Reviewer's report

Title: Risk factors of acute and overuse musculoskeletal injuries among young conscripts: a population-based cohort study.

Version: 3
Date: 24 November 2014

Reviewer: Rasmus Gottschalk Nielsen

Reviewer's report:

Comments for the author

The topic itself is interesting and worthwhile. In general, the manuscript reads well and the quality of written English is good (from a non-native perspective) and the coherence is excellent. The design is strong and the sample size appropriate.

Even though the paper is interesting and well written, I do, however, have some major issues (primarily statistical), which need to be addressed. Since these comments may lead the authors to reanalyze their data, specific comments are provided at a later stage in the review process.

Major comments:

1) Statistics, page 8, Line 9-15: A Forward Likelihood Ratio method to identify the variables to be included in the multivariate adjusted Cox regression model. However, this approach may lead to estimates in the final model, which are biased away from the null, confidence intervals that tend to be narrow, and P-values to be small (For instance, see Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52(10):935-942.) Although the referenced article focus on logistic regression some parallels to cox regression may be drawn. I see it as a strong limitation that the final result is mathematically / statistically driven (P < 0.20) and not based on a clinically-oriented / research-oriented hypothesis.

The authors must justify their use of stepwise procedures to compute their final result based on a multivariate regression (with references to relevant supportive statistical / epidemiological literature) OR reanalyze their data and base their hypothesis on a non-mathematical approach.

2) The authors interpret the hazard rate ratio as a relative risk (for instance, page 12, line 4-6 or page 2, line 17-20). Since the outcome was frequent (27% sustained acute injury and 51% suffered from overuse), the risk is overestimated if a hazard ratio is interpreted as a risk. Please revise throughout the manuscript. (For additional information, please refer to Symons MJ, Moore DT, Hazard rate ratio and prospective epidemiological studies, Journal of Clinical Epidemiology 2002, 55: 893-988).

3) Statistical analyses, page 8, line 1-7: Please describe the Cox regression in a
more detailed manner:
- What was the time-scale?
- How did the authors validate the assumption about proportionality - log-minus-log plots; Observed versus expected plots; other? Was the assumption for proportionality checked for ALL explanatory variables? If the rates are proportional for all variables, the use of cox regression is appropriate. If the rates are not proportional, the authors may consider using other regression models (for example: a generalized regression model using the pseudo values method, Klein et al. Analyzing survival curves at a fixed point in time. Stat med. 2007; 26(24): 4505-4519.) rather than the log rank method traditionally used if the assumption of proportionality is violated.
- It remains unclear how the authors took into account the possibility for sustaining AI or OI amongst the non-injured. From my perspective, all conscripts were at risk for AI and OI at baseline and, as a consequence, the data should be analyzed using competing risks (See Putter H, Fiocco M and Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med 2007; 26: 2389-2430). Did the authors do this?
- Did the authors analyze time-to-first injury or did they include injury reoccurrences?

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests