Social Outcomes Contracting (SOC) in Social Programmes and Public Services: A Mixed-Methods Systematic Review Protocol

Vanessa Picker[1], Eleanor Carter*[11], Mara Airoldi[1], James Ronicle[2], Rachel Wooldridge[2], Jo Llewellyn[2], Lilly Monk[2], Sophia Stone[2], Michael Gibson[1], Franziska Rosenbach[1], Tanyah Hameed[1], Harry Bregazzi[1], Felix-Anselm van Lier[1]

[1]Blavatnik School of Government, University of Oxford, 120 Walton Street, Oxford, OX2 6GG, UK
[2]Ecorys, Queen Elizabeth House, 4 St Dunstan's Hill, Billingsgate, London EC3R 8AD, UK

ABSTRACT

Background: Across a range of policy areas, governments and philanthropists are increasingly adopting a Social Outcomes Contracting (SOC) approach. Under this model, an independent service provider must achieve specific, measurable social and/or environmental outcomes, and payments are only made when these outcomes are achieved. The growing interest in SOC has been accompanied by research on specific programmes, policy domains, and geographies, but there has not been a systematic attempt to synthesise this emerging evidence. To address this gap, this systematic review aims to surface the best evidence on when and where effects have been associated with SOC.

Methods/Design: A mixed-methods systematic review will be undertaken, using a participatory research process involving a Policy Advisory Group (PAG). Twelve bibliographic databases will be searched, alongside a comprehensive search of grey literature. Studies will be screened independently by two reviewers in Covidence. We will conduct risk of bias and quality assessment using recommended tools. Data synthesis will involve meta-synthesis and/or narrative synthesis for quantitative studies, thematic content analysis for qualitative studies, and a cross-study synthesis. If possible, we will also analyse the available economic data to understand the costs and benefits associated with SOC.

Discussion: We will use the systematic review findings to produce accessible and reliable empirical insights on whether, when, and where (and if possible, how) SOC approaches deliver improved impact when compared to more conventional funding arrangements. The outputs will support policymakers to make informed decisions in relation to commissioning and funding approaches.

Systematic review registration: This systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO), on 20th November 2020 and was last updated on 21 January 2021: (registration number PROSPERO CRD42020215207).

Keywords: social outcomes contracting; payment-by-results; outcomes-based contracting; outcomes-based commissioning; social impact bonds; development impact bonds; pay for success

[1] Corresponding author: Eleanor Carter, Blavatnik School of Government, University of Oxford, 120 Walton Street, Oxford, OX2 6GG, UK. Email: eleanor.carter@bsg.ox.ac.uk

1. Background

1.1 The rise of social outcomes contracting

A fundamental shift has taken hold of welfare state governance and the coordination of social programmes since the 1980s with the increasingly widespread use of independent non-governmental delivery organisations, private firms and market-like methods for the delivery of public services and social programmes (Deakin & Walsh, 1996). Public service commissioning and philanthropic performance management systems have become increasingly focused on outcomes – the positive results that services produce in the lives of service users and citizens – rather than the volume or quality of inputs or outputs (Bovaird & Davies, 2011). The term 'social outcomes contracting' (SOC) captures a range of mechanisms, including payment-by-results (PbR) and Social Impact Bonds (SIBs). SOC is pursued by governments and other philanthropic and development agencies on the understanding that it can deliver a range of benefits including: improved cost effectiveness, innovation, accountability, systems-level planning, and responsiveness, with risk transferred to the private sector (Albertson et al., 2018; Brown, 2013; Morse, 2015).

While these outcome-oriented models are lauded by some as being highly-effective and uniquely innovative, attempts to reengineer public service decision-making around outcomes have been pursued since the 1980s (Pollitt & Talbot, 2003). In practice however, shifting commissioning focus from inputs and outputs to outcome is challenging (Bovaird & Davies, 2011) and payment for outcomes is still seen as a relatively new way of giving development aid (Clist, 2019). Nevertheless, SOC continues to be promoted with increasing vigour, with the practice taking on increasing international significance at the leading edge of public service reform (Cabinet Office, 2011; Farr, 2016).

The compelling logic within SOC – that specifying and steering services on the basis of social outcomes will deliver better outcomes – is appetisingly straightforward and aligns comfortably with much of the literature underpinning the advancement of performance measurement and management in public service spaces. At the same time, evidence supporting the effectiveness of schemes operating under SOC is alarmingly limited (Carter et al., 2018; Fox & Morris, 2019; Fraser et al., 2018; Lagarde et al., 2013).

1.2 Aim of the review

This study will examine the effects associated with different SOC approaches, including PbR, SIBs, and their synonyms. It will be global in reach, encompassing studies from low, middle, and high-income contexts. While this study is primarily focused on understanding the effectiveness (or otherwise) of SOC approaches, we intend to produce several outputs over a number of years, exploring the option of focusing on more contextualised, realist work in future pieces. Through each of these outputs, we will aim to offer accessible and reliable empirical insights so that organisations responsible for funding social programmes can make evidence-informed decisions on the most appropriate form of outcome contract or financing model to adopt in different contexts.

To aid with this objective, we are establishing a Policy Advisory Group (PAG). A list of PAG members is provided at Appendix 4. The PAG will help steer the study, ensuring it produces the insights and outputs that are of most value to policymakers. They will be involved at the beginning to set the scope of the study; in the middle to help with sourcing and refining the documents to review; and at the end to shape outputs and consider future activity. While the role of the PAG will be to advise the researchers, the research team will have the discretion over how to apply (or not) the advice.

1.3 Review question

The overarching question guiding this systematic review is: *"What are the impacts of social outcomes contracts, often referred to as payment-by-results or impact bonds, on person-level and system-level outcomes when compared to more conventional funding approaches?"* We propose to use a series of sub-questions to direct research phases and additional review activities. Early work to support evidence mapping will be guided by the sub-question: *"What is the nature, quality and coverage of the existing literature focused on SOC approaches and how does this differ by type of funding instrument, policy area and/or country"?*

2. Methods/Design

This review will be undertaken using the Preferred Reporting Items and Meta-analysis (PRISMA). Consequently, this protocol has been prepared using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol (PRISMA-P) guidelines (Shamseer et al., 2010) (Appendix 3).

2.1 Data sources and search strategy

2.1.1 Electronic bibliographic databases

The following bibliographic databases will be searched (1990 – current), with no language restrictions applied: ABI/INFORM Global, Applied Social Sciences Index & Abstracts (ASSIA), Scopus, International Bibliography of the Social Sciences (IBSS), PAIS Index, PolicyFile Index, Proquest Dissertations and Theses, ProQuest Social Science, Social Services Abstracts, Web of Science, Worldwide Political Science Abstracts and PsycINFO. The search strategy will primarily include terms relating to the intervention. Where the search would otherwise be too broad, we will include terms related to study design. In addition to identifying search terms from relevant literature, search terms (English and non-English) recommended by PAG members will be tested and included. The search strategy for Scopus is provided in Appendix 1. The terms and syntax will be adjusted across databases to accommodate independent indexing systems. Prior to search execution, a second reviewer will appraise the strategy, taking into account the considerations outlined by McGowan et al. (2016). We will search for relevant systematic reviews within the Cochrane Library, the Joanna Briggs Institute (JBI) and PROSPERO.

2.1.2 Searching trial registers

We will search the following registers for relevant ongoing, recently completed and unpublished clinical studies trials: World Health Organization International Clinical Trials Registry Platform (global); National Institutes of Health clinical trials registry (US); National Institute for Health Research Clinical Research Network Portfolio Database (UK); National Research Register Projects Database Archive (UK); and Current Controlled Trials (includes the International Standard RCT Number Register).

2.1.3 Grey literature search

There is a considerable body of research and evaluation on SOC that has been produced beyond a formal academic context. Consequently, we will not restrict based on publication status and we will conduct a comprehensive grey literature search. We have devised a four-pronged approach, including: 1) searching grey literature databases (e.g. Open Grey); 2) running 192 specified search terms on Google and reviewing the first 100 results for each search term; 3) issuing a 'Call for Evidence' to gain input from 'content experts' with

experience in SOC; and 4) searching websites of relevant organisations/bodies involved in SOC. The PAG will assist with components three and four.

2.1.4 Searching reference lists and contacting authors

We will search the reference lists of relevant studies and reviews to locate additional studies. While far from objective, this may enable us to locate some studies that have not been indexed in academic databases (Higgins et al., 2019). As this is a relatively new field of research, we will also contact authors of relevant studies via email to ask whether they could provide additional published or unpublished work.

2.2. Study inclusion criteria

2.2.1 Study eligibility criteria

This review intentionally includes a diverse range of study designs, which will then by synthesised through tailored approaches in an ongoing series of research outputs. To determine the effectiveness of SOC approaches, we will include studies that use an appropriate experimental or quasi experimental design. This may include: individually randomised controlled trials; cluster-randomised controlled trials; and non-randomised studies. Non-randomised studies can include designs such as: controlled before and after studies; historically controlled studies; and retrospective or prospective cohort studies that include a control group. Other quantitative research designs (e.g., pre-post test, regression analysis, or descriptive statistics) may be analysed as part of a more flexible strand synthesis work, depending on the form and prevalence of identified studies.

Studies that offer original, independent synthesis may also be included. Examples of independent syntheses include systematic reviews and formal literature reviews; these will be included only if they describe the review method used.

Qualitative evidence is crucial for maximising the value of a systematic review in informing policy and decision-making (Fraser et al., 2009; Tong et al., 2007). We will therefore also include evidence from qualitative studies that explore perceptions of practitioners and/or service recipients on the application of SOC. This will include studies which use qualitative methods for data collection and analysis, such as in-depth interviews, focus groups, observation, reflective diaries, and/or case studies. The qualitative evidence may be contained within studies included in the effectiveness part of the review and/or in stand-alone qualitative studies.

As the review also aims to better understand the costs and benefits associated with SOC approaches, we will include studies that contain economic data. Examples of relevant economic analysis include any form of cost effectiveness or cost-benefit analysis and may also include information on transaction costs.

2.2.2 *Population*

This review has no restrictions on study participants in terms of characteristics such as age, gender, ethnicity, or morbidities. This is reflective of the fact that SOC approaches have been used to address a wide array of policy issues, thereby involving a diverse set of participants. We also have no restrictions on study setting or context, as we are aiming to identify studies from low, middle, and high-income contexts.

2.2.3 Intervention

The intervention being studied is the use of SOC as a funding structure for programmes which pursue social and/or environmental outcomes. Included studies need to evaluate social

interventions that have been funded in total or in part by a SOC approach. We have defined SOC arrangements as the provision of any public service or social programme on behalf of a commissioner (i.e., a government outcomes payer) by non-governmental service providers where payment to providers is contingent (either in full or partly) on the achievement of pre-specified, measurable outcomes. Key components therefore include: independent, non-governmental delivery agents; contracted provision; and payment contingent on outcomes performance/results achieved.

To be included, the 'unit of incentivisation' within the intervention must be appropriate. For example, payment for outcomes where the incentivised agent is a government (e.g., some forms of Results Based Financing) or an individual person or household (performance-related pay for teachers or conditional cash transfers) will be excluded. The incentivised organisation(s) must be service providers from the private or not-for-profit sectors or in the case of impact bond type models, investment managers/special purpose vehicles where returns are contingent on the achievement of specified outcomes.

2.2.4 Comparator

For the conventional effectiveness component of the review, SOC structures will be compared to other, more conventional funding structures. Thus, quantitative studies must include a comparator enabling a comparison between a SOC approach and at least one other, more conventional funding approach. Examples of conventional funding structures that are relevant for inclusion include grants, in-house public sector provision, and/or fee-for-service contracts with independent providers. Each study will likely refer to a different form of 'business as usual' or conventional service payment structure and the review will be sensitive to these alternate comparator arrangements. While an appropriate comparator will be crucial for the traditional effectiveness review, quantitative studies that lack a comparator may be analysed as part of a more flexible strand of work. The qualitative studies will not necessarily make explicit reference to a single or direct comparator funding approach.

2.2.5 Outcomes

To be included in the effectiveness review, the study must include social and/or environmental outcomes. It is expected that a range of outcome measures and indicators will serve as the underlying 'payable' outcome (or the main outcome of interest in performance clauses). Studies may report individual, person-level outcomes (e.g., changes in recidivism rates experienced by people leaving prison), and/or contract- or system-level outcomes (e.g., effect of the funding model on service quality or access). Where possible, we will include economic outcomes related to costs incurred by governments and the associated benefits. The qualitative studies must include perceptions from practitioners and/or service recipients on the application of SOC.

