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

TITLE (PROVISIONAL)	Evaluating the endometabolic and bone health effects of Tyrosine Kinase Inhibitors in Chronic Myeloid Leukemia: A systematic review protocol
AUTHORS	Balakumaran, Janatani; Birk, Tanisha; Golemiec, Breanne; Helmeczi, Wryan; Inkaran, Jeyanth; Kao, Yun-ya; Leigh, Jennifer; Saliba, Sarah; Sharma, Rishi; Spatafora, Laura; Wright, Kristin; Yao, William; Hillis, Christopher; Banfield, Laura; Thabane, Lehana; Athale, Uma; Samaan, M. Constantine

VERSION 1 – REVIEW

REVIEWER	Jason C. Hsu School of Pharmacy and Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Taiwan
REVIEW RETURNED	31-Mar-2019

GENERAL COMMENTS	1. This study aimed to conduct a systematic review to investigate the endometabolic and bone health effects of TKI therapy in Chronic Myeloid Leukemia. 2. Data analysis: “A meta-analysis will be performed if there are at least two studies reporting similar populations, study design, methods and outcomes.” How can you define the “similar” populations, study design, methods and outcomes? Especially, study design. 3. What are possible causes of heterogeneity between articles in this issue? (For example, different inclusion and exclusion criteria…) 4. Even though there might be heterogeneity between articles, the authors should come up with some solutions for dealing with heterogeneity, instead of just “give up” the meta-analysis. For example, using random effect models, sub-group analysis, meta regression, transforming outcome measurements… 5. Have you considered using GRADE to evaluate the quality of meta-analysis? If so, how?

REVIEWER	Alexandra Smith University of York UK
REVIEW RETURNED	02-Apr-2019

GENERAL COMMENTS	The authors describe a strategy to undertake a systematic review of endometabolic and bone health effects of tyrosine kinase inhibitors in chronic myeloid leukaemia (CML). It is well described, my only concern is have there been sufficient studies examining the long-term effects, as defined by the endpoints in the protocol,

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

	especially for paediatric cases? It may be helpful to perform an initial search and include the number of studies identified to reassure the readers that it is possible. I have some minor comments around the references quoted in the abstract and introduction: The reference quoted for what percentage of CML diagnoses comprise adult and paediatric leukaemia requires updating to an original source, i.e. estimates from national cancer registrations systems or those reported by SEER, HAEMACARE. Especially, as the figure of 5% of paediatric cases seems high. The second paragraph of the introduction where it states “Prior to their widespread use, the median survival of CML patients was 5 - 7 years [3].” This figure is derived from a clinical trial, but may not be applicable to the general patient population at this time. This should be made more explicit or a figure should be quoted from a population-based series.

REVIEWER	Derek Smith Vanderbilt University Medical Center, USA
REVIEW RETURNED	16-May-2019

GENERAL COMMENTS	This manuscript is a well-written protocol for a proposed systematic review and meta-analysis of glucose tolerance among CML patients treated with tyrosine kinase inhibitors including several secondary outcomes. It is clear the authors have a knowledge of the appropriate reporting guidelines and facility with guidelines for conducting a systematic review put forth in the Cochrane Handbook. As such, my comments are largely a request for clarity in some parts. In the data analysis section, the authors state “We aim to define the point estimate of the adverse outcomes…”. There is a fair amount of ambiguity in this statement. I assume the authors are intending to arrive at a point estimate of the incidence of this complication. If this is the case, then I believe it would be advisable to add more clarity to the study eligibility criteria. As it stands, the criteria list case-control studies as eligible for inclusion. Any study that oversamples specific groups will not necessarily have the same incidence of glucose intolerance (or the respective secondary outcomes) as the general population of CML patients on TKIs. The authors should consider 1) making it more clear what is being estimated and 2) either being very clear about what kind of case-control studies will be allowed (what the cases are) or consider excluding them altogether (under the presumption that incidence is the goal). The other item that I thought would merit more explanation was the discussion of heterogeneity. The authors stated that they are “Expecting a high level of heterogeneity across studies…,” but there is no discussion of how heterogeneity will be addressed beyond the employ of the random-effects model. It is often possible to explore sources of heterogeneity (i.e. is heterogeneity induced by how the outcome is assessed, by the specific drug, or some other tangible differences between the study). Understanding the source of heterogeneity can often lend clarity to the eventual analysis beyond what pooling over it can achieve. In addition, it is mentioned that “An I2 >75% … will represent considerable heterogeneity,” which is consistent with the Cochrane Handbook’s definition, but no comment is made on what

