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)	Do social protection programs have a causal effect on suicide mortality? A protocol for a systematic review and meta-analysis
AUTHORS	Chum, Antony; Kim, Chungah; Azra, Karanpreet; Teo, Celine; Nielsen, Andrew; Bellows, Zachary; Young, Thomas

VERSION 1 – REVIEW

REVIEWER	Gertner, Alex University of North Carolina
REVIEW RETURNED	09-Aug-2021

GENERAL COMMENTS	The manuscript under review is a protocol for a systematic review and meta-analysis of the effect of “income security programs” on suicide mortality. This is an important topic. A systematic review would make a valuable contribution to the literature. The paper states that the review began in June 2021, so I will presume that the methods are largely not modifiable. I will therefore focus comments on the presentation of methods. The protocol is remarkably broad in its inclusive definition of “income security programs,” even including housing support and food aid. The protocol cites the “ILO guidelines” for its interpretation of income security programs, but I couldn’t find a clear definition in the cited ILO report. The closest to a definition I found was: “income security. Protection of income through minimum wage machinery, wage indexation, comprehensive social security, progressive taxation to reduce inequality and to supplement those with low incomes, etc.” I find this definition difficult to interpret since it includes several specific measures as well as “comprehensive social security.” The protocol also cites a paper by Sun and colleagues that defines cash transfers as “transfers of cash or cash equivalents (e.g., tax benefits) and vouchers for cash or cash-like in-kind assistance vouchers (e.g., Supplemental Nutrition Assistance Program benefits, known as food stamps).” The protocol seems to use a combination of these two definitions. A clearer justification of the types of programs included in the review based in a theoretical or practical reasoning would strengthen the review. A clearer justification for theorizing all these forms of social assistance as “income security” would also be helpful. I wonder if this protocol is not better defined as focusing broadly on social security programs rather than merely income security. I worry that reducing these varied forms of social assistance to “income security” will elide much of heterogeneity in program characteristics and effects. A description of a plan for presenting results from such varied programs would also be helpful. Though the Sun and colleagues’ papers conflates cash transfer and in-kind transfer programs under the umbrella of cash

	transfer, I would oppose such a conflation in the current project, since these two types of programs have markedly different theoretical implications. On a related note, the protocol will include studies of individual-level interventions, such as cash transfer RCTs, as well as policy-level effects, such as adoption of minimum wage. This seems to me a possibly perilous venture. It would be like doing a systematic review of the effect of diet on diabetes outcomes and reviewing RCTs of diet programs alongside public policies to reduce calory intake like limiting soda beverage size. The targeted individual level intervention is obviously different than a population-level policy, though the essential hypothesized mechanism may be the same. The studies evaluating these two categories of programs may also produce different types of treatment effect. I would find it helpful if the protocol discussed the specific hypothesized mechanism(s) through which income security affects suicide to clarify the inclusion of different program and study types. I would also find it helpful if the protocol discussed how it will synthesize or interpret studies of individual level and population level effects, with some consideration of whether studies produce average treatment effects or treatment effects on the treated or so on. I worry that it may be difficult to meaningfully summarize the variety of study designs and effects in the S1 table that the protocol presents. It's not clear to me why the study is restricted to government programs or how government programs are defined. I imagine that there are a non-trivial number of cash transfer experiments that were conducted by and through non-governmental organizations and academic institutions. The exclusion of such program would be a major limitation to the study. More explanation of this would be beneficial. One reason for exclusion that appears in the protocol documents but not in the text of the protocol is "Specific intervention in specific setting." It's not clear to me what this exclusion reason represents since it doesn't seem to be discussed in the protocol. Excluding studies of "specific interventions in specific setting" would seem to represent a broad exclusion that would weaken the study. Explanation and discussion of this exclusion would be helpful. The inclusion of studies in English, French, Spanish, German, Chinese, Korean, and Japanese is a significant potential strength of the study. The protocol mentions these languages are "familiar" to the review team, but it is not clear what this means specifically. Do all members of the review team possess advanced knowledge of all these languages that will permit nuanced analyses of academic research papers? If so, this should be clearly stated. If not, then the protocols should state how studies will be reviewed in all languages by all review members without compromising the quality of the reviews and without biasing the reviews. For instance, it's difficult to imagine how one author might translate a paper to other authors without introducing bias regarding their assessment. I would encourage the authors to err on the side of including fewer languages without compromising the quality of reviews. The language used to describe quasi-experimental studies is imprecise. For instance, the protocol states that quasi experimental studies "allow for treatment to be randomly assigned," which is not

