

What policies and interventions have been strongly associated with changes in in-country income inequality?

A Systematic Review Protocol

Protocol written by Edward Anderson, Lucio Esposito, Maren Duvendack, Maria Ana Jalles d'Orey, School of International Development (University of East Anglia)

EPPI-Centre
Social Science Research Unit
UCL Institute of Education
University College London

December 2014

PROTOCOL

The authors are part of the School of International Development, University of East Anglia and were supported by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre).

This protocol should be cited as: Anderson E, Esposito L, Duvendack M, Jalles d'Orey MA (2014) What policies and interventions have been strongly associated with changes in in-country income inequality? A Systematic Review Protocol. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London

© Copyright

Authors of the systematic reviews on the EPPI-Centre website (<http://eppi.ioe.ac.uk>) hold the copyright for the text of their reviews. The EPPI-Centre owns the copyright for all material on the website it has developed, including the contents of the databases, manuals, and keywording and data extraction systems. The centre and authors give permission for users of the site to display and print the contents of the site for their own non-commercial use, providing that the materials are not modified, copyright and other proprietary notices contained in the materials are retained, and the source of the material is cited clearly following the citation details provided. Otherwise users are not permitted to duplicate, reproduce, re-publish, distribute, or store material from this website without express written permission.

Contents

1. Background.....	3
1.1 Description of the problem	3
1.2 Description of the interventions	4
1.3 How the interventions might work	4
1.4 Importance of the review	10
2. Objectives of the review.....	12
3. Methods.....	14
3.1 Criteria for including studies in the review	14
3.2 Search methods	20
3.3 Data collection and analysis	24
3.4 Data synthesis	30
4. Timeline	34
References	35
Appendices	41
Appendix 1: Authorship of the review	41
Appendix 2: Country classifications	44
Appendix 3: Concepts for search strategy in English, Portuguese and Spanish	47
Appendix 4: Search strategy checklist (provided by DFID)	48
Appendix 5: Data extraction tools.....	50

1. Background

1.1 Description of the problem

The issue of inequality has been a key issue in development for several decades now. Since the 1970s, a large literature has emerged which documents the many adverse effects of inequality on socio-economic outcomes, including investment and economic growth, poverty, health and well-being, crime, conflict and social cohesion - see for example Easterlin (1974), Williams (1984), Alesina and Perotti (1996), Ravallion (1997), Barber (2001), Luttmer (2005), Eibner and William (2005), Veenstra (2005), Subramanian and Kakawi (2006), Clark et al. (2008), Gravelle and Sutton (2009), Wilkinson and Pickett (2009), Stiglitz (2013) and Ostry et al. (2014). In addition, all societies share a basic, intrinsic concern for equity and justice, and high levels of inequality often conflict with those notions - as for example when life chances or opportunities differ significantly between groups defined by gender, inherited wealth, ethnicity or other accidents of birth (World Bank 2005).

Inequalities are observed in various dimensions, including income, wealth, health, education and access to employment, in each case within and between countries. This systematic review focuses in particular on inequalities in income within countries. Income is of course only one dimension of welfare, and does not always correlate highly with other important dimensions such as health and well-being. However, it is sufficiently important to merit detailed examination in its own right. In addition, although globalisation makes inequality across countries more and more relevant, within-country inequality is arguably still the most important because a person's own country represents the main social milieu that they interact with and compare to. It is also the dimension of inequality that national governments typically have more direct influence on.

Evidence on the extent of income inequality within low and middle income countries has become increasingly available since the 1970s, with the increasing availability of good quality, nationally-representative household surveys measuring income, or its commonly used proxy, consumption expenditure). Currently, there is no clear overall trend: there are almost as many countries where income inequality is increasing as there are countries where it is decreasing (Olinto and Saavedra 2012). One of the biggest rises in income inequality has occurred in China, where the Gini coefficient increased from around 30 in the early 1980s to around 45 in 2005; income inequality also increased in Vietnam, Indonesia and Mongolia (*ibid*).¹ But in Latin America, the dominant trend over the past decade has been one of falling income inequality - including Brazil, Argentina, Mexico, Peru and Ecuador (Lustig et al. 2013a). Note however that the verdict of falling inequality in Brazil does not hold if an absolute notion of inequality is adopted, where actual income gaps between individuals are taken into account (Anderson and Esposito 2014).

Nevertheless, income inequality remains high in a large number of developing countries. Over 50 low and middle income countries have Gini coefficients of income inequality exceeding 40, above which the potential to undermine progress in key development

¹ The (relative) Gini coefficient is one of the most widely used measures of income inequality; it varies between 0, indicating perfect equality of income, to 100, indicating perfect inequality. Other measures of income inequality are discussed further below.

outcomes, and/or to conflict with basic notions of equity and fairness, is considered significantly greater (e.g. Birdsall 2007). For this reason, there is a clear demand from policy-makers in national governments and international organisations for accurate, reliable and up-to-date evidence as regards which policies and interventions can be used to reduce income inequality, and also which policies and interventions may (in the absence of complementary, offsetting measures) raise income inequality. The overall aim of this systematic review is to identify and synthesise this evidence in order to meet these demands.

1.2 Description of the interventions

Many of the systematic reviews that have been carried out in the field of international development to date have focused on quite specific policies or interventions; examples include microcredit programmes, conditional cash transfers, school-feeding programmes, and farmer field schools (White and Waddington 2012, Table 1). Others have focused on fairly narrowly-defined sets of policies or interventions, defined by the sector of the intervention and/or by the specific purpose of the interventions; examples include land property rights interventions for increasing productivity, and water, Sanitation and Hygiene (WSH) interventions to combat childhood diarrhoea (ibid).

This systematic review, by contrast, does not focus on a specific policy or intervention, nor on a narrowly-defined set of policies and interventions. Instead, it includes *any* government policy or intervention which has been shown to have had an impact on income inequality. The sorts of policies and interventions which may affect income inequality are broad, including land reform; social policy; trade, industrial and agricultural policy; macroeconomic policy; government spending on education, health and infrastructure; taxes and transfers; and anti-discrimination legislation and affirmative action (see Section 1.3 below).

The broad nature of our review gives rise to two main dangers. The first is that the amount of literature relevant to the review will be too large, and not possible to review and synthesise adequately within the time available. The second is that the range of policies and interventions covered by the review will be too diverse, preventing meaningful and interesting comparisons of the effects of similar types of policies and interventions across different countries and contexts. We aim to respond to these dangers by mapping the relevant literature, prior to carrying out the synthesis. The mapping exercise will identify all of the evidence relevant to the review question, but the synthesis will then be applied to a sub-group of the studies identified in the mapping stage. This is discussed further in Section 3 below.

1.3 How the interventions might work

This section sets out the conceptual framework (or ‘theory of change’) that will be used to guide the review. We begin by distinguishing five different concepts of income: market income, net market income, disposable income, post-fiscal income and final income. These concepts are taken from Lustig (2011) and closely match those used by other researchers in this field, including Hemming and Hewitt (1991), Goni et al. (2011),

What policies and interventions have been strongly associated with changes in in-country income inequality?

Castelletti (2013) and Ostry et al. (2014).² They are shown diagrammatically in Figure 1, and can be described as follows:

- **Market income** refers to private income from all sources *before* government taxes and transfers. It is equal to the sum of a household's ownership of each productive asset (e.g. land, labour and human capital) multiplied by the return to each asset, plus private transfers (e.g. remittances);
- **Net market income** is equal to market income minus personal income taxes and employee contributions to social security;
- **Disposable income** is equal to net market income plus government direct transfers, e.g. pensions, unemployment insurance, and social cash transfers;
- **Post-fiscal income** adjusts disposable income for the effects of indirect taxes (e.g. VAT) and indirect subsidies (e.g. subsidised food or fuel items);
- **Full income** is equal to post-fiscal income plus the value of in-kind transfers and public services received by the household (e.g. health and education, water and sanitation), minus any payments made for the use of these services.

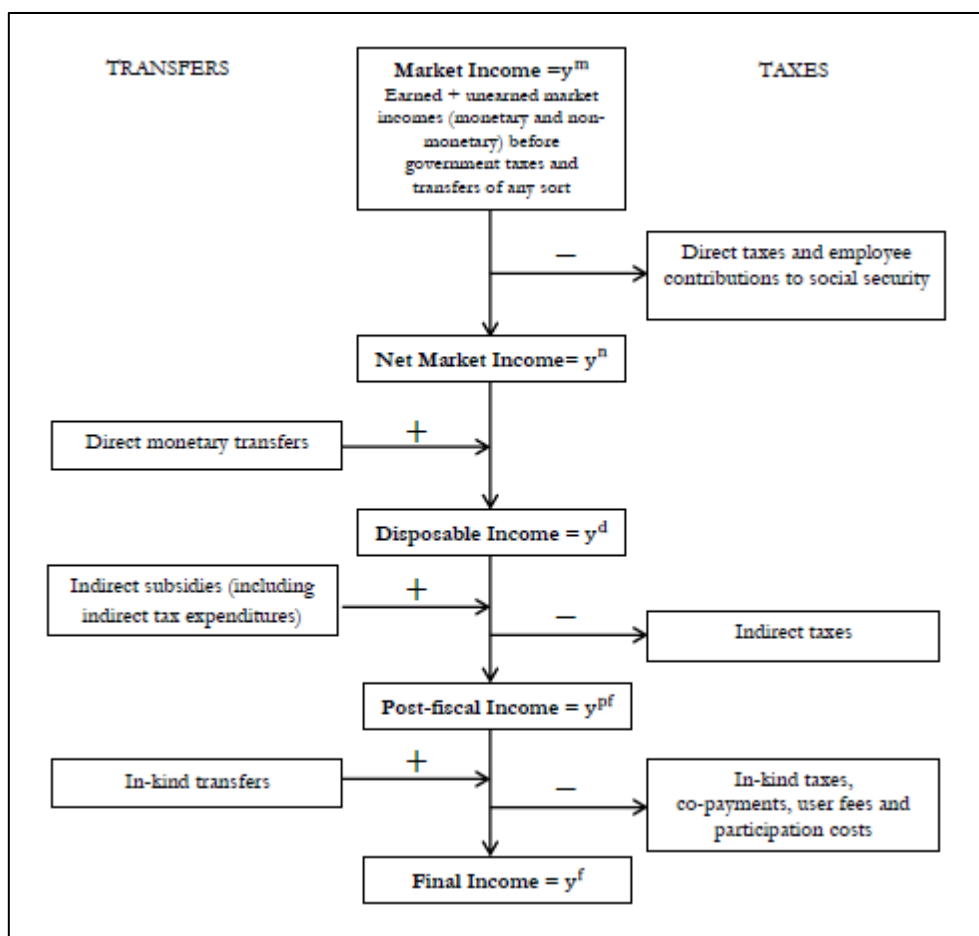
As we will see, different studies use different concepts of income and we do not restrict our attention to any one concept. The concept of 'full income' is probably the most complete from a theoretical point of view but in practice it is the most difficult to calculate. Typically, this is done calculated by combining information in household surveys about the usage of public services - for example, whether children are enrolled in school, whether household members use public health services - with information in the public accounts regarding the average cost of providing such services (Demery 1997, Schwarz and Ter-Minassian 2000). This raises significant challenges: for example, how to take into account differences in the quality of public services, and also the extent to which households themselves value the services being provided. It also fails to captures only some publicly-provided goods and services, and ignores various others - such as law and order or environmental amenities.

Most studies therefore tend to focus on one of the other concepts of income - in particular, market, disposable and post-fiscal income. Most cross-country datasets on income inequality (e.g. Deninger and Squire 1994) tend to include some estimates of inequality based on market income and some based on disposable income. The latter corresponds most closely to a household's purchasing power over private goods and services. It is a fairly comprehensive measure, which includes income from all sources, e.g. wages, profits, both private and government transfers, and adjusts for both direct and indirect taxes. Note also that market income should in theory include items that are both produced and consumed by the household, e.g. staple food items and housing,

² The specific terms used to describe each concept sometimes differ, but the definitions remain the same. For example, Hemming and Hewitt (1991) use the term 'original income' rather than 'market income', and 'post-tax income' rather than 'disposable income'; while Goni et al (2011) use the term 'post-tax' income rather than 'post-fiscal' income. Note also that Ostry et al (2014) distinguish between 'market inequality', which refers to inequality in market income, and 'net inequality', which refers to inequality in disposable income.

particularly relevant for many low and lower middle-income countries - even though this also raises practical challenges.

