

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Cohort profile: A nationwide population-based retrospective assessment of oesophageal cancer in the Finnish National Esophago-Gastric Cancer Cohort (FINEGO)
<b>AUTHORS</b>	Söderström, Henna; Räsänen, Jari; Saarnio, Juha; Toikkanen, Vesa; Tyrväinen, Tuula; Rantanen, Tuomo; Valtola, Antti; Ohtonen, Pasi; Pääaho, Minna; Kokkola, Arto; Kallio, Raija; Karttunen, Tuomo; Pohjanen, Vesa-Matti; Ristimäki, Ari; Laine, Simo; Sihvo, Eero; Kauppila, Joonas

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Wanqing Chen National Cancer Center, China
<b>REVIEW RETURNED</b>	19-May-2020

<b>GENERAL COMMENTS</b>	<p>This is a well-organized manuscript reporting the esophageal cancer patient characteristics and survival using FINEGO data. The results are straightforward. My comments to improve this manuscript are listed below.</p> <ol style="list-style-type: none"> <li>1. An obvious problem with this paper is the lack of method part. Although this is descriptive analysis, the method of survival analysis should be provided in detail.</li> <li>2. The results were too simple. The authors might want to add the number of incident cases and deaths of esophageal cancer by sex into figure1.</li> <li>3. Page 16, second line: "All-cause and cancer specific mortality are depicted in Figures 4. and 5". Does cancer in this sentence refer to esophageal cancer or all cancer combined? I suggest add esophageal cancer mortality here.</li> <li>4. Figures 1-3 do not conform to standards, i.e., missing label of x-axis and y-axis in the figures.</li> <li>5. There is at least one spelling error in the manuscript, such as, in page 14, line 47, "6,3%" would be "6.3%". Please check the manuscript carefully.</li> </ol>
-------------------------	---

<b>REVIEWER</b>	Ulrike Haug Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany
<b>REVIEW RETURNED</b>	04-Jun-2020

<b>GENERAL COMMENTS</b>	<p>The article „Cohort profile: A nationwide population-based retrospective assessment of oesophageal cancer in the Finnish National Esophago-Gastric Cancer Cohort (FINEGO) by Söderström et al. describes a database with information on all patients who had surgery for esophageal cancer between 1987 and 2016 in Finland (2,045 patients). The</p>
-------------------------	--

	<p>database includes data from different sources (patient registry, cancer registry) and thus includes a lot of information relevant to describe the characteristics (e.g. distribution by type of esophageal cancer) as well as the course and treatment of these patients including temporal trends.</p> <ul style="list-style-type: none"> <li>- The article provides some information on these patients and the database and highlights the potential of the database for future studies, but it is not very clear regarding its aims and research questions. So in my eyes, it is a kind of mixture between a narrative article on this database and an original article on the database. The authors call the article a “cohort profile” which “describes the 2,045 oesophageal cancer patients included in the FINEGO-cohort”. If the journal supports this kind of articles, it is fine. I find the format somewhat unusual and would suggest to include the information in the first study that will be conducted using this database (with part of the information, for example, in the Appendix).</li> <li>- The authors should be consistent regarding the limitations of the database. Although they clearly state in the discussion that the sample is restricted to surgically treated patients and does thus not represent all esophageal cancer patients in Finland, they often call it an unselected cohort or say that the database includes “all oesophageal cancer patients” (page 7, line 9). They should be clear throughout the manuscript that unselected only refers to the total group of surgically treated patients but not to the total group of patients.</li> <li>- I would also suggest to describe the database in a more balanced way. There are a couple of limitations (lack of information on lifestyle, missing information on education in the majority of patients, missing information on stage in one third of patients, limited sample size limited) and discuss in a more differentiated way for which type of research questions the database is more useful or less useful in view of these limitations.</li> <li>- Page 5, line 33: these are not prevalences, but the proportions; same on page 14: better say proportion was higher rather than more prevalent</li> </ul>
--	---

## VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 [SEP] Reviewer Name: Wanqing Chen [SEP] Institution and Country: National Cancer Center, China [SEP] Competing Interests: None declared

Q1: An obvious problem with this paper is the lack of method part. Although this is descriptive analysis, the method of survival analysis should be provided in detail.