2.3 Study screening and selection

After executing our search strategy, we will upload all studies into 'Covidence' for review. Duplicates will automatically be removed upon import. A team of eleven trained reviewers will screen titles/abstracts. Two reviewers will independently screen each title/abstract of all retrieved studies, using our eligibility criteria to remove any obviously irrelevant articles. Two reviewers will examine full-text versions of the remaining, potentially relevant articles, for eligibility. See Appendix 2 for our full-text hierarchical exclusion criteria tool. We will resolve disagreements regarding study eligibility through discussion, referring to an independent arbiter. We will document this process using the PRISMA flow diagram.

2.4 Data extraction

We will extract data from the included studies using modified and piloted versions of standardised data extraction forms (e.g. JBI 2014). Separate forms will be used for differing types of studies, including: 1) studies with a control group; 2) studies without a control group; 3) studies with a qualitative design; 4) studies using systematic review methods; and 5) studies providing economic analysis. The forms will be selected and adapted after conducting an interactive workshop with several academics with relevant methods and domain expertise. We will extract data on key study components relating to the research question, including features of the study design, population characteristics, funding instrument characteristics, study outcomes, benefits, and adverse events and perceptions. We are also exploring the suitability of machine learning tools to facilitate data management and extraction. For the qualitative component, recognising pragmatic constraints, we will use an adapted version of Patton's 16 purposeful sampling strategies for qualitative research synthesis, to determine which qualitative studies to extract data from (Patton 1990, 2002; Suri 2011). Missing data will be requested by the study authors via email. Where data remain unavailable, we will analyse the available data and discuss the potential impact of the missing data.

2.5 Assessing risk of bias and study quality

Reviewers will assess the methodological quality of the included studies, with a sample of the studies being selected for double review. Disagreements between reviewers will be resolved by discussion; a third reviewer will be consulted where necessary. As the evidence base is still developing, studies will not be excluded on the basis of quality. Regardless, the quality appraisal step will be integral for assessing any potential bias and considering the extent to which it is possible to have confidence in study findings (Gough et al., 2012).

For quantitative studies, we will use Cochrane recommended risk of bias tools. We will assess the risk of bias in each randomised study using a tool such as Version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) (Sterne et al., 2019). We may also use a tool such as the Outcome Reporting Bias in Trials (ORBIT) classification system for assessing publication bias and outcome reporting bias in Randomised Controlled Trials (RCT's) (Kirkham et al., 2010). For non-randomised studies, we will use a different risk of bias tool, such as the Risk Of Bias in Non-randomised Studies - of Interventions (ROBINS-I) tool for assessing risk of bias in non-randomised studies of interventions (Sterne et al., 2016). For both randomised and non-randomised studies, we will compare outcomes reported in the study protocol (where available) and the published report to further assess publication bias and outcome reporting bias. Where a protocol is unavailable, we will compare the outcomes reported in the methods and results sections. If meta-analysis is possible, we may also use the Risk Of Bias due to Missing Evidence (ROB-ME) tool to assess the risk of bias resulting from missing evidence (Page et al., 2018). We will present risk-of-bias judgements using a table and/or forest plot (if meta-analysis is possible), along with brief text justification (Higgins et al., 2019; Page et al., 2018).

For the qualitative component, we will conduct quality assessment using a critical appraisal tool developed specifically for qualitative research studies, as recommended by Hannes (2011). Reviewers will carry out quality assessment using a tool such as the Critical Appraisal Skills Program (CASP) Checklist for Qualitative Research (CASP 2020). If extra guidance is required when filling out the main qualitative tool, the reviewers may refer to the Consolidated Criteria for Reporting Qualitative Research (COREQ), as required (Tong et al., 2007). To qualitatively assess publication bias and outcome reporting bias, we will consider the extent to which: funders and/or researchers have a vested interest in the results; the authors' interpretations are consistent with actual results; and only positive effects in support of the intervention are reported.

2.6 Analysis

2.6.1 Overview of descriptive analysis

If meta-synthesis cannot be conducted, we will carry out a descriptive narrative synthesis, categorised by type of intervention, type of outcome/s, population characteristics and/or policy sector. We will provide summaries of intervention effects for each included study by calculating standardised mean differences (for continuous outcomes) and risk ratios (for dichotomous outcomes), using the data provided in the studies and/or the data obtained by contacting authors.

2.6.2 Overview of statistical analysis

We anticipate that there will be limited scope for conducting meta-analysis as the evidence base is still developing and initial scoping indicated an absence of experimental and quasi-experimental studies that contain an appropriate comparator. Further, it is likely that there will be significant variation in the types of outcomes measured and the intervention, settings and populations are also likely to differ considerably. The heterogeneity across studies may make it inappropriate to pool the results. Only where studies have used a similar intervention (programme and funding contract), along with the same outcome measure, will we pool the results.

Specifically, we will pool the results of randomised controlled trials, if possible, using a random-effects meta-analysis. Non-randomised studies may be included in the meta-analysis if we identify fewer than three randomised or quasi-randomised trials. We will use RevMan 5 software to conduct the meta-analyses.

For continuous outcomes, we will calculate standardised mean differences and for binary outcomes, we will calculate risk ratios (RRs). We will calculate 95% confidence intervals (CIs) and two-sided P values for each outcome. We will consider a two-sided P value of <0.05 to be significant. If the effects of clustering have not been taken into account in a study, we will adjust the standard deviations (SDs) for the design effect, using intra-class coefficients, if they are provided in the study reports. If they are not provided, we will use external estimates obtained from similar studies. Where required, we will log transform skewed data.

2.6.3 Overview of thematic analysis

The qualitative data will describe service user and/or practitioner experiences in developing, implementing, and/or receiving interventions that have been funded using a SOC approach. The data may also describe barriers to and/or facilitators of SOC. To identify common categories and themes, we will undertake thematic content analysis. This will involve the following three stages outlined by Thomas and Harden (2008): 1) coding text; 2) developing descriptive themes; and 3) generating broader analytical themes. Each of these steps will be independently undertaken prior to consolidating content; a sample will be validated by a second reviewer with any disagreements to be discussed and resolved by a third reviewer. The approach to coding and analysing qualitative research will be refined in response to recommendations from the PAG and the nature and scope of identified studies.