	true. Quasi-experimental studies make use of exogeneous variation, typically in the absence of random assignment and typically with reliance of many assumptions. While the language can be easily corrected, I'm concerned that it may belie a lack of expertise on quasi-experimental methods that could undermine the quality of study reviews. The protocol should describe the expertise of reviewers in quasi-experimental methods, including their past use of quasi-experimental techniques in published research. PRISMA-P states that the search strategy should be presented in a way that it could be repeated. I don't think the current protocol reaches that level of detail and clarity. I suggest providing in the paper or in a supplement the specific text that will be used in searches, including AND/OR statements and the like, as well as the parameters of the search, including where in records terms will be searched for. The rationale for inclusion of both "suici*" and "suicide complete" in the search is unclear since the first would presumably include the second. It's unclear to me what the protocol offers a definition of suicide mortality. Is this to mean that the protocol will only include studies that use this definition of suicides? That would represent a weakness of the study in my view. Rather, I expect the protocol will identify multiple approaches to defining suicide in literature. Indeed, presenting and comparing different approaches to defining suicide could be a valuable contribution of the review. The term "medicare-eligibility" may not be legible to non-Canadian readers and should be replaced with a more generic terms such as "public insurance eligibility" or some such. How are relevant studies being defined in Stage 1? What are the criteria? How will disagreements be handled? What is the point of stage 2, especially since all authors are involved in stage 1? Page 9 says all team members will assess study quality at stage 3 and stage 10 says only two team members will. Clarification would be helpful. The protocol states that it will use the ROBINS-E tool, but the University of Bristol website cited states that the ROBIN-E tool does not yet exist: http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-e/. Clarification would be helpful. The paper states that the review will provide evidence to support implementation of income security programs. I'm not sure what's meant by this or how the review will provide evidence to support implementation. More explanation would be helpful. The abstract states, "Our review aims to evaluate the causal relationship between income security programs and suicide mortality by examining all relevant experimental and quasi-experimental studies between January 1980 and May 2021." Perhaps it would be more accurate to say that this review aims to evaluate the evidence of a causal relationship, since the review does not directly evaluate the causal relationship.

REVIEWER	Pak, Tae-Young Sungkyunkwan University, Consumer Science
REVIEW RETURNED	13-Aug-2021

GENERAL COMMENTS	This paper addresses an important and timely issue that could have implications beyond social epidemiology. I believe this review will amount to a well-respected summary of what we know about the health effects of government income support and the role the government needs to play in suicide prevention. Below, I suggest a few points the authors may need to consider along with the project. * The scope of income security programs may need to be restricted to cash transfer programs. I believe the mechanism through which cash transfer program affects suicidal behaviors is very different from how other types of government intervention (e.g., social insurance, minimum wage, in-kind transfer, and any program that offers coupons and discounts) relate to suicidal outcomes. I encourage the authors to examine these possible mechanisms and limit attention to a few, more homogeneous income support programs. * Following the previous point, the review needs to focus on a particular age group (retirees or non-retirees) and examine the policy intervention appropriate to that group. From what I understand, it has been mostly retirees the previous research has examined to evaluate this type of research question. * It will be great if the authors could present a brief summary of research on (non-clinical) suicidal outcomes like suicidal ideation. This broader scope involving both clinical and non-clinical outcomes will help readers grasp the progress of suicidal risks and what role the government could play to minimize such risks. * Lastly, the review needs to be split by the developmental context of the study site (developing vs. developed countries) and present a number of key underlying mechanisms specific to developing or developed countries.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Alex Gertner, University of North Carolina
Comments to the Author:
The manuscript under review is a protocol for a systematic review and meta-analysis of the effect of "income security programs" on suicide mortality. This is an important topic. A systematic review would make a valuable contribution to the literature. The paper states that the review began in June 2021, so I will presume that the methods are largely not modifiable. I will therefore focus comments on the presentation of methods.