Authors' response:

We are very thankful for this observation, this is an obvious oversight on our part. We have added statistical methods to the manuscript, describing survival analysis and clarifying how survival was calculated and reported. [SEP]-----

Q2: The results were too simple. The authors might want to add the number of incident cases and deaths of esophageal cancer by sex into figure 1.

Authors' response:

We have updated Figure 1. to show incident cases and deaths by sex. To keep the figure as simple as possible, we excluded total incident cases and deaths.

Q3: Page 16, second line: "All-cause and cancer specific mortality are depicted in Figures 4. and 5". Does cancer in this sentence refer to esophageal cancer or all cancer combined? I suggest add esophageal cancer mortality here.

Authors' response:

We thank the reviewer for this relevant question that deserves clarification. In Sweden, it has been shown that a significant misclassification between esophageal and gastric (cardia) cancers exists in the registry data, and we have no reason to believe that this would not happen in Finland as well. Therefore, we chose to define cancer-specific mortality as mortality due to esophageal or gastric cancer. This way, we are more confident that all deaths relevant to the study are included. As these cancers are relatively uncommon in Finland, there were very few patients having both gastric and surgically treated esophageal cancer in the cohort. We clarified this sentence by changing it to "esophago-gastric cancer specific mortality". We also added this fact to the "Methods" chapter.

Q4: Figures 1-3 do not conform to standards, i.e., missing label of x-axis and y-axis in the figures.

Authors' response:

Axis labels have been added to Figures 1-3.

Q5: There is at least one spelling error in the manuscript, such as, in page 14, line 47, "6,3%" would be "6.3%". Please check the manuscript carefully.

Authors' response:

We thank the reviewer for this observation, and we have corrected the abovementioned spelling error and three others.

Reviewer: 2 Reviewer Name: Ulrike Haug Institution and Country: Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany Competing interests: None declared

Q1: The article provides some information on these patients and the database and highlights the potential of the database for future studies, but it is not very clear regarding its aims and research questions. So in my eyes, it is a kind of mixture between a narrative article on this database and an original article on the database. The authors call the article a "cohort profile" which "describes the 2,045 oesophageal cancer patients included in the FINEGO- cohort". If the journal supports this kind of articles, it is fine. I find the format somewhat unusual and would suggest to include the information in the first study that will be conducted using this database (with part of the information, for example, in the Appendix).

Authors' response:

We thank the reviewer for this interesting observation. BMJ Open has published cohort profiles since 2014. According to BMJ, the indication for doing so is to provide information on a cohort's establishment and describe the methodology more thoroughly than can be expected in the research papers. This serves to clarify any questions readers may have about the methodology, and also advises other researchers of existing datasets and opportunities for collaboration. BMJ states that they hope to generate an ongoing database for answering many different research questions.

-----

Q2: The authors should be consistent regarding the limitations of the database. Although they clearly state in the discussion that the sample is restricted to surgically treated patients and does thus not represent all esophageal cancer patients in Finland, they often call it an unselected cohort or say that the database includes “all oesophageal cancer patients” (page 7, line 9). They should be clear throughout the manuscript that unselected only refers to the total group of surgically treated patients but not to the total group of patients.

Authors' response:

We are thankful to the reviewer for pointing out this potentially misleading factor. We have reviewed the manuscript thoroughly and changed the wording to “surgically treated” in all sentences where it applies.

-----

Q3: I would also suggest to describe the database in a more balanced way. There are a couple of limitations (lack of information on lifestyle, missing information on education in the majority of patients, missing information on stage in one third of patients, limited sample size limited) and discuss in a more differentiated way for which type of research questions the database is more useful or less useful in view of these limitations.