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

	consequence this would have for your review. Is there a level of heterogeneity beyond which the meta-analysis proposed would lose meaning? Is there a chance that you might decide to do the narrative review instead of the meta-analysis in the face of high heterogeneity? If there is, it might be worth specifying this here. In the section on sensitivity analyses, I recommend caution when excluding small studies because they are small. Especially in the case of high heterogeneity the number of well-conducted studies included becomes much more important relative to the size of the studies included. While smaller studies are more susceptible to publication bias, you may be able to assess this in cases where you have a sufficient number of studies, which would allow you to conduct (or not conduct) a more informed sensitivity analysis. To be clear, this is advice and not a recommendation for a change to the protocol as written.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Jason C. Hsu

- Data analysis: "A meta-analysis will be performed if there are at least two studies reporting similar populations, study design, methods and outcomes." How can you define the "similar" populations, study design, methods and outcomes? Especially, study design.

Response: We thank the reviewer for their comment. Our goal is to combine similar studies to provide the direction of effect, effect size, effect consistency, and this will be combined with an assessment of the strength of the evidence for the effects of TKI on endometabolic and bone health in CML patients. We have updated the paper to clarify our plans for the meta-analysis in the abstract (Page3, Paragraph 3, Lines 62-64) and the methods section (Page 17, Paragraph 4, Lines 252-256; Page 18, Paragraph 1-2, Lines 257-270):

Abstract: "We will perform a meta-analysis if there are at least two studies reporting similar populations and interventions, implementing similar methods, and tracking the same outcome measures. The studies should also have similar age and sex distributions".

Main paper:

"A meta-analysis will be performed for each of the primary and secondary outcomes if two or more studies have reported on populations with a similar age and sex distribution, have used the same TKI with similar dosing and duration, have comparable follow-up periods, and have been studied using a similar study design. For example, we will combine randomized controlled trial data that fulfill the above criteria together to assess the direction and magnitude of the effect as well as the effect consistency. Even if a minimum of two studies per outcome were available, if there are different treatment and comparison groups reported, diversity of treatment outcomes, or a high risk of bias then a meta-analysis may not be feasible. In this case, a narrative summary of results with tabulated data will be presented.

For studies reporting outcomes as dichotomous outcome, we will use risk ratio, odds ratio, and risk difference to assess the event rate differences between groups. For continuous outcomes, we will use mean difference to report the differences between groups. For studies with repeated measures of continuous outcomes over time, we will use mean difference as a summary measure. In the presence of a control group, the mean difference is the comparison of change from baseline to follow-up between the two groups. The cases are patients with CML treated with TKIs. The controls are either non-cancer healthy individuals or patients with CML without receiving TKIs. These two control groups

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

will be analyzed separately. In the absence of a control group, the mean difference will represent the difference between baseline and follow-up measures."

- What are possible causes of heterogeneity between articles in this issue? (For example, different inclusion and exclusion criteria…)

Response: We thank the reviewer for this important comment. We have added the following paragraph to list possible causes of heterogeneity between studies (Page 18, Paragraph 3, Lines 271-278; Page 19, Paragraph 1, Lines 279-286):

"We predict that a high level of heterogeneity will likely to be due to several sources, including age and sex differences of participants between different studies. Age may be a factor in the development of adverse outcomes and children may be more susceptible than adults to certain side effects when receiving TKIs such as delayed growth.[1] It has also been suggested that there are sex differences in response to TKI in patients with CML, with females being more susceptible to some side effects including anemia and giving birth to infants with congenital abnormalities.[2-3] Therefore, we will analyze our data based on the male and female distribution in the studies."

"There are other potential sources of clinical heterogeneity including the different medications used, dosing, and duration of treatment when the outcomes measures were tested. Furthermore, there are three phases of CML – chronic, accelerated, and blast phase. Patients who are on treatment at different phases may experience variable adverse endometabolic and bone health outcomes.[4] Differences in study design, inclusion and exclusion criteria, variation in outcome measurement, and differences in the risk of bias assessment among studies may be other sources of heterogeneity. Statistical heterogeneity may result from heterogeneity in clinical and methodological variation.[5]"

The following references were added for this paragraph:

1. Samis J, Lee P, Zimmerman D, et al. Recognizing Endocrinopathies Associated With Tyrosine Kinase Inhibitor Therapy in Children With Chronic Myelogenous Leukemia. Pediatr Blood Cancer 2016;63(8):1332-8. doi: 10.1002/pbc.26028 [published Online First: 2016/04/22]
2. Abruzzese E, Trawinska MM, Perrotti AP, et al. Tyrosine kinase inhibitors and pregnancy. Mediterr J Hematol Infect Dis 2014;6(1):e2014028. doi: 10.4084/mjhid.2014.028 [published Online First: 2014/05/08]
3. Moura MS, Benevides TCL, Delamain MT, et al. Evaluation of anemia after long-term treatment with imatinib in chronic myeloid leukemia patients in chronic phase. Hematol Transfus Cell Ther 2019 doi: https://doi.org/10.1016/j.htct.2019.03.006
4. Chereda B, Melo JV. Natural course and biology of CML. Annals of hematology 2015;94 Suppl 2:S107-21. doi: 10.1007/s00277-015-2325-z [published Online First: 2015/03/31]
5. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of intervneions version 510 (updated March 2011): The Cochrane Collaboration 2011. Available from www.handbook.cochrane.org.

- Even though there might be heterogeneity between articles, the authors should come up with some solutions for dealing with heterogeneity, instead of just "give up" the meta-analysis. For example, using random effect models, sub-group analysis, meta regression, transforming outcome measurements…

Response: We thank the reviewer for this comment.
In addition to the above expansion of the assessment of heterogeneity, we have added the following paragraph to describe the methods to handle heterogeneity (Page 19, Paragraph 2, Lines 287-292):

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

"Heterogeneity will be evaluated with the chi-square test for homogeneity and inconsistency index (I2). A p-value of <0.10 and an I2 □75% will represent the thresholds for detecting heterogeneity. To deal with heterogeneity, a random-effects model will be used for the meta-analysis. If the above thresholds for heterogeneity are met, we will attempt to perform sub-group analyses to identify sources of heterogeneity including age, sex, medication used, and CML phase."

- Have you considered using GRADE to evaluate the quality of meta-analysis? If so, how?

Response: We thank the reviewer for raising this question. We will use GRADE guidelines to evaluate the confidence in the recommendations drawn from the results. The following statement has been added (Page 17, Paragraph 2, Lines 242-245):
"The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines will be used to evaluate the overall confidence in the evidence for each outcome analyzed based on eligible studies. The GRADE guidelines evaluate risk of bias, inconsistency, indirectness, imprecision, and publication bias.[6]"
6. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328(7454):1490. doi: 10.1136/bmj.328.7454.1490
Reviewer 2: Alexandra Smith

- The authors describe a strategy to undertake a systematic review of endometabolic and bone health effects of tyrosine kinase inhibitors in chronic myeloid leukaemia (CML). It is well described, my only concern is have there been sufficient studies examining the long-term effects, as defined by the endpoints in the protocol, especially for paediatric cases? It may be helpful to perform an initial search and include the number of studies identified to reassure the readers that it is possible.

Response: We thank the reviewer for their comment. The main reason for performing this systematic review is that we are trying to summarize the current evidence base available for our topic. If there is insufficient evidence base for the endometabolic and bone health effect of TKI use in CML, this is a gap in knowledge that will be specifically highlighted.
Importantly, it is not unusual for systematic reviews not to find sufficient evidence for their question, and this routinely includes Cochrane reviews. We do not want to bias our conclusions by having a superficial search of the literature and would like to perform thorough and systematic searches for the systematic reviews.

- The reference quoted for what percentage of CML diagnoses comprise adult and paediatric leukaemia requires updating to an original source, i.e. estimates from national cancer registrations systems or those reported by SEER, HAEMACARE. Especially, as the figure of 5% of paediatric cases seems high.

Response: We thank the reviewer for their feedback. We have replaced these statistics with ones from the national statistics registration system.

The following statement was updated in the abstract (Page 3, Paragraph 1 Lines 48-50) and the introduction (Page 6, Paragraph 1, Lines 117-119):

Abstract:
"Chronic Myeloid Leukemia (CML) constitutes 15% of new adult leukemia cases as well as 2-3% of leukemias in children under 15 and 9% of leukemias in adolescents 15-19 years of age annually."

Introduction:

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

"Chronic Myeloid Leukemia (CML) accounts for approximately 15% of all newly diagnosed adult leukemias as well as 2-3% of leukemias in children under 15 and 9% of leukemias in those15-19 years of age.[7, 8]"

The following citations were updated/added to the manuscript:

7. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2019. CA: a cancer journal for clinicians 2019, 69(1):7-34.
8. Ries LAG, Smith MA, Gurney JG, Linet M, Tamra T, Young JL, Bunin GR (eds). Cancer incidence and survival among children and adolescents: United States SEER Program 1975-1995, National Cancer Institute, SEER Program. NIH Pub. No. 99-4649. Bethesda, MD, 1999.

