scores showed closely similar coefficients for men and women, 0.71 (95% CI 0.62 - 0.80) and 0.76 (95% CI 0.67 – 0.86), respectively. The R squares for these models were 0.50 for men and 0.58 for women.

Scatterplots of the actual L3 SMI and SMD plotted against the L3 SMI and SMD recomputed by Th4 SMI by z-scores (Figure 3 B and C) showed a substantial spread of the actual values when compared to the estimated values.

**Discussion**

In this study comparing muscle measures from CT images at both Th4 and L3 levels, using widely accepted methodology, we found that the muscle area was larger at the thoracic level in both genders. There was also a substantial difference between the Th4 SMD and L3 SMD, with higher SMD in the thoracic muscle.

Furthermore, the agreement between SMD and SMI at the two levels was only moderate, and for SMI there was also less agreement between Th4 and L3 among the women than among the men. According to
regression analyses, z-scores at the Th4 level were not strongly related to L3 z-scores. The agreement between actual L3 SMI and SMD and the measures recomputed by means of Th4 z-scores was moderate. We are aware of only one other study comparing muscle measures at the thoracic- and lumbar levels in cancer patients. Kim et al. analysed 90 patients with both limited and extensive small-cell lung cancer, and found poor agreement between pectoral muscle mass at the level above the aortic arch (which is approximately at the Th4-level) and cross sectional muscle area at the L3 level. Though there are differences in patient populations, software for assessing muscle area, the thoracic level for muscle assessment, and muscle groups measured, their study support our findings.

Body composition analyses were not a pre-planned part of the RCTs we collected data from. CT images of the thorax and upper abdomen were mandatory for trial inclusion, but specific requirements for the CT protocols were not defined in the study protocols. Adequate CT-images at both levels were available for only 38% of the patients. We anticipated that muscle measures at the Th4 level would be available for the majority of patients, whereas images at the L3 level would be missing in more cases. As it turned out, a large number of the Th4 level images were insufficient for muscle analyses. This was mostly due to “cutting of edges”, i.e. the outer circumference of the muscle mass was missing, or the image quality was not satisfactory for quantification of muscle mass. Thus, future studies of LBM in cancer patients should include specific instructions to radiology departments to ensure that body composition can be assessed.

A strength of our study is the large sample size of patients with similar diagnosis and stage of disease, though the cohort was too small to allow for subgroup analyses. None of the patients had received any former systemic cancer treatment, and the study sample included a relatively large proportion of elderly and PS 2 patients. Thus, although muscle measures could be obtained for only a minority of the targeted population, we find it reasonable to believe that our findings are representative for advanced NSCLC patients eligible for first-line palliative chemotherapy. For generalisation of our results, confirmation from other studies and other cancer populations is, however, necessary.

CT images at the L3 level include core muscles, such as the rectus abdominis, external and internal oblique and erector spinae, which are assumed to initiate most full-body functional movement and are fundamental for stabilizing the body in dynamic movements. Although some of these muscles (erector spinae) extend into the Th4 level, the major muscles captured at Th4, such as the pectoralis muscles, have other functions, mainly related to arm and shoulder movements. Their volume and strength may therefore to a larger extent depend on specific manual activities, and activities that more often apply to men than women. These functional differences between the muscle groups might contribute to the only moderate agreement between Th4 SMI/SMD and the L3 SMI/SMD, although the reasons may be more complex. We have not
found any good explanations in the literature, but a substantial difference in SMD between muscle groups has formerly been reported.\textsuperscript{19} We are not aware of any studies investigating whether there is a different impact of cancer-related muscular depletion between muscle groups.

The gold standard for measuring LBM is analysing whole body CT or MRI scans. Analyses of single slices may not predict the LBM correctly, especially in longitudinal studies,\textsuperscript{29} but is currently the most feasible approach in larger and multicentre studies of cancer patients. Whole body CT scans are seldom available unless it is part of specific studies. Thus, such scans were not available from our patients, and it was not possible for us to investigate whether the Th4 or L3 SMI is in best agreement with the whole body muscle mass. Further studies are needed to investigate the relationship between Th4 muscle measures and whole body skeletal muscle mass, and the clinical role of Th4 muscle measures. Until such studies are conducted, we believe that adequate CT images at the L3 level remains the recommended approach in studies of the clinical role of muscle measures in cancer patients.