Authors' response:

We thank the reviewer for this very important observation and comment. The dataset in the current cohort has many limitations, mainly due to the method of data gathering from databases. For the upcoming sub-studies, some of the data needs to be manually checked from the patient records for verification and to add missing data. For example, we believe the stages are erroneously reported, as we state in our manuscript, mainly due to the high proportion of advanced disease. Verification for this, and clarification of the large number of missing TNM stages, can and will be found in patient records after manual review. This is an arduous process, which is why it has not been done in the initial data gathering and creation of the FINEGO cohort. However, there are some data that can not be verified or added to the cohort even with manual review as they are not reliably recorded. These include, but are not limited to, smoking and alcohol consumption, dietary and exercise habits, lifestyle and socio-economic status. The limited amount on information on years of education limits its usability, as well as the fact that we can not draw conclusions on lifestyle based on this data. Considering these obvious limitations, the cohort is most suitable for studies on cancer treatment methods, effect of national centralization and tumour characteristics, and these factors' effects on patients' survival. We have added these observations to the “Strengths and Limitations” chapter.

-----

Q4: Page 5, line 33: these are not prevalences, but the proportions; same on page 14: better say proportion was higher rather than more prevalent

Authors' response:

We thank the reviewer for pointing out this flaw in our manuscript. We have corrected the suggested sentences as well as the abstract in this regard.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Wanqing Chen National Cancer Center/ Cancer Hospital, Chinese Academy of Medical Sciences, China
<b>REVIEW RETURNED</b>	02-Jul-2020

<b>GENERAL COMMENTS</b>	<p>The point-by-point resubmission letter shows a general agreement with the suggestions in the previous review; however, the actual implementation of some of those suggestions appears not to be made. Listed below are my two suggestions meant to highlight those modifications which are still necessary.</p> <p>(1) In this revised manuscript, some critical information still did not present clearly. For example, aims and research questions of the FINEGO, which mentioned in the Q1 of reviewer 2 in the previous comment; the consideration and calculation of samplings; the specific variables, et.al. The authors should provide a clear and complete description of the study design of the FINEGO, which is essential as a study protocol.</p> <p>(2) Regarding the statistical analysis, it should be provided more detail descriptions, served for the FINEGO, not merely for this paper.</p>
-------------------------	--

<b>REVIEWER</b>	Ulrike Haug Leibniz Institute for Prevention Research and Epidemiology
<b>REVIEW RETURNED</b>	21-Jul-2020

<b>GENERAL COMMENTS</b>	<p>Overall, the comments have been adequately considered, except for the following:</p> <p>Q2: The point has not been considered consistently across the manuscript.</p> <p>See for example</p> <ul style="list-style-type: none"> <li>- "Because of FINEGO, we now have a population based, unselected cohort of patients" or</li> <li>- "The main strengths include the population-based design with complete and accurate ascertainment of all oesophageal cancer patients"</li> </ul> <p>Please clarify throughout the manuscript that the database is limited to surgically treated patients</p> <p>There are some mistakes regarding spelling / grammer , e.g. "In this cohort profile, we describe the oesophageal cancer patients include in the FINEGO-cohort."</p>
-------------------------	--

## VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Wanqing Chen

Institution and Country: National Cancer Center, China

Competing Interests: None declared

General comment: The point-by-point resubmission letter shows a general agreement with the

suggestions in the previous review; however, the actual implementation of some of those suggestions appears not to be made. Listed below are my two suggestions meant to highlight those modifications which are still necessary.

Authors' response: Thank you for your insightful review. On the first round, we tried to address the comments as perfectly as we could. We might have misunderstood some of your comments, and now have strived to polish these points more thoroughly.

