Protocol: Effects of El Niño and positive Indian Ocean Dipole on health, food security, migration, economics, and conflicts in the Indo-Pacific: A systematic review

Andrea Floridi
International Initiative for Impact Evaluation (3ie)

Tomasz Kozakiewicz
3ie

Daniela Anda Leon
3ie

Anilkrishna Thota
3ie

Megha Bhattacharyya
3ie

Shannon Shisler
3ie

Zahra Premji
Independent Information Specialist

Systematic Review Protocol

April 2024

International Initiative for Impact Evaluation
UK International Development
Partnership | Progress | Prosperity
UNIVERSITY OF BIRMINGHAM

About the Research Commissioning Centre

The Foreign, Commonwealth and Development Office (FCDO) Research Commissioning Centre (RCC) has been established to commission and manage research to enhance development and foreign policy impact. Led by the International Initiative for Impact Evaluation (3ie), the University of Birmingham, and an unmatched consortium of UK and global research partners, the RCC aims to commission different types of high-quality research in FCDO's key priority areas.

About systematic reviews

Systematic reviews appraise and synthesise the available high-quality evidence to answer specific research questions in a rigorous, objective and stepwise process. . These reviews follow scientifically recognised review methods, and they are peer-reviewed and quality-assured according to internationally accepted standards.

About this systematic review protocol

This report presents the protocol for a systematic review to gather and synthesize the evidence on the effects of El Niño and the positive Indian Ocean Dipole on health, economics, conflicts, migration, and food security across low- and middle-income countries in the Indo-Pacific region.

Funding

Funding for this systematic review has been provided by the UK's Foreign, Commonwealth & Development Office (FCDO) through the Research Commissioning Centre. The content of this report does not necessarily reflect the views of FCDO or 3ie and its partners who manage the Research Commissioning Centre. The authors bear sole responsibility for the content of this report, and any errors and omissions are the authors' sole responsibility. Please direct any comments or queries to the corresponding author, Andrea Floridi, at afloridi@3ieimpact.org.

Suggested citation: Floridi, Andrea, Tomasz Kozakiewicz, Daniela Anda Leon, Anilkrishna Thota, Megha Bhattacharyya, Shannon Shisler and Zahra Premji. 2024. Protocol: *Effects of El Niño and positive Indian Ocean Dipole on health, food security, migration, economics, and conflicts in the Indo-Pacific*: A systematic review. New Delhi: International Initiative for Impact Evaluation (3ie).

Contents

List of figures and tables

1. Background

1.1. El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD)

Climate drivers and teleconnections such as El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), can impact multiple sectors in various geographies. ENSO is the combination of two interrelated phenomena: El Niño, that involves changes in the water surface temperature along the coasts of Eastern and Western Pacific Ocean; and the Southern Oscillation, consisting of changes in the air pressure. During El Niño phase the pressure become lower over the Eastern Pacific Ocean coasts (with more abundant rainfalls and flooding) and higher over the Western Pacific Ocean coasts (with less rainfalls and droughts). Bjerknes (1969) postulated that El Niño and the Southern Oscillation occur in close connection and are two aspects of the same phenomenon (hence, ENSO).

El Niño is a large-scale oceanic warming event that occurs in the eastern tropical Pacific Ocean, whilst the Southern Oscillation is characterized by an interannual seesaw in tropical sea level pressure (SLP) between the western and eastern Pacific, consisting of a weakening and strengthening of the easterly trade winds over the tropical Pacific (Wang et al., 2017). Usually, trade winds blow toward the west, bringing warmer water to the coasts of Oceania and East Asia, and facilitating the upwelling of cold and nutrient water from the bottom of the South American Pacific coasts. However, during El Niño the trade winds change direction, hindering the upwelling of more nutrient (and colder) water in South American coasts and preventing warmer water from reaching Oceanian and East Asian coasts with consequential drier weather and colder surface water temperature. The concomitant Southern Oscillation lowers the air pressure over the South American Ocean coasts, leading to more abundant rainfalls and flooding; and higher air pressure over the East Asian and Oceanian coasts (with less rainfall and droughts).

A peculiar type of El Niño is the El Niño Modoki (EM). This teleconnection is characterized by the upwelling of colder water in both the west and east coasts of the Pacific Ocean, with a concentration of humid warm air and warmer water in the central part of the Pacific Ocean. This facilitates colder and drier weather with reduced rainfall and droughts in the west coast of the Pacific Ocean, and in the west Indian Ocean (Salimum et al., 2014; Marathe and Karumuri, 2021). The cascading effects of EM on climate patterns, monsoon seasons, and tropical storms affect both the Pacific Ocean and Indian Ocean regions (Feba et al., 2021).

The IOD is the difference in temperatures of the surface water between the Western and the Eastern poles of the Indian Ocean. Positive phase of the IOD (+IOD) is characterized by warmer surface water in the Western Indian Ocean region, whilst the Negative IOD (-IOD) sees warmer surface water in the Eastern Pole. Like ENSO, the IOD brings drastic changes to weather patterns with warmer and more humid conditions resulting in abundant monsoons and increased risk of flooding. Although considered distinct phenomena (Ashok et al., 2003), IOD is affected by the ENSO, and the two often occur in temporal proximity (Stuecker et al., 2017).

The effects of ENSO and IOD at the global level materialize as changes in the seasonal cycle, increased global temperature, and more frequent climate disasters, such as droughts, flooding, and fires. Furthermore, cascading effects include crop loss, food insecurity, infectious diseases, and cholera pandemics due to floodings, respiratory and cardiovascular disease due to hotter weather, collective and idiosyncratic economic shocks, displacement and land loss, migration, and conflicts.

The way that ENSO and +IOD are related to climate change is still unclear, however these climate drivers are increasing in duration (Cai et al., 2009; Yeh et al., 2009). On one side, ENSO waves are associated with higher global temperatures (Cai et al., 2021), and +IOD events account for the reduction in rainfall and consequential hotter and drier weather over the eastern coasts of the Pacific Ocean (Sanji et al., 2003). On the other side, global warming has reduced the frequency of canonical ENSO events over the East Pacific Ocean coasts in favour of ENSO events in the central Pacific Ocean (Yeh at al., 2009).

1.2. The considered teleconnections

This review takes stock of the evidence on the effects of El Niño and +IOD in the tropical Indo-Pacific region (South Asia, East Asia, South-East Asia, and Oceania). Given the complex nature of the two considered phenomena and the differences between El Niño and +IOD, we conceived the latter phenomena as two distinct treatments.

The effects of El Niño and +IOD on weather vary considerably (Zheng et al., 2014); consequentially, the respective effects on health and socio-economic outcomes are equally widely divergent. In Western Pacific Ocean countries, El Niño typically leads to hotter and drier weather conditions that increase cardiovascular and respiratory issues, undermine harvests with repercussions on income and food security, and lead to increased migration (especially internal migration) and conflicts. On the other hand, the effects of +IOD in the Western Indian Ocean coasts include more humid weather and abundant rainfalls that lead to increased risk of flooding; the latter eases the spread of diseases such as cholera and malaria and leads to displacement of people and conflicts. Positive IOD equally damages crop yields with negative effects on income, food security, and employment.

1.3. Effects of El Niño and +IOD: outlining a theory of change

This section outlines the theory of change underpinning the effects of El Niño and +IOD in the western Pacific Ocean region (Oceania, South-East Asia, and East Asia) and on the western Indian Ocean region. The theoretical framework of this review is drawn from the broader literature on the effects of El Niño and +IOD at local and regional levels.

1.3.1. El Niño

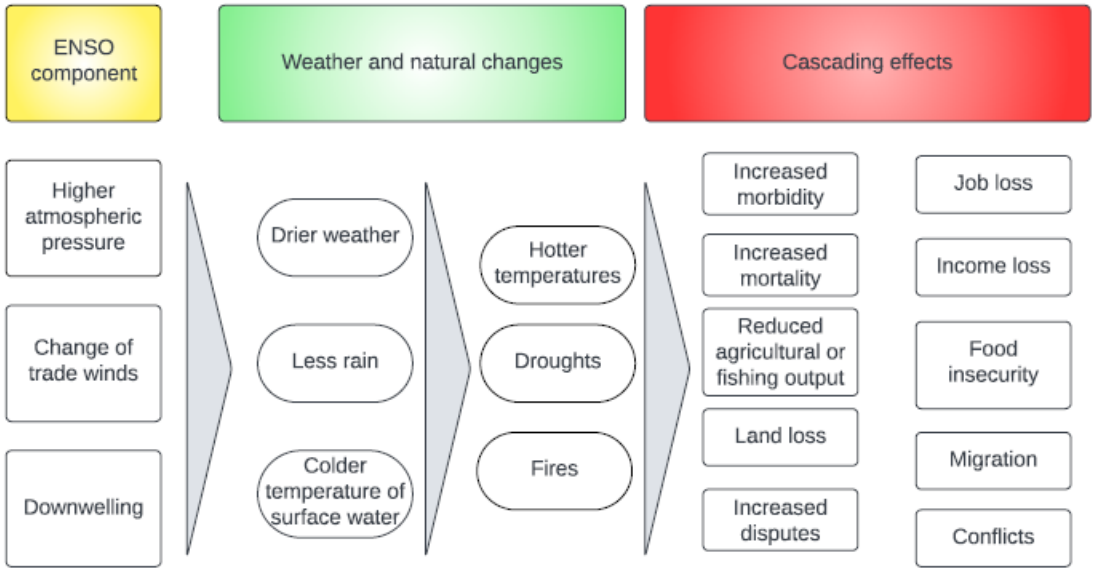
Countries in the Pacific and South-East Asian region are expected to experience cooling of surface water combined with higher atmospheric pressure. The combination of these two factors leads to drier weather and less rainfall, which in turn can lead to warmer temperatures. In these weather conditions, droughts and fires are more likely to occur with disruptive effects on the yields. In turn, the crop loss undermines income and food security, eventually spreading unemployment, and poverty. The simultaneous cooling down of the surface water temperature has equally negative effects on fish catches and aquaculture with cascading effects on income and conflicts (Hendrix et al., 2022).

In this context, adverse effects on health are likely to intensify. For example, drier and hotter weather can favour respiratory and cardiovascular disease especially among most vulnerable groups. Equally important are the effects on nutrition induced by income and employment loss.

Lastly, evidence suggests that conflicts are more frequent in the years of El Niño. For instance, Hendrix and colleagues (2022) report that fisheries disputes in the South Chinese Sea increased during the years of El Niño due to the reduction of fishing yields. In a similar vein, Hsiang, and

colleagues (2011) found that between 1950 and 2004, climate changes induced by El Niño increased the likelihood of civil conflicts. Figure 1 illustrates the cascading effects of El Niño in the Western tropical Pacific Ocean.

Figure 1: Flow diagram of cascading effects of El Niño



fa

1.3.2. Positive Indian Ocean Dipole (+IOD)

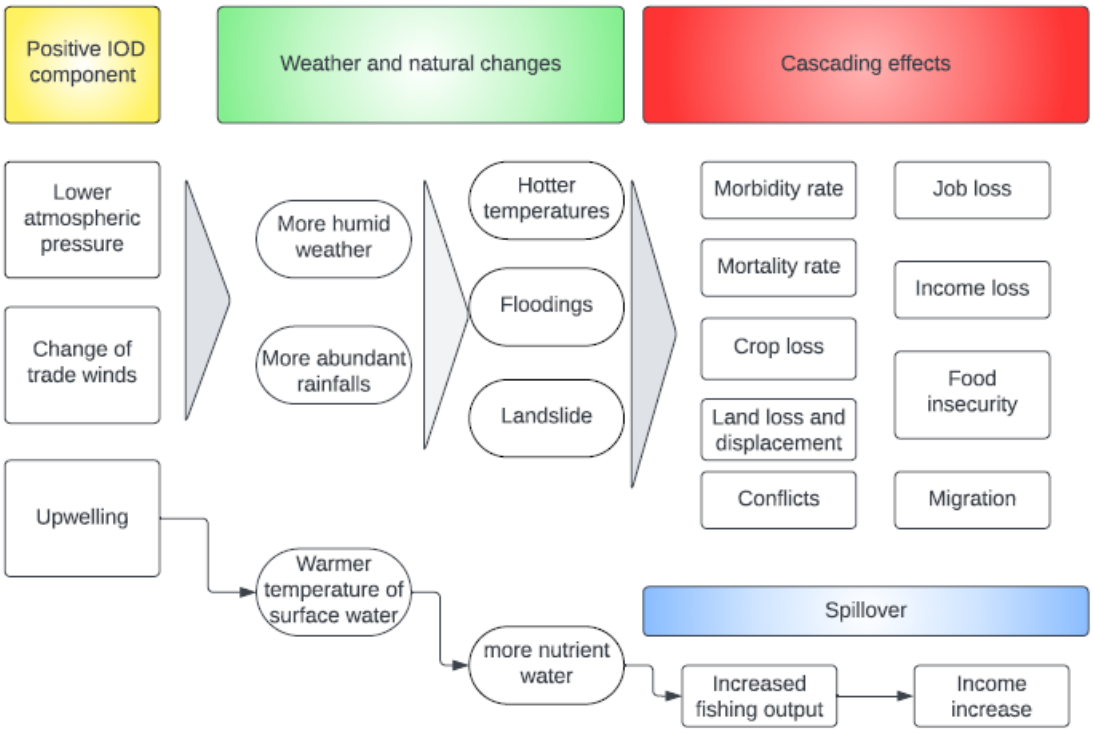
Positive IOD is expected to increase air pressure and water surface temperature of countries in the Western Indian Ocean pole. This leads to warmer water surface temperature and more humid weather, causing more intense monsoon rainfalls (Ashok et al., 2007). The excessive rainfalls increase the risks of flooding, which in turn facilitate the spreading of malaria, cholera pandemics, and other disease. For instance, Pascal and colleagues (2000) found evidence of higher incidence of cholera in Bangladesh during the years of El Niño over the period between 1980 and 1998.

The increased moisture in the soil caused by excessive rainfalls weakens its stability, making slopes more susceptible to landslides. The displacement of soil and obstruction of watercourses, coupled with the already intense rainfall, results in higher risks of flooding along with waterborne diseases such as cholera (Levy at al., 2016). The consequences of floodings and landslides go beyond the effects on health as they can destroy entire neighbourhoods, villages, and towns and lead to displacements of people. For instance, Zhou and colleagues (2021) bring evidence that the +IOD of 2019 is among the main causes of the Yangtze flooding of 2020 in China, which led to the displacement of millions of people in the middle of the Covid-19 pandemic emergency.

Positive IOD is also considered among the key factors affecting harvest and yields, particularly of rice, with negative consequences on income and food security (Ghose et al., 2021). For instance, Ghose and colleagues (2021) documented a sharper decline in rice production in correspondence with the +IOD months in Bangladesh. Despite causing damage to harvest and yields, during +IOD, warmer surface water temperatures of Western Indian Ocean coasts increase the nutrient level and attract more fish, thus resulting in increased fish captures (Chen

et al., 2023), thus improving income in the fishery sector. Figure 2 illustrates the cascading effects of +IOD over the northern and north-eastern Indian Ocean region.

Figure 2: Flow diagram of cascading effects of +IOD



1.4. Why it is important to do this review

Predictability of seasonal scale climate drivers such as ENSO and IOD, and their associated weather variations is increasingly important. There are already rapidly manifesting consequences of teleconnections in a changing climate. For instance, last year (2023) has witnessed recorded weather anomalies associated with the occurrence of the El Niño 2023-24. In this context, demand for more and better evidence of direct and indirect impacts at the local, national, and regional levels is required. In our scoping of the literature, we found literature reviews examining the impact of ENSO on health outcomes (Kovats et al., 2003, McGregor et al., 2018) and one systematic review of ENSO on diarrheal disease (Demissie et al., 2017). The latter included 30 studies but only a handful from Indo-Pacific countries.

Understanding compounding and cascading[1] socio-economic impacts from weather, seasonal climate variability, and associated hazards holds the potential to inform policy and enable actionable outcomes to minimise or optimise impacts.

This systematic review aims to identify the effects of El Niño and +IOD events on health, food security, migration, conflicts, and socio-economic outcomes in the Indo-Pacific. The research aims to draw insights to inform live policy discussions and future policy and actions regarding the current 2023 El Niño, future impacts from similar seasonal climate drivers (e.g. La Niña, and -IOD), as well as highlighting near-term climate security implications.

[1] https://www.chathamhouse.org/2021/09/climate-change-risk-assessment-2021/04-cascading-systemic-risks

2. Objectives

The scope of this systematic review is to gather, assess, and synthesize the available evidence on the impacts of El Niño and +IOD in the Indo-Pacific region over the past 34 years. Thus, the review aims to offer a synthesis of evidence to inform current El Niño and +IOD related response and diplomacy. To this end, the review explores the direct and indirect health, food security, socio-economic, migration, and conflict impacts of El Niño, +IOD and the combined El Niño and +IOD effects across Indo-Pacific countries.

The review will address the following questions:

1. What does the available evidence indicate about direction and magnitude of the effects of El Niño and +IOD across Indo-Pacific countries over the past 34 years?
2. How do effects vary by type of outcome (i.e., health, food security, conflicts, economics, and migration outcomes)?
3. How do effects vary with respect to a stronger or weaker El Niño and +IOD, and how does it vary between teleconnection types, (e.g., Canonical El Niño or Modoki El Niño)?
4. What are the main factors of variation accounting for the heterogeneity of findings (e.g., geographical area, and type of measurement)?
5. What is the risk of bias (or quality) of the available evidence?
6. What are the evidence gaps and how can future research address these?

3. Methods

3.1. Criteria for eligibility

The following sections describe the inclusion and exclusion criteria applied to determine whether a study can be included in the review. Drawing on the PICOS protocol, we identified the eligibility criteria to be employed in the review.

3.1.1. Type of population and setting

We will include studies of the effects of El Niño and +IOD on populations residing in Indo-Pacific countries classified as low and middle-income countries (L&MICs) by the World Bank. The geographical area of the studies will be classified based on the countries where the Indo-Pacific Directorate-General in the FCDO has lead responsibility for relations[2]. Namely, the review will include studies providing evidence on one or more of the following L&MICs (low and middle income countries) and territories: Indian subcontinent and Indian Ocean (Bangladesh, Bhutan, India, Maldives, Nepal, and Sri Lanka with the exception of Pakistan), South East Asia (Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Thailand, Timor Leste, Vietnam) and Oceania (Fiji, Kiribati, Marshall Islands, Micronesia, Nauru, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, and Vanuatu).

Studies including impacts from multiple countries will be included if results are provided separately for the selected Indo-Pacific countries outlined above.

3.1.2. Type of "Intervention"

This review covers two main "treatments" (or "interventions"); namely, El Niño and +IOD.

Considering the differences between El Niño and the +IOD, we conceived the phenomena as two distinct interventions. This consideration is motivated by the fact that the mechanisms underpinning the impacts of El Niño and +IOD are entirely different, making comparisons

[2]See: https://publications.parliament.uk/pa/cm5803/cmselect/cmfaff/172/report.html#heading-9

between the two phenomena trivial. On one hand, El Niño manifests in the Western Tropical Pacific Ocean with cooler temperatures of the surface water and less frequent rainfalls with consequential drier weather and droughts A peculiar type of El Niño is the so-called El Niño Modoki which is characterized by drier and colder climates with less rainfalls across the whole Indo-Pacific region and anomalies in the monsoon season and on tropical storms. On the other hand, +IOD is characterized by warmer temperatures of the surface water and more abundant rains resulting seldom in flooding and landslides.

The review will include studies providing evidence on the two types of teleconnections. Table 1 provides details on the mechanisms considered in the review.

Table 1: List of included teleconnections

Teleconnection Type	Mechanism	Description
Canonical El Niño	Colder surface water temperature	Reduction of surface water temperature
	Drier weather	Reduction of humidity of air
	Hotter air temperature	Increase in registered temperature of air in either or both coasts and inland
	Less rainfalls	Reduction of rainfalls in South-East Asia and Pacific regions
El Niño Modoki	Drier weather and less rainfalls	Reduction of humidity of air and reduction of rainfalls in the whole Indo-Pacific region
	Monsoon seasons and tropical storms	Anomalies in monsoon and storms patterns
Positive Indian Ocean Dipole	Copious rainfalls	More abundant rainfalls in Western Ocean Indian region
	Warmer surface water temperature	Increase of surface water temperature in Western Ocean Indian region

Eligible studies shall provide evidence on the impact of either El Niño or +IOD, or their joint effects. The review will also include studies assessing the combined effects of El Niño or with other weather patterns and teleconnections (e.g., global warming).

3.1.3. Type of Comparison

Given the topic area, we will include studies with and without a comparison group. Comparison groups as pipeline, waitlist, and other interventions are not applicable in this field, but the comparison group could be constructed a posteriori using units not affected by the teleconnection or observations before or after its occurrence.

3.1.4. Type of Outcome

The review will include studies providing evidence on five main outcome categories: health, food security, migration, conflicts, and socioeconomics (Table 2). The review embraces an iterative approach, which consists of leaving open the chance of adding relevant sub-categories that are not identified yet in this stage. These will be clearly identified in the final output.

Gender Equality and Social Inclusion risks and impacts should be considered across these four outcome categories. Outcomes are summarized in the table below.

Table 2: List of included outcomes

Outcome group	Outcome type	Description
Health	Direct injuries or fatalities	This includes any measures of direct harm, injury, casualties, and fatalities caused by flooding, storms, and wildfires.
	Disruption of health services	Any measures of disruption or decreased access, physical or financial, to health services
	Morbidity and mortality	Morbidity or mortality rate of enteric infectious diseases, vector-borne diseases and zoonotic diseases, respiratory infections and ailments, and heat stress.
	Cholera	This includes but is not limited to incidence, case load or relative risks of contracting cholera.
	Enteric infectious diseases	This includes but is not limited to incidence, case load or relative risks of contracting water-borne or food-borne infections and diseases such as dysentery, viral hepatitis (hepatitis E), and typhoid.
	Malaria	This includes but is not limited to incidence, case load or relative risks of contracting malaria.
	Zoonotic and vector-borne diseases	This includes but is not limited to incidence, caseload, or relative risk of contracting vector-borne or zoonotic diseases (carried by rodents and animal hosts) such as dengue, Japanese encephalitis, and avian influenza. Measures of Malaria are excluded.
	Respiratory infections and ailments	This includes but is not limited to incidence, case load or relative risks of contracting respiratory ailments such as acute rhinitis, influenza, pneumonia, or any other acute illnesses in the upper respiratory tract caused by adverse weather and environmental changes such as air pollution from forest fires.
	Mental health and psychological effects	Any measures of incidences of conditions requiring mental health and psychosocial support such as hospitalisations due to deteriorating mental health or any other psychological impacts from of livelihood and food insecurity arising from adverse climatic events.
	Heat stress	Any measures of heat stress including but limited to incidences of heat exhaustion and heat strokes.
	Other communicable diseases	This includes but is not limited to incidence, caseload, or relative risk of contracting any other communicable diseases.
Conflict and violence	Local conflict	Any measures of conflict, instability, crime, or extremist behaviour, including but not limited to violent conflict and violence at the local level (e.g., heat effect, sexual and gender-based violence, extremist and factionist behaviour, theft, other misdemeanours, and criminal activity).

Outcome group	Outcome type	Description
	Trans-border conflicts	Any measures of conflict, instability, or extremist behaviour, including but not limited to conflict and violence at the inter-state level (e.g., trans-border river disputes, wars over territorial or maritime borders).
	Domestic abuse/IPV	Measures related to incidences/risk perceived of experiencing domestic abuse or intimate partner violence.
	Civic unrest	Any measures of civic unrest including but not limited to political mobilization, strikes and demonstration
	Disputes	This includes but does not limit to land disputes, local trans-river disputes, fishery disputes.
Economics	Total income and wealth	This includes measures of total household income and other measures of socio-economic status (such as total household expenditure and asset or wealth indices).
	Aggregated production	This includes measures of GDP, GNP, GRP, or any other measure of the total aggregated economic output such as GVA.
	Production	This includes any measure of the disaggregated economic output. Would typically include fish captures, total volume of production or outputs, (share of) land/area cultivated or harvested.
	Productivity	This includes measures of business productivity, agricultural productivity (yields).
	Employment	This includes measures of employment, amount of time worked (e.g., hours and days), employment incidence, unemployment, rate, or number of new employments created.
	Trade	Including measures of trade activities, trade balance, import, export, taxes on exports and imports.
	Consumption and expenditures	Total amount or portion of income spent by individuals, households.
	Prices	This includes changes in prices of goods and services. Measures can be obtained from manifold sources including but not limited to indices of prices, market surveys, and self-reported prices.
	Investments	Measures of total amount or changes in investments held by businesses, corporations, and individuals.
	Economic supply chains	Disruption to economic activities or the reliability of connections between hubs, ports, routes, warehouses, factories, and commercial centres.
	Tourism	Any measure of tourism and tourism-related activities, including but not limited to domestic or international tourism propensity, number of tourists, tourism sector contribution to employment and income.
	Inequalities	Any measures of income inequality, disparities of wealth and well-being, index of poverty and other measures of poverty.

Outcome group	Outcome type	Description
	IT	Any measures of IT usage such as internet connections, telephones, TVs, and radios. We will also include measures of IT literacy.
	Empowerment	Any measures and indices of political, social, and economic empowerment, representations, and access to services of women and marginalized groups.
Migration	Internal/domestic	The number or rate of movement of persons (individuals/households) from their place of usual residence and within the borders of a country.
	International/cross-border	The number or rate of movement of persons (individuals/households) from their place of usual residence and across international borders to a country of which they are not nationals.
	Transhumance	Seasonal movement of livestock, such as sheep, goats, or cattle, between higher and lower elevations in search of better grazing and climate conditions.
	Economic/labour	Movement of persons (individuals/households) from their place of usual residence for economic reasons, which may include better job opportunities, higher wages, improved living standards, or to escape economic hardship in their place of usual residence.
Food and nutrition security	Food and nutrition security	Indices of food and nutrition security, composite scores of the extent to which households have food to meet basic dietary needs, measures of nutritional intake and food consumption, and outcomes based on whether households report they have sufficient food.
	Malnutrition	Nutrition deficiency among children and adults arising from food insecurity, diarrhoea, or other illnesses. Measures include but are not limited to wasting and stunting among children, indices of nutrient intake, indices of dietary diversity etc.

Outcomes that are not socio-economic outcomes will be excluded. This concerns primarily a large body of literature on proximate-level outcomes related to meteorology and atmospheric sciences. Examples include air temperatures, precipitation, or biological processes such as chlorophyll-a, which is a key indicator of phytoplankton biomass and primary productivity (amount of organic material produced per unit area per unit time). Similarly, intermediate outcomes related to algal proliferation, fish and shellfish poisoning would also be out of scope for this review.

3.1.5. Type of Studies

The review will include studies providing quantitative evidence on the effects of El NIÑO and +IOD. No restriction will be applied based on the publication status of the studies; implying that the review will include studies published in peer-review journals and studies from the grey literature, such as working papers, conference papers, and policy reports.

Eligible studies shall provide evidence on the changes in outcomes and use quantitative techniques to attribute such changes to the included teleconnections. Recalling that teleconnections such as El NIÑO and +IOD are 'natural' treatments that cannot be manipulated, studies employing randomized control trials for measuring their impacts are not likely to be found. However, the teleconnection can be considered a natural experiment given that exposure happens "by chance" and cannot be deliberately assigned to units.

We will include studies using a range of quasi-experimental designs. Such studies retrospectively construct the comparison group which was not affected by the teleconnection using methods such as synthetic control, instrumental variables, statistical matching, difference-in-difference, regression discontinuity, and interrupted time series. We will also include studies using other statistical methods such as regression analysis, time series models, spatial correlation, fixed and random effects models, and other methods to measure the association between the teleconnections, as natural experiments, and the outcomes of interest.

We recognize the value of qualitative methods to unravel complex human-environment interactions though due to resource limitations we will only be able to include a subset of qualitative studies that aim to infer causation or shed light on the causal chain of events underpinning the observed change. Further information on the included study designs can be found in Appendix 1.

3.1.6. Other inclusion and exclusion criteria

Language: No restriction will be applied based on the language of study. However, the search terms used in the search stage will be in English only.

Publication date: We will include studies published in 1990 or later. This restriction is necessary to keep the review's search feasible within project resources.

3.2. Search methods for identification of studies

We will search for studies in online electronic databases and specialist websites and repositories (Table 3). In addition to the electronic search, we will conduct citation tracking of included studies (forward and backward) and of existing systematic reviews and contact key experts and organisations.

3.2.1. Electronic searches

The search strategy will be developed in collaboration with an information specialist. The search of studies will use a set of relevant English terms drawn from the selection criteria, along with any matching index terms found in each source. The keywords will be enhanced with source-specific syntax such as truncation and proximity operators and will be combined with index terms using Boolean operators (AND and OR). The search strategy will be adapted to each electronic database and website searched in the retrieval stage (Table 3). For a list of search terms applied to one database refer to Appendix 2

Table 3: List of electronic databases and websites

Type	Sources
Databases and search engines	Scopus
	CAB Abstracts (Ovid)

	Web of Science Core Collection[3]
	MEDLINE (Ovid)
	EMBASE (Ovid)
	Global Health (Ovid)
	Agricola (Ovid)
	Econlit (Ovid)
	BIOSIS Citation Index (Web of Science)
	Academic Search Complete (EBSCOhost)
	EBSCO Discovery Service[4]
	ProQuest dissertations and theses database (Web of Science)
Institutional websites and repositories	Google scholar – manual search
	AgEcon Search
	Social Science Research Network
	National Bureau of Economic Research (NBER) – Working Papers
	3ie Development Evidence Portal
	Development Experience Clearinghouse (USAID)
	Food and Agriculture Organization (FAO)
	International Fund for Agricultural Development (IFAD)
	International Food Policy Research Institute (IFPRI)
	Oxfam Policy and Practice
	Research for Development: FCDO
	IPCC — Intergovernmental Panel on Climate Change
	Prevention Web
	ReliefWeb

The list of electronic databases and relevant websites will be refined in consultation with experts. An information specialist will provide support throughout the search process.

3.2.2. Citation tracking

The search process involves one round of both backward and forward citation tracking. Backward tracking consists of checking the references lists for identifying other eligible studies. Forward tracking implies retrieving articles that cite the initially included study or a

[3] Including Web of Science Core Collection (Social Sciences Citation Index (SSCI), Science Citation Index Expanded (SCI-EXPANDED), Conference Proceedings Citation Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH), Emerging Sources Citation Index (ESCI).

[4] Including GreenFILE, Science Direct, AGRIS, RePEc, World Bank e-Library.

relevant synthesis. We will automate backward and forward citation tracking using the citationchaser Shiny app (Haddaway et al., 2021). It sources records from the Lens.org API. While it contains records from a range of sources such as PubMed, PubMed Central, CrossRef, Microsoft Academic Graph and CORE, the relationships are not necessarily complete, as they rely on open-source resources. For studies not retrieved using citationchaser we will attempt to track their citations using Web of Science or Scopus. If not found, we will use the Publish or Perish software (Harzing, 2010) using the Google scholar API for forward tracking only. We anticipate that the number of studies identified will be large. We will de-duplicate them, and upload to EPPI reviewer where a machine learning model will be built to classify them based on their likelihood of being included (ten buckets will be created: 90-99%, 80-89% and so on). We will keep screening records that are most likely to be included until a bucket contains no includable studies or 100 in a row are found not to be includable.

3.2.3. Contacting experts

We will contact key experts and organizations for providing inputs to the search process. Experts will validate that relevant electronic databases and websites used for the search and confirm that the search process did not omit any eligible study. Further, experts will provide feedback on the final reports and corroborate the interpretation of the results.

3.2.4. Other sources

We will identify existing reviews, meta-analyses, and gap maps. Those will be included in the citation tracking workflow.

3.3. Data collection

Following the search stage, we will deduplicate and screen the retrieved studies through a two-stage selection process for identifying the eligible studies that can be included in the review.

3.3.1. Selection of studies

De-duplication

Following the search process, the retrieved studies will be pooled in a unique dataset and go through a de-deduplication process. The first round of de-duplication will be performed using Covidence[5] which is expected to pick up most duplicates. Afterwards, studies will be imported into the EPPI-Reviewer software (Thomas et al., 2022) which will perform a second round of de-duplication.

The deduplication will be followed by the selection of studies following a two-stage process. Both stages of the screening will be implemented using EPPI-Reviewer software.

Stage 1: Title and abstract screening

The first selection stage consists of screening studies based on the information available in title and abstract. Reviewers will independently screen in pairs a random sample of records until 85% interrater reliability is achieved. This will help us to establish the baseline inclusion rate to train the priority screening function in EPPI Reviewer (Thomas et al., 2022). At the end of each round of screening, disagreements in the decisions of excluding/including will be discussed and reconciled. Afterwards, reviewers will screen the rest of the studies (in the order

[5] https://www.covidence.org/

provided by the priority screening tool) using single screening mode until the search is saturated (if out of 500 studies screened in a row less than 1% end up being included on title and abstract). We will then screen 100 of the remaining studies picked at random, and if none of these 100 studies qualify for inclusion, then we will stop screening.

The screening will follow a sequential approach, implying that studies will be screened against a defined hierarchical list of inclusion/exclusion criteria (Table 4 in Appendix 4) This facilitates a homogeneous decision-making process across reviewers and ensures comparability of decisions.

Stage 2: Full text screening

After the full-text retrieval of the studies identified as potentially relevant through title and abstract screening, the full texts will be screened against the review's inclusion criteria. The full-text screening will follow a double-blind selection process, implying that pairs of reviewers will independently screen studies and reconcile any disagreement in the decisions. When necessary, a third reviewer will be involved in the reconciliation process.

Reviewers will initially screen 'training batches' following an iterative process until 85% inter-rater reliability rate is achieved (on include/exclude decisions) and discuss any disagreements within each pair of reviewers. After reaching the desired consistency rate, batches of studies will be allocated across pairs of reviewers and screened following a sequential list of inclusion/exclusion criteria (Table 5 in Appendix 4).

3.3.2 Data extraction and management

Following the selection process, we will start the data extraction of the included body of evidence. We will extract descriptive, methodological, quantitative, and qualitative information about the included studies. The following information will be extracted:

- Descriptive information including bibliographic characteristics such as title, names of authors, publication status, outlet, and year of publication; country of the study; year when the analysed teleconnection occurred.
- Methodological information about study design, employed method, type of comparison, unit of analysis.
- Quantitative information about outcome means and standard deviations and *t* statistics (or other tests such as F test), sample size, and time span of analysed data
- Qualitative information about types of outcome measure, reported findings, mitigating factors and facilitators

For a provisional data extraction codebook, refer to Appendix 3. Information will be extracted from each study by reviewers and double-checked by other reviewers.

Two team members will independently extract the data from each study using MS Excel, drawing on both the included paper as well as any additional papers identified. Any differences that cannot be reconciled between them will be addressed through discussion with a third reviewer who is a senior team member.

3.3.3. Assessment of risk of bias in included studies

The risk of bias assessment of included studies will be independently carried out by pairs of reviewers, who will discuss and reconcile disagreements with the collaboration of a third reviewer when necessary.

We will use an adapted version of the 3ie's risk of bias assessment tools suitable for assessing the internal validity of experimental and quasi-experimental studies. We are not expecting to find any randomized studies, so we plan to adapt the tool for experimental studies to assess the quality of studies using other statistical methods considering a teleconnection as a natural experiment. Further, we will also revise the tool for quasi-experimental designs to consider the particularities of the type of treatment in this review (exposure to a climate phenomenon) and to methods such as interrupted time series, which may be more common in this sector. We include both tools in their original version for reference (Appendices 5 and 6).

The risk of bias assessment will evaluate the quality of extracted estimates in relation to factors such as confounding bias, missing data, outcome measurement bias, and reporting bias. Each criterion will be coded as to whether they are *free from the bias*, using a response scale of "Yes," "Probably Yes," "Probably No," "No" and "Unclear." Based on the rating of individual criteria we will assign the overall rating of each study as either "high risk of bias," "some concerns" or "low risk of bias." The rating will be assigned as follows:

- "High risk of bias": if any of the domains were assessed as "No" or "Probably No."
- "Some concerns": if one or several domains were assessed as "Unclear," and none were "No" or "Probably No."
- "Low risk of bias": if all of the domains were assessed as "Yes" or "Probably Yes."

The results of the risk of bias assessment of each study will be provided for the overall rating and for each domain.

3.3.4. Measures of treatment effect

We will extract treatment effects, or effect sizes, from each study where sufficient data is provided. Effect sizes indicate the magnitude and direction of the difference in outcomes between treatment and comparison groups. We will compute standardized effect sizes using a single metric to allow for cross-study comparisons.

We will explore the presence of unit analysis errors, which occur when the unit of analysis is located at individual level whilst the treatment is located at cluster level. Studies that do not account for this issue are likely to over-estimate the effects of treatment, and in turn can have greater weight when included in meta-analyses (Donner et al., 2001). For this reason, we will appraise included studies against the persistence of this issue, and where necessary adjust the reported SEs.

3.3.5. Independent findings

To minimize the redundancy of the extracted information we will try to avoid double counting of studies. This is considered as a good practice for dealing with between-study dependency of the extracted estimates (Borenstein et al., 2009). Papers presenting identical evidence will be linked to a main paper and used in case further information needs to be extracted. In cases where multiple studies evaluate the same event exploiting the exact same dataset we will prioritize peer-reviewed articles; or in the case of multiple unpublished studies, the most recent paper.

We will extract one estimate effect per outcome per study. To this end, we will deal with each specific challenge by following different criteria explained hereafter:

- Where studies report on the same teleconnection event but using different samples, both studies will be separately included.

- Where studies report multiple mechanisms, we will include them in separate (meta)analyses.
- Where studies report on multiple time points, we will synthesize effects and present the average effects.
- Where studies report effects from multiple models, we will follow the general rule of including estimates from the authors' preferred model specification. In case the preferred model is unclear, we will include the most precise estimate (i.e. the effects with the highest t-value).
- Where studies report effects from multiple estimators, we will include estimates from the authors' preferred specification. If the preferred specification is not clear, we will include the specification that is most robust to falsification tests
- Where studies report different measures of the same outcome, we will prioritize according to the most recurrent outcome measurement adopted by the included studies.

3.3.6. Dealing with missing data

When carrying out the full-text screening and the data extraction, we might find studies that are omitting some key information. In this case we will contact the corresponding author to request the data necessary to compute the effect sizes. If the author does not respond, we will try to estimate the missing data where possible or exclude the study from any quantitative meta-analysis.

3.3.7. Data synthesis

The data synthesis will rely on a combination of narrative analysis with descriptive statistics. If the number of eligible studies would allow it, a meta-analysis will be performed for synthesizing effect sizes referred to the same type of intervention and using a similar outcome category. Where feasible, effects will be pooled using a random-effects inverse variance weighted meta-analysis. (Borenstein et al., 2009). Narrative synthesis will be conducted where effect sizes are too heterogeneous and if they are derived from few studies and will be accompanied by descriptive statistics about individual effect sizes. Namely, we will narratively discuss effect sizes alone and highlight underpinning methodological and contextual aspects such as the method employed for estimating the effects, type and intensity of teleconnection, country, or geographical area of the study. The narrative analysis will be integrated by descriptive statistics about median and interquartile ranges.

3.3.8. Assessment of publication biases

If a meta-analysis containing at least 10 studies is conducted, we will test the presence of publication bias with both a rank correlation test (see Begg and Mazumdar,1994) and a regression test using the standard error of the observed outcomes as predictor (Sterne and Egger, 2005), to test the presence of funnel plot asymmetry.

3.3.9. Subgroup analysis and harvest plots

The systematic review intends to explore how the reported effects vary by type of teleconnections, intensity of teleconnection, geographical region, and outcome. Further, the review will explore factors related to publication characteristics, and methodological features such as methods employed to estimate the effects, and measurement method adopted to measure the intensity of the teleconnection.

Given the heterogeneity of considered treatments and outcomes, we will carry out sub-group analysis by sorting effects by type of treatment and by type of outcome. If meta-analysis will

be feasible, we will conduct a moderator analysis using meta-regression (where feasible) and perform statistical tests to analyse the heterogeneity of the effect sizes analysed in the model using Q-test, $I2$, and $\tau 2$.

In the eventuality that the included body of evidence is limited to a few heterogeneous studies, we will synthesize findings using harvest plots. The use of harvest plots in systematic reviews is particularly suitable when meta-analysis is not possible due to the above-mentioned heterogeneity (Ogilvie at al., 2008). Harvest plots allow the display of quantitative data for all studies and for the outcome categories of interest when it would not be possible to combine them in one single forest plot.

4. Contributions of authors

Content: Andrea Floridi, Anil Thota, Shannon Shisler, Tomasz Kozakiewicz, María Daniela Anda León, and Megha Bhattacharya.

Systematic review methods: Andrea Floridi, Shannon Shisler, Anil Thota, Tomasz Kozakiewicz, María Daniela Anda León, and Megha Bhattacharya.

Information retrieval: Zahra Premji, Tomasz Kozakiewicz, Anil Thota, Andrea Floridi, María Daniela Anda León

Screening and data extraction: Andrea Floridi, Tomasz Kozakiewicz, María Daniela Anda León, Megha Bhattacharya, and Anil Thota, and Shannon Shisler

Analysis: Andrea Floridi, María Daniela Anda León, Tomasz Kozakiewicz, Megha Bhattacharya, Anil Thota, and Shannon Shisler

5. Preliminary timeframe

We aim to complete the systematic review by the end of March 2024.

6. Conflicts of interest

There are no conflicts of interest.

7. Acknowledgements

We wish to acknowledge the generous support of FCDO. We wish to thank the experts who review this protocol: Dr Jan Selby, Dr Luca Tasciotti, Dr Peter Burt, Roufa Khanum, Dr Farzana Misha, Tesse de Boer, Catalina Jaime, and Dr Christopher Jack.

References

Ashok, K., Guan, Z., & Yamagata, T. (2003). A look at the relationship between the ENSO and the Indian Ocean dipole. *Journal of the Meteorological Society of Japan. Ser. II*, *81*(1), 41-56.

Ashok, K., & Saji, N. H. (2007). On the impacts of ENSO and Indian Ocean dipole events on sub-regional Indian summer monsoon rainfall. *Natural Hazards*, *42*, 273-285.

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly weather review*, *97*(3), 163-172.

Borenstein, M, Hedges, LV, Higgins, JPT and Rothstein, HR, 2009. I*ntroduction to Meta-Analysis*. Chichester, U.K: John Wiley & Sons, Ltd.

Cai, W., Sullivan, A., & Cowan, T. (2009). Climate change contributes to more frequent consecutive positive Indian Ocean Dipole events. *Geophysical Research Letters*, *36*(23).

Cai, W., Santoso, A., Collins, M., Dewitte, B., Karamperidou, C., Kug, J. S., ... & Zhong, W. (2021). Changing El Niño–Southern oscillation in a warming climate. *Nature Reviews Earth & Environment*, *2*(9), 628-644.

Demissie, S., and Bezatu M. (2017). The impact of El Niño on diarrheal disease incidence: A systematic review." Science 5, no. 6: 446-451.

Donner, A., Piaggio, G., & Villar, J. (2001). Statistical methods for the meta-analysis of cluster randomization trials. *Statistical methods in medical research*, *10*(5), 325-338.

Feba, F., Govardhan, D., Tejavath, C. T., & Ashok, K. (2021). ENSO Modoki teleconnections to Indian summer monsoon rainfall—A review. *Indian Summer Monsoon Variability*, 69-90.

Ghose, B., Islam, A. R. M. T., Salam, R., Shahid, S., Kamruzzaman, M., Das, S., ... & Mallick, J. (2021). Rice yield responses in Bangladesh to large-scale atmospheric oscillation using multifactorial model. *Theoretical and Applied Climatology*, *146*, 29-44.

Haddaway, N. R., Grainger, M. J., & Gray, C. T. (2021). citationchaser: An R package and Shiny app for forward and backward citations chasing in academic searching. *Zenodo, Feb*, *16*. doi: 10.5281/zenodo.4543513

Hendrix, C. S., Glaser, S. M., Lambert, J. E., & Roberts, P. M. (2022). Global climate, El Nino, and militarized fisheries disputes in the East and South China Seas. *Marine Policy*, *143*, 105137.

Hsiang, S. M., Meng, K. C., & Cane, M. A. (2011). Civil conflicts are associated with the global climate. *Nature*, *476*(7361), 438-441.

Kovats, R. S., Bouma, M. J., Hajat, S., Worrall, E., & Haines, A. (2003). El Niño and health. *The Lancet*, *362*(9394), 1481-1489.

Levy, K., Woster, A. P., Goldstein, R. S., & Carlton, E. J. (2016). Untangling the impacts of climate change on waterborne diseases: a systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought. *Environmental science & technology*, *50*(10), 4905-4922.

Marathe, S., & Karumuri, A. (2021). The El Niño Modoki. In *Tropical and Extratropical Air-Sea Interactions* (pp. 93-114). Elsevier.

McGregor, G. R., & Ebi, K. (2018). El Niño Southern Oscillation (ENSO) and health: An overview for climate and health researchers. Atmosphere, 9(7), 282.

Ogilvie, D., Fayter, D., Petticrew, M., Sowden, A., Thomas, S., Whitehead, M., & Worthy, G. (2008). The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC medical research methodology*, *8*, 1-7.

Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., & Bouma, M. J. (2000). Cholera dynamics and El Nino-southern oscillation. *Science*, 289(5485), 1766-1769.

Salimun, E., Tangang, F., Juneng, L., Behera, S. K., & Yu, W. (2014). Differential impacts of conventional El Niño versus El Niño Modoki on Malaysian rainfall anomaly during winter monsoon. *International journal of climatology*, *34*(8), 2763-2774.

Saji, N. H., & Yamagata, T. J. C. R. (2003). Possible impacts of Indian Ocean dipole mode events on global climate. *Climate research*, *25*(2), 151-169.

Stuecker, M. F., Timmermann, A., Jin, F. F., Chikamoto, Y., Zhang, W., Wittenberg, A. T., ... & Zhao, S. (2017). Revisiting ENSO/Indian Ocean dipole phase relationships. *Geophysical Research Letters*, *44*(5), 2481-2492.

Thomas, J, Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., & Bond, M. Koryakina A, 2022. *EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis*. London: EPPI-Centre, UCL Social Research Institute, University College London

Wang, C., Deser, C., Yu, J. Y., DiNezio, P., & Clement, A. (2017). El Niño and southern oscillation (ENSO): a review. *Coral reefs of the eastern tropical Pacific: Persistence and loss in a dynamic environment*, 85-106.

Yeh, S. W., Kug, J. S., Dewitte, B., Kwon, M. H., Kirtman, B. P., & Jin, F. F. (2009). El Niño in a changing climate. *Nature*, *461*(7263), 511-514.

Zheng, F., Fang, X. H., Yu, J. Y., & Zhu, J. (2014). Asymmetry of the Bjerknes positive feedback between the two types of El Niño. *Geophysical Research Letters*, *41*(21), 7651-7657.

Zhou, Z. Q., Xie, S. P., & Zhang, R. (2021). Historic Yangtze flooding of 2020 tied to extreme Indian Ocean conditions. *Proceedings of the National Academy of Sciences*, *118*(12), e2022255118.

Appendices

Appendix 1: Further information on included study designs

We will include studies using quantitative techniques to find an association between the teleconnection and the outcomes of interest. This includes quasi-experimental study designs and other statistical methods to find a relation between the observed changes in the outcomes and the independent variable, in this case a weather phenomenon, which can be considered a natural experiment. Natural experiments exploit the natural randomness in treatment assignment (exposure to a teleconnection) and measure its impact through a comparison between the treatment and control group.

Quasi-experimental designs:

a. Regression discontinuity designs (RDD) or fuzzy-RDD

b. Instrumental variables (IV)

b. Endogenous treatment-effects models, endogenous switching regression, and other methods synonymous to the Heckman two step model.

d. Difference-in-differences (DID), two-way fixed-effects (TWFE), and two-way Mundlak regressions (TWM).

i. DID models will include an interaction term between a time and intervention variable in a regression model. They may also regress an intervention variable

on an outcome variable measuring the changes in outcomes over time or present a t-test comparing changes in outcomes over time between the intervention and control group.

ii. TWFE regressions must include time fixed-effects and unit fixed-effects at the level of the intervention (or lower). For example, if the intervention varies at a village level, it must include either village fixed-effects or fixed-effects of a smaller unit, such as households.

iii. TWM models should be synonymous with the approach described by Wooldridge (2021). This includes correlated random-effects and pooled OLS regression models that control for unit-specific time averages and time-period specific cross-sectional averages.

e. Interrupted time series (ITS) models, with or without a contemporaneous comparison group. This includes segmented regressions, where the time-period is divided into pre- and post-intervention segments, and separate intercepts and/or slopes are estimated for each segment.

f. Weighting and matching approaches that control for observable confounding, including non-parametric approaches (e.g., statistical matching, covariate matching, coarsened-exact matching, propensity score matching) and parametric approaches (e.g., propensity-weighted multiple regression analysis).

g. The synthetic control method creates a synthetic or counterfactual control group that closely mimics the characteristics of the treated unit before the intervention by assigning weights to the units that were not exposed to the treatment (exposure to a teleconnection) and comparing the outcomes of both groups over time.

Other statistical methods to assess the impact of natural experiments:

i. Regression analysis uses statistical techniques to estimate the coefficients of a mathematical model that explains the relationship between a dependent variable and one (ANOVA) or more (ANCOVA) independent variables known as covariates. Regression analysis further tests if the coefficients are statistically different from zero.

j. Time series models are statistical models that analyse and forecast data points collected over successive, evenly spaced intervals of time.

k. Spatial correlation refers to the degree to which the values of a variable at nearby locations in a geographical space are similar or related.

l. Fixed and random effects models are used to analyse panel data and vary in the way they account for the units' unobserved heterogeneity.

m. Other quantitative models using mathematical, statistical, spatial methods to estimate a relationship between two variables and its significance.

n. Qualitative methods that investigate a causal inference question: employ at least one method among realist evaluation, general elimination methodology, process tracing or contribution analysis; or use other methods but provide a theory of change for explaining the underpinning logic and theoretical links.

Appendix 2: Search terms

CAB Abstracts (Ovid) <1990 to 2024 Week 03>

#	Query	Results from 29 Jan 2024	Annotations
1	"El Nino-Southern Oscillation".sh.	5,413	ENSO subject heading
2	("El Niño*" or "El Nino*" or "El Ni˜no*" or "Oceanic Niño Index" or "Oceanic Nino Index" or "southern oscillation index" or (("sea surface temperature" or "SST*") adj3 "anomal*") or "SOI index" or "IOD?" or "+IOD" or "IOD+" or "pIOD?" or "ENSO?" or "SSTA?" or "MENSOI" or ("Indian Ocean" adj5 "Dipole") or (("nino*" or "Niño*") adj3 (pacific or "3 4" or "modoki" or "canonical" or "conventional" or "cold tongue" or "warm pool" or "dateline")) or (("dipole mode index" or "DMI" or "walker cell" or "walker circulation" or "surface sea temperature gradient") and "indian ocean") or "Delayed Oscillator" or "Recharge Oscillator" or "Western Pacific Oscillator" or "Advective-Reflective Oscillator" or "Unified Oscillator" or "La Niña*" or "La Nina*").ti,ab,ot,hw,gl.	11,948	ENSO free text terms searched in title, abstract, original title, heading word, and geographic location fields
3	1 or 2	11,948	ENSO combination line
4	("bangladesh" or "bhutan" or "borneo" or "burma" or "cambodia" or "kampuchea" or "khmer republic" or "cook islands" or "fiji" or "guam" or "india" or "indonesia" or "kiribati" or "laos" or "marshal island?" or "malaysia" or "malay? federation" or "maldives" or "melanesia" or "micronesia" or "myanmar" or "nauru" or "nepal" or "Pleasant Island" or "northern mariana islands" or "new guinea" or "Oceania" or "palau" or "philippines" or "philipines" or "phillipines" or "phillippines" or "pilipinas" or "pacific islands" or "polynesia" or "samoa" or "samoan islands" or "navigator island" or "navigator islands" or "solomon island?" or "norfolk island?" or "santa cruz island?" or "sri lanka" or "ceylon" or "thailand" or "siam" or "timor" or "tonga" or "tahiti" or "tuvalu" or "ellice islands" or "vanuatu" or "vietnam" or "viet nam" or "west indies" or "bangladeshi?" or "bangalees" or "bajan?" or	865,114	Relevant Countries free text terms searched in title, abstract, original title, heading word, and geographic location fields

	"bhutanese" or "bornean?" or "burmese" or "cambodian?" or "cook islander?" or "fijian?" or "guamanian?" or "indonesian?" or "kirabatian?" or "lao" or "laotian?" or "malaysian?" or "maldivian?" or "marshallese" or "melanesian" or "micronesian?" or "myanma" or "nepali?" or "nepalese" or "northern mariana islander?" or "mariana?" or "chamorros" or "nauruan?" or "norfolk islander?" or "oceanian" or "palauan?" or "papua new guinean?" or "philippine?" or "philipine?" or "phillipine?" or "phillippine?" or "filipino?" or "filipina?" or "pacific islander?" or "polynesian?" or "samoan?" or "solomon islander?" or "sri lankan?" or "ceylonese" or "tahitian?" or "thai" or "timorese?" or "tongan?" or "tuvaluan?" or "vanuatuan?" or "vietnamese").ti,ab,ot,hw,gl.		
5	("north* pacific ocean*" or "tropical pacific ocean*" or "equatorial pacific ocean*" or "south* pacific ocean*" or "mekong delta?" or "ganges delta?" or "ayeyarwady delta?" or "arabian sea" or "andaman sea" or "bay of Bengal" or "north indian ocean" or "southwestern indian ocean" or "south china sea?" or "indian subcontinent*" or ("countries" adj10 ("the world" or "worldwide" or "global*")) or ("southeastern" adj2 "asia") or ("south eastern" adj2 "asia") or "southeast asia" or "south east asia" or "south asia" or (("indian ocean" or "pacific ocean" or "indo pacific" or "indo-pacific" or "indopacific*" or "north pacific" or "tropical pacific" or "equatorial pacific" or "south pacific" or "indochina" or "indochinese") adj4 ("adjacent" or "border*" or "country" or "countries" or "region" or "regions" or "island" or "islands" or "nation" or "nations" or "economies"))).ti,ab,ot,hw,gl.	796,581	Relevant regions and oceans free text terms searched in title, abstract, original title, heading word, and geographic location fields
6	4 or 5	1,021,147	Region concept combination line
7	(("match*" adj2 ("propensity" or "coarsened" or "covariate" or "co-variate" or "neighbo?r")) or "propensity score" or "difference* in difference*" or "difference-in-difference*" or "differences-in-difference*" or "double difference*" or "quasi-experiment*" or "quasi experiment*" or ("estimat*" and "evaluat*") or "instrumental variable*" or ("IV" adj3 ("estimation" or "approach*")) or ("Heckman" adj4 ("model*" or "approach*")) or (("two-stage" or "two stage") adj4	680,196	Study designs and analyses concept

	("control*" or "function*" or "least squares")) or "regression discontinuity" or "time series" or "counterfactual" or "segment* regression" or "coefficient of variation" or ("non" adj3 "participant*") or (("control" or "comparison") adj3 ("group*" or "condition*" or "area*" or "village*" or "household*" or "intervention")) or ("panel*" adj3 ("data" or "household*" or "model*")) or (("exploit*" or "tak* advantage") adj4 ("variation*" or "variety" or "exogen*" or "heterogen*")) or ("econometric" adj3 ("model*" or "adjust*")) or ("select*" adj3 ("bias*" or "self")) or ("experiment*" adj3 ("design" or "study" or "research" or "evaluation" or "evidence" or "vary" or "varies" or "variation")) or (("random" or "randomi?ed" or "randomly") adj3 ("trial" or "assign*" or "treatment" or "control*" or "allocat*" or "experiment*" or "vary" or "varies" or "variation" or "choose" or "chose*" or model*))).ti,ab,ot,hw,gl.		
8	((("impact?" or "effect*") adj6 ("evaluat*" or "assess" or "assessing" or "assessment*" or "analyze*" or "analyse*" or "analyzing" or "analysing" or "analysis" or "analyses" or "analytical" or "estimate*" or "estimating" or "estimation*" or "examin*" or "quantif*" or "investigat*" or "cause" or "causes" or "causal" or "causation" or "causatively" or "association?" or "associate*" or "hypothesi*" or "produce*" or "production*" or "food" or "crop" or "crops" or "disease" or "infection*" or "health" or "economic" or "price" or "prices" or "markets" or "socioeconomic" or "migration")) or ("association*" adj6 ("evaluat*" or "assess" or "assessing" or "assessment*" or "analyze*" or "analyse*" or "analyzing" or "analysing" or "analysis" or "analyses" or "analytical" or "estimate*" or "estimating" or "estimation*" or "examin*" or "quantif*" or "investigat*" or "cause" or "causal" or "causation" or "causatively" or "hypothesi*")) or "spatial correlation" or "spatial temporal" or "inciden* rate?" or "inciden* ratio?" or "rate ratio?" or (("quant*" or "effect?" or "pattern?") adj6 ("association" or "associated")) or (("composite" or "component*" or "spatial*" or "variabilit*" or "function" or "wavelet" or "correlation*" or "statistical*" or "economi*" or "macroeconomi*" or "empirical*") adj6 ("analys*" or "analyz*"))).ti,ab,ot,hw,gl.	1,733,197	

9	7 or 8	2,161,134	Analysis concept combination line
10	3 and 6 and 9	1,383	Final search combination line
11	limit 10 to yr="1990 - 2024"	1,383	Date limit applied

Appendix 3: Initial sample of studies used for search strategy development

Adams, Nicholas, Meghnath Dhimal, Shifali Mathews, Veena Iyer, Raghu Murtugudde, Xin-Zhong Liang, Muhiuddin Haider, et al. 2022. "El Niño Southern Oscillation, Monsoon Anomaly, and Childhood Diarrheal Disease Morbidity in Nepal." Edited by Sandro Galea. *PNAS Nexus* 1 (2): pgac032. https://doi.org/10.1093/pnasnexus/pgac032.

Ajithkumar, B, and Arjun Vysakh. 2018. "Impact of El Niño on Crop Productivity in Kerala." *Research Journal of Agricultural Sciences* 9 (5): 1041–45.

Bhuvaneswari, K., V. Geethalakshmi, A. Lakshmanan, R. Srinivasan, and Nagothu Udaya Sekhar. 2013. "The Impact of El Niño/Southern Oscillation on Hydrology and Rice Productivity in the Cauvery Basin, India: Application of the Soil and Water Assessment Tool." *Weather and Climate Extremes* 2 (December): 39–47. https://doi.org/10.1016/j.wace.2013.10.003.

Cashin, Paul, Kamiar Mohaddes, and Mehdi Raissi. 2015. "Fair Weather or Foul? The Macroeconomic Effects of El Niño." Working Paper WP/15/89. IMF Working Paper. International Monetary Fund.

Chapman, Ross, James Cock, Marianne Samson, Noel Janetski, Kate Janetski, Dadang Gusyana, Sudarshan Dutta, and Thomas Oberthür. 2021. "Crop Response to El Niño-Southern Oscillation Related Weather Variation to Help Farmers Manage Their Crops." *Scientific Reports* 11 (1): 8292. https://doi.org/10.1038/s41598-021-87520-4.

Cobon, David H., Maureen Ewai, Kasis Inape, and R. Michael Bourke. 2016. "Food Shortages Are Associated with Droughts, Floods, Frosts and ENSO in Papua New Guinea." *Agricultural Systems* 145 (June): 150–64. https://doi.org/10.1016/j.agsy.2016.02.012.

Dorosh, Paul A., Xinshen Diao, James Thurlow, Pankaj Koirala, Krishna Timsina, and Timothy J. Krupnik. 2023. "What Might Be at Stake? El Niño, Global Price Shocks and Food Security in Nepal." 0 ed. Washington, DC: International Food Policy Research Institute. https://doi.org/10.2499/p15738coll2.137060.

Fajri, Halimah Cahyaning, Hermanto Siregar, and Sahara Sahara, eds. 2019. "IMPACT OF CLIMATE CHANGE ON FOOD PRICE IN THE AFFECTED PROVINCES OF EL NINO AND LA NINA PHENOMENON: CASE OF INDONESIA." *International Journal of Food and Agricultural Economics (IJFAEC)*. https://doi.org/10.22004/ag.econ.296761.

Ferris, John N. 1999. "AN ANALYSIS OF THE IMPACT OF ENSO (EL NINO/SOUTHERN OSCILLATION) ON GLOBAL CROP YIELDS." https://doi.org/10.22004/AG.ECON.11741.

Generoso, Couharde, Damette, and Mohaddes. 2020. "The Growth Effects of El Niño and La Niña: Local Weather Conditions Matter." *Annals of Economics and Statistics*, no. 140: 83. https://doi.org/10.15609/annaeconstat2009.140.0083.

Harzing, Anne-Wil. 2010. "The publish or perish book". Melbourne, Australia: Tarma Software Research Pty Limited. https://www.harzing.info/download/popbook12.pdf

Hashizume, Masahiro, Luis Fernando Chaves, A. S. G. Faruque, Md Yunus, Kim Streatfield, and Kazuhiko Moji. 2013. "A Differential Effect of Indian Ocean Dipole and El Niño on Cholera Dynamics in Bangladesh." Edited by Dongsheng Zhou. *PLoS ONE* 8 (3): e60001. https://doi.org/10.1371/journal.pone.0060001.

Hashizume, Masahiro, A.S.G. Faruque, Toru Terao, Md Yunus, Kim Streatfield, Taro Yamamoto, and Kazuhiko Moji. 2011. "The Indian Ocean Dipole and Cholera Incidence in Bangladesh: A Time-Series Analysis." *Environmental Health Perspectives* 119 (2): 239–44. https://doi.org/10.1289/ehp.1002302.

Heaney, Alexandra K., Jeffrey Shaman, and Kathleen A. Alexander. 2019. "El Niño-Southern Oscillation and under-5 Diarrhea in Botswana." *Nature Communications* 10 (1): 5798. https://doi.org/10.1038/s41467-019-13584-6.

Hendrix, Cullen S., Sarah M. Glaser, Joshua E. Lambert, and Paige M. Roberts. 2022. "Global Climate, El Niño, and Militarized Fisheries Disputes in the East and South China Seas." *Marine Policy* 143 (September): 105137. https://doi.org/10.1016/j.marpol.2022.105137.

Iyer, Veena, Ayushi Sharma, Divya Nair, Bhavin Solanki, Pradeep Umrigar, Raghu Murtugudde, Chengsheng Jiang, Dileep Mavalankar, and Amir Sapkota. 2021. "Role of Extreme Weather Events and El Niño Southern Oscillation on Incidence of Enteric Fever in Ahmedabad and Surat, Gujarat, India." *Environmental Research* 196 (May): 110417. https://doi.org/10.1016/j.envres.2020.110417.

Liao, Xiaomei, Jinfeng Ma, and Haigang Zhan. 2012. "Effect of Different Types of El Niño on Primary Productivity in the South China Sea." *Aquatic Ecosystem Health & Management* 15 (2): 135–43. https://doi.org/10.1080/14634988.2012.687655.

Malau, L R E, N A Ulya, R Anjani, and M Rahmat. 2021. "Study of ENSO Impact on Agricultural Food Crops Price as Basic Knowledge to Improve Community Resilience in Climate Change." *IOP Conference Series: Earth and Environmental Science* 874 (1): 012008. https://doi.org/10.1088/1755-1315/874/1/012008.

Pascual, Mercedes, Xavier Rodó, Stephen P. Ellner, Rita Colwell, and Menno J. Bouma. 2000. "Cholera Dynamics and El Niño-Southern Oscillation." *Science* 289 (5485): 1766–69. https://doi.org/10.1126/science.289.5485.1766.

Pramanik, Malay, Poonam Singh, Gaurav Kumar, V. P. Ojha, and Ramesh C. Dhiman. 2020. "El Niño Southern Oscillation as an Early Warning Tool for Dengue Outbreak in India." *BMC Public Health* 20 (1): 1498. https://doi.org/10.1186/s12889-020-09609-1.

Rodó, Xavier, Mercedes Pascual, George Fuchs, and A. S. G. Faruque. 2002. "ENSO and Cholera: A Nonstationary Link Related to Climate Change?" *Proceedings of the National Academy of Sciences* 99 (20): 12901–6. https://doi.org/10.1073/pnas.182203999.

Uyen, Huynh Thi Kim, Yoko Saito, and Nguyen Thi Minh Hoa. 2022. "Natural Disasters and Migration Choice in Vietnam: Estimating the Impact of EL NIÑO Using Household Data." *Journal of Agricultural Development Studies* 32 (3): 28–34.

Villoria, Nelson B., Michael Delgado, Nelson B. Villoria, and Michael Delgado. 2017. "Worldwide Crop Supply Responses to El Niño Southern Oscillation." https://doi.org/10.22004/AG.ECON.258564.

Appendix 4: Sequential screening criteria

Stage 1 – Title and abstract screening

Table 4: Inclusion/exclusion criteria at the title and abstract screening stage

Priority order	Question	Excluded if the answer is
1.	Is the study a duplicate?	Yes
2.	Has the study been published prior to the year 1990?	No
3.	Does the study include an independent variable that is relevant?	No
4.	Does the study evaluate the effect of the teleconnection by using quantitative or qualitative methods to establish a link between the climate event and at least one outcome?	No
5.	Does the study include data from at least one country of interest?	Yes
6.	Is the study only concerned with mechanisms and effects related to meteorology, atmospheric sciences, or biological sciences?	No

Notes: If insufficient information is available to confidently answer a question, reviewers will proceed to the next question without excluding the study.

Stage 2 – Full-text screening

Table 5: Inclusion/exclusion criteria at the full-text screening stage

Priority order	Question	Excluded if the answer is
1.	Is the study excludable at title and abstract?	Yes
2.	Is this study a duplicate?	Yes
3.	Does the study include an independent variable that is relevant?	No
4.	Was the study published prior to the year 1990?	Yes

5.	Does the study include data from at least one country of interest?	No
6.	Does the study evaluate the effect of the teleconnection by using quantitative or qualitative methods to establish a link between the climate event and at least one outcome?	No
7.	Does the study include an outcome consistent with the review's inclusion criteria? (See Table 2)	No
8.	Does the design meet the minimum criteria for inclusion?	No
9.	Does the study mention or provide details of the data used to quantify a graphical relationship between the dependent and independent variables?	No
10.	Does the study only provide results of a simulation with no ex-post analysis?	Yes
11.	Does the study only provide visual or spatial analyses but no quantitative estimates or coefficients?	Yes

Notes: If insufficient information is available to confidently answer a question, reviewers will proceed to the next question without excluding the study.

Appendix 5: Provisional data extraction form

Variable group	Variable	Description
Publication Information	Study ID	The unique ID code that is assigned to each included study
	Estimate ID	The unique ID code that assigned to each individual estimate
	Study status	Select one of the following: i) Completed; ii) Protocol; iii) Ongoing
	Author Name	Authors last names [Open Answer]
	Year of Publication	Year published (publication date, not preprint or first online publication dates)
Teleconnection Information	Teleconnection code	Choose one or more teleconnection code(s) for each corresponding effect size: i)
	Teleconnection	Choose one or more intervention sub-group code(s) for each corresponding effect size: ● Canonical El Niño Southern Oscillation ● El Niño Modoki ● Positive Indian Ocean Dipole ● Not specified
	Country	Country for which effects are measured (select more than one if applicable)
	Exposure to teleconnection (in months)	For how long are the observations exposed to the teleconnection?
	Evaluation period (in months)	The total number of months elapsed between the end of a teleconnection and the point at which an outcome measure is taken post teleconnection, or as a follow-up measurement. If less than one month, use decimals (e.g., measurement immediately after the intervention end would be coded as 0, one week would be .25, etc.)

	Teleconnection description	Provide detailed description of the intervention and its different components such that a reader could easily understand what happened. Include page numbers for quick reference. If two or more teleconnections are being evaluated, please provide descriptions for each teleconnection arm under separate rows.
	Type of measurement	Select one or more out of the following: Oceanic Niño Index, Southern Oscillation Index. Sea Surface Temperature, El Niño Years, Other
	Teleconnection year	For time series designs it is fine to list the interval corresponding to the dataset (e.g. 1950-2000)
Method information	Evaluation Design	Select one of the options below: 1. Experimental (defined as prospective randomised assignment, where randomisation is implemented by researchers (or by decision makers in the context of an evaluation study) 2. Quasi-experimental (including natural experiments and non-randomised studies). 3. Observational (typically longitudinal time series designs)
	Evaluation Method	If Experimental, then select: Randomised controlled trial If Quasi-experiment or natural