In conclusion, there is a large variation between the skeletal muscle areas at the Th4 and L3 levels in patients with advanced non-small-cell lung cancer, and muscle measures at the L3 level cannot be reliably estimated by transformation of measures at the Th4 level using z-scores.

**Conflicts of interest and source of funding**

The study was funded by the South-Eastern Norway Regional Health Authority. The collection of CT scans was supported by unrestricted grants from Pierre Fabre, Norway. The Canadian participation in the body composition analyses was supported by the Canadian Institute of Health Research and Alberta Cancer Foundation. None of the authors have any conflicts of interests to declare.

**Acknowledgements**

We want to thank Rachel Murphy (PhD) and Nina Esfandiari (B.Sc) both at Dept of Oncology, University of Alberta, Canada, for their participation in the body composition analyses. Thanks to Nina Helbakkmo and Ingrid Sandstad for handling the collection of CT scans from the RCT1. We also thank the Departments of Radiology at Haukeland University Hospital, Bergen, St Olav University Hospital, Trondheim and Hamar, Gjøvik and Kongsvinger Hospital, Innlandet Hospital HF, as well as the following and their respective Departments of Radiology for their participation in the collection of CT scans: Bjørn Wembstad, Hammerfest Hospital and Kåre Hansen, Kirkenes Hospital; Helse Finnmark HF. Nina Helbakkmo, University Hospital North Norway Tromsø and Nada Zafran Groh, University Hospital North Norway Harstad; University Hospital North Norway HF.
References


Figure 1: Patient selection

Figure 2: Scatterplot illustrating the agreement between measures at the Th4 and L3 for muscle area (cm$^2$), skeletal muscle index (SMI) (cm$^2$/m$^2$) and skeletal muscle radiodensity, for men and women separately. A line for perfect agreement has been added to all plots.

Figure 3: A) Bland Altman plot for the agreement between Th4 SMI and L3 SMI z scores (with loess curves for each gender). B) Bland Altman plot for the agreement between Th4 SMD and L3 SMD z scores (with loess curves for each gender). C) Scatter plot showing actual L3 SMI values and L3 SMI values recomputed from Th4 SMI-scores (by z-scores) (linear fit line for overall sample with 95% CI and loess curves for each gender). D) Scatter plot showing actual L3 SMD values and L3 SMD values recomputed from Th4 SMI-scores (by z-scores) (linear fit line for overall sample with 95% CI and loess curves for each gender).

Table 1: Baseline characteristics

Table 2: Body composition measures at the Th4 and L3 levels
Table 1  Baseline characteristics

<table>
<thead>
<tr>
<th></th>
<th>All patients (n=401)</th>
<th>Men (n=220)</th>
<th>Women (n=181)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Mean (range)</td>
<td>66 (37-90)</td>
<td>68 (37-90)</td>
</tr>
<tr>
<td></td>
<td>≥ 75 years</td>
<td>79 19.7%</td>
<td>48 21.8%</td>
</tr>
<tr>
<td>Histology</td>
<td>Squamous cell carcinoma</td>
<td>92 22.9%</td>
<td>64 29.1%</td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma</td>
<td>217 54.1%</td>
<td>104 47.3%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>92 21.0%</td>
<td>52 23.7%</td>
</tr>
<tr>
<td>Disease stage</td>
<td>IIIB</td>
<td>85 22.9%</td>
<td>47 21.4%</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>316 78.8%</td>
<td>173 78.6%</td>
</tr>
<tr>
<td>Performance status</td>
<td>0</td>
<td>80 20.0%</td>
<td>46 20.9%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>232 57.9%</td>
<td>122 55.5%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>89 22.2%</td>
<td>52 23.6%</td>
</tr>
<tr>
<td>Body weight, kg, mean (SD)</td>
<td>69.0 (13.8)</td>
<td>73.7 (11.9)</td>
<td>65.1 (13.1)</td>
</tr>
<tr>
<td>Body Mass Index, kg/m², mean (SD)</td>
<td>23.9 (3.9)</td>
<td>23.8 (3.4)</td>
<td>23.9 (4.5)</td>
</tr>
<tr>
<td>Appetite loss</td>
<td>Yes</td>
<td>211 52.6%</td>
<td>113 51.4%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>190 47.4%</td>
<td>107 48.6%</td>
</tr>
</tbody>
</table>
Table 2  Body composition measures at the Th4 and L3 levels