Figure 1: Definition of income concepts



Source: Lustig (2011)

Given these definitions of income, the conceptual framework or ‘theory of change’ that will be used to guide the review is shown in the form of a phase diagram (Figure 2). Government policies and interventions can affect inequality in market income via three main channels, namely:

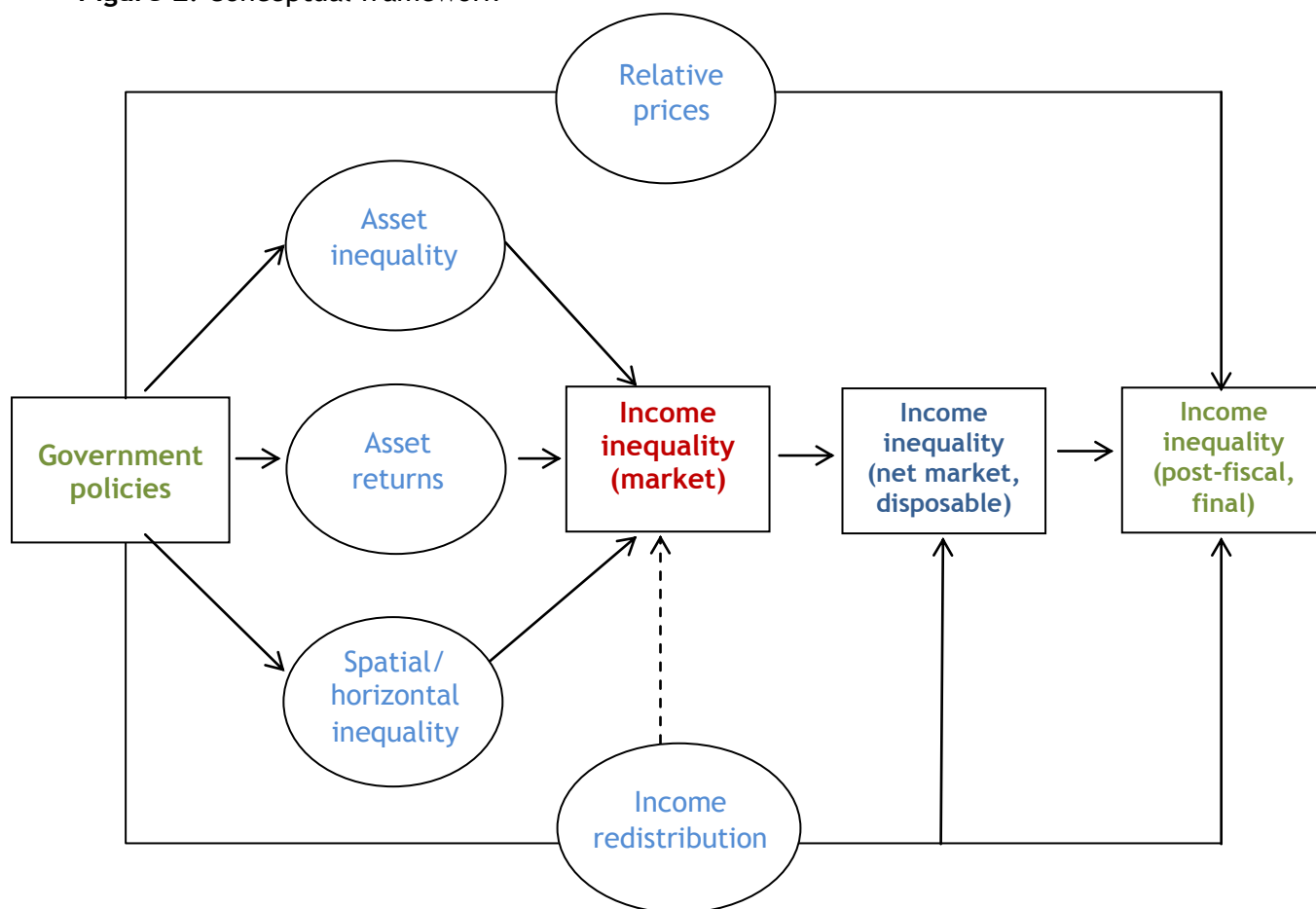
- 1) by altering the distribution among individuals and households in the ownership of the assets that generate income, e.g. labour, human and physical capital, land
- 2) by altering the relative returns to the assets that generate income, e.g. the ratio of skilled relative to unskilled wages
- 3) by altering inequalities in the returns to an asset across regions or population sub-groups, e.g. differences in the wages of men and women for any given level of skill

Government policies can also affect inequality in net market, disposable, post-fiscal and/or final income directly, by redistributing income via taxes and transfers. To the extent that taxes and transfers affect behaviour, and therefore have second-round, ‘knock-on’ effects, they may also affect inequality in market income. Government policies

What policies and interventions have been strongly associated with changes in in-country income inequality?

can also affect inequality in net income by altering the prices of goods and services which are consumed in different combinations by households of different income levels.

Figure 2: Conceptual framework



Note: Inequality in final income can be thought of as the final outcome of interest, although data on this outcome are not always available. Inequality in market, net market, disposable, and post-fiscal income are intermediary outcomes for which data availability tends to be higher.

a. Asset inequality

High income inequality often reflects high inequality in the ownership of important productive assets. There is for example a close correlation across countries between inequality in income and inequality in land ownership (Goni et al. 2011); there is a similar close correlation with inequality in years of schooling (ibid). Asset inequality tends to have deep roots in history, and often changes only gradually over time, except during large scale social upheavals (e.g. wars, revolutions). Governments can however take various steps to redistribute assets and broaden asset ownership.

So-called 'static' asset redistribution seeks to redistribute an asset whose overall supply is relatively fixed. An obvious example here is land reform, including state-centred, compulsory redistribution as well as voluntary, 'market assisted' programmes. By contrast, 'dynamic' asset redistribution aims to promote faster asset accumulation individuals and groups who are relatively asset-poor, leading to lower inequality in asset ownership over time. A wide range of interventions in the areas of health, education and social policy

have this aim. One example is a cash transfer programme, which aims to promote investment in education and human capital by children in low-income households who might otherwise be prevented from going to school, for financial reasons. By increasing access by the relatively poor to financial services (e.g. credit, savings), microfinance programmes can also broaden the ownership of business and financial assets.

Note that policies aimed at asset redistribution will have different impacts on income inequality, depending on how important the asset is in the functional distribution of income.³ A land redistribution programme which is highly effective in terms of reducing inequality in land ownership may only have a limited impact on income inequality, if the returns to land account for only a small share of national income.

b. Asset returns

The ownership of some assets tends to be distributed much more evenly among the population than others. This means that changes in the returns to assets can impact income inequality, even when inequality in the ownership of those assets does not change. For example, ‘human capital’ is distributed more unequally than ‘labour’, so a rise in the return to human capital (or skilled labour) relative to (unskilled) labour will tend to increase income inequality, and vice versa. There is recent evidence for this in practice - in particular, a decline in the earnings of more educated labour is considered to be a key driver behind the fall in income inequality in many Latin American countries during the 2000s (Lustig et al. 2013a,b).

Government policies can affect the returns to different assets via both demand-side and supply-side channels. On the demand side, significant economy-wide reforms in the areas of trade, industrial, agricultural and macro-economic policy can all affect the demand for different assets. For example, the process of trade liberalisation (reductions in tariffs and other barriers to trade) in many middle income countries during the 1980s and 1990s (e.g. Mexico, Chile, Colombia) is considered to have raised the demand, and hence the wages, of skilled relative to unskilled labour (e.g. Wood 1997, Robertson 2000, Gindling and Robbins 2001). Labour market policies can also have an effect - employment subsidies or public works programmes, for example, can increase the demand for less-skilled labour, increasing their wages relative to other workers, and tending to narrow income inequality. Minimum wage legislation can also have this effect, although in this case the risk is that wage gains are offset by reductions employment.

On the supply side, government steps to increase the supply of assets also affect inequality. The most obvious example here is government investment in education and training programmes, which tends to increase the supply of skilled relative to unskilled labour, tending to reduce the relative wages of skilled labour. This is considered to have been one factor behind the recent decline in the relative earnings of more educated labour in Latin America for example (Lustig et al. 2013b). Government investment in health can have a similar effect, by encouraging individuals to invest more in their education - since the return to investment in education tends to be higher, the longer is a person’s life expectancy.

³ The functional distribution of income indicates the share of total national income made up by each productive asset.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Note that when assessing the return to an asset, particularly labour, allowance should also be made for the risk (and duration) of unemployment. In cases where labour markets are subject to significant institutional rigidities (e.g. minimum wages, trade union wage bargaining), changes in the demand for skilled relative to unskilled labour will be reflected mainly in terms of changes in unemployment rates rather than wages.

c. Spatial and horizontal inequalities

Economic theory often assumes that the returns to an asset (of given quality) will be equalised within countries, because of factor mobility. Since workers move in search of employment opportunities, and firms move in search of lower wages, wages and profits should gradually be equalised. However, there are various natural barriers to mobility (e.g. relocation costs for businesses and migrants) and governments also impose their own artificial barriers - e.g. in China where the *hukou* system has long been regarded as a major disincentive to rural-urban labour migration.

Because of these barriers, the returns to each factor of production tend to differ within countries, and these 'spatial inequalities' add to overall income inequality. Governments can however take steps to reduce spatial inequalities, either by investing in domestic transport infrastructure, thereby reducing the natural barriers to the mobility of factors and goods and services, and by liberalising existing legal or policy barriers to factor mobility.

The returns to an asset (of given quality) may also differ between population sub-groups, because of discrimination. Such 'horizontal' inequalities also add to overall income inequality. Governments can take various steps to tackle discrimination, and the horizontal inequalities in income, including anti-discrimination legislation and affirmative action.

d. Income redistribution

The fourth way in which governments can reduce inequality is to redistribute income via taxes and transfers. This can be achieved by a progressive income tax and transfer structure, under which a) people with higher incomes pay a greater share of their income in tax than those on lower incomes, and b) people with lower incomes receive a greater share of their income in transfers than those on higher incomes. It can also be achieved by the use of indirect taxes and subsidies, by lower rates of tax (or subsidies) on goods which are consumed more intensively by lower-income households (e.g. basic food items).

The effect of fiscal redistribution on income inequality has been estimated for many developing countries (e.g. Devarajan and Hossain 1998; Engel et al. 1999; Baer and Galvao 2008, Goni et al. 2011). The evidence appears to suggest that income redistribution via *direct* taxes is not very significant, one of the reasons being that direct tax revenues tend to be smaller relative to GDP in developing countries, particularly low income countries. Institutional and administrative tax reforms, which aim to increase overall revenue from direct taxes, can therefore make an important contribution to income redistribution, by increasing the impact of a progressive income tax structure. *Indirect* taxes, which account for a much larger share of total revenue for most developing countries, appear to slightly regressive in terms of their impact (Goni et al. 2011). Targeted transfer programmes (e.g.

conditional cash transfers) appear to be strongly progressive, but despite their expansion in recent years they typically remain small relative to GDP, so that their overall redistributive impact remains limited (Lindert et al. 2006).

An important issue in this context is the behavioural effects of taxes and transfers. There is by now a lot of evidence that taxes and transfers do impact behaviour; for example, Sahn and Alderman (1996) find that rice subsidies reduced labour supply in Sri Lanka, while Cox and Jimenez (1995) find that government transfers displaced private transfers in the Philippines. Typically, some form of economic modelling is required to take such effects into account when assessing the impacts of taxes and transfers on inequality. These range from relatively simple models which take into account a limited number of behavioural responses considered to be most important, to more complex, computable general equilibrium (CGE) models. In theory, CGE models should provide more accurate estimates; however, they rely on the accuracy of the equations and parameters which are used to construct them. They are also computationally complex and the results are often sensitive to modelling choices.

e. Relative prices

Indirect taxes and subsidies affect households by altering the real purchasing power of their disposable income (referred to here as ‘post-fiscal’ income), rather than disposable income itself. Many other government policies may affect income inequality in this way. More generally, any policy which reduces the prices of goods which are consumed more intensively by lower-income households (e.g. basic food items) will tend to reduce inequality in post-fiscal income, and vice versa. This will be in addition to the effects that a policy has on market income.

For example, Porto (2006) distinguishes two main channels by which the 1992 MERCOSUR trade agreement affected income inequality in Argentina. The first, referred to as the ‘income effect’, is the effect of the agreement on wage levels, which affects inequality in market income. The second, referred to as the ‘consumption effect’, is the effect of the agreement on consumer prices, which affects inequality in post-fiscal income. According to the results of the analysis, these two effects were offsetting, although in overall terms the agreement reduced income inequality.

1.4 Importance of the review

As noted in Section 1.1, inequalities in income and other dimensions are a key policy issue in a large number of low and middle-income countries around the world. There is a clear demand from policy-makers for accurate, reliable and up-to-date evidence as to the sorts of policies and interventions that can reduce income inequality, combined with better understanding of the way in which these policies work, in different contexts.

There is of course already a very large literature on income inequality and development. A large and still increasing number of studies have addressed both the determinants of inequality and the effects of inequality on development outcomes (see references in Section 1.1 above). There have also been various reviews of this literature, and various landmark publications by the World Bank and other international organisations, which have sought to identify and summarise key findings from the literature, and highlight their

What policies and interventions have been strongly associated with changes in in-country income inequality?

implications for policy. Examples include the World Bank report, *Redistribution with Growth*, published in 1974, to the more recent World Development Report on the theme of *Equity and Development*, published in 2005.

Despite this very large literature, there is still arguably no clear consensus about the sorts of policies and interventions that governments can use to reduce income inequality - particularly when taking into account the indirect, knock-on effects, e.g. of fiscal redistribution via taxes and transfers. There is also relatively little understanding of the extent to which different policies work more effectively in some contexts rather than others. In the words of Killick (2002: 5):

“A substantial number of interventions [for reducing inequality] have been identified ... but none are new and not all are of great proven potency. But it would be wrong to conclude that there is little that can be done, as there are many examples of diminishing income disparities.”

For these reasons, there appears to be a clear need for a systematic review that directly addresses the issue of income inequality - and the particular policies and interventions that have been associated with changes in income inequality, in low and middle-income countries.

2. Objectives of the review

The objectives of our review are as follows:

- 1) to map the available evidence that seeks to evaluate or better understand the effects of government policies and interventions on income inequality, in low and middle income countries;
- 2) to establish whether any particular types of policies or interventions tend to reduce or increase income inequality on average: in other words, whether there are any consistent and generalisable findings or results across contexts and methods;
- 3) to explain heterogeneity in the estimated effect of such policies or interventions, across countries, regions or over time ('structural' heterogeneity) or research methods used ('method' heterogeneity);
- 4) to understand better the processes and mechanisms through which government policies and interventions affect income inequality.

The first aim is to map the research field. As noted in Section 1.2, mapping is an important component of this review, given the very broad nature of the underlying question. The mapping exercise will be used to identify all of the evidence relevant to the review question, and to categorise the evidence according to key descriptive information, namely:

- the country (or countries) of focus
- the type of policy or intervention
- the method(s) used to assess the impact of the policy or intervention on income inequality.