- We thank Reviewer 1 for their comments and feedback, and have addressed them as follows:

1. The protocol is remarkably broad in its inclusive definition of "income security programs," even including housing support and food aid. The protocol cites the "ILO guidelines" for its interpretation of income security programs, but I couldn't find a clear definition in the cited ILO report. The closest to a definition I found was: "income security. Protection of income through minimum wage machinery, wage indexation, comprehensive social security, progressive taxation to reduce inequality and to supplement those with low incomes, etc." I find this definition difficult to interpret since it includes several specific measures as well as "comprehensive social security." The protocol also cites

a paper by Sun and colleagues that defines cash transfers as “transfers of cash or cash equivalents (e.g., tax benefits) and vouchers for cash or cash-like in-kind assistance vouchers (e.g., Supplemental Nutrition Assistance Program benefits, known as food stamps).” The protocol seems to use a combination of these two definitions. A clearer justification of the types of programs included in the review based in a theoretical or practical reasoning would strengthen the review. A clearer justification for theorizing all these forms of social assistance as “income security” would also be helpful. I wonder if this protocol is not better defined as focusing broadly on social security programs rather than merely income security. I worry that reducing these varied forms of social assistance to “income security” will elide much of heterogeneity in program characteristics and effects. A description of a plan for presenting results from such varied programs would also be helpful. Though the Sun and colleagues’ papers conflates cash transfer and in-kind transfer programs under the umbrella of cash transfer, I would oppose such a conflation in the current project, since these two types of programs have markedly different theoretical implications.

We have reworked the conceptualization of the main interventions to be focused on “social protection programs”. This is a concept that is grounded in a widely recognized definition from Norton et al (2001), which includes public actions that address “the deprivation and vulnerabilities of the poor, and also with the needs of the non-poor for security in the face of shocks and the particular demands of different stages of the life cycle” (p.22). We also drew on a synthesized report (funded by the UK Department for International Development) aimed at summarizing the evidence base on the use of social protection programs to minimize negative shocks in the global context (cite). Based on the report, we created Figure 1 (pg 8) to show the 4 types of social protection programmes (i.e. social assistance, social care, social insurance, and labour market policies/interventions) and examples that fall under each type.

We have included additional search terms on pg 11 and 12 (Table 1 and Table 2) based on the social protection programs which are presented in Figure 1 (pg. 8). We also now include Supplementary File 1 to show how these terms are operationalized in each database. The new terms included are maternity benefits, active labour market, employment service, wage subsidy, vocational training, job-search services, work sharing, injury compensation, and family support. We believe these changes have made our target interventions more interpretable and grounded in prior theory and literature.

The rationale behind using a broad range of search terms is due to the dearth of available literature for our evaluation (based on preliminary work, we know that there are few studies using a quasi-experimental approach that evaluated suicide outcomes). Thus, we believe that casting a wide net to include all social protection programs will allow us to capture relevant research studies where our study team will further evaluate the fit of these studies in our final review. While there are concerns about the heterogeneity of our listed programs and the impact of the various mechanisms on our study findings, we will discuss each category of identified programs (Table 1) separately which will allow us to summarize our results in a coherent manner for a clear takeaway to inform policymakers. In other words, results of studies with heterogeneous programs will not be conflated in the review. We have included the following sentence in our manuscript for clarification:
“We will consider each program’s economic contexts (e.g. low- or middle-income vs. high-income countries), study design (e.g. use of individual- or population-level data), types of program (e.g. universality, delivery, conditionality), and underlying mechanisms, and use this information to analytically categorize these programs. The results will be summarized separately for each program category.” (pg.13”)

In addition, to justify our approach, we have identified prior systematic reviews that utilized broad search terms where prior research on certain topics have been sparse:

- Courtin E, Kim S, Song S, Yu W, Muennig P (2020) Can social policies improve health? A systematic review and meta-analysis of 38 randomized trials. Milbank Q 98(2):297–371

● Simpson, J, Bambra, C, Bell, Z, et al. Effects of social security policy reforms on mental health and inequalities: a systematic review of observational studies in high-income countries. Soc Sci Med 2021;272:113717.