-----

Q1: In this revised manuscript, some critical information still did not present clearly. For example, aims and research questions of the FINEGO, which mentioned in the Q1 of reviewer 2 in the previous comment; the consideration and calculation of samplings; the specific variables, et.al. The authors should provide a clear and complete description of the study design of the FINEGO, which is essential as a study protocol.

Authors' response: We thank you for your comments. The study protocol of FINEGO was published previously in BMJ open (<https://bmjopen.bmj.com/content/9/1/e024094>) where we provided sample size calculations and list of variables.

The author guidelines for cohort profiles define the cohort profile as follows: "The cohort profile is an article type set up in BMJ Open to fill the space between a study protocol and a results paper. Cohort profiles should describe the rationale for a cohort's creation, its methods, baseline data and its future plans. Cohorts described should be long-term, prospective projects and not time-limited cohorts established to answer a small number of specific research questions."

The cohort was created to allow a wide variety of research questions which may be formed in detail only after the study database has been collected. We thought that providing an exhaustive list of study hypotheses with description of specific background, rationale, sample size calculations, and analysis plans might obscure the original purpose of this manuscript of describing the baseline data of these patients. We have, however, described the rationale of the cohort creation in the introduction, methods for obtaining the baseline data in the methods section, and our future plans in that section. Furthermore, the general study design, data collection details and sample size calculation have been described in the abovementioned study protocol paper in more detail, and in this manuscript we tried to avoid repeating the information described in the protocol.

For describing the clinical data collection from the patient records, which is still on-going, we feel that it is difficult to include in the methods of this manuscript, as the patient records data is not presented in the manuscript.

If still deemed necessary, we would be interested in learning which specific points of the study design would require more attention and where to describe this information in order to be able to define each issue in enough detail.

Revisions: We have now revised the methods section as related to data presented in this manuscript and added a reference to the study protocol publication so that the readers can find this information if they want.

-----

Q2: Regarding the statistical analysis, it should be provided more detail descriptions, served for the FINEGO, not merely for this paper.

Authors' response: We thank you for this idea.

For describing the statistical analysis outside the analyses for this cohort profile manuscript i.e. for research studies in FINEGO, we could either present a very general analysis plan (for example: Mortality outcomes will be assessed using cox regression, providing HRs and 95% CIs for the

exposure variables, adjusted for the relevant confounders, et cetera), or a detailed analysis plan of a single outcome. The former would be applicable in many studies, but we feel it would not provide any relevant information to the reader. On the other hand, the latter would be only applicable for one single study, and would be more suitable to a study protocol of a prospective or a randomized study with a single research question. We feel that presenting many detailed analysis plans for multiple studies would obscure the original purpose of this cohort profile manuscript.

Therefore, we feel that the current statistical analysis section is adequate.

Revisions: We have added a reference to the original study protocol stating that the protocol includes more complete information.

-----

Reviewer: 2

Reviewer Name: Ulrike Haug

Institution and Country: Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany

Competing interests: None declared

Q1: Overall, the comments have been adequately considered, except for the following: Q2: The point has not been considered consistently across the manuscript. See for example

- "Because of FINEGO, we now have a population based, unselected cohort of patients" or
- "The main strengths include the population-based design with complete and accurate ascertainment of all oesophageal cancer patients"

Please clarify throughout the manuscript that the database is limited to surgically treated patients

Authors' response: Thank you for pointing out this inconsistency. We have now specified in multiple instances, including the abstract-, strengths and weaknesses-, and methods sections that this cohort is solely a surgical cohort.

-----

Q2: There are some mistakes regarding spelling / grammar, e.g. "In this cohort profile, we describe the oesophageal cancer patients include in the FINEGO-cohort."

Authors' response: We have fixed spelling errors and one error related to text being out of place.

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Wanqing Chen National Cancer Center / National Clinical Research Center for Cancer / Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
<b>REVIEW RETURNED</b>	19-Sep-2020
<b>GENERAL COMMENTS</b>	All comments have been addressed adequately.