Reviewer's report

Title: Population-based SEER trend analysis of overall and cancer-specific survival in 5,138 patients with gastrointestinal stromal tumor

Version: 2 Date: 6 April 2015

Reviewer: Adam Olszewski

Reviewer's report:

Guller et al. describe a survival analysis of patients diagnosed with GIST between 1998 and 2011 using the US SEER registry data base. Their primary question was well defined as the impact of introduction of imatinib and further TKI's on survival in GIST as measured by population statistics. They conclude a significant improvement in overall and cancer-specific survival in the study period, which they ascribe to the effect of new systemic therapies.

BMC Review questions:
1. Is the question posed by the authors well defined? Yes.
2. Are the methods appropriate and well described? Generally, yes, with some caveats discussed further.
3. Are the data sound? Generally, yes, with some caveats discussed further.
4. Do the figures appear to be genuine, i.e. without evidence of manipulation? Yes.
5. Does the manuscript adhere to the relevant standards for reporting and data deposition? Yes.
6. Are the discussion and conclusions well balanced and adequately supported by the data? Yes, with some reservations.
7. Are limitations of the work clearly stated? Yes, with some reservations.
8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished? Generally, yes, with some caveats discussed further.
9. Do the title and abstract accurately convey what has been found? Yes
10. Is the writing acceptable? For the most part, although the text needs further editing.

Discretionary Revisions:
1. I would recommend harmonizing the nomenclature of all genes (KIT, PDGFR, BCR, ABL) according to HUGO Nomenclature. Correct “brc-abl” in particular (it should be bcr). Generally, human GENES should be capitalized and italicized, while PROTEINS should be capitalized.
2. Names of generic drugs (imatinib) need not be capitalized.
3. Some language errors make certain sentences incomprehensible. For
example: “For selected pairs, size respectively mitotic rate serving as the cut-off value are shown”. I would also not refer repeatedly to metastatic or adjuvant “situation”.

4. Results (Page 6) and Table 1. Consider using either cm or mm for the size consistently to avoid confusion. Listing size statistics to accuracy of 0.1 mm is unlikely to be of any value.

5. Table 1 lists N=10 patients with “unknown primary surgery”, but then N=25 patients with “unknown surgery” and M1 disease. Please clarify this discrepancy.

6. Figures and Results (Page 6, Line 167, 169): Please name the panels on Figures 2, 3, 4, 5 and refer to specific panels in the text (rather than “Figure 2”).

Minor Essential Revisions:

1. Please explain in the Methods (Page 4, Line 108) the actual primary site and histology codes used. This is important because of known issues with identification of GIST cases in the US registries, which the authors did not discuss as a limitation of the study. As reviewed e.g. in http://seer.cancer.gov/seerinquiry/index.php?page=view&id=20100014&type=q only “malignant GIST tumors” are reportable to the SEER registry, and it is unclear whether the registry captures all GIST cases appropriately, and how this had worked out historically, before the specific ICD-O-3 code for malignant GIST was introduced in 2001. Perez et al. (J Am Coll Surg. 2006 Apr;202(4):623-9) showed a 25x increase in reported GIST incidence from 1992 to 2002 in the SEER data, which is almost certainly related to reclassification of various tumors as GIST. This may affect survival rates and make survival trends in this analysis unreliable.

2. The Methods state that “Chi-square statistics and t-tests were used to compare proportions and continuous variables”, but I see no analyses that would be using those tests in the paper. Moreover, the authors state that “Continuous data are expressed as median and interquartile range”, but then (Page 6, Line 152) describe age as mean and SD. Please explain or correct.

3. This analysis shows no difference in prognosis between gastric and small intestinal GIST, which is different than in prior studies. For example, the Gold criteria show significantly worse prognosis for tumors of the small intestine. Please discuss this discrepancy.

4. Figure 2: all graphs are cut off at 5 years, although the 5-year survival rates are the main endpoints reported in the results. Please extent the KM graphs to the entire available follow up, or at least to 7 years so that the reader can assess stability of KM curves around the 5-year time point.

5. Figure 2: Please harmonize the Y-axis for clarity – the axis in the two top panels are cut off at 50%, which is confusing and exaggerates the survival differences. The same pertains to top panels in Figure 5.

6. Page 7, Line 182. The authors state that “For GIST size, the statistically optimal cut-off is 8 cm”. Please explain this statement. What was the purpose of this analysis and for what purpose is this result optimal? If this is “optimal”, why did you decide NOT to use it in your own analysis (you use the 5 cm and 10 cm
cutoffs instead). The ROC curves show the trade-off of sensitivity and specificity and the “optimal” assignment depends on many factors, e.g. number of categories desired, whether the high-risk or low-risk group is more important to identify etc.