2. On a related note, the protocol will include studies of individual-level interventions, such as cash transfer RCTs, as well as policy-level effects, such as adoption of minimum wage. This seems to me a possibly perilous venture. It would be like doing a systematic review of the effect of diet on diabetes outcomes and reviewing RCTs of diet programs alongside public policies to reduce calory intake like limiting soda beverage size. The targeted individual level intervention is obviously different than a population-level policy, though the essential hypothesized mechanism may be the same. The studies evaluating these two categories of programs may also produce different types of treatment effect. I would find it helpful if the protocol discussed the specific hypothesized mechanism(s) through which income security affects suicide to clarify the inclusion of different program and study types. I would also find it helpful if the protocol discussed how it will synthesize or interpret studies of individual level and population level effects, with some consideration of whether studies produce average treatment effects or treatment effects on the treated or so on. I worry that it may be difficult to meaningfully summarize the variety of study designs and effects in the S1 table that the protocol presents.

The meta-analysis will be separately performed for studies at the individual-level and population-level designs. If we do not have enough studies to undertake meta-analyses, then only a narrative systematic review will be conducted. For the narrative systematic review, we will present the study design as a part of study characteristics in a summary table. This is now reflected in the manuscript. (see pg. 14) Therefore, since these study designs will be interpreted separately in our review, there should not be any difficulties when summarizing the study effects.

3. It's not clear to me why the study is restricted to government programs or how government programs are defined. I imagine that there are a non-trivial number of cash transfer experiments that were conducted by and through non-governmental organizations and academic institutions. The exclusion of such program would be a major limitation to the study. More explanation of this would be beneficial.

Now that the main focus of the review is based on social protection programs, our definition of social protection programs includes both governmental and non-governmental services. Reflecting the reviewer's comment, we are not going to exclude non-governmental interventions for inclusion. (see pg. 7)

4. One reason for exclusion that appears in the protocol documents but not in the text of the protocol is "Specific intervention in specific setting." It's not clear to me what this exclusion reason represents since it doesn't seem to be discussed in the protocol. Excluding studies of "specific interventions in specific setting" would seem to represent a broad exclusion that would weaken the study. Explanation and discussion of this exclusion would be helpful.

"Specific intervention in specific setting" implied the exclusion of non-governmental action. However, now that we have expanded our criteria to include non-government services and programs, this criterion will be removed from our data extraction form (in supplementary files).

5. The inclusion of studies in English, French, Spanish, German, Chinese, Korean, and Japanese is a significant potential strength of the study. The protocol mentions these languages are "familiar" to the review team, but it is not clear what this means specifically. Do all members of the review team possess advanced knowledge of all these languages that will permit nuanced analyses of academic research papers? If so, this should be clearly stated. If not, then the protocols should

state how studies will be reviewed in all languages by all review members without compromising the quality of the reviews and without biasing the reviews. For instance, it's difficult to imagine how one author might translate a paper to other authors without introducing bias regarding their assessment. I would encourage the authors to err on the side of including fewer languages without compromising the quality of reviews.

Thank you for your suggestion. We have decided to limit our study inclusion to only the English language because of the need for multiple reviewers to understand the same language.

6. The language used to describe quasi-experimental studies is imprecise. For instance, the protocol states that quasi experimental studies "allow for treatment to be randomly assigned," which is not true. Quasi-experimental studies make use of exogeneous variation, typically in the absence of random assignment and typically with reliance of many assumptions. While the language can be easily corrected, I'm concerned that it may belie a lack of expertise on quasi-experimental methods that could undermine the quality of study reviews. The protocol should describe the expertise of reviewers in quasi-experimental methods, including their past use of quasi-experimental techniques in published research.

The error was made by accident, we meant to say "as if there was random assignment to establish exchangeability like the RCT... which it does by exploiting exogenous variation in the exposure". We understand that quasi-experimental studies do not include random assignment of treatment. We admit this was a poor choice of words, and have since changed it to:
"We have corrected the sentence to the following: "Where manipulation to the exposure is not an option, quasi-experimental studies (i.e. natural experiments) can be a viable alternative for causal inference since exogeneity can be established through other means such as through nature, policy, and practice [12]." (p.5)

After checking a number of recent BMJ Open protocols from recent years (see i to v below), we determined that the expertise of reviewers was not information that is typically included. However, we would like to reassure the reviewer that we have experience conducting quasi-experimental studies in our team (e.g. using propensity score matching and difference-in-difference analysis to investigate the impact of the UK bedroom tax on mental health).