- The second paragraph of the introduction where it states "Prior to their widespread use, the median survival of CML patients was 5 - 7 years [3]." This figure is derived from a clinical trial, but may not be applicable to the general patient population at this time. This should be made more explicit or a figure should be quoted from a population-based series.

Response: We thank the reviewer for pointing out this issue. We have replaced this statement with one informed by the SEER cancer registry data, with the associated supporting citation below (Page 6, Paragraph 2, Lines 123-125):

"As TKI use became more widespread, the estimated five-year survival rate has more than doubled from 31% in the early 1990s to around 70% in 2015.[9]"

The following citations was used:

9. Howlader N, Noone A, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2016 Bethesda, MD: National Cancer Institute; 2019 [Available from: https://seer.cancer.gov/csr/1975_2016/] accessed July 2019.

Reviewer 3: Derek Smith

- In the data analysis section, the authors state "We aim to define the point estimate of the adverse outcomes…". There is a fair amount of ambiguity in this statement. I assume the authors are intending to arrive at a point estimate of the incidence of this complication. If this is the case, then I believe it would be advisable to add more clarity to the study eligibility criteria. As it stands, the criteria list case-control studies as eligible for inclusion. Any study that oversamples specific groups will not necessarily have the same incidence of glucose intolerance (or the respective secondary outcomes) as the general population of CML patients on TKIs. The authors should consider 1) making it more clear what is being estimated and 2) either being very clear about what kind of case-control studies will be allowed (what the cases are) or consider excluding them altogether (under the presumption that incidence is the goal).

Response: We thank the reviewer for their feedback. We will be reporting the pooled prevalence of adverse outcomes secondary to the use of TKIs in patients with CML. We added the following statements in response to the comment to clarify what is being done and to define the cases and controls more clearly (Page 17, Paragraph 3, Lines 247-249; Page 18, Paragraph 2, Lines 261-270):

Page 17, Paragraph 3, Lines 247-249:

"The primary goal of this systematic review is to determine the pooled prevalence of the primary and secondary endometabolic and bone health outcomes in TKI-treated patients with CML."

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

Page 18, Paragraph 2, Lines 261-270
"For studies reporting outcomes as dichotomous outcome, we will use risk ratio, odds ratio, and risk difference to assess the event rate differences between groups. For continuous outcomes, we will use mean difference to report the differences between groups. For studies with repeated measures of continuous outcomes over time, we will use mean difference as a summary measure. In the presence of a control group, the mean difference is the comparison of change from baseline to follow-up between the two groups. The cases are patients with CML treated with TKIs. The controls are either non-cancer healthy individuals or patients with CML without receiving TKIs. These two control groups will be analyzed separately. In the absence of a control group, the mean difference will represent the difference between baseline and follow-up measures."

- The other item that I thought would merit more explanation was the discussion of heterogeneity. The authors stated that they are "Expecting a high level of heterogeneity across studies…," but there is no discussion of how heterogeneity will be addressed beyond the employ of the random-effects model. It is often possible to explore sources of heterogeneity (i.e. is heterogeneity induced by how the outcome is assessed, by the specific drug, or some other tangible differences between the study). Understanding the source of heterogeneity can often lend clarity to the eventual analysis beyond what pooling over it can achieve. In addition, it is mentioned that "An I2 >75% … will represent considerable heterogeneity," which is consistent with the Cochrane Handbook's definition, but no comment is made on what consequence this would have for your review.

Is there a level of heterogeneity beyond which the meta-analysis proposed would lose meaning? Is there a chance that you might decide to do the narrative review instead of the meta-analysis in the face of high heterogeneity? If there is, it might be worth specifying this here.

Response: We thank the reviewer for their comment.

If the heterogeneity is above 75%, we will still perform subgroup analyses. We are not sure if there will be no opportunity to perform the meta-analysis at this stage, but we believe that a subgroup analysis will be a viable option, so we do not think that we will end up with a narrative review at this point. In addition, we will be performing GRADE to assess our overall confidence with the recommendations and conclusions from our analyses.

However, if we end up with significant heterogeneity that preclude a meta-analysis, then we will proceed with a narrative summary of the data in addition to data tabulations to describe the interventions.