Major Compulsory Revisions:

1. Methods (Page 5, Line 141). Please explain how the survival rate was “extrapolated” by Cox regression. Cox regression is a semi-parametric model for (log-)hazard ratio, not a survival probability at some time point, so this kind of modeling is unusual. Additionally, please clarify that you did not include your extrapolated rates in the trend analysis, because this would be fallacious.

2. Methods: (page 5, line 128): please explain for the reader how size (radiology? gross pathology) and mitotic count is recorded in the SEER data. Is mitotic count a continuous factor like size, or is it categorized? Data from Table 1 look like there were no missing values for mitotic count. Is that truly the case? Most “site-specific factors” in cancer registry data are known to have a significant proportion of missing values. Consider generating panels on Figure 3 for mitotic count like you did for tumor size so that the reader can understand the distribution of the variable and the relationship with survival.

3. Methods: The endpoints of the study need better description and definition. The title advertises analysis of OS and CSS and both are discussed in the abstract. The Methods do not define any endpoint. How was cancer-specific survival determined? The entire section “Multivariable analysis” makes no comment about CSS analysis at all, although it refers to the table that presents the results. It would be sound to have a symmetric analysis of both endpoints (OS and CSS) also with regard to the ROC tumor size/mitotic count analysis and the trend—are they prognostic for CSS too? At similar or different cutoffs? At minimum the authors should discuss their conclusions from OS and CSS analysis. The entire trend analysis

4. The section on pediatric GIST in my opinion should be removed. Discussing mortality of a single patient allows for individual patient identification which is against SEER Data Use Agreement, which states “I will not present or publish data in which an individual patient can be identified. I will not publish any information on an individual patient, including any information generated on an individual case by the case listing session of SEER*Stat. In addition, I will avoid publication of statistics for very small groups”. Describing statistics like mean age and SD in a group specifically defined by age range cutoffs is anyway of dubious value.

5. The trend analysis, the primary objective of the study, needs a more consistent presentation and discussion. The title of Figure 5 specifies CSS, while the manuscript text states that the figure represents OS (Page 7, Line 195). Furthermore, hazard ratios from Table 2 and graphs from Figure 5 indicate that any survival improvement occurred only between the 1998-2002 and 2003+ cohort. There is no evident significant improvement between the three subcohorts 2003/5, 2006/8 and 2009/11. This unfortunately is not easy to see because the authors chose to employ a very general likelihood-ratio and
Spearman’s correlation tests, which produce P-values, but no actual measures of the trend. In a very large dataset like SEER, P-values are commonly spuriously “significant”, and a more informative epidemiologic measure would be better in order (such as, for example, annual percentage change). The authors correctly discuss that the survival improvement occurred in the period of 1998-2001, which is before imatinib use, but considering no evident improvement afterwards, the entire speculation about the effect of imatinib is dubious. Although the effect of imatinib on survival in clinical trials is without doubt, concluding from these data (which have no indication of any imatinib use) that it was a “cornerstone in the history of cancer care” is arguably a gross exaggeration. The entire effect on survival could be due to reclassification of GIST cases according to new molecular diagnostic criteria developed in 2001.

6. Discussion, Page 9 Line 241: please correct the erroneous statement that regorafenib “resulted in an overall survival benefit” in “third line treatment”. The referenced GRID study showed no significant difference in overall survival (admittedly with some twists in the recent updated analysis).

7. Additional limitations of the study warrant discussion:

a. The authors exclude all cases of GIST which were not patient’s first malignancy, likely in order to use the “cause-specific survival” as defined in the SEER (although they do not motivate this exclusion at all in the paper). However, this exclusion eliminates 1,067 cases, almost 17%, and may skew the study population by preferentially excluding older patients (who have a higher cumulative chance of having an incidental other screening-detected cancer). This might result in a systematic bias in the survival estimates and trends, especially if the exclusion correlates with calendar years.

b. The authors discuss trends in survival by stage of disease (metastatic vs. non-metastatic), but provide no analysis of stage migration. However, PET scans have become a standard tool in evaluation of GIST in the early/mid-2000’s, potentially leading to stage migration which would bias trend analysis because of the Will Rogers phenomenon.

c. The authors should at least acknowledge the study by Woodall et al. (JAMA Surgery 2009), which analyzed GIST tumors using the same data source and had several similar conclusions: size cutoff for poor prognosis of 7 cm,

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:

I declare that I have no competing interests