Example protocols:
i) Cheloni, R., Gandolfi, S. A., Signorelli, C., & Odone, A. (2019). Global prevalence of diabetic retinopathy: protocol for a systematic review and meta-analysis. BMJ open, 9(3), e022188. https://bmjopen.bmj.com/content/bmjopen/9/3/e022188.full.pdf
ii) Bai, X., Feng, Y., Li, L., Yang, K., Wang, T., Luo, J., ... & Jiao, L. (2020). Treatment strategies for asymptomatic carotid artery stenosis in the era of lipid-lowering drugs: protocol for a systematic review and network meta-analysis. BMJ open, 10(7), e035094. https://bmjopen.bmj.com/content/10/7/e035094.abstract
iii) Pozza A, Coluccia A, Kato T, et al. The 'Hikikomori' syndrome: worldwide prevalence and co-occurring major psychiatric disorders: a systematic review and meta-analysis protocol. BMJ Open 2019;9:e025213. doi: 10.1136/bmjopen-2018-025213 https://bmjopen.bmj.com/content/9/9/e025213.abstract
iv) Fortier C, Obeid H, Paré M, et al. Changes in arterial stiffness indices during a single haemodialysis session in end-stage renal disease population: a systematic review and meta-analysis protocol. BMJ Open 2021;11:e045912. https://bmjopen.bmj.com/content/11/9/e045912\

v) Shah SGS, Nogueras D, van Woerden H, et al. Effectiveness of digital technology interventions to reduce loneliness in adults: a protocol for a systematic review and meta-analysis. BMJ Open 2019;9:e032455. doi: 10.1136/bmjopen-2019-032455 https://bmjopen.bmj.com/content/9/9/e032455.abstract

7. PRISMA-P states that the search strategy should be presented in a way that it could be repeated. I don't think the current protocol reaches that level of detail and clarity. I suggest providing in the paper or in a supplement the specific text that will be used in searches, including AND/OR statements and the like, as well as the parameters of the search, including where in records terms will be searched for. The rationale for inclusion of both "suici*" and "suicide complete" in the search is unclear since the first would presumably include the second.

We thank the reviewer for pointing this out, we have revised our search strategy to provide more detail and clarity. We've added the exact search strategy required to replicate our search results in the supplementary materials (see Supplementary File S1)

8. It's unclear to me what the protocol offers a definition of suicide mortality. Is this to mean that the protocol will only include studies that use this definition of suicides? That would represent a weakness of the study in my view. Rather, I expect the protocol will identify multiple approaches to defining suicide in literature. Indeed, presenting and comparing different approaches to defining suicide could be a valuable contribution of the review.

We present the ICD-10 codes in our protocol to identify suicide mortality since it is a international coding system that is most often used in this field. However, researchers may identify suicides without using ICD codes, and in those cases, we will try our best to match the authors identification of suicide mortality in the paper ICD definition in our protocol (i.e. X60-X84, Y10-Y34, Y87.0). If we cannot match the definitions ourselves, we will contact the author to confirm whether the deaths included in the study matches with the definitions we used above. We reflected the change in the manuscript (pg 10-11).

We used ICD codes to illustrate that we would include deaths that are officially determined to be suicides (X60-X84), and we are going to include studies that use additional codes for "likely suicides" (Y87.0, subsets of Y10-Y34), since a narrow definition of suicides that excludes potential suicide deaths may produce results that are biased towards the null. Following prior literature (see below), it has been recommended that studies use an extended definition of suicide, which includes undetermined deaths (Y10-34), sequelae of intentional self-harm, assault and events of undetermined intent (Y87.0).