We have added the following paragraph to list possible causes of heterogeneity between studies (Page 18, Paragraph 3, Lines 271-278; Page 19, Paragraph 1, Lines 279-286):

"We predict that a high level of heterogeneity will likely to be due to several sources, including age and sex differences of participants between different studies. Age may be a factor in the development of adverse outcomes and children may be more susceptible than adults to certain side effects when receiving TKIs such as delayed growth.[1] It has also been suggested that there are sex differences in response to TKI in patients with CML, with females being more susceptible to some side effects including anemia and giving birth to infants with congenital abnormalities.[2-3] Therefore, we will analyze our data based on the male and female distribution in the studies."

"There are other potential sources of clinical heterogeneity including the different medications used, dosing, and duration of treatment when the outcomes measures were tested. Furthermore, there are three phases of CML – chronic, accelerated, and blast phase. Patients who are on treatment at

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

different phases may experience variable adverse endometabolic and bone health outcomes.[4] Differences in study design, inclusion and exclusion criteria, variation in outcome measurement, and differences in the risk of bias assessment among studies may be other sources of heterogeneity. Statistical heterogeneity may result from heterogeneity in clinical and methodological variation.[5]"

The following references were added for this paragraph:

1. Samis J, Lee P, Zimmerman D, et al. Recognizing Endocrinopathies Associated With Tyrosine Kinase Inhibitor Therapy in Children With Chronic Myelogenous Leukemia. Pediatr Blood Cancer 2016;63(8):1332-8. doi: 10.1002/pbc.26028 [published Online First: 2016/04/22]
2. Abruzzese E, Trawinska MM, Perrotti AP, et al. Tyrosine kinase inhibitors and pregnancy. Mediterr J Hematol Infect Dis 2014;6(1):e2014028. doi: 10.4084/mjhid.2014.028 [published Online First: 2014/05/08]
3. Moura MS, Benevides TCL, Delamain MT, et al. Evaluation of anemia after long-term treatment with imatinib in chronic myeloid leukemia patients in chronic phase. Hematol Transfus Cell Ther 2019 doi: https://doi.org/10.1016/j.htct.2019.03.006
4. Chereda B, Melo JV. Natural course and biology of CML. Annals of hematology 2015;94 Suppl 2:S107-21. doi: 10.1007/s00277-015-2325-z [published Online First: 2015/03/31]
5. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of intervneions version 510 (updated March 2011): The Cochrane Collaboration 2011. Available from www.handbook.cochrane.org.
In addition to the above expansion of the assessment of heterogeneity, we have added the following paragraph to describe the methods to handle heterogeneity (Page 19, Paragraph 2, Lines 287-292):

"Heterogeneity will be evaluated with the chi-square test for homogeneity and inconsistency index (I2). A p-value of <0.10 and an I2 □75% will represent the thresholds for detecting heterogeneity. To deal with heterogeneity, a random-effects model will be used for the meta-analysis. If the above thresholds for heterogeneity are met, we will attempt to perform sub-group analyses to identify sources of heterogeneity including age, sex, medication used, and CML phase."

We have also added the use of GRADE guidelines to evaluate the confidence in the recommendations drawn from the results. The following statement has been added (Page 17, Paragraph 2, Lines 242-245):
"The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines will be used to evaluate the overall confidence in the evidence for each outcome analyzed based on eligible studies. The GRADE guidelines evaluate risk of bias, inconsistency, indirectness, imprecision, and publication bias.[6]"
6. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328(7454):1490. doi: 10.1136/bmj.328.7454.1490

- In the section on sensitivity analyses, I recommend caution when excluding small studies because they are small. Especially in the case of high heterogeneity the number of well-conducted studies included becomes much more important relative to the size of the studies included. While smaller studies are more susceptible to publication bias, you may be able to assess this in cases where you have a sufficient number of studies, which would allow you to conduct (or not conduct) a more informed sensitivity analysis. To be clear, this is advice and not a recommendation for a change to the protocol as written.

Response: We thank the reviewer for this suggestion. The primary goal is to determine the pooled prevalence of each adverse outcome in CML patients treated with TKIs. If one patient in a small study happened to develop an adverse outcome, the prevalence of this outcome will be large due to the small denominator. This will highly skew the pooled prevalence when compared to other studies with

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.

large sample sizes, especially in rare outcomes. We hope to exclude small studies as part of the sensitivity analysis to examine their influences on the results. However, we agree that some small studies may hold importance especially when our population of interest is unique.

BMJ Open: first published as 10.1136/bmjopen-2019-030092 on 11 September 2019. Downloaded from http://bmjopen.bmj.com/ on April 15, 2021 by guest. Protected by copyright.