<table>
<thead>
<tr>
<th></th>
<th>Measures at the Th4 level</th>
<th>Measures at the L3 level</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All (n=401)</td>
<td>Men (n=220)</td>
<td>Women (n=181)</td>
</tr>
<tr>
<td>Measured muscle area, cm²</td>
<td>Mean 176.4 (SD 39.6)</td>
<td>Mean 200.7 (SD 31.7)</td>
<td>Mean 147.0 (SD 25.8)</td>
</tr>
<tr>
<td>Skeletal muscle index (SMI), cm²/m²</td>
<td>Mean 60.1 (SD 10.9)</td>
<td>Mean 65.0 (SD 10.1)</td>
<td>Mean 54.1 (SD 8.8)</td>
</tr>
<tr>
<td>Skeletal muscle radiodensity (SMD), HU</td>
<td>Mean 41.5 (SD 6.9)</td>
<td>Mean 42.0 (SD 6.8)</td>
<td>Mean 40.8 (SD 6.9)</td>
</tr>
</tbody>
</table>

*p-value for the comparison between men and women
Figure 1  Patient selection

All patients (n=1305)
- RCT 1 (n=432)
- RCT 2 (n=436)
- RCT 3 (n=437)

CT images not received (n=186)
- RCT 1 (n=174)
- RCT 2 (n=0)
- RCT 3 (n=12)

CT images collected (n=1119)
- RCT 1 (n=258)
- RCT 2 (n=436)
- RCT 3 (n=425)

CT images at the Th4 level not analysable* (n=340)
- RCT 1 (n=85)
- RCT 2 (n=122)
- RCT 3 (n=133)

CT images at the L3 level not analysable ** (n=130)
- RCT 1 (n=37)
- RCT 2 (n=51)
- RCT 3 (n=42)

CT images at both Th4 and L3 not analysable *** (n=218)
- RCT 1 (n=59)
- RCT 2 (n=97)
- RCT 3 (n=62)

TH4 and L3 analyses conducted, (n=428)
- RCT 1 (n=77)
- RCT 2 (n=166)
- RCT 3 (n=188)

Eligible for analyses (n=401)
- RCT 1 (n=73)
- RCT 2 (n=147)
- RCT 3 (n=181)

Missing data on muscle radiodensity at either L3 (n=7) or Th4 (n=13) or both (n=4) and relevant baseline data (3)
- RCT 1 (n=4)
- RCT 2 (n=17)
- RCT 3 (n=6)

*Whole cross sectional area not included; or too poor image quality
**Lack of images at the L3-level; whole cross sectional area not included in the images; or image-quality too poor
*** Either of the above
Figure 2  Scatterplots illustrating the agreement between measures at the TH4 and L3 for muscle area (cm²), skeletal muscle index (SMI) (cm²/m²) and skeletal muscle radiodensity, for men and women separately. A line for perfect agreement has been added to all plots.
Figure 3  
A) Bland Altman plot for the agreement between Th4 SMI and L3 SMI z-scores (with loess curves for each gender).  
B) Bland Altman plot for the agreement between Th4 SMD and L3 SMD z-scores (with loess curves for each gender).  
C) Scatter plot showing actual L3 SMI values and L3 SMI values recomputed from Th4 SMI-scores (by z-scores) (linear fit line for overall sample with 95% CI and loess curves for each gender).  
D) Scatter plot showing actual L3 SMD values and L3 SMD values recomputed from Th4 SMI-scores (by z-scores) (linear fit line for overall sample with 95% CI and loess curves for each gender).