Conner KR, Langley J, Tomaszewski KJ, Conwell Y. Injury hospitalization and risks for subsequent self-injury and suicide: a national study from New Zealand. Am J Public Health. 2003;93(7):1128-1131. doi:10.2105/AJPH.93.7.1128

Burrows S, Auger N, Gamache P, St-Laurent D, Hamel D. Influence of social and material individual and area deprivation on suicide mortality among 2.7 million Canadians: A prospective study. BMC Public Health. 2011;11(1):577. doi:10.1186/1471-2458-11-577

Rhodes AE, Links PS, Streiner DL, Dawe I, Cass D, Janes S. Do hospital E-codes consistently capture suicidal behaviour? Chronic Dis Can. 2002;23(4):139-145.

9. The term "Medicare-eligibility" may not be legible to non-Canadian readers and should be replaced with a more generic terms such as "public insurance eligibility" or some such.

Reflecting the comment, we changed this in the manuscript. We changed the example to be clearer, and now we say “For example, in South Korea, a medical aid program, which provides medical service for the bottom 3-4% of households of income, is often used as a means-testing criterion for social protection programs.” (pg. 10)

10. How are relevant studies being defined in Stage 1? What are the criteria? How will disagreements be handled? What is the point of stage 2, especially since all authors are involved in stage 1? Pg 9 says all team members will assess study quality at stage 3 and stage 10 says only two team members will. Clarification would be helpful.

To provide better clarification of our study selection stages, we explain in detail each stage below. We have also clarified each stage based on the reviewer’s suggestions in the protocol manuscript (pg. 13):

1. Stage 1: All authors screened titles and abstracts at this stage, and each title and abstract required two authors to determine its relevance. If a disagreement arose, all authors would discuss the relevance of the title and abstract and decide collectively on its inclusion/exclusion.
2. Stage 2: We did not mean that all authors are going to review all the titles repeatedly. In the case that relevant studies are excluded with no discrepancy (i.e. was not subject to team discussion and consensus), stage 2 is included.
3. Stage 3: All authors will be screening full-text and assess eligibility. Discrepancies are dealt with in the same manner as in stage 1.

Regarding the issue about the assessment of quality by all authors on pg 9 vs. two authors on pg 10, we have now corrected pg 12 so that quality assessment is being conducted by all authors.

11. The protocol states that it will use the ROBINS-E tool, but the University of Bristol website cited states that the ROBIN-E tool does not yet exist: http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-e/. Clarification would be helpful.

Reflecting the reviewer’s comment, we will use Risk Of Bias In Non-randomised Studies of Interventions (ROBINS-I), instead of ROBINS-E.

12. The paper states that the review will provide evidence to support implementation of income security programs. I’m not sure what’s meant by this or how the review will provide evidence to support implementation. More explanation would be helpful.

We have now re-worded the sentence to reflect that the findings may or may provide evidence in support of these programs: “to provide evidence to support the decision making process with regards to the implementation of social protection programs as a core part of suicide prevention strategy.” (pg. 7)

13. The abstract states, “Our review aims to evaluate the causal relationship between income security programs and suicide mortality by examining all relevant experimental and quasi-experimental studies between January 1980 and May 2021.” Perhaps it would be more accurate to say that this review aims to evaluate the evidence of a causal relationship, since the review does not directly evaluate the causal relationship.

We have changed the sentence in the abstract as suggested by the reviewer.

Reviewer: 2
Dr. Tae-Young Pak, Sungkyunkwan University
Comments to the Author:

This paper addresses an important and timely issue that could have implications beyond social epidemiology. I believe this review will amount to a well-respected summary of what we know about the health effects of government income support and the role the government needs to play in suicide prevention. Below, I suggest a few points the authors may need to consider along with the project.
- We thank Reviewer 2 for their comments and feedback, and have responded to each point as follows:

14. The scope of income security programs may need to be restricted to cash transfer programs. I believe the mechanism through which cash transfer program affects suicidal behaviors is very different from how other types of government intervention (e.g., social insurance, minimum wage, in-kind transfer, and any program that offers coupons and discounts) relate to suicidal outcomes. I encourage the authors to examine these possible mechanisms and limit attention to a few, more homogeneous income support programs.

We intentionally chose to broaden the program definition because only very few programs have been tested to have impacts on suicide mortality. From through reading of previous literature reviews, we found that including only cash-transfer programs may prevent us from identifying a number of studies on social protection programs and suicide that are important in answering our research question. Also, even though the mechanism can be different for heterogeneous programs, given the preliminary stage of the research question in the area, the proposed systematic review will contribute to the existing knowledge synthesis on what social protection programs have been examined and what the provisional mechanisms can be.