2.6.4 Overview of economic analysis

We anticipate that the types of economic analysis will be diverse, and we will therefore adopt a pragmatic approach to synthesis. Our approach is informed by (Gomersall et al., 2015) and the most recent guidance for synthesising evidence from primary economic evaluation research from the JBI. We do not expect to generate an average generalisable incremental cost-effectiveness measure. We instead aim to identify the range and quality of available studies

related to SOC contracting resource use/cost and/or cost-effectiveness. We propose to adopt the JBI critical appraisal checklist for studies reporting economic evaluations and the three-by-three dominance ranking matrix tool for synthesising and interpreting findings from economic evaluations.

2.6.5 Overview of cross-study synthesis

For the cross-study synthesis, we will bring together the findings from the effectiveness review, economic review, and qualitative review, informed by the approach adopted by Hannes and Lockwood (2012). We will compare the findings from the qualitative review against the findings of the quantitative review and the findings of the economic review, using a conceptual and methodological matrix. This will allow us to identify synergies, conflicts, and research gaps.

2.6.6 Future analysis

Recognising that the literature on SOC approaches may not conform neatly to traditional systematic review methods, we also intend to produce future pieces of research that incorporate methods designed to work with complex social interventions. We will explore a range of methods that go beyond measuring and reporting on programme effectiveness. This could include incorporating the 'realist' approach to evaluation, which aims to understand *"what works for whom, in what circumstances, in what respects and how"* (Pawson et al., 2015). The realist approach to research synthesis is aligned to many questions of importance to policymakers, as it aims to explain the link between the intervention context, the mechanisms by which it works and the outcomes produced (Pawson et al., 2015).

2.7 Other considerations

2.7.1 Subgroup and sensitivity analysis

We will conduct sensitivity analysis to ensure that the overall result and conclusions are not affected by the different decisions made during the review (Higgins et al., 2019). We will repeat the primary statistical analysis or meta-analysis, substituting alternative decisions or ranges of values for decisions that were arbitrary or unclear. This may involve excluding non-randomised studies, excluding studies rated as high risk of bias, and/or assessing the impact of selective reporting on meta-analytic results. We will present the results of the sensitivity analysis in a summary table. If possible, we will also conduct subgroup analysis, to compare the effectiveness of SOC mechanisms between subgroups such as different regions, policy domains, populations, and/or intervention types. This may involve using plots with outcome measures and/or text descriptions to complement and explain any subtle differences that have been identified but which are not explained by the statistical findings.

2.7.2 Unit of analysis issues

The unit of analysis may vary between studies, as studies may report individual person-level outcomes and/or contract- or system-level outcomes. For studies reporting individual person-level outcomes, we will treat the individual study participants as the unit of analysis. If any multi-arm studies meet the inclusion criteria, we will combine the groups in order to create a single pairwise comparison (Higgins et al., 2019). Where this is not possible, we will select the treatment group receiving the most intense level of intervention, and the control group receiving the least intervention, from each study. Studies reporting contract or system level outcomes will include a comparator of at least one other, more conventional funding approach, allowing for a comparison of the effect of the different funding mechanisms.

2.7.3 Assessing confidence in evidence

We will use a tool such as Grading of Recommendations Assessment, Development and Evaluation (GRADE) to assess the confidence in the evidence arising from the quantitative studies (Mustafa et al., 2013). We will present the assessments in a summary of findings table. Additionally, we will use an approach such as the Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach to assess confidence in the qualitative evidence (Guyatt et al., 2011).

3. Discussion

This mixed-methods systematic review will provide a detailed synthesis of the existing evidence on the effectiveness of SOC approaches, thereby informing future commissioning decisions and improving a broad range of individual and system level outcomes. It will provide empirical insights on whether, when, and where (and if possible, how) SOC approaches deliver improved impact when compared to more conventional funding arrangements. This will support policymakers to make more informed decisions in relation to commissioning and funding approaches.

List of Abbreviations

ASSIA: Applied Social Sciences Index & Abstracts
CASP: Critical Appraisal Skills Programme
CERQual: Confidence in the Evidence from Reviews of Qualitative Research
CI: Confidence Interval
Consolidated Criteria for Reporting Qualitative Research: COREQ
GRADE: Grading of Recommendations Assessment, Development and Evaluation
GRADE-CERQual: Confidence in the Evidence from Reviews of Qualitative research
IBSS: International Bibliography of the Social Sciences
JBI: Joanna Briggs Institute
ORBIT: Outcome Reporting Bias in Trials
PbR: Payment by Results
PAG: Policy Advisory Group
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol
PROSPERO: The International Prospective Register of Systematic Reviews
RoB 2: Version 2 of the Cochrane risk-of-bias tool for randomized trials
ROB-ME: Risk Of Bias due to Missing Evidence
ROBINS-I: Risk Of Bias in Non-randomized Studies - of Interventions tool
RR: Risk Ratio
SD: Standard Deviation
SIB: Social Impact Bond
SOC: Social Outcomes Contracting

Additional Files
Additional File 1: Example Search Strategy (Scopus)
Additional File 2: Hierarchical Exclusion Critera
Additional File 3: Completed PRISMA-P Checklist
Additional File 4: List of Policy Advisory Group (PAG) Members

Declarations

Availability of data and material: The outputs from this project will be made available on the Government Outcomes Lab (GO Lab) website and Ecorys website, as well as being published in an appropriate academic journal.

Competing interests: Senior members of the review team are regularly invited to present at international conferences and events on the application of SOC, including SIBs and PbR. Ecorys conducts research and evaluations internationally in the field of SOC and holds contracts with a range of public and private organisations to undertake this work. Similarly, the GO Lab has been funded by government departments and independent foundations to investigate the application of impact bonds and outcome contracts. Some members of the review's PAG, who are invited to make recommendations on the review scope and reporting, are previous or existing clients and some could potentially be clients for future opportunities.

Funding: This review is funded by UK Department for Digital, Culture, Media and Sport (A1953 - Programme for Effective Partnerships); the Blavatnik School of Government, University of Oxford and the Netherlands Economic Institute. The funders have not been involved in developing this protocol. Dr Eleanor Carter is supported by a UKRI Future Leaders Fellowship.