Mapping the research field, as a prior stage to synthesis, is an important part of many systematic reviews. In the words of Gough et al. (2013: 16):

“The studies contained within a research field may be too numerous or heterogeneous for meaningful synthesis; it might be methodologically too difficult or just take too much time. The map provides an opportunity to select a sub-group of studies for synthesis. This can involve undertaking a single synthesis based on a narrowed review question and set of inclusion criteria; or undertaking a series of syntheses. ... Syntheses can also be restricted to studies employing specific research methods.”

Mapping is also a useful output in its own right:

“Systematic maps of research fields can also highlight gaps in research. [They] can be used to compare policy and practice on the ground with what has been studied in research; they may reveal that only a specific sub-set of policy and/or practice has been studied.”(ibid).

What policies and interventions have been strongly associated with changes in in-country income inequality?

The second and third aims both relate to the synthesis of the evidence. We aim to establish whether any specific types of policies tend to reduce or increase income inequality on average, and to explain any heterogeneity in the estimated effects of particular policies on income inequality, by structural characteristics or by research method. This will be done using a combination of meta-analysis and meta-regression, applied to studies which use an appropriate counterfactual in assessing the impact on income inequality - namely ex-post quasi-experimental studies and ex-ante simulation studies. Of course, meta-analysis and meta-regression methods can only be applied if there is a sufficiently large body of comparable studies which all relate to a particular type of policy or intervention. The results of the mapping, which are described in a separate document attached to this Protocol, indicate that there is a sufficiently large body of comparable studies on the effects on income inequality of fiscal policy interventions (i.e., government tax and spending policies) and trade policy interventions (e.g. import tariffs and quotas), which can be subjected to meaningful synthesis using meta-analysis and meta-regression methods.

The final aim of the review is to understand better the processes and mechanisms through which government policies and interventions affect income inequality. This will involve synthesising the results of detailed case studies of income inequality in particular low or middle income countries, or particular regions within such countries. These studies allow us to explore in detail the various assumptions in our conceptual framework about the ways in which government policies affect inequality, and to identify and explore any unanticipated effects.

3. Methods

3.1 Criteria for including studies in the review

3.1.1 Types of studies

Study designs: We will include studies using any one of four main research methods:

- a) ex-post quasi-experimental studies, e.g. cross-country econometric analysis
- b) ex-ante simulation studies, e.g. CGE modelling
- c) quantitative case studies using decomposition analysis
- d) qualitative case studies, which draw on primary data, e.g. focus group discussions, semi-structured interviews.

The vast majority of *ex-post observational studies* are econometric studies, in which a measure of income inequality is the dependent variable and the explanatory variables include one or more policy variable. Some studies just use correlation analysis rather than regression analysis; we do not exclude these studies from the review although they are awarded a higher risk of bias in the meta-analysis.

Ex-ante simulation studies encompass a variety of methods and approaches. They all have in common that they analyse the impact of government policies or government spending on the distribution of income. This is done via a counterfactual, and a set of assumptions about whether and if so how economic agents respond to policies and/or spending - combined with actual empirical data for a particular country or economy. Ex-ante simulation studies include **fiscal incidence analysis** - this is the general term used for research that tries to understand or assess how government fiscal policies - in particular, those related to the revenue and expenditure sides of the budget - affect the distribution of income (see Martinez-Vasquez 2004). This includes:

- **tax incidence analysis** - i.e. analysis of who ultimately bears the burden of government taxes. The burden (also called 'economic incidence') of a tax refers to who finally experiences a decrease in real income as a result of the tax, not necessarily who is required by law to pay the tax (referred to as 'statutory incidence')
- **benefit incidence analysis** is the analysis of who benefits from government spending, and by how much. This includes analysis of direct government transfers (e.g. cash transfers), as well as in-kind transfers (e.g. subsidised public education and health services).

Both forms of fiscal incidence analysis encompass a variety of methods, from the simple to the more complex, depending mainly on how the likely behavioural responses of economic agents are dealt with (Martinez-Vasquez 2004; van de Walle 1998). Standard fiscal incidence analysis assumes there are no behavioural responses to taxes and government spending. Households and individuals are assumed to have perfectly inelastic supplies of the factors of production that they own (e.g. labour, human capital), and consumers are assumed to have perfectly inelastic demand for commodities (van de Walle 1998). It is

What policies and interventions have been strongly associated with changes in in-country income inequality?

often justified as a reasonable ‘first approximation’ to the real results that would be obtained if behavioural responses were included. General equilibrium analysis studies the incidence of taxes or spending in the context of a model of the whole economy, which allows for (some, not necessarily all) behavioural responses. In theory, GE analysis should provide more accurate estimates of incidence. However, they rely on the accuracy of the equations and parameters which are used to construct them; they are also computationally complex and the results are often sensitive to modelling choices.

The literature on fiscal incidence analysis is considered by experts to be vast (Martinez-Vasquez 2004). Not all fiscal incidence analysis is relevant to the current review however - only those studies which calculate impacts on one of the measures of inequality referred to in Section 3.1.5 below. Many incidence studies do not do this, and are therefore excluded from this review.

Decomposition analysis is a term used to explain a method of quantitative analysis whereby a measure of income inequality (e.g. the Gini coefficient) is broken down into its constituent parts or components. The two main types of decomposition analysis are (Cowell 2000):

- **decomposition by sub-group:** this involves decomposing a measure of inequality in a population of households or individuals to a part representing inequality between sub-groups of the population, and another part representing inequality within each sub-group;
- **decomposition by income source:** this involves decomposing a measure of inequality in total income to components relating to inequality of wages and salaries (income from work), rents (income from property), government transfers, and so on.

Decomposition analysis is typically used to when seeking to explain or ‘account for’ observed levels of inequality at a point in time or trends in income inequality over time.⁴ For example, Lustig et al. (2013b) show that one of the major sources of the decline in income inequality in many Latin American countries during the 2000s has been a fall in the inequality of labour incomes. This is helpful in that it allows further research to focus on the factors that may have contributed to the decline in labour inequality. Decomposition analysis differs from econometric (regression) analysis however, in that it is only able to identify the proximate sources of inequality, not the factors that determine inequality - including the effects of government policies and interventions. Thus the fact that much of the fall in inequality in Latin America can be explained by a fall of inequality in labour earnings does not tell us whether the latter was driven by policy or not. This limits the relevance of decomposition analysis to this review.

Some forms of decomposition analysis are relevant to this review however. One example is using decomposition analysis by income source, in which income from government

⁴ In the words of Azevedo et al (2013: 3): “Although ... decompositions do not allow for the identification of causal effects, they are a useful tool to identify empirical regularities and, as an accounting tool, can be useful to focus attention on the elements which are quantitatively more important in describing distributional changes.”

transfers as a separate source of income. In this case, decomposition analysis provides an estimate of the impact of those transfers on income inequality - which is directly comparable with estimates derived from ex-ante simulations under the assumption of no behavioural effects. Therefore, studies which use decomposition analysis by income source and which report income from government transfers as a separate category are included. However, studies using decomposition analysis by income source but which do not treat income from transfers as a separate income source are excluded.

Qualitative case studies make use of primary data collected by the researcher, e.g. small-scale household surveys, focus group discussions, and semi-structured interviews. Examples include Adger (1999), Silva (2007, 2013), and Copestake (2008). Unlike the quantitative study designs, they do not attempt to establish a counterfactual and are not therefore used to assess impact; they are instead used to shed light on the processes and mechanisms through which government policies and interventions affect income inequality.

Publication status: We will include published and unpublished studies, including refereed and non-refereed journal articles, working papers, conference proceedings, book chapters, government reports, NGOs reports and other technical reports. We will exclude comments and media briefings, review articles, and dissertations (PhD and MA). The exclusion of dissertations is mainly due to time and budgetary constraints: although our searches did identify a number of dissertations which are potentially relevant to the review, these are on the whole not available electronically. The financial and opportunity costs of obtaining hard copies of each dissertation for full text screening would therefore be very high, and detract from the review and synthesis of the other publication types.

Timeframe: We will restrict the review to studies published since 1990. This is mainly on the grounds that reliable, cross-country data on income inequality have only been available since the early 1990s, so that any studies before this date would not meet basic requirements in terms of data quality. In addition, studies published before 1990s are generally not available electronically; this again drives up the financial and opportunity costs of the screening process.

Language: We will include studies published in English, Portuguese, and Spanish.

3.1.2 *Types of participants (population)*

The review will be restricted to studies of low income countries (LICs) and middle income countries (MICs) at the time of the government intervention; studies of high income countries will be excluded. The World Bank definitions of LICs and MICs will be used in applying this criterion.⁵ The World Bank classifications of low, middle and high income countries have been in operation since 1988. Appendix 2 lists three groups of countries:

- a) those that have always been low or middle income since classifications began
- b) those which have been low or middle income in some years but not all
- c) those which have always been high income.

⁵ The official World Bank definitions of low and middle income countries are based on GNI per capita, and date back to 1989. Classification systems for earlier years are based on the Bank's operational lending categories; see <http://data.worldbank.org/about/country-classifications/a-short-history> for more details.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Studies of countries in group (a) are always included while studies of countries in group (c) are always excluded. Studies of countries in group (b) are included if the intervention being studied took place while the country was low or middle income.

Some studies relevant to our review question do not focus on specific countries but instead focus on groups of countries. In particular, many cross-country econometric studies of income inequality include countries from all income groups in the analysis, in the interests of sample size. We include such studies, on the grounds that they typically contain a significant proportion of low and/or middle income countries.⁶ However, we exclude studies which focus on groups of countries which consist mainly of high income countries: in particular, “developed countries”, “OECD countries”, “advanced industrial countries”, “Western Europe”, “North America”, or the European Union. Although some of these groups do sometimes contain one or two countries that are (or have at times been) middle income - for example, Mexico (an MIC) has been an OECD member since 1994 - they are overwhelmingly made up of high income countries.

3.1.3 Types of interventions

As discussed in Section 1.2, the review is not restricted to any one type of policy or intervention; all government policies and interventions are relevant to the review. We will also include policies and interventions by any level of government, including local, state and national. We will however exclude any studies of interventions by non-government and private sector organisations.

However, it is important to be clear what we mean by government policies and interventions. For this review, we define an intervention or ‘policy reform’ as a change in a variable that is directly controlled by the government, which we call a ‘policy variable’. By directly controlled, we mean that the variable is determined by parliamentary law, official (presidential) decree, bureaucratic decision, and so on; it is not determined by anything other than the government’s own decision-making. Some examples are shown in Table 1.

Table 1: Examples of policy variables and associated policy reforms

Policy variables (broad types)	Policy variables (examples)	Policy reforms (examples)
Tax and subsidy rates	The rate of VAT	A reduction in VAT
Transfers	Government spending on transfers	A new cash transfer programme
The supply of publicly-provided goods and services	Government spending on roads	An expanded road building programme
The price charged for publicly-provided goods and services	School tuition fees	The removal of school tuition fees
Official price floors and price ceilings	Official minimum wage	An increase in the minimum wage
Quantity restrictions and prohibitions	Restrictions on the use of child labour; anti-discrimination legislation	New legislation which bans the use of child labour

⁶ Later, at the data extraction stage, we seek to assess precisely what proportion of the countries included in a cross-country study are low or middle income.

Notes: The list is not designed to be exhaustive, it is simply designed to explain the definition of the terms ‘policy’ and intervention used in the review, as stated above in the main text. Any policy or intervention meeting the above definition is relevant to the review.

There is a lot of evidence on the effects of specific policy reforms of the type shown in Table 1 on income inequality. However, this evidence is made up mainly by ex-ante simulation approaches rather than ex-post quasi-experimental studies. The latter tend to focus on the effects of more aggregate indicators that are clearly influenced by policy, but are not in themselves policy variables. For the purposes of this review, this feature of the literature using ex-post analysis is a clear drawback. Nevertheless, to exclude such evidence altogether would also be problematic, since it would imply relying on one particular methodological approach.

As a result, we seek to include econometric studies which look at the effect of indicators which are *clearly and closely* influenced by policy, and not just those that look at the effect of specific policy reforms. We will exclude studies that only look at broader determinants of income inequality which are not clearly and closely influenced by policy. Similarly, we will exclude any ex-ante simulation studies which focus only on the effects of external or internal shocks on income inequality (e.g. a deterioration in a country’s terms of trade, a collapse in demand for exports, or a decline in productivity caused by a natural event), as opposed to the effects of government policy changes.

3.1.4 *Types of comparison groups*

The control or comparison group for assessing the impact of government policies and interventions will be constructed using either an ex-post quasi-experimental approach or an ex-ante simulation-based approach. The former involves comparisons of inequality across countries and over time, using panel data. The latter involves comparisons of the observed level of inequality in a country before a particular intervention, and the simulated level of inequality after the intervention.

We will also include studies focusing on income inequality in regions or states within a country, as well as those that focus on income inequality at the national level. Thus the unit of analysis may be the country as a whole, or a region or state within the country.

3.1.5 *Types of outcome measures*

We will include studies that focus on inequality in a comprehensive measure of income that includes income from all sources (e.g. wages and salaries, business profits, investment earnings, rental income, transfers); we will exclude any studies that focus on inequality in one source of income (e.g. wages). We also require that data on income or expenditure be drawn from a representative household survey covering all of the relevant population. We will exclude any estimates which are derived from the National Accounts, or from household surveys that cover only a subset of the relevant population. Note however that the relevant population need not be the country as a whole; it may also be the state or locality within the country.

We will include studies that focus on any of the five definitions of income set out in Section 1.3, namely market, net market, disposable, post-fiscal and final income. We will also include studies which focus on inequality in total consumption expenditure, since the

What policies and interventions have been strongly associated with changes in in-country income inequality?

latter is often considered to be a more reliable indicator when data on income are difficult to collect.