15. Following the previous point, the review needs to focus on a particular age group (retirees or non-retirees) and examine the policy intervention appropriate to that group. From what I understand, it has been mostly retirees the previous research has examined to evaluate this type of research question.

From our initial search, the previous studies have examined programs that targeted a wide range of populations, including the working-age population or the general population, rather than mostly retirees. For example, both Alves (2019) and Christian (2019) include both non-retirees and retirees. Additionally, our preliminary search has found a number of labour market policy studies, and restricting our focus to a particular age group may exclude the working-age population. Given the limited number of quasi-experimental studies on this topic, we do not believe that we should restrict the review to a certain age group. In the narrative portion of the review, we can highlight potential different effects of these policies on different population groups.

16. It will be great if the authors could present a brief summary of research on (non-clinical) suicidal outcomes like suicidal ideation. This broader scope involving both clinical and non-clinical outcomes will help readers grasp the progress of suicidal risks and what role the government could play to minimize such risks.

We would like to only focus our protocol on mortality since individual-level socioeconomic positions have a differential impact on non-fatal (e.g. suicidal ideation and attempts) and fatal suicidal events. More specifically, prior literature has provided evidence that trends in socioeconomic inequalities in suicide mortality, attempts, and ideation have different patterns over time, which suggests that predictors of socioeconomic inequalities for fatal and non-fatal suicidal behaviour are different.

Even though those who died by suicides are more likely to attempt and have suicidal ideation, only a very small proportion lead to suicide deaths (see #1 below), and often the demographic patterns differ (e.g. women are more represented in suicidal ideation and attempts while men die by suicide more -- see #2 below). Therefore, it is possible that social protection programs have different impacts on different types of suicidal behaviour, and our focus is to evaluate the programs that have effects for mortality, rather than other non-fatal suicidal behaviours.

At a methodological level, the measure of suicide ideation and attempts may have greater variability in how it is measured from country to country. Prior literature has noted a lack of an international data-collection system for the reporting of suicide ideation and attempts (see #3 below), and the inclusion of these non-clinical outcomes may introduce a significant level of bias. Additionally, there are cross-cultural differences to the notion of suicidal ideation that will require additional research to ensure they have comparability across countries/cultures. For example, discussion of suicide ideation is discouraged in some countries (e.g. Singapore, India), and therefore the reporting of ideation may not accurately represent these populations (see #4). Where deaths by suicide can be identified through mortality with determined or probable cause, the current state of data collection across countries makes the comparison of studies on suicide ideation and attempts especially difficult.

1. Klonsky ED, May AM, Saffer BY. Suicide, suicide attempts, and suicidal ideation. Annual Review of Clinical Psychology. 2016;12(1) doi: 10.1146/annurev-clinpsy-021815-093204. annurev-clinpsy-021815-093204.
2. Kim MH, Jung-Choi K, Jun HJ, Kawachi I. Socioeconomic inequalities in suicidal ideation, parasuicides, and completed suicides in South Korea. Social Science and Medicine. 2010;70(8) doi: https://doi.org/10.1016/j.socscimed.2010.01.004
3. Silverman MM, Leo DD. Why There Is a Need for an International Nomenclature and Classification System for Suicide. Crisis. 2016;37 83-87 https://doi.org/10.1027/0227-5910/a000419
4. Biswas T et al. Global variation in the prevalence of suicidal ideation, anxiety and their correlates among adolescents: A population based study of 82 countries. EClinicialMedicine. 2020;24 100395. doi: https://doi.org/10.1016/j.eclinm.2020.100395

17. Lastly, the review needs to be split by the developmental context of the study site (developing vs. developed countries) and present a number of key underlying mechanisms specific to developing or developed countries.