Authors' contributions: EC is the guarantor. MA, EC, and JR developed the overarching objectives and research questions for the review. VP drafted the manuscript drawing on background notes prepared by EC. VP developed the search strategy, the inclusion/exclusion criteria and the strategy for the key steps of the systematic review (screening, data extraction, risk of bias assessment etc). EC, MA, JR, and RW provided feedback on each of these decisions. EC provided guidance on various components of the manuscript, as required and contributed to the background section and economic analysis section. EC reviewed the search strategy. RW developed the strategy for identifying grey literature. JR led the establishment of the PAG. VP, EC, MA, and RW provided assistance with identifying potential PAG members. MG, TH, JL, LM, FR, RW, SS, and VP piloted the abstract screening tool, prepared notes to support the PAG and made recommendations on the inclusion criteria and scope. EC, MG, HB, and FvL developed and tested the hierarchical exclusion criteria. All authors read, provided feedback, and approved the final manuscript.

Acknowledgements: This review is jointly conducted by the Blavatnik School of Government at the University of Oxford and Ecorys (research and consultancy company). We thank all of the PAG members who have agreed to contribute to this review. We also thank Professor Carolyn J. Heinrich, Professor Sir Julian Le Grand, Professor Nicholas Mays, and Professor Stéphane Saussier for agreeing to participate in a workshop and to provide methodological feedback.

References

Albertson, K., Fox, C., O'Leary, C., & Painter, G. (2018). *Payment by results and social impact bonds—Outcome-based payment systems in the UK and US*. Policy Press.

Bovaird, T., & Davies, R. (2011). Outcome-based service commissioning and delivery: Does it make a Difference? In S. Groeneveld, & S. Van De Walle (Eds.), *New steering concepts in public management* (Vol. 21, pp. 93–114). Emerald Group Publishing Limited. https://doi.org/10.1108/S0732-1317(2011)0000021011

Brown, A. (2013, March 5). *Disappointment by results*. Institute for Government. https://www.instituteforgovernment.org.uk/blog/disappointment-results

Büchter, R., Weise, A., & Pieper, D. (2020). Development, testing and use of data extraction forms in systematic reviews: A review of methodological guidance. *BioMed Central* (*BMC) Medical Research Methodology, 20*(259), 1-14. https://doi.org/10.1186/s12874-020-01143-3

Carter, E., FitzGerald, C., Dixon, R., Economy, C., Hameed, T., & Airoldi, M. (2018). *Building the tools for public services to secure better outcomes: Collaboration, prevention, innovation*. Government Outcomes. https://golab.bsg.ox.ac.uk/documents/176/BSG-GOLab-EvidenceReport.pdf

Clist, P. (2018). Payment by results in international development: Evidence from the first decade. *Development Policy Review*, *37*(6), 719-734. https://doi.org/10.1111/dpr.12405

Critical Appraisal Skills Program. (2020). *CASP qualitative studies checklist*. https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf.

Deakin, N., & Walsh, K. (1996). The enabling state: The role of markets and contracts. *Public Administration*, *74*(1), 33-47. https://doi.org/10.1111/j.1467-9299.1996.tb00856.x

Eijkenaar, F. (2012). Pay for performance in health care: An international overview of initiatives. *Medical Care Research and Review*, *69*(3), 251-276. https://doi.org/10.1177/1077558711432891

Farr, M. (2016). Co-production and value co-creation in outcome-based contracting in public services. *Public Management Review*, *18*(5), 654-672. https://doi.org/10.1080/14719037.2015.1111661

Fox, C., & Morris, S. (2021). Evaluating outcome-based payment programmes: Challenges for evidence-based policy. *Journal of Economic Policy Reform*, *24*(1), 61-77. https://doi.org/10.1080/17487870.2019.1575217

Fraser, A., Tan, S., Lagarde, M., & Mays, N. (2018). Narratives of promise, narratives of caution: A review of the literature on social impact bonds. *Social Policy & Administration*, *52*(1), 4-28. https://doi.org/10.1111/spol.12260

Fraser, M. W., Richman, J. M., Galinsky, M. J., & Day, S. H. (2009). *Intervention research: Developing social programs*. Oxford University Press.

Gomersall, J., Jadotte, Y., Xue, Y., Lockwood, S., Riddle, D., & Preda, A. (2015). Conducting systematic reviews of economic evaluations. *International Journal of Evidence-Based Healthcare*, *13*(3),170-178. https://doi.org/10.1097/XEB.0000000000000063

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). An introduction to systematic reviews. Sage Publications.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., deBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemanna. H. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Hannes, K. (2011). Critical appraisal of qualitative research. In J. Noyes., A. Booth., K. Hannes., A. Harden., J. Harris., S. Lewin & C Lockwood (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions.* Cochrane Collaboration Qualitative Methods Group.

Hannes, K., & Lockwood, C. (2012). *Synthesizing qualitative research*. John Wiley & Sons, Ltd.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane Handbook for systematic reviews of interventions* (2nd ed.). John Wiley & Sons.

HM Government (2011). Open public services white paper. Cabinet Office. https://www.gov.uk/government/publications/open-public-services-white-paper

Kirkham, J., Dwan, K., Altman, D., Gamble, C., Dodd, S., Smyth, R., Williamson, P., (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *The British Medical Journal (BMJ), 340*(7747), 637-640. https://doi.org/10.1136/bmj.c365

Lagarde, M., Wright, M., Nossiter, J., & Mays, N. (2013). *Challenges of payment-for-performance in health care and other public services – Design, implementation and evaluation*. London School of Hygiene and Tropical Medicine Research Online. https://researchonline.lshtm.ac.uk/id/eprint/2478790/1/Challenges%20of%20payment%20for%20performance%20in%20healthcare%20and%20other%20public%20services%20final.pdf

Mason, P., Fullwood, Y., Singh, K., & Battye, F. (2015). *Payment by results: Learning from the literature*. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2015/06/Payment-by-Results-Learning-from-the-Literature.pdf

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology*, *75(2016),*40-46. https://doi.org/10.1016/j.jclinepi.2016.01.021

Morse, A. (2015). *Outcome-based payment schemes: Government's use of payment by results*. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2015/06/Outcome-based-payment-schemes-governments-use-of-payment-by-results.pdf

Mustafa, R. A., Santesso, N., Brozek, J., Akl, E. A., Walter, S. D., Norman, G., Kulasegaram, M., Christensen, R., Guyatt, G., Falck-Ytter, Y., Chang, S., Murad, M., Vist, G., Lasserson, T., Gartlehner, G., Shukla, V., Sun, X., Whittington, C., Post, P., … Schunemann, H. (2013). The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *Journal of Clinical Epidemiology*, *66*(7), 736-742. http://doi.org/10.1016/j.jclinepi.2013.02.004

Oxman, A. D., & Fretheim, A., (2009). Can paying for results help to achieve the millennium development goals? Overview of the effectiveness of results-based financing. *Journal of Evidence Based Medicine*, *2*(2), 70-83. https://doi.org/10.1111/j.1756-5391.2009.01020.x

Page, M., Sterne, J., & Higgins, J. (2018). The ROB-ME (Risk Of Bias due to Missing Evidence) tool: A new tool for assessing reporting biases in evidence syntheses. *Cochrane Database of Systematic Reviews 2018*(9 Suppl 1). https://doi.org/10.1002/14651858.CD201801

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage Publications, Inc.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage Publications, Inc.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *The Journal of Health Services Research & Policy, 10*(1), 21-34. https://doi.org/10.1258/1355819054308530

Perrin, B. (2013). *Evaluation of payment by results (PBR): Current approaches, future needs: final report.* Department for International Development. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213938/payment-results-current-approaches-future-needs.pdf

Pollitt, C., & Talbot, C. (Eds.). (2003). *Unbundled government: A critical analysis of the global trend to agencies, quangos and contractualisation*. Routledge.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *The British Medical Journal* (*BMJ), 349*(7647). https://doi.org/10.1136/bmj.g7647

Sterne, J., Hernán, M., Reeves, B., Savović, J., Berkman, N., Viswanathan, M., Henry, D., Altman, D., Ansari, M., Boutron, I., Carpenter, J., Chan, A-W., Churchill, R., Deeks, J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y., Pigott, T., … Higgins, J. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *The British Medical Journal* (*BMJ), 355*(4919), 1-7. https://doi.org/10.1136/bmj.i4919

Sterne, J., Savović, J., Page, M., Elbers, R., Blencowe, N., Boutron, I., Cates, C., Cheng, H-Y., Corbett, M., Eldridge, S., Hernán, M., Hopewell, S., Hróbjartsson, A., Junqueira, D., Jüni, P., Kirkham, J., Lasserson, T., Li, T., McAleenan, A., . . . Higgins, J. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *The British Medical Journal* (*BMJ), 366*(l4898), 1-8. https://doi.org/10.1136/bmj.l4898

Suri, H (2011). Purposeful sampling in qualitative research synthesis, *Qualitative Research Journal*, *11*(2), 63-75. https://doi.org/10.3316/qrj1102063

The Joanna Briggs Institute (Ed.) (2014). *The Joanna Briggs Institute reviewers' manual.* The Joanna Briggs Institute.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*(45). https://doi.org/10.1186/1471-2288-8-45

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349-357. http://doi.org/10.1093/intqhc/mzm042

Appendices

Appendix 1

Draft Search Strategy. (Used for SCOPUS – searched from 1990 – present)

1. TITLE-ABS-KEY ("outcome*-base* contract*")
2. TITLE-ABS-KEY ("outcome*-contract*")
3. TITLE-ABS-KEY ("Result*-base* contract*")
4. TITLE-ABS-KEY ("outcome*-base* commissioning")
5. TITLE-ABS-KEY ("Outcome*-base* procurement")
6. TITLE-ABS-KEY ("social outcome*" W/5 "contract*")
7. TITLE-ABS-KEY ("social outcome*" AND "contract*")
8. TITLE-ABS-KEY ("performance-base* incentive*")
9. TITLE-ABS-KEY ("Result*-base* financ*")
10. TITLE-ABS-KEY ("Performance-base* financ*")
11. TITLE-ABS-KEY ("Sustainability-linked loan*")
12. TITLE-ABS-KEY ("Performance-base* contract*")
13. TITLE-ABS-KEY ("Performance-base* aid")
14. TITLE-ABS-KEY ("output-base* aid")
15. TITLE-ABS-KEY ("result*-base* aid")
16. TITLE-ABS-KEY ("Outcome*-base* financ*")
17. TITLE-ABS-KEY ("Performance base* transfer")
18. TITLE-ABS-KEY ("social" W/3 "public-private partnership*")
19. TITLE-ABS-KEY ("social" AND "outcome*" W/3 "public-private partnership*")