We will also include any measure of inequality in income or consumption expenditure. This includes:

- global measures, i.e. measures which seek to capture the dispersion across the whole distribution and utilize all values of the underlying indicator (e.g. income) in its evaluation. Examples include the coefficient of variation, the relative mean deviation, the variance, the Gini coefficient, the Atkinson family of measures and the Generalized Entropy class of measures (e.g. Theil index, mean log deviation);
- partial summary measures, including i) percentile ratios, e.g. the ratio of the 90th to the 10th percentile of the income or expenditure distribution, the ratio of the 90th percentile to the median of the distribution; ii) income shares, e.g. the share of the poorest 10%, 20% or 40% of households in national income; and iii) income share ratios, e.g. the ratio of the income share of the richest 20% of households to that of the poorest 40% (the so-called Palma index).

In each case, inequality may be measured across households or individuals; in the former case, average household income or expenditure may be expressed per capita or per adult equivalent. There are many studies which look at income inequality across groups as opposed to inequality across households or individuals. Examples include studies of income inequality between spatial units (e.g. states, provinces, regions) within countries, studies of income gaps between urban and rural areas, and studies of income gaps between demographic groups, as defined for example by gender or ethnicity. These 'between-group' inequalities do of course often account for a large proportion of the 'overall' amount of inequality observed across households and individuals. Moreover, many government policies are likely to affect income inequality across households precisely by affecting one or more between-group inequality.

Nevertheless, in this particular review we only include studies estimating the impact of government policies on measures of 'overall' inequality across individuals or households. The review is already very broad, and to include studies of between-group inequality would widen it considerably, making it excessively broad in our view. There however is a case for a further systematic review or reviews looking at the effects of government policies on the sorts of between group inequalities most likely to be affected by that policy: for example, a study of the effect of regional policies on inter-regional inequalities within countries, or the effect of anti-discrimination legislation on inequalities between ethnic groups.

In addition, the focus of this review is on the 'size' distribution of income, not the so-called 'functional' distribution of income. While the former refers to the distribution of income across individuals, the latter refers instead to the share of national income received by each factor of production, i.e. labour, capital, land and so on. We exclude any studies which only report impacts of government policies on the functional distribution of income.

Finally, certain measures of the size distribution of income are related to inequality but do not themselves constitute measures of inequality. Examples include measures of:

- relative poverty, e.g. percentage of the population with incomes less than 50% of mean or median income
- the size of the middle class, e.g. percentage of the population within a certain range of median income, the share of the middle three quintiles in national income
- income polarisation, which refer to the amount of ‘between-group’ dispersion of income relative to ‘within-group’ dispersion, with the groups themselves defined by income - see for example Duclos et al. (2004).

We exclude any studies which focus only on one of these indicators and none of the global or partial summary measures of inequality listed above.

3.2 Search methods

3.2.1 Electronic searches

In order to select appropriate databases for this review we followed the Campbell Collaboration guide on key online databases for systematic reviews in International Development (Campbell Collaboration 2012). This list was complemented with additional databases and websites used by other systematic reviews on questions relevant to this review. The electronic databases that will be searched for relevant studies are shown in Table 2.

Table 2: Electronic databases and search strings

GENERAL BIBLIOGRAPHIC DATABASES	SEARCH
EBSCO EJS Science Direct Scopus JSTOR	long version long version long version short version
SUBJECT SPECIFIC DATABASES & WEBSITES	
Social Sciences	
Web of Knowledge (Social Science Citation Index) IBSS (International Bibliography of the Social Sciences) ASSIA (Applied Social Sciences Index and Abstract) SSRN (Social Science Research Network)	short version long version long version short version
Economics	
IDEAS NBER Econlit (EBSCO)	short version long version long version
International Development	
3IE Impact Evaluation Database British Library of Development Studies Eldis	short version short version short version

Notes: Searches will be filtered to abstract, title and key words. Whenever possible search strings will be filtered to social sciences. Please see Appendix 4 for the detailed check-list on the search strategy.

Each database will be searched using a combination of the search terms indicated in Table 3. The search terms were identified by 1) reviewing the literature for relevant and appropriate terms and 2) extracting key words from a sample of relevant literature. Table 3 shows three sets of concepts (A, B and C), each of them containing a list of associated

What policies and interventions have been strongly associated with changes in in-country income inequality?

terms or synonyms that will be used in our search, derived from the conceptual framework of this systematic review. When using foreign language databases, each of these terms will be translated into the appropriate language, i.e., Portuguese or Spanish (see Appendix 3 for the search terms in these languages).

Due to the fact that some search engines only allow a limited number of operators, two search query strings are used: a long version and a short version. The long version follows the equation:

$$A + (B W/n C)$$

Thus the terms within columns A, B or C will be combined with ‘OR’; columns B and C will be combined with the proximity operator ‘W/n’, where n is the number of words that separate the terms from two columns; and column A will be combined with the combination of B and C using the ‘AND’ command. Our strategy is to use n=1 to capture concepts such as ‘distribution of income’, ‘inequality of income’, as well as ‘income distribution’ and ‘income inequality’.

Table 3: Concepts for search strategy

A Policy	B Income	C Inequality
Polic* Intervention* Program* Instrument* Tool* Reform* Legislation* Govern*	Income* Expenditure*	*Equal* *Distribut* Disparit* Differen* Gap* *Equit* Share* Ratio* Gini

Notes: * is included as a truncation symbol to capture automatically conjugated forms of each word; thus *equal* captures “inequality” as well as “inequalities”; *distribut* captures “distribution” as well as “redistribution”.

The short version uses only one term from each column at a time. Different short version strings will be used, including A+B+C, BC, B+C and A+BC. Table 4 shows examples of the short version search strings that will be used, depending on the database. Information on the specific search strings used for each database will be included in the final report. Note that to capture a concept such as income inequality, quotation marks are used. Thus the search BC will give the same results as B+C; however the reverse does not apply.

Table 4: Search strings

LONG VERSION¹	STRING
A + (B W/n C)	((polic* OR intervention* OR program* OR instrument* OR tool* OR reform* OR legislation* OR govern*) AND ((income* OR expenditure*) W/1 (*equal* OR *distribut* OR disparit* OR differen* OR gap* OR *equit* OR share* OR ratio* OR gini))
SHORT VERSION	
A + B + C	polic* AND income AND inequal*
B+C	income AND inequal*
BC	“income inequal*”
A+BC	polic* AND “income inequal*”

3.2.2 Other searches

We will also review relevant institutional websites of key institutions and conference proceedings (see table 5).

What policies and interventions have been strongly associated with changes in in-country income inequality?

Table 5: Other searches

DEVELOPMENT ORGANIZATIONS LIBRARIES & WEBSITES	
General	
World Bank Open Knowledge Repository	short version
OECD iLibrary	short version
International Labour Organization	short version
Chronic Poverty Research Center	short version
Overseas Development Institute	short version
Center for Global Development	short version
International Policy Center for Inclusive Growth	short version
JOLIS (IMF and World Bank databases)	short version
African Development Bank Evaluation Reports	short version
Asian Development Bank Evaluation Resources	short version
Inter-American Development Bank	short version
Impact Evaluation/ Effectiveness	
Millennium Challenge Corporation Independent Evaluations	short version
Research for Development (R4R)-DFID	short version
USAID Development Experience Clearing House	short version
OTHERS	
Grey Literature	
OpenGrey	short version
Foreign Language Databases	
CLASE (Citas Latinoamericana en Ciencias Sociales y Humanidades)	short version
e-Revistas (Plataforma Open Access de Revistas Cientificas Electronicas Espanolas y	short version
REDALyC (Red de Revistas Cientificas de America Latina et el Caribe, Espana e Portugal)	short version
Scielo	short version
Others	
Google Scholar	short version

Beyond searching the databases and websites listed in Tables 2 and 5 we will also search:

- reference lists of review articles and included articles adopting a snowballing approach,
- track citations of included studies
- contact key authors, experts and practitioners to enquire about unpublished, forthcoming and/or ongoing studies,
- draw on our advisory group to check for any studies we might have missed.

In addition to English language publications, we also search the Portuguese and Spanish literature to address any potential language bias. These other searches will use the search terms outlined above, translated into Portuguese and Spanish (see Appendix 3). Our completed search strategy checklist is contained in Appendix 4.

3.3 Data collection and analysis

3.3.1 Selection of studies

A PRISMA diagram will be produced in order to keep track of the search process (see Figure 2). Following the removal of duplicate studies, the results of the searches will initially be screened in terms of title and abstract by a research officer or assistant. This will be used to remove studies which clearly do not meet the inclusion criteria. This process will be checked and monitored by 2 lead reviewers. Any studies for which uncertainty exists about the criteria will be referred for a second opinion, or retained for full text analysis.

Once a certain number of studies have been excluded on the basis of abstract, the remaining studies will be obtained in full text. Each of these will then be assessed independently in duplicate by two reviewers using inclusion forms developed for this review. This same approach will be used for both quantitative as well as qualitative studies. Once an additional number of studies have been excluded on the basis of the full text, or due to unavailability of full text, the remaining studies will all be included in the mapping exercise.

Finally, a selection of the studies included in the mapping stage will be selected for inclusion in the synthesis stage. As discussed in Section 1, this is designed to avoid the problems stemming from the very broad question of this systematic review - namely a) that the amount of relevant literature will be too large, and b) that the policies and interventions will be too diverse, preventing meaningful and interesting comparisons of the effects of similar types of policies and interventions across different countries and contexts. In the words of Gough et al. (2013: 16):

“The studies contained within a research field may be too numerous or heterogeneous for meaningful synthesis; it might be methodologically too difficult or just take too much time. The map provides an opportunity to select a sub-group of studies for synthesis.”

The preliminary results of the mapping stage have now been completed and are provided in a separate document submitted with this updated Protocol. So far we have identified a total of 194 studies which meet our inclusion criteria. These studies cover a wide range of different policies and interventions, ranging from fiscal policy (government tax and spending policies), trade policy (e.g. import tariffs, export quotas), macroeconomic policies (e.g. exchange rates, monetary policy, financial reforms, land reforms, labour market reforms, and energy sector reforms. By far the most common type of policy is however fiscal policy, which is covered by 144 studies out of the 194.

We will restrict the synthesis to studies which focus on the effects of a) fiscal policy interventions or b) trade policy interventions, but exclude from the synthesis studies of other intervention types. We believe that an analysis of the effects of fiscal and trade policy interventions on income inequality would be a meaningful task for the synthesis, for which meta analysis is possible (see Section 3.4 below). Focusing on fiscal and trade policy interventions also allows us to compare the results of different study designs - for

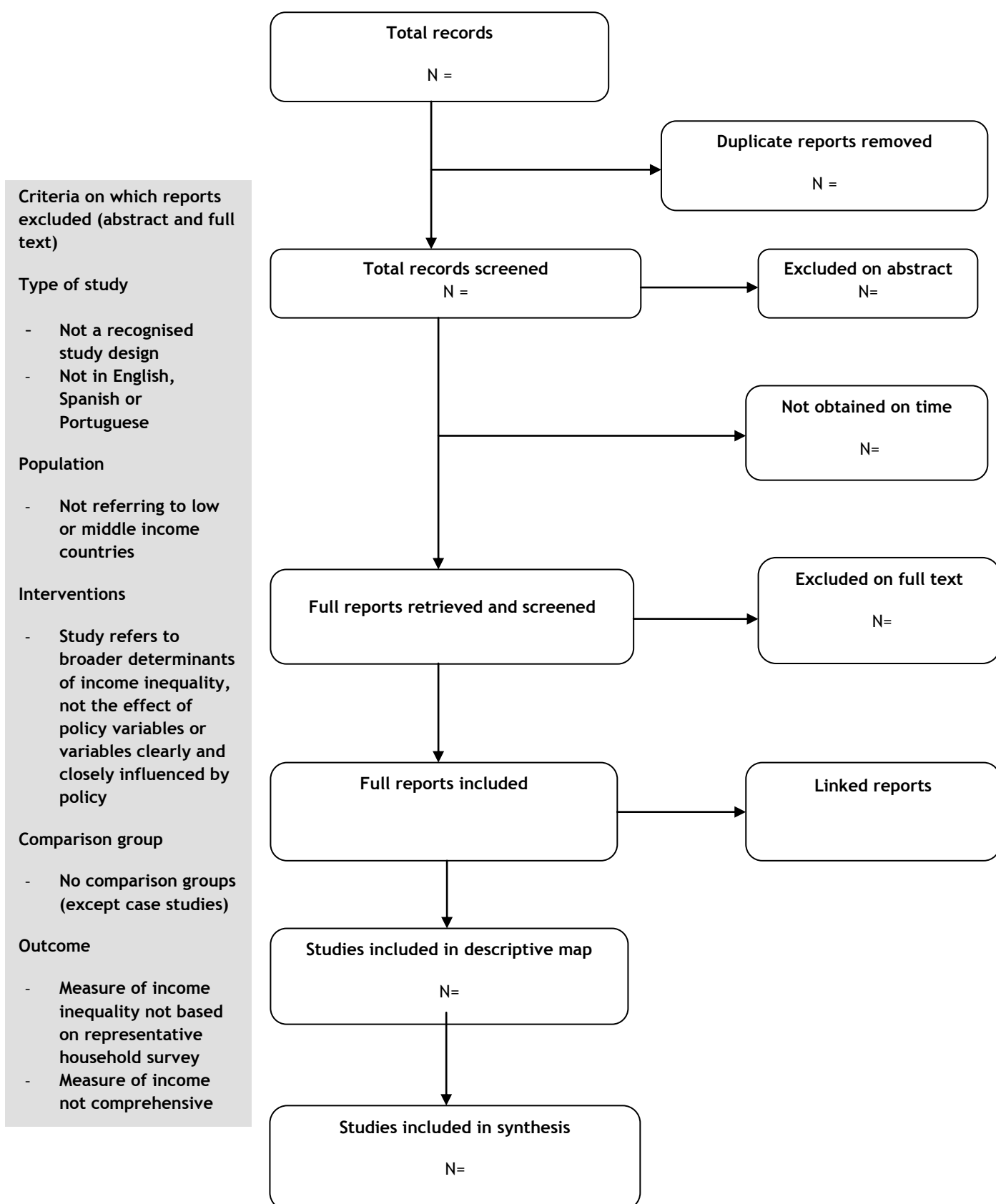
What policies and interventions have been strongly associated with changes in in-country income inequality?

example, the results of econometric studies vs. CGE models -potentially with separate meta analysis for each study design.

By contrast, for most other intervention types (e.g. finance or labour market reforms), the number of studies would be too small to allow meta analysis, and the evidence is often restricted to one study design. It is worth stressing however that studies of the other types of interventions would remain in the mapping, and we will comment briefly on these studies in the final report - so that they serve as a resource to other users, and an indication of the policy areas where evidence of impacts on income inequality is relatively scarce.

We will restrict the synthesis by focusing only on studies which look at income inequality at the national level. The majority of studies do this, but some look only at inequality at the level of regions within a country (e.g. states or provinces), or in urban or rural areas, or sometimes at a very localised level (e.g. the village). Inequality at these sub-national levels is not directly comparable with inequality at the national level (nor are they comparable among themselves).

Figure 2: Flow of literature through the review: the PRISMA diagram



What policies and interventions have been strongly associated with changes in in-country income inequality?

3.3.2 Data extraction and management

Data extraction will take place in two main stages. The first stage will extract descriptive information about all studies meeting the inclusion criteria, in the following three main areas:

1. context and population
2. type of intervention
3. study design and methods used
4. outcome measures

Data extracted in this first stage will feed directly into the research mapping, allowing us to provide a descriptive survey of all the relevant evidence relating to the question, categorising and cross-tabulating the available evidence in interesting ways, e.g. the overall balance of studies between intervention types, outcome indicators, country groupings and study designs.

A further extraction stage will extract additional information required for the quality appraisal and synthesis, in particular:

5. study results and findings
6. quality of research methods

Table 6 lists examples of the type of information that will be extracted from each study under each of the above six headings.

Table 6: Data extraction form (template)

Data extraction items	
1. Context and population	4. Outcome measures
Source	Measure of inequality (e.g. Gini, Theil index)
Author	Unit of measurement (household, individual)
Publication year	Income or expenditure
Single country or multi-country	Measure of income (market, disposable or net)
Country or countries studied	Source of information
Country categories (region and income level)	
Unit of analysis (national, regional or local)	5. Study results and findings
2. Type of policy/intervention	<i>For each outcome of interest</i>
Broad category	Sample size
Detailed sub-category	Effect sizes for meta-analysis if possible such as p-values, standard errors, t-values etc.
Indicators used to measure intervention	
Year (period) of intervention(s)	6. Quality of study
Level of government	Clarity of research question
Source of funding	Description of population
3. Study design and methods	Quality of research methods
Study design (main category)	Researcher bias
Study design (sub-category)	Any other validity problems?
Control variables (ex-post studies)	
Modelling choices (ex-ante studies)	

3.3.3 Assessment of relevance and quality

Once studies have been judged as meeting the inclusion criteria and therefore included in the descriptive map, the next step will be to assess their relevance to the review question and their overall quality (Gough et al. 2013: 17).

As discussed above, following the mapping we will restrict the synthesis to studies which focus on the effects of fiscal or trade policy interventions on income inequality at the national level. This is not to say that the studies of other intervention types are not of interest, but instead that it would not be feasible to try to synthesise a wide variety of different intervention types simultaneously (given that we are also considering different study designs). In addition, there is typically much less evidence with which the impact of these other intervention types can be usefully compared or synthesised.

As a next step, studies included in the synthesis will be assessed for their quality - otherwise referred to as 'risk of bias'. The risk of bias tools developed by Duvendack et al. (2011 and 2012) and IDCG⁷ are based on criteria adapted from the Cochrane Handbook (Higgins and Green 2008, 2011) and EppiCentre (Gough 2007 and EppiCentre 2010). The

⁷ See Appendix 4 for details of the IDCG tool.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Cochrane Collaboration suggests that the key components of bias in any study are selection bias, performance bias, attrition bias, detection bias, and reporting bias. The EppiCentre formulates the risk of bias as being composed of factors such as the trustworthiness of results (or methodological quality), including transparency, accuracy, accessibility and specificity of the methods; the appropriateness of focus for answering the review question (topic relevance, including relevant answers and legal and ethical propriety); and the overall weight of evidence.

We will begin the quality assessment by categorising each study by its proclaimed research design and analytical method. Following Duvendack et al. (2011 and 2012), each study will be scored depending on its design and analytical approach. In a next step each of these scores will be combined into an index. An arbitrary threshold of 2 will be applied, i.e. a study with a score of less than 2 is classified as low risk of bias while a study with a score of 2 and above is classified as high risk of bias (Duvendack et al., 2011, 2012 and 2014).

We will adapt the table below to our particular situation and rank studies by research design and analytical method using scores 1 - 5, where 1 implies low risk of bias and 5 high risk of bias (3 in the case of analytical method).

Table 7: Distribution of studies by research design and analytical method

		Statistical Methods of Analysis		
		IV,PSM,2SLS/LIML,DID, RD	Multivariate	Tabulation
Research Design	Scores	1	2	3
RCT	1			
Pipeline	2			
Panel or before/after & with/without	3			
Either before/after & with/without	4			
Natural Experiment	5			
Legend				
	Low score		High score	
	Medium score			

Source: Duvendack et al. (2011 & 2014, 2012 for an adaptation).

Based on the initial mapping exercise we expect that most of the ex-post observational studies will be scored 3 for research design and 1 or 2 for analytical method. Many of these studies adopted cross-country regression approaches which have been criticized widely (see for example Beck et al. 2000; Graff, 2001). Jalilian and Kirkpatrick (2002) summarise Beck et al.'s critique of cross-country approaches as follows:

“(i) time series dimension of data is generally ignored; (ii) parameter estimates may be biased because of omission of cross country differences; and (iii) no control for endogeneity of regressors. An additional shortcoming of this approach is that it cannot be used for causal inference” (p.99).

This implies that many of our included studies are likely to be classified as medium to high risk of bias. However, Jalilian and Kirkpatrick (2002) further argue that advances in analytical approach such as dynamic panel estimations can correct for the drawbacks of cross-country approaches outlined earlier. This point further motivates the use of the Duvendack et al. tool as it assesses risk of bias by providing a combined score for research design as well as analytical technique. E.g. a study might get a score of 3 when using cross-country panel data but can considerably improve its score when using a sophisticated analytical approach. A combined score per study will reflect this and provide an overall risk of bias score.

Please note that ideally we should only be including low risk of bias studies in the synthesis stage but this might leave us with a rather small sample, hence we will include all studies irrespective of their risk of bias score and then conduct sub-group analysis to tease out differential impacts by risk of bias classifications (see Duvendack et al. 2014 for an example).

We are aware that the Duvendack et al. tool is subjective (see Duvendack et al. 2014, footnote 7 for an explanation) and hence we complement this approach with the risk of bias tool developed by IDCG which also includes risks due selection bias and confounding, spill-overs/contamination, outcome and analysis reporting as well as other risk of biases. See Duvendack et al. (2014) for an example of how these 2 tools have been applied in combination. We will have to further adapt the IDCG tool for our particular context as some of its checklist items are not applicable to the studies we have included.

We intend to include qualitative as well as mixed methods studies in the review, but the results of the mapping exercise indicate that only a very small number of such studies meet our inclusion criteria. From the searches carried out to date we have only identified one qualitative study. As a result of this we do not see the need to develop a separate risk of bias assessment tool to assess the quality of the qualitative and mixed methods studies.

3.4 Data synthesis

Systematic reviews in the social sciences are increasingly drawing on evidence from both quantitative and qualitative studies and thus a number of synthesis methods are available. Table 8 below provides a summary of the most commonly used synthesis methods. Given the limited number of qualitative studies we identified so far, we anticipate to largely draw on quantitative synthesis methods.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Table 8: Synthesis methods

Quantitative evidence	<i>Meta-analysis</i>
	<i>Meta-regression</i>
Qualitative and/or mixed methods evidence	<i>Meta-ethnography</i>
	<i>Narrative synthesis</i>
	<i>Meta-narrative mapping</i>
	<i>Realist synthesis</i>
	<i>Thematic synthesis</i>
	<i>Framework synthesis</i>

3.4.1 Quantitative synthesis

For the ex-post observational studies, we will attempt meta-analysis following the approach taken by Abdullah et al. (2013). Their study examines the impact of education on income inequality using a meta-regression approach. From the results outlined in the mapping document we have seen that the majority of ex-post observational studies are multi-country regression-based approaches using Gini coefficients as the main outcome variable and a measure of government spending as the main policy variable of interest. We should have a sufficient number of studies that meet these characteristics and could be synthesized using a quantitative approach.

It is argued that meta-analysis is only possible for studies that can be meaningfully compared, i.e. they need to be comparable on a conceptual level which means that similar constructs and relationships are used and they need to follow similar statistical approaches (Lipsey and Wilson, 2001). In other words, for econometric studies to be included in our meta-analysis they should:

- have a common measure of income inequality (i.e. this can be the Gini, income shares or similar - we will convert these different income inequality measures into a comparable measure using partial correlation measures) as the dependent variable and a relevant policy variable among the explanatory variables, such as government spending, taxation, or average import tariffs.
- pursue a regression-based approach which will allow us to convert the numerous measures for dependent and explanatory variables into a comparable measure. Estimation of effect sizes from regression results appears to be less well developed and more problematic than for mean based results (Fritz, Morris and Richler, 2011). However, recent literature proposes to use partial correlation measures which can be calculated from regression estimates. Aloe and Thompson (2013) provide guidance on how best to estimate and use partial correlation measures for synthesis and we will follow their advice.
- be published or unpublished. In the course of our meta-analysis we will account for publication bias using funnel plots to examine whether it potentially distorts the effects of government spending and tax on income inequality.

Despite attempting to synthesize studies that are as similar as possible in terms of conceptual framing as well as analytical approach, we suspect that an element of diversity among our included studies will remain suggesting the so-called “apples and oranges” problem is likely to arise where studies which are distinctly different are pooled (Lipsey and Wilson, 2001: 2). As mentioned earlier, studies that are methodologically flawed or of low quality should not be included in the same meta-analysis as other studies, since this could adversely affect the overall results (Slavin, 1986). However, we argued above that we will explore the quality aspect among our studies with subgroup analysis to tease out differential effects by risk of bias grouping.

Given the “apples and oranges” problem might be an issue we will be exploring potential sources of heterogeneity across included studies and describe what this implies for meta-analysis. E.g. effect size estimates can be biased by non-normality and heteroscedasticity (Wilcox, 2008), which are generally not reported in our studies. Studies with low or negative effects may be under-reported, not find their way into the included studies, and hence meta-analysis would be upward biased. There might also be some sort of heterogeneity among the main explanatory variables (e.g. government spending and tax) which we will explore further.

An interesting component of the synthesis will be to compare the results from the ex-post quasi-experimental studies (e.g. cross-country econometrics) with ex-ante simulation studies (e.g. CGE models). While ex-post studies are often preferred since it represents external data validation, cross-country econometric studies unavoidably work at a high level of aggregation and as a result provide little evidence on the effects of specific policies on poverty or inequality - a change in a key tax rate or import tariff for example. By contrast, CGE and other simulation-based studies are able to analyse much more specific policy changes. We are certainly not the first to include ex-ante simulation studies (e.g. CGE models) in a systematic review. For example, the DFID-funded Systematic Review on the effects of trade liberalisation on employment and fiscal revenue by Cicera et al. (2011) specifically included such studies, alongside more conventional econometric analysis; the study by McCorrison et al. (2013) on trade liberalisation and food security also took this approach. We aim to show how the results of studies using ex-ante simulation compare to those using ex-post quasi-experiments, and to discuss the likely reasons for any systematic differences between these two different research approaches.

We anticipate however that meta-analysis will need to be done separately for each approach, as was the case of Cirera et al. (2011) and McCorrison et al. (2013). In terms of the specific methods for meta-analysis of the ex-ante simulation studies, we will follow the broad approach used by Hess and von Cramon-Taubadel (2008), who synthesise the results from ex-ante simulation studies (including CGE models) on the impacts of multilateral trade reform. These authors use meta-regression to show how the results of simulations vary according to the different modelling assumptions used.

3.4.2 *Qualitative synthesis*

The qualitative synthesis will be used to improve and develop understanding of the processes and mechanisms through which government policies and interventions affect income inequality. This will involve synthesising the results of detailed case studies of income inequality in particular low or middle income countries, or particular regions

What policies and interventions have been strongly associated with changes in in-country income inequality?

within such countries. The synthesis of these studies will allow us to explore in detail the various assumptions in our conceptual framework about the ways in which government policies affect inequality, and to identify and explore any unanticipated effects. In the selection of the qualitative synthesis approach we will be guided by the precise nature of the available case study evidence. We anticipate however relying mainly on the range of narrative summary techniques suggested in Arai et al. (2007) and Rodgers et al. (2009).

4. Timeline

4.1 Estimates of the start and end dates for the following stages:

	Start date	End date
Registration of title with DFID	1 st January 2014	31 st January 2014
Preparation of protocol	1 st January 2014	14 th March 2014
DFID and External Review of protocol	17 th March 2014	25 th April 2014
Study search	28 th April 2014	16 th May 2014
Mapping and assessment of relevance	19 th May 2014	31 st July 2014
Synthesis and/or statistical analysis	1 st August 2014	1 st December 2014
Preparation of draft report	1 st December 2014	13 th February 2015
DFID and External review of draft report	16 th February 2015	27 th March 2015
Revision of draft report	30 th March 2015	14 th May 2015
Preparation of Evidence Brief for Policy	1 st May 2015	14 th May 2015
Publication of Final Report and Evidence	14 th May 2015	12 th June 2015

4.2 Deliverables (nature and due date):

	Due date
Title	31 st January 2014
Protocol	14 th March 2014
Mapping report*	31 st July 2014
Draft report	13 th February
Final report and Evidence	14 th May 2015

*A short report detailing the results of the research mapping exercise, including the updated protocol (see Section 3.3.1)

References

- Abdullah, A., Doucouliagos, H. and Manning, E. (2013). Does education reduce income inequality? A meta-regression analysis. *Journal of Economic Surveys*, 00(0), 1-16.
- Adger, W. Neil. (1999). Exploring Income Inequality in Rural, Coastal Vietnam. *Journal of Development Studies*, 35, 96-119.
- Alesina, A. and Perotti, R. (1996). Income distribution, political instability, and investment, *European Economic Review*, 40, 1203-1228.
- Aloe, A. and Thompson, C. (2013). The Synthesis of Partial Effect Sizes. *Journal of the Society for Social Work and Research*, 4(4), 390-405.
- Anderson, E. and L. Esposito (2014). On the Joint Measurement of Absolute and Relative Deprivation, *Journal of Economic Inequality*, forthcoming.
- Arai, L, Britten, N, Popay, J, Roberts, H, Petticrew, M, Rodgers, M and Sowden, A (2007). Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions. *Evidence & Policy: A Journal of Research, Debate and Practice*, 3(3), 361-383.
- Azevedo, J., Inchaust, G. and Sanflice, V. (2013). Decomposing the recent inequality decline in Latin America. Policy Research Working Paper 6715, World Bank, Washington D.C.
- Baer, W. and Galvao, A. Jr. (2008). Tax burden, government expenditures and income distribution in Brazil. *Quarterly Review of Economics and Finance*, 48, 345-358.
- Barber J. G. (2001). Relative misery and youth suicide. *Australian Journal of Psychiatry*, 35, 49-57.
- Beck, T., Levine, R. and Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics* 58, 261-300.
- Birdsall, N. (2007). Income distribution: effects on growth and development. *International Handbook of Development Economics*, 2, 45.
- Campbell Collaboration (2012). Systematic Reviews in International Development: Key Online Databases. Available at:
http://www.campbellcollaboration.org/artman2/uploads/1/Database_Guide_for_SRs_in_International_Development_1.pdf
- Castelletti, B. (2013). How redistributive is fiscal policy in Latin America? The case of Chile and Mexico. Working Paper 318, OECD Development Centre, OECD, Paris.
- Centre for Reviews and Dissemination (2009) Systematic Reviews: CRD's guidance for undertaking reviews in health care. CRD, University of York.

- Chenery, H., Ahluwalia, M. S., Duloy, J. H., Bell, C. L. G. and Jolly, R. (1974). *Redistribution with growth; policies to improve income distribution in developing countries in the context of economic growth*. Oxford University Press.
- Cirera, X., Willenbockel, D. and Lakshman, R. (2011). What is the evidence of the impact of tariff reductions on employment and fiscal revenue in developing countries? Technical report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
- Clark, A.E., Frijters, P. and Shields, M.A. (2008). Relative income, happiness, and utility: an explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46, 95-144.
- Copstake, J. (2008). Multiple dimensions of social assistance: The case of Peru's 'glass of milk' programme. *Journal of Development Studies*, 44, 4, 545-561.
- Cowell, F. A. (2000) "Measurement of Inequality" in Atkinson, A.B. and Bourguignon, F. (ed.) *Handbook of Income Distribution*, North Holland, Amsterdam.
- Cox, D. and Jimenez, E. (1995). Private Transfers and the Effectiveness of Public Income Redistribution in the Philippines. In *Public Spending and the Poor: Theory and Evidence*, ed. Dominique van de Walle and Kimberley Nead. Baltimore, Md.: John Hopkins University Press.
- Critical Appraisal Skills Programme (CASP). (2006). Qualitative research: appraisal tool. 10 questions to help you make sense of qualitative research. In. Oxford: Public Health Resource Unit, p. 1-4.
- Devarajan, S. and Hossain, S. I. (1998). The combined incidence of taxes and public expenditures in the Philippines. *World Development*, 26(6), 963-977.
- Dixon-Woods M, Shaw, R.L, Agarwal, S and Smith, J.A. (2004). The problem of appraising qualitative research. *Quality and Safety in Health Care*, 13, 223-5.
- Duclos, J-Y., Esteban, J. and Ray, D. (2004). Polarisation: concepts, measurement, estimation. *Econometrica* 72(6), 1737-1772.
- Duvendack, M., Palmer-Jones, R., Copstake, J. G., Hooper, L., Loke, Y. and Rao, N. (2011). "What is the Evidence of the Impact of Microfinance on the Well-being of Poor People?" London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Duvendack, M., Hombrados, J., Palmer-Jones, R. and Waddington, H. (2012). Assessing 'What Works' in International Development: Meta-Analysis for Sophisticated Dummies. *Journal of Development Effectiveness*, 4(3), 456-471.
- Duvendack, M., Palmer-Jones, R. and Vaessen, J. (2014). Meta-analysis of the Impact of Microcredit on Women's Control over Household Decisions: Methodological Issues and Substantive Findings. *Journal of Development Effectiveness*, 6(2), 73-96.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Easterlin, R.A. (1974). Does economic growth improve the human lot? Some empirical evidence. In: David, R., Reder, R. (Eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Academic Press, New York.

Eibner, C. E. and William N. E. (2005). Relative Deprivation, Poor Health Habits and Mortality, *Journal of Human Resources*, 40, 591-620

Engel, E., Galetovic, A., & Raddatz, C. (1999). Taxes and income distribution in Chile: Some unpleasant redistributive arithmetic. *Journal of Development Economics*, 59, 155-192.

EPPI-Centre website: Quality and relevance appraisal, Available at: <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=177> (accessed July 2010).

Fritz, C.O., Morris, P.E. and Richler, J.J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology General*, 141(1), 2-18.

Gindling, T. H., & Robbins, D. (2001). Patterns and sources of changing wage inequality in Chile and Costa Rica during structural adjustment. *World Development*, 29(4), 725-745.

Goni, E., Lopez, J.H. and Serven, L. (2011). Fiscal Redistribution and Income Inequality in Latin America. *World Development*, 39(9), 1558-1569.

Gough, D. (2007). Weight of Evidence: A Framework for the Appraisal of the Quality and Relevance of Evidence. In Furlong, J. & Oancea, A., eds. *Applied and Practice-based Research. Special Edition of Research Papers in Education*, 22, (2), 213-228.

Gough, D., Oliver, S. and Thomas, J. (2013). Learning from research: systematic review for informing policy decisions. A quick guide. A paper for the Alliance for Useful Evidence. London: Nesta.

Graff, M. (2001). Causal links between financial activity and economic growth: Empirical evidence from cross country analysis, 1970-1990. Mimeo

Gravelle, H., and Sutton, M. (2009). Income, relative income, and self-reported health in Britain 1979-2000, *Health Economics*, 18, 125-145.

Hemming, R. and Hewitt, D. (1991). The distributional impact of public expenditure. In Chu, K. and Hemming, R., eds. *Public Expenditure Handbook: A Guide to Public Policy Issues in Developing Countries*. Washington D.C.: IMF.

Hess, S. and von Cramon-Taubadel, S. (2008). A meta-analysis of general and partial equilibrium simulations of trade liberalisation under the Doha Development Agenda. *World Economy* 31(6), 804-840.

Higgins, J.P.T. and Green, S. (2008) eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.0 [updated February 2008]. Available at: www.cochrane-handbook.org.

Jalilian, H. and Kirkpatrick, C. (2002). Financial Development and Poverty Reduction in Developing Countries. *International Journal of Finance and Economics*, 7, 97-108.

- Killick, T. (2002). Responding to Inequality. Briefing Paper No 3, March 2002, Overseas Development Institute.
- Lindert, K., Skoufias, E. and Shapiro, J. (2006). Redistributing Income to the Poor and the Rich: Public Transfers in Latin America and the Caribbean, SP Discussion Paper No. 0605, The World Bank.
- Lipsey, M. P. and Wilson, D. (2001). Practical Meta-Analysis. Sage Publications.
- Long, A. F. and Godfrey, M. (2004). An evaluation tool to assess the quality of qualitative research studies. *International Journal of Social Research Methodology*, 7, (2), 181-196.
- Lustig, N. (2011). Fiscal policy and income redistribution in Latin America: challenging the conventional wisdom. Working Paper 2011-227, Society for the Study of Economic Inequality.
- Lustig, N., Lopez-Calva, L. and Ortiz-Juarez, E. (2013a). Deconstructing the decline in inequality in Latin America. Working Paper 1314, Department of Economics, Tulane University.
- Lustig, N., Lopez-Calva, L. and Ortiz-Juarez, E. (2013b). Declining Inequality in Latin America in the 2000s: The Cases of Argentina, Brazil, and Mexico. *World Development*, 44, 129-141.
- Luttmer, E. (2005). Neighbours as negatives: Relative earnings and wellbeing. *Quarterly Journal of Economics*, 120, 963-1002.
- Martinez-Vasquez, J. (2004). The impact of fiscal policy on the poor: fiscal incidence analysis. ISP Working Paper 01-10, Andrew Young School of Policy Studies, Georgia State University.
- McCorrison, S., Hemming, D.J., Lamontagne-Godwin, J.D., Parr, M.J, Osborn, J. and Roberts, P.D. (2013). What is the evidence of the impact of agricultural trade liberalisation on food security in developing countries? A systematic review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Olinto, P. and Saavedra, J. (2012). Inequality in Focus. The World Bank. Poverty Reduction and Equity Department. Volume 1. Number 1. April 2012.
- Ostry, J.D., Berg, A. and Tsangarides, C. G. (2014). Redistribution, Inequality and Growth. IMF Staff Discussion note. SDN/14/02.
- Popay, J., Rogers, A. and Williams, G. (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*, 8, (3), 341-351.
- Porto, G. (2006), Using survey data to assess the distributional effects of trade policy. *Journal of International Economics*, 70(1), 140-160.
- Ravallion, M. (1997). Can high-inequality developing countries escape absolute poverty? *Economics Letters*, 56, 51-57.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Robertson, R. (2000). Trade liberalisation and wage inequality: Lessons from the Mexican experience. *The World Economy*, 23(6), 827-849.

Rodgers, M, Sowden, A, Petticrew, M, Arai, L, Roberts, H, Britten, N and Popay, J (2009). Testing methodological guidance on the conduct of narrative synthesis in systematic reviews: effectiveness of interventions to promote smoke alarm ownership and function. *Evaluation*, 15 (1), 49-73.

Sahn, D. E., & Alderman, H. (1996). The effect of food subsidies on labor supply in Sri Lanka. *Economic Development and Cultural Change*, 45(1), 125-145.

Slavin, R.E. (1986). Best-Evidence Synthesis: An Alternative to Meta-Analytic and Traditional Reviews. *Educational Researcher* 15(9), 5-11.

Silva, J. A. (2007). Trade and Income Inequality in a Less Developed Country: The Case of Mozambique. *Economic Geography*, 83, 111-136.

Silva, J.A. (2013). Rural income inequality in Mozambique: National dynamics and local experiences, *Review of Regional Studies*, 43, 1, 23-50.

Spencer, L., Ritchie, J., Lewis, J. and Dillon, L. (2003). Quality in qualitative evaluation: a framework for assessing research evidence [monograph online]. London: Cabinet Office. Available from: www.gsr.gov.uk/evaluating_policy/era_papers/qual_eval.asp

Stiglitz, J. (2013). *The Price of Inequality*. Penguin.

Subramanian, S. V. and I. Kawaki (2006). Being well and doing well: on the importance of income for health. *International Journal of Social Welfare*, 15, (Suppl. 1), S13-S22.

Van de Walle, D. (1998). Assessing the welfare impacts of public spending. *World Development* 26(3): 365-379.

Veenstra, G. (2005). Social status and health: absolute deprivation or relative comparison, or both? *Health Sociologic Review*, 14, 121-134.

Walsh, D. and Downe, S. (2006). Appraising the quality of qualitative research, *Midwifery*, 22, 108-119.

White, H. and Waddington, H. (2012). Why do we care about synthesis? An introduction to the special issue on systematic reviews. *Journal of Development Effectiveness*, 4(3), 351-358.

Wilcox, R.R. (2008). Sample Size and Statistical Power, in: A. M. Nezu and C. M. Nezu (eds.) *Evidence Based Outcome Research : A Practical guide to conducting randomized controlled Trials for Psychosocial Interventions*, Oxford University Press, Oxford.

Wilkinson, R. and Pickett, K. (2009). *The Spirit Level: Why More Equal Societies Almost Always Do Better*. Allen Lane.

Williams, K. R. (1984). Economic Sources of Homicide: Reestimating the effects of Poverty and Inequality. *American Sociologic Review*, 49, 283-289.

Wood, A. (1997). Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom. *The World Bank Economic Review*, 11(1), 33-57.

World Bank (2005). *World Development Report 2006: Equity and Development*. Oxford University Press.

Appendices

Appendix 1: Authorship of the review

Advisory Group

Andy McKay

Professor of Development Economics, University of Sussex

Email: A.Mckay@sussex.ac.uk

Andy McKay researches on development economics, especially in relation to poverty/inequality and how these are impacted by policy (trade, fiscal etc.); on pro-poor growth; on agriculture; and on international trade. His geographic area of expertise is predominantly Africa, especially East and West Africa. He is main supervisor or co-supervisor of five DPhil students. In Sussex he co-organises the ESRC Development Economics conferences, and organised the first one on economic growth in Sussex in September 2008 (<http://www.sussex.ac.uk/economics/1-4-4.html>). Within Sussex he co-organises internal workshops on development economics with the Institute for Development Studies (<http://www.ids.ac.uk/>). Andy is an associate director of the DFID-funded Chronic Poverty Research Centre (www.chronicpoverty.org), within which he is an active researcher. He acts as a resource person for the biannual workshops of the African Economic Research Consortium (www.aercafrica.org), a role he has played since 2005, and serves on the steering committee of a collaborative research project on growth-poverty reduction linkages in Africa. Andy has extensive experience of giving policy advice to bilateral donors including DFID, international organisations and governments of developing countries. He also has significant experience of giving short courses, especially for government employees, in the north and south. He has previously worked at the Universities of Nottingham (1992 to 2003) and Bath (2003 to 2006); as well as the Overseas Development Institute (2001 to 2005). He has also worked on many consultancy assignments.

Birte Snilstveitm

Evaluation Specialist, 3ie

Email: bsnilstveit@3ieimpact.org

Birte assists in coordinating and developing 3ie's Systematic Reviews Programme in London. She works with Hugh Waddington in coordinating the Reviews commissioned by 3ie and the Campbell Collaboration International Development Coordinating Group. She is a co-author for Reviews on "Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries" and "Interventions to promote social cohesion in sub-Saharan Africa" as well as a scoping study on "Impact evaluation and interventions to address climate change". Birte started working for 3ie after completing her MA in Political Economy of Development from University of Birmingham, UK. She also holds a BSc in Politics with Sociology from Aston University.

Hugh Waddington

Senior Evaluation Specialist, 3ie

Email: hwaddington@3ieimpact.org

Hugh is an Economist by training and manages 3ie's Systematic Reviews Programme and the Campbell Collaboration International Development Coordinating Group. Before joining 3ie,

Hugh was employed as an Overseas Development Institute (ODI) Fellow in the Ministry of Finance of the Government of Rwanda, where he worked on the elaboration of Rwanda's Economic Development and Poverty Reduction Strategy (PRSP2). He has previously worked with the Economist Intelligence Unit, UK National Audit Office, Save the Children UK and the World Bank's Operations Evaluation Department. Hugh holds an MA in Development Economics from the University of Sussex.

Peter Lloyd- Sherlock

Professor of Social Policy and International Development, University of East Anglia

Email: p.lloyd-sherlock@uea.ac.uk

Peter's main area of research looks at social protection, health and the wellbeing of older people in developing countries. He is also interested in the economic and social effects of non-communicable diseases, such as stroke, heart disease and Alzheimer's Disease. He has a more general interest in social policy, particularly in Latin America. Contrary to popular belief, more older people live in the developing world than in the rich north. Despite this, the condition of older people and the wider effects of population ageing are still seen as peripheral concerns in development policy. He has been involved in studies of older people's wellbeing and vulnerability in Argentina, Brazil, South Africa and Thailand. He is currently involved in two funded research projects examining the wellbeing of older people and their families in eight different countries. He works closely with a wide range of international development agencies. Between 2011 and 2012 he was a Senior Research Fellowship at the UK Government's Department for International Development, providing advice on social development and social protection. Between 2010 and 2011 he was seconded to the World Health Organisation's Ageing and Lifecourse Programme, as lead planner for a new WHO Programme on Primary Healthcare for Older People. He has also worked with the UN Secretary General's Office to promote national capacity for mainstreaming age into development policy. As part of this, he was primary author for a United Nations Report "Guide to the National Implementation of the Madrid International Plan of Action on Ageing" (February 2008).

Richard Morgan

(formerly) Senior Advisor, Executive Office, UNICEF; Save the Children UK (current from 2014)

Email: richgmorgan@hotmail.com

Richard Morgan was recently the Senior Advisor on the Post-2015 Development Agenda at the United Nations Children's Fund (UNICEF), responsible for promoting children's rights and equity through participating and engaging in the multi-stakeholder processes leading up to 2015 and the expiry of the current Millennium Development Goals (MDGs). Richard was appointed to this position on 1 March 2012 by UNICEF's Executive Director, Anthony Lake. Previous to this appointment, Richard was the Director of Policy and Practice at UNICEF Headquarters in New York from (2009-2012), responsible for organizational standards and practice in the areas of gender, children's human rights, child and youth participation and communication for development. Richard also led UNICEF's work on child statistics and advocacy for child-focused social and economic policies. Earlier, Richard worked in Africa for more than two decades as a development planner, rural development specialist, economic and social sector advisor and humanitarian programme manager. He was a civil servant for several years with the Government of Botswana at central and local government levels, before joining UNICEF. Richard is a national of the United Kingdom of Britain. He is a graduate of the University of Oxford, UK (BA in Economics, Politics and Philosophy) and obtained his Masters' Degree in Development Economics at the University of East Anglia. Most recently he has joined Save the Children UK as a senior advisor on social protection.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Conflict of interest

There is no conflict of interest arising from financial sources.

Sources of funding

UK's Department for International Development (DFID)

Contact Details

Name	Email	Address
Dr Edward Anderson	e.anderson@uea.ac.uk	School of International Development, University of East Anglia Norwich NR4 7TJ UK
Dr Lucio Esposito	l.esposito@uea.ac.uk	
Dr Maren Duvendack	m.duvendack@uea.ac.uk	
Dr Maria Ana Jalles d'Orey	m.jalles-dorey@uea.ac.uk	

Appendix 2: Country classifications

A1. Countries that have always been low or middle income

Afghanistan	Guatemala	Panama
Albania	Guinea	Papua New Guinea
Algeria	Guinea-Bissau	Paraguay
Angola	Guyana	Peru
Argentina	Haiti	Philippines
Armenia	Honduras	Romania
Azerbaijan	India	Rwanda
Bangladesh	Indonesia	Samoa
Belarus	Iran, Islamic Rep.	São Tomé and Príncipe
Belize	Iraq	Senegal
Benin	Jamaica	Serbia
Bhutan	Jordan	Serbia and Montenegro (former)
Bolivia	Kazakhstan	Seychelles
Bosnia and Herzegovina	Kenya	Sierra Leone
Botswana	Kiribati	Solomon Islands
Brazil	Korea, Dem. Rep.	Somalia
Bulgaria	Kosovo	South Africa
Burkina Faso	Kyrgyz Republic	South Sudan
Burundi	Lao PDR	Sri Lanka
Cambodia	Lebanon	St. Lucia
Cameroon	Lesotho	St. Vincent and the Grenadines
Cape Verde	Liberia	Sudan
Central African Republic	Libya	Suriname
Chad	Macedonia, FYR	Swaziland
China	Madagascar	Syrian Arab Republic
Colombia	Malawi	Tajikistan
Comoros	Malaysia	Tanzania
Congo, Dem. Rep.	Maldives	Thailand
Congo, Rep.	Mali	Timor-Leste
Costa Rica	Marshall Islands	Togo
Côte d'Ivoire	Mauritania	Tonga
Cuba	Mauritius	Tunisia
Czechoslovakia (former)	Mexico	Turkey
Djibouti	Micronesia, Fed. Sts.	Turkmenistan
Dominica	Moldova	Tuvalu
Dominican Republic	Mongolia	Uganda
Ecuador	Montenegro	Ukraine
Egypt, Arab Rep.	Morocco	USSR (former)
El Salvador	Mozambique	Uzbekistan
Eritrea	Myanmar	Vanuatu
Ethiopia	Namibia	Venezuela, RB
Fiji	Nepal	Vietnam
Gabon	Nicaragua	West Bank and Gaza
Gambia, The	Niger	Yemen, Rep.
Georgia	Nigeria	Yugoslavia (former)
Ghana	Pakistan	Zambia
Grenada	Palau	Zimbabwe

What policies and interventions have been strongly associated with changes in in-country income inequality?

A2. Countries that have sometimes been high income

	High income during the following years
American Samoa	1987-89
Antigua and Barbuda	2002, 2005-8, 2012-
Aruba	1987-90, 1994-
Bahrain	1987-89, 2001-
Barbados	1989, 2000, 2002, 2006-
Chile	2012-
Croatia	2008-
Cyprus	1988-
Czech Republic	2006-
Equatorial Guinea	2007-
Estonia	2006-
Gibraltar	2009-
Greece	1996-
Guam	1987-89, 1995-
Hungary	2007-11
Isle of Man	1987-89, 2002-
Korea, Rep.	1995-97, 2001-
Latvia	2009, 2012-
Lithuania	2012-
Macao SAR, China	1994-
Malta	1989, 1998, 2000, 2002-
Mayotte	1990
Netherlands Antilles (former)	1994-
New Caledonia	1994-
Northern Mariana Islands	1995-2001, 2007-
Oman	2007-
Poland	2009-
Portugal	1994-
Puerto Rico	1989, 2002-
Russia	2012-
Saudi Arabia	1987-89, 2004-
Slovak Republic	2007-
Slovenia	1997-
St. Kitts and Nevis	2011-
Trinidad and Tobago	2006-
Uruguay	2012-

A3. Countries that have always been high income

Andorra	Italy
Australia	Japan
Austria	Kuwait
Bahamas, The	Liechtenstein
Belgium	Luxembourg
Bermuda	Monaco
Brunei Darussalam	Netherlands
Canada	New Zealand
Cayman Islands	Norway
Channel Islands	Qatar
Curaçao	Singapore
Denmark	Sint Maarten (Dutch part)
Faeroe Islands	Spain
Finland	St. Martin (French part)
France	Sweden
French Polynesia	Switzerland
Germany	Taiwan, China
Greenland	Turks and Caicos Islands
Hong Kong SAR, China	United Arab Emirates
Iceland	United Kingdom
Ireland	United States
Israel	Virgin Islands (U.S.)

Source: <http://data.worldbank.org/about/country-classifications/country-and-lending-groups>⁸

⁸ This list includes the group of LIMC countries as defined today by the World Bank. We will however take into account any country listed as part of the LMIC category at the time when specific interventions took place.

Appendix 3: Concepts for search strategy in English, Portuguese and Spanish

A Policy/ Política/ Política	B Income/ Rendimento/ Ingreso ¹	C Inequality/ Desigualdade/ Desigualdad
Polic* / Política*/ Política*	Income* / Rendimento* / Ingreso*	*Equal*/ *Iguar*/ *Iguar*
Intervention* / Intervenç*/ Intervenc*	Expenditure / Despesa* / Gasto*	*Distribut*/ *Distribu*/ *Distribu*
Program* / Programa* / Programa*		Disparit*/ Disparidade*/ Disparidad*
Instrument* / Instrumento* / Instrumento*		Differen*/ Diferen*/ Diferen*
Tool* / Ferramenta* / Ferramenta*		Gap / Divisão / División
Reform* / Reforma* / Reforma*		*Equit*/ *Equidade*/ *Equidad*
Legislation* / Legislaç*/ Legislac*		Ratio* / Rácio*/ Ratio*
Govern* / Governo* / Govern*		Gini / Gini / Gini

Note: The terms appear in the following order: English/Portuguese/Spanish. Other Portuguese and Spanish synonyms of the words presented in this table might be considered. However, after consulting a large number of studies, these are the terms that consistently come out of the literature. ¹“Renda” (PT) or “Renta” (SP) are another common words used to refer to “Rendimento” or “Ingreso”, respectively.

Appendix 4: Search strategy checklist (provided by DFID)⁹

Aspect of search		Actions and notes
Section A Search Sources	Are the following used: - Bibliographic databases - Library catalogues - Specialised registers - Regional databases - Search engines - Websites	✓ (See Table 2) ✓ (See Table 2 & 5) Not applicable ✓ (See Table 5) ✓ (See Table 5) ✓ (See Table 5)
	What disciplines does the topic cover, and are these reflected in the search strategy?	Social Sciences in general. Specific disciplines: economics & international development. Yes, the search strategy reflects these.
	Does the choice of search sources reflect any geographical focus and/or study design?	Geographical focus reflected by using databases in Portuguese and Spanish. No choice based on study design.
	Are the types of publication sought reflected in the search strategy? (e.g. conference proceedings, government publications, dissertations, books)	✓ (See Table 2 & 5)
	List additional database search sources:	
	Are relevant websites, organisations, search engines to be searched?	✓ (See Table 5)
	List additional websites and organisations:	
	Are other search methods described? - handsearching - reference checking - forward citation searching - author and expert contact	(See 3.2.2/Other searches) ✓ ✓ ✓ ✓
Section B Search concepts	Does the search strategy match the research question?	✓
	Are the search concepts clear?	✓
	Are there too many search concepts?	Moderate (See Table 3 for a list of concepts)
	Are the search concepts too narrow or too broad?	Broad
	Does the search appear to retrieve too many or too few records?	Too many records (an initial search in SCOPUS using a long string version over 9000 hits).
	Are the concepts combined with appropriate Boolean logic?	✓
	If NOT is used, are there likely to be any unintended consequences?	We do not expect to be using the operator "AND NOT".
	How will the search be adapted for each database?	2 search strings: long and short versions (See Table 4). Search strings tailored to each database according to the available boolean operators.
Section C Search Terms		
a) Controlled terms	Are relevant controlled vocabulary terms used for each concept?	Depends on the database. A few websites only allow for controlled index language.
	Are they appropriately exploded for narrower terms?	✓

⁹ Checklist prepared by: *C Stansfield*, EPPI-Centre (2011). Adapted from PRESS Sampson, M; McGowan, J; Lefebvre, C; Moher, D; Grimshaw, J (2008) *PRESS: Peer Review of Electronic Search Strategies*. Ottawa: Canadian Agency for Drugs and Technologies in Health.

What policies and interventions have been strongly associated with changes in in-country income inequality?

	Are any headings too broad or too narrow?	Generally broad
	Any other suitable controlled terms?	Not aware of.
	Could sub-headings be used instead of subject headings or vice versa (not always applicable)	✓ This is applicable in development organizations websites where sub-headings/sub-topics might be considered.
	Do any terms appear irrelevant?	No
b) Natural language/Free text terms	Are relevant free-text field terms used for each concept?	✓ See Table 3 for a list of all free-text terms used
	Are there other suitable terms?	The terms presented in Table 2 pretend to cover a wide range of words for policy, income and inequality. Other terms could complement these, but after some initial searches they seem irrelevant.
	Are there any irrelevant or excessively broad terms?	Broad terms are limited by using proximity operators.
	Are Boolean terms nested within brackets correct?	✓
	If AND is used, could precision be improved by proximity searching (adjacent, near, within), or phrase searching?	✓ A proximity operator W/n with n=1 is used to connect the terms "income" and "inequality". Please see Tables 3 & 4.
	Are spellings correct?	✓
	Are there any variants of spellings (e.g. UK/US spellings) that need to be considered?	✓ e.g. program/programme. Program* will be used to capture both versions.
	Are there other synonyms?	A comprehensive list of synonyms is used (See Table 3)
	Is truncation used correctly?	✓
	Any language or technical jargon terms that are relevant (even where outdated, but within the timescale of the search)?	✓ (See Table 3) Terms will also be searched in Portuguese and Spanish (See Appendix 2)
	If there are any acronyms or abbreviations, are these also given in full format, and vice versa?	Not aware of.
	Are there some terms that are redundant?	No
c) Syntax	Are there any errors in the system syntax or line numbers?	No.
	Limits and filters	Yes. See below.
	Are there existing limits or filters that may be useful? E.g. (human, date Do any limits seem unwarranted? limits, publication type, study design)	Filtered for Social Sciences, when that option is available. Filter for studies after 1945. Filter for publication type on websites of development organizations. No filter based on country or study design.
Section D Special considerations	Is the time period for the literature search defined?	✓ Yes. After 1945.
	Is the search in line with the type of review? (e.g. scoping review, rapid review, full systematic review)	✓

Appendix 5: Data extraction tools

1. Mechanism of assignment: was the allocation or identification mechanism able to control for selection bias?

a) For Randomized assignment (RCTs),

Score “YES” if:

- a random component in the sequence generation process is described (e.g. referring to a random number table);
- and if the unit of allocation was at group level (geographical/ social/ institutional unit) and allocation was performed on all units at the start of the study;
- or if the unit of allocation was by beneficiary or group and there was some form of centralised allocation mechanism such as an on-site computer system;
- and if the unit of allocation is based on a sufficiently large sample size to equate groups on average.

Score “UNCLEAR” if:

- the paper does not provide details on the randomization process, or uses a quasi-randomization process for which it is not clear has generated allocations equivalent to true randomization.

Score “NO” if:

- the sample size is not sufficient or any failure in the allocation mechanism could affect the randomization process.

b) For discontinuity assignment (Regression Discontinuity Designs)

Score “YES” if:

- allocation is made based on a pre-determined discontinuity on a continuous variable (regression discontinuity design) and blinded to participants or;
- if not blinded, individuals reasonably cannot affect the assignment variable in response to knowledge of the participation decision rule;
- and the sample size immediately at both sides of the cut-off point is sufficiently large to equate groups on average.

Score “UNCLEAR” if:

- the assignment variable is either non-blinded or it is unclear whether participants can affect it in response to knowledge of the allocation mechanism.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Score “NO” if:

- the sample size is not sufficient or;
- there is evidence that participants altered the assignment variable prior to assignment.

c) For assignment based non-randomized programme placement and self-selection (studies using a matching strategy or regression analysis, excluding IV),

Score “YES” if:

- participants and non-participants are either matched based on all relevant characteristics explaining participation and outcomes, or;
- all relevant characteristics are accounted for.

Score “UNCLEAR” if:

- it is not clear whether all relevant characteristics (only relevant time varying characteristics in the case of panel data regressions) are controlled.

Score “NO” if:

- relevant characteristics are omitted from the analysis.

d) For identification based on an instrumental variable (IV estimation),

Score “YES” if:

- an appropriate instrumental variable is used which is exogenously generated: e.g. due to a ‘natural’ experiment or random allocation.

Score “UNCLEAR” if:

- the exogeneity of the instrument is unclear (both externally as well as why the variable should not enter by itself in the outcome equation).

Score “NO” otherwise.

2. Group equivalence: was the method of analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding?

a) For randomized control trials (RCTs) and quasi-RCTs,

Score “YES” if:

- baseline characteristics of the study and control/comparisons are reported and overall similar based on t-test or ANOVA for equality of means across groups;
- or covariate differences are controlled using multivariate analysis;
- and the attrition rates (losses to follow up) are sufficiently low and similar in treatment and control, or the study assesses that loss to follow up units are random draws from the sample (e.g. by examining correlation with determinants of outcomes, in both treatment and comparison groups);
- and problems with cross-overs and drop outs are dealt with using intention-to-treat analysis or in the case of drop outs, by assessing whether the drop outs are random draws from the population;
- and, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g. weather, infrastructure, community fixed effects, etc.) through multivariate analysis.

Score “UNCLEAR” if:

- insufficient details are provided on covariate differences or methods of adjustment;
- or insufficient details are provided on cluster controls.

Score “NO” otherwise.

b) For regression discontinuity designs (RDDs),

Score “YES” if:

- the interval for selection of treatment and control group is reasonably small;
- or authors have weighted the matches on their distance to the cut-off point;
- and the mean of the covariates of the individuals immediately at both sides of the cut-off point (selected sample of participants and non-participants) are overall not statistically different based on t-test or ANOVA for equality of means;
- or significant differences have been controlled in multivariate analysis;
- and, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g. weather, infrastructure, community fixed effects, etc.) through multivariate analysis.

Score “UNCLEAR” if:

- there are covariate differences across individuals at both sides of the discontinuity which have not been controlled for using multivariate analysis, or if insufficient details are provided on controls;
- or if insufficient details are provided on cluster controls.

What policies and interventions have been strongly associated with changes in in-country income inequality?

Score “NO” otherwise.

c) For non-randomized trials using difference-in-differences methods of analysis,

Score “YES” if:

- the authors use a difference-in-differences (or fixed effects) multivariate estimation method;
- the authors control for a comprehensive set of time-varying characteristics;
- and the attrition rate is sufficiently low and similar in treatment and control, or the study assesses that drop-outs are random draws from the sample (e.g. by examining correlation with determinants of outcomes, in both treatment and comparison groups);
- and, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g. weather, infrastructure, community fixed effects, etc.) through multivariate analysis.

Score “UNCLEAR” if:

- insufficient details are provided;
- or if insufficient details are provided on cluster controls.

Score “NO” otherwise.

d) For statistical matching studies including propensity scores (PSM) and covariate matching,

Score “YES” if:

- matching is either on baseline characteristics or time-invariant characteristics which cannot be affected by participation in the programme; and the variables used to match are relevant (e.g. demographic and socio-economic factors) to explain both participation and the outcome (so that there can be no evident differences across groups in variables that might explain outcomes);
- in addition, for PSM Rosenbaum’s test suggests the results are not sensitive to the existence of hidden bias;
- and, with the exception of Kernel matching, the means of the individual covariates are equated for treatment and comparison groups after matching;
- and, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g. weather, infrastructure, community fixed effects, etc.) through multivariate or any appropriate analysis.

Score “UNCLEAR” if:

- relevant variables are not included in the matching equation, or if matching is based on characteristics collected at endline;
- or if insufficient details are provided on cluster controls.

Score “NO” otherwise.

e) For regression-based studies using cross sectional data (excluding IV)

Score “YES” if:

- the study controls for relevant confounders that may be correlated with both participation and explain outcomes (e.g. demographic and socio-economic factors at individual and community level) using multivariate methods with appropriate proxies for unobservable covariates;
- and a Hausman test with an appropriate instrument suggests there is no evidence of endogeneity;
- and none of the covariate controls can be affected by participation;
- and either, only those observations in the region of common support for participants and non-participants in terms of covariates are used, or the distributions of covariates are balanced for the entire sample population across groups;
- and, for cluster-assignment, authors control particularly for external cluster-level factors that might confound the impact of the programme (e.g. weather, infrastructure, community fixed effects, etc.) through multivariate analysis.

Score “UNCLEAR” if:

- relevant confounders are controlled but appropriate proxy variables or statistical tests are not reported;
- or if insufficient details are provided on cluster controls.

Score “NO” otherwise.

f) For instrumental variables approaches,

Score “YES” if:

- the instrumenting equation is significant at the level of $F \geq 10$ (or if an F test is not reported, the authors report and assess whether the R-squared (goodness of fit) of the participation equation is sufficient for appropriate identification);
- the identifying instruments are individually significant ($p \leq 0.01$); for Heckman models, the identifiers are reported and significant ($p \leq 0.05$);

What policies and interventions have been strongly associated with changes in in-country income inequality?

- where at least two instruments are used, the authors report on an over-identifying test ($p \leq 0.05$ is required to reject the null hypothesis); and none of the covariate controls can be affected by participation and the study convincingly assesses qualitatively why the instrument only affects the outcome via participation;
- and, for cluster-assignment, authors particularly control for external cluster-level factors that might confound the impact of the programme (e.g. weather, infrastructure, community fixed effects, etc.) through multivariate analysis.

Score “UNCLEAR” if:

- relevant confounders are controlled but appropriate statistical tests are not reported or exogeneity of the instrument is not convincing;
- or if insufficient details are provided on cluster controls (see category f) below).

Score “NO” otherwise.

3. Hawthorne and John Henry effects: was the process of being observed causing motivation bias?

Score “YES” if either:

- a) For data collected in the context of a particular intervention trial (randomized or non-randomized assignment), the authors state explicitly that the process of monitoring the intervention and outcome measurement is blinded, or argue convincingly why it is not likely that being monitored in ways that could affect the performance of participants in treatment and comparison groups in different ways.
- b) The study is based on data collected in the context of a survey, and not associated with a particular intervention trial, or data are collected in the context of a retrospective (ex post) evaluation.

Score “UNCLEAR” if:

- it is not clear whether the authors use an appropriate method to prevent Hawthorne and John Henry Effects (e.g. blinding of outcomes and, or enumerators, other methods to ensure consistent monitoring across groups).

Score “NO” otherwise.

4. Spill-overs: was the study adequately protected against performance bias?

Score “YES” if:

- the intervention is unlikely to spill-over to comparisons (e.g. participants and non-participants are geographically and/or socially separated from one another and general equilibrium effects are unlikely).

Score “UNCLEAR” if:

- spill-overs are not addressed clearly.

Score “NO” if:

- allocation was at individual or household level and there are likely spill-overs within households and communities which are not controlled for in the analysis;
- or if allocation at cluster level and there are likely spill-overs to comparison clusters.

5. Selective outcome reporting: was the study free from outcome reporting bias?

Score “YES” if:

- there is no evidence that outcomes were selectively reported (e.g. all relevant outcomes in the methods section are reported in the results section).

Score “NO” if:

- some important outcomes are subsequently omitted from the results or the significance and magnitude of important outcomes was not assessed.

Score “UNCLEAR” otherwise.

6. Selective analysis reporting: was the study free from analysis reporting bias?

Score “YES” if:

- authors use ‘common’ methods of estimation and the study does not suggest the existence of biased exploratory research methods.

Score “NO” if:

- authors use uncommon or less rigorous estimation methods such as failure to conduct multivariate analysis for outcomes equations where it has not been established that covariates are balanced.

See also the following for particular estimation methodologies.

For PSM and covariate matching, score “YES” if:

What policies and interventions have been strongly associated with changes in in-country income inequality?

- where over 10% of participants fail to be matched, sensitivity analysis is used to re-estimate results using different matching methods (Kernel Matching techniques);
 - for matching with replacement, no single observation in the control group is matched with a large number of observations in the treatment group.
- Where not reported, score “UNCLEAR”. Otherwise, score “NO”.

For IV (including Heckman) models, score “YES” if:

- the authors test and report the results of a Hausman test for exogeneity ($p \leq 0.05$ is required to reject the null hypothesis of exogeneity);
 - the coefficient of the selectivity correction term (Rho) is significantly different from zero ($P < 0.05$) (Heckman approach).
- Where not reported, score “UNCLEAR”. Otherwise, score “NO”.

For studies using multivariate regression analysis, score “YES” if:

- authors conduct appropriate specification tests (e.g. reporting results of multicollinearity test, testing robustness of results to the inclusion of additional variables, etc).
- Where not reported or not convincing, score “UNCLEAR”. Otherwise, Score “NO”.

7. Other: was the study free from other sources of bias?

Important additional sources of bias may include: concerns about blinding of outcome assessors or data analysts; concerns about blinding of beneficiaries so that expectations, rather than the intervention mechanisms, are driving results (detection bias or placebo effects); concerns about courtesy bias from outcomes collected through self-reporting; concerns about coherence of results; data on the baseline collected retrospectively; information is collected using an inappropriate instrument (or a different instrument/at different time/after different follow up period in the comparison and treatment groups).

Score “YES” if:

- the reported results do not suggest any other sources of bias.

Score “UNCLEAR” if:

- other important threats to validity may be present

Score “NO” if:

- it is clear that these threats to validity are present and not controlled for.

First produced in 2014 by:

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR

Tel: +44 (0)20 7612 6397

<http://eppi.ioe.ac.uk/>

<http://www.ioe.ac.uk/ssru/>

The **Evidence for Policy and Practice Information and Co-ordinating Centre** (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

The views expressed in this work are those of the authors and do not necessarily reflect the views of the EPPI-Centre or the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:

telephone: +44 (0)20 7947 9556 email: info@ioe.ac.uk