We have planned to consider the heterogeneous economic contexts in the final review. In the case of conducting a meta-analysis, we can test for effect modification by economic context. We will specifically test for subgroup effects by low, middle, vs high income countries using World Bank's definition (see #1 below). In the case of a narrative systematic review, we will present the summary table that includes economic context as well and present the possible underlying mechanisms separately. We added the following into the manuscript to clarify:

"We will provide a summary table of the included studies with effect sizes and details on program specifications. We will consider each program's economic contexts (e.g. low- or middle- or high-income countries), study design (e.g. use of individual- or population-level data), types of program (e.g. universality, delivery, conditionality), and underlying mechanisms, and use this information to analytically categorize these programs. The results will be summarized separately for each program category. Based on these factors, if we have at least three studies with similar programs, we will perform a meta-analysis. Otherwise, only a systematic narrative review will be performed." (p. 14)

1. The World Bank. World Bank Country and Lending Groups. 2021. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

VERSION 2 – REVIEW

REVIEWER	Gertner, Alex University of North Carolina
REVIEW RETURNED	22-Nov-2021

GENERAL COMMENTS	The revised manuscript largely addresses my previous comments. I'm hopeful the review will make a valuable contribution to the literature. I'm still a bit confused about the review process. The manuscript states that all authors will screen all titles and abstracts, but then states that each title and abstract must be screened by at least two authors. If all 7 authors are screening all abstracts, how wouldn't an abstract be screened by at least 2? Perhaps what's meant is that all authors will participate in the screening, with each abstract screened by at least two authors? Having all authors participate in reviewing the quality of all papers seems quite laborious. Will all authors do this together collaboratively or do this independently and then compare their findings? I'm concerned that the authors may find studies that do not fully describe how suicide deaths were identified, such as including ICD codes. I'd encourage the authors to include these studies if they arise and make any variability they find in identification of suicide deaths a part of their results.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Dr. Alex Gertner, University of North Carolina
Comments to the Author:

The revised manuscript largely addresses my previous comments. I'm hopeful the review will make a valuable contribution to the literature.

4. I'm still a bit confused about the review process. The manuscript states that all authors will screen all titles and abstracts, but then states that each title and abstract must be screened by at least two authors. If all 7 authors are screening all abstracts, how wouldn't an abstract be screened by at least 2? Perhaps what's meant is that all authors will participate in the screening, with each abstract screened by at least two authors?

To clarify the review process, while all reviewers will participate in the screening of titles and abstracts, each abstract is not required to be screened by all 7 authors. The program that our team is using (Covidence) will present each abstract to two reviewers, who will determine whether the abstract is included or excluded. For example, if both reviewers agree on inclusion of an abstract, it

will be included in the full text screening process. In our protocol, we have clarified this to avoid confusion:
“At stage 1, all authors (AC, CK, CT, KA, AN, ZB, and TY) will participate in screening of titles and abstracts to identify relevant studies by checking whether the target program, outcome and methods were used. Each title and abstract are required to be screened by two authors, and any discrepancies that arise will be resolved through a discussion between all authors on its relevance based on the inclusion/exclusion criteria.”

5. Having all authors participate in reviewing the quality of all papers seems quite laborious. Will all authors do this together collaboratively or do this independently and then compare their findings?

We (all authors) will be working collaboratively throughout the review process. Therefore, the quality assessment will be conducted by the team discussion. We have now clarified the sentences (pg. 13).

“At stage 3, for the chosen studies screened through titles and abstracts, all team members will be working collaboratively to review the full-texts (comparing results throughout the process), assess the eligibility of the texts and then appraise the quality of the included studies where results are determined by consensus.”

6. I'm concerned that the authors may find studies that do not fully describe how suicide deaths were identified, such as including ICD codes. I'd encourage the authors to include these studies if they arise and make any variability they find in identification of suicide deaths a part of their results.

Under “suicide mortality” on page 9, we now include more detail on if we find studies that do not use ICD codes: “we will not exclude a study if ICD codes were not used”. Additionally, as per your suggestion, we mention that “variability in the identification of suicides will be noted in the results of the review”. See below for the updated section:
“We will not exclude a study if ICD codes were not used. If a study does not use ICD or other standardized diagnostic codes at the full text review stage, we will try our best to match what is written in the paper to the above ICD definition (e.g. contacting the author to confirm whether the deaths included in the study matches with the definitions we used above). Variability in the identification of suicides will be noted in the results of the review.”

Reviewer: 1
Competing interests of Reviewer: none