20. TITLE-ABS-KEY ("public-private mix")
21. TITLE-ABS-KEY ("social" W/3 "PPP")
22. TITLE-ABS-KEY ("social" AND "outcome*" W/3 "PPP")
23. TITLE-ABS-KEY ("contract* for health*")
24. TITLE-ABS-KEY ("payment-by-result*")
25. TITLE-ABS-KEY ("pay-by-result*")
26. TITLE-ABS-KEY ("Payment-by-results contract*")
27. TITLE-ABS-KEY ("pay*-by-outcome*")
28. TITLE-ABS-KEY ("Program-for-result*")
29. TITLE-ABS-KEY ("pay for success")
30. TITLE-ABS-KEY ("Payment-by-results contract*")
31. TITLE-ABS-KEY ("Pay*-for-outcome*")
32. TITLE-ABS-KEY ("Outcome-based payment*")
33. TITLE-ABS-KEY ("Outcome-base* financ*")
34. TITLE-ABS-KEY ("Social impact invest*")
35. TITLE-ABS-KEY ("Impact auction*")
36. TITLE-ABS-KEY ("impact bond" OR "impact bonds") OR TITLE-ABS-KEY ("social" W/4 ("impact bond" OR "impact bonds"))
37. TITLE-ABS-KEY ("social impact bond*")
38. TITLE-ABS-KEY ("Bonos de Impacto Social")
39. TITLE-ABS-KEY ("Development Impact Bond*")
40. TITLE-ABS-KEY ("Health Impact Bond*")
41. TITLE-ABS-KEY ("Social Impact Project*")
42. TITLE-ABS-KEY ("social impact partnership")
43. TITLE-ABS-KEY ("social impact" W/4 "partnership")
44. TITLE-ABS-KEY ("social success note*")
45. TITLE-ABS-KEY ("Contract" W/4 "service provider")
46. TITLE-ABS-KEY ("Alliance contract*")
47. TITLE-ABS-KEY ("performance-base*-grant*")
48. TITLE-ABS-KEY ("Outcome*-base*-grant*")
49. TITLE-ABS-KEY ("Targeted grant*")
50. TITLE-ABS-KEY ("Disbursement-linked-indicator*")
51. TITLE-ABS-KEY ("Performance-base*-condition*")
52. Combined search 1 – 51 (OR)
53. TITLE-ABS-KEY ("randomized controlled trial*")
54. TITLE-ABS-KEY ("randomised controlled trial*")
55. TITLE-ABS-KEY ("RCT*")
56. TITLE-ABS-KEY ("controlled clinical trial*")
57. TITLE-ABS-KEY ("interrupted time series analysis")
58. TITLE-ABS-KEY ("interrupted time series analyses")
59. TITLE-ABS-KEY ("randomis*" OR "randomiz*" OR "randomly")
60. TITLE-ABS-KEY ("pretest-posttest study" OR "pretesting" OR "pre-post tests" OR "quasi-experimental design" OR "quasi-experimental study" OR "quasi-experimental study design" OR "repeated measurement" OR "repeated measurements" OR "repeated measures" OR "time series")
61. TITLE-ABS-KEY ("meta analysis")
62. TITLE-ABS-KEY ("meta analyses")
63. TITLE-ABS-KEY (("systematic reviews"))
64. TITLE-ABS-KEY ("trial" OR "multicenter" OR "multi center" OR "multicentre" OR "multi centre")

65. TITLE-ABS-KEY ("evaluat*" OR "assess*" OR "impact" OR "measur*" OR "experience*" OR "perception*" OR "learning" OR "performance" OR "program*" OR "random*" OR "experiment*" OR "control*" OR "ebp" OR "evidence based practice" OR "outcome*")
66. TITLE-ABS-KEY ("intervention" OR "effect" OR "impact" OR "controlled" OR "control group" OR ("before" W/5 "after") OR ("pre" W/5 "post") OR ("pretest" OR "pre test")) AND TITLE-ABS-KEY ("posttest" OR "post test" OR "quasiexperiment*" OR "quasi experiment" OR "evaluat*" OR "time series" OR "time point" OR ("repeated measure" OR "repeated measurement" OR "repeated measurements" OR "repeated measures") OR "generalized estimating equation" OR "generalised estimating equation")
67. TITLE-ABS-KEY ("pragmatic clinical trial*")
68. TITLE-ABS-KEY ("non-randomized control* trial*")
69. TITLE-ABS-KEY ("non-randomised control* trial*")
70. TITLE-ABS-KEY ("controlled before-after studies")
71. TITLE-ABS-KEY ("cost*" OR "value for money" OR "VFM" OR "economic analysis" OR "economic analyses" OR "cost benefit analysis" OR "CBA" OR "transaction cost*")
72. TITLE-ABS-KEY ("interview*" OR "focus group*" OR "observation*" OR "case stud*" OR "reflective diar*")
73. Combined search 53 – 72 (OR)
74. Combined search (52 AND 73)

Appendix 2

Full text review – hierarchical exclusion criteria

Circulated version 19th July 2021

Best-practice review guidance suggests that at full-text stage an 'exclusion reason' is provided by team members.[2] To ensure papers are tagged with an appropriate and consistent exclusion reason the following questions should be asked, in chronological sequence, selecting the first appropriate 'exclusion reason'.

Step 1: Open Full Text of paper under consideration. Click "view full text" to see the PDF/s that have been attached to the relevant record.[3]

Practical check: Are you able to read the paper? Is it in a language that you are comfortable reading? If no, add a note explaining the limiting feature and if possible, write to a member of the review team who is able to review the paper. A separate document lists language skills of review team members.

Step 2: Work through the following questions chronologically, selecting the first appropriate response as the reason for excluding the study.

[2] A basic overview of common inclusion/exclusion criteria is available here: https://unimelb.libguides.com/c.php?g=492361&p=3368110

[3] If the record shows "Add Full Text", it means that the full text has not yet been added. Sort by 'Author' to see the most available list of full text papers.

1. **Wrong study design – i.e., paper does not provide an empirical contribution**

To be included, studies need to present primary research, original analysis of secondary data sources OR offer original, independent synthesis. Examples of relevant quantitative designs include experimental or quasi-experimental designs, pre-post test, regression analysis or descriptive statistics. Examples of appropriate qualitative designs include in-depth interviews, focus groups, observation, reflective diaries and/or case study methodologies that explore perceptions of participants on the application of social outcomes contracting. Examples of relevant economic analysis include any form of cost effectiveness or cost-benefit analysis and may also include information on transaction costs. Examples of independent syntheses include systematic reviews and formal literature reviews; these should be included *only if they describe the review method used.*

Q1: Does the paper provide an empirical contribution? An evidence contribution must describe the ***primary research or secondary data analysis*** *(quantitative, qualitative or economic) or provide an* ***independent synthesis****.*

For papers that <u>do not provide an empirical contribution</u>

Click Exclude > Exclusion reason > 'Absent empirics'.

Advice on what qualifies as 'independent synthesis': As stated above, independent syntheses must provide a *description of the method* to meet the empirical criteria of this review. General discussions of outcomes-based contracts may well include some synthesis of existing data – but if no method is described, these should be excluded as 'absent empirics'.

Papers that present purely ex ante models (i.e., based on forecasts) which feature <u>*no observed data*</u> associated with outcomes contracting and papers that describe themselves as 'feasibility studies' should be excluded under 'wrong study design'.

Click Exclude > Exclusion reason > 'Wrong study design'.

> NOTE: *At this screening stage we do not need to pass comment on the quality or appropriateness of the methods used (provided that they make a contribution on the lines described above).*

For papers that describe systematic reviews, coherence between the inclusion criteria of the paper under consideration and our own review should be assessed. Where a systematic review describes its method and where all included studies meet the inclusion criteria for our SOC review (as far as can be assessed on the basis of the paper), the paper meets our empirical requirement, and you should proceed to question 2. For example, a systematic review of health impact bonds would meet our empirical requirement. Where a systematic review includes one or more paper(s) that would *not* meet our inclusion criteria the paper should be excluded (since at least some of the studies are out of scope for our review). For example, a systematic review of Results based Financing in education, where some of the studies investigate country government level incentives, should be excluded.

Click Exclude > Exclusion reason > 'Misaligned systematic review'

2. **Wrong intervention – Contracted agent**

To be included, papers need to investigate a form of social outcome contracting at an organisational level. The contracted agent must be an organisation from the not-for-profit or private sector. Therefore, e*xclude public sector*; exclude country governments; local governments; individual people. By contract we also mean grant agreements.

*Q2. Is an **independent organisation** from the not-for-profit or private sector party to a **contract?***

Agent: The agent is the party being paid to deliver results by the principal. Supporting questions to guide full text review: *Who gets paid? Is it possible to name the organisation or type of organisation that is delivering under the outcomes contract?*

If it is not possible to identify i) the presence of a contract; or ii) a delivery 'agent' organisation in the not-for-profit or private sector…

Then Click Exclude > Exclusion reason > 'Wrong intervention – CONTRACTED AGENT'.

3. **Wrong intervention – Outcome Measure with financial incentive**

To be included, papers need to investigate an outcomes contract with financial implications for the non-achievement of social or environmental outcomes. This could be in the form of an 'outcomes payment'; a bonus for the achievement of outcomes or negative financial implications (e.g., clawback) for poor performance against outcome indicators.

*Q.3 Is a **financial incentive** attached to the achievement of a pre-agreed social or environmental **outcome measure**?*

Measure: The measure describes the indicator or set of metrics that underpin the payment-by-results contract. Supporting questions: *What is the pre-agreed measure? Are the measures described as 'outcome' measures? And is a financial incentive attached?*

Make sure the outcomes discussed are actually included in the contract and are not solely the researchers' chosen measurements. Use a generous interpretation of 'outcome' (e.g. these can be individual level care quality. Equity 'outcome' indicators should be included). But <u>be strict on ensuring that a financial incentive is attached to performance against a specified measure.</u>

At a person level (i.e., programme participant) there is potential to classify outcomes against a theory of change e.g., outputs are tangible goods and services that are delivered by the project – e.g. how many children are vaccinated, how many textbooks are distributed etc. the implementing agency has direct control over these outputs. According to Duvendack (2017, p. 17): "Outcomes build on outputs, they are realised once beneficiaries have used the project outputs".

Classification of system level 'outcome' indicators is more challenging. Indicators relating to service quality and equity should be included.

ACTION: Write the outcome measure(s) that is incentivised in the contract in the notes field.

(Independent syntheses may discuss a number of incentivised outcomes from different contracts. In such cases, please describe at least one of the outcome measures represented in the study.)

If the paper discusses Social Outcomes Contracting in general, but a specific nameable example of an 'outcome measure' is not identified...

Then Click Exclude > Exclusion reason '**Wrong intervention – broad SOC'**

If outcome measures are not linked to financial incentives...

Then Click Exclude > Exclusion reason '**Wrong intervention – No Financial Incentive'**

If you get to this stage without excluding the study – that is, 1) it offers empirics; 2) has an independent organisation as the contracted agent and 3) a financially incentivised outcome measure in the contract then…

We would also like to add tags to identify different types of study to help us allocate for full text data extraction and critical appraisal:

- Quantitative analysis with relevant non-SOC comparator group
- Quantitative analysis without comparator group (includes any regression analysis, descriptive statistics)
- Qualitative analysis (any qualitative method including ethnographic participant observation, interviews, documentary analysis)
- Economic analysis (any stated costs ($); value for money or cost-benefit analysis)
- Independent synthesis, literature review or systematic review
- More than one form of research design (Note – more than 1 tag can be applied to each paper. Apply as many as needed to comprehensively describe the nature of the paper and also click 'more than one research design')

And finally, click 'include'.

Appendix 3

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item	Location where item is reported
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	Page 1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	N/A
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	Page 1
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	Page 1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	Page 10
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	N/A
Support:			
Sources	5a	Indicate sources of financial or other support for the review	Page 10
Sponsor	5b	Provide name for the review funder and/or sponsor	Page 10
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	Page 10
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	Page 2
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	Pages 4-5
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	Pages 4-5
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	Pages 3-4

Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	Appendix 1
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	Page 5
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	Page 5
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	Pages 5-6
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	Pages 4, 5, 6
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	Page 5
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	Page 6
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	Pages 7-8
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	Pages 7-8
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	Page 8
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	Pages 7
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	Page 6
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	Page 9

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

Appendix 4

List of Policy Advisory Group (PAG) Members

Name	Organisation	Country
Radana Crhova	Foreign, Commonwealth & Development Office (FCDO)	United Kingdom
Simon Marlow	Department for Work & Pensions (DWP)	United Kingdom
Dominique Be	European Commission	Belgium
Vera Barracho	Portugal Social Innovation	Portugal
Juliana Sancehz Calderon	Prosperidad Social	Colombia
Jose Ramon Romero Pineda	Prosperidad Social	Colombia
Frank Tulus	Global Affairs Canada	Canada
Inga Afanasieva	World Bank	United States
Elaine Tinsley	World Bank	United States
Karine Bachongy	International Finance Corporation	Austria
Jonathan Ng	U.S. Agency for International Development	United States
Lars Stein	Foreign Department of Foreign Affairs FDFA	Switzerland
Elina Järvelä	United Nations Development Programme (UNDP)	Finland
Samantha Magne	The National Lottery Community Fund	United Kingdom
Miranda Lee	Office of Social Impact Investment (OSII), NSW Treasury	Australia
Rajnish Prasad	UN Women	Thailand
Outi Valkama	City of Tampere – Government Agency	Finland
Naomi Ishida	Cabinet Office, Government of Japan	Japan
Jari Pekuri	Hämeenlinnan kaupunki - The City of Hämeenlinna	Finland
Tarja Keltto	Vantaan kaupunki - Vanda stad - City of Vantaa	Finland
Giulio Pasi	European Commission	Spain
Jeffrey Matsu	The Chartered Institute of Public Finance & Accountancy	United Kingdom

*The list above only contains the details of PAG members who provided their consent to be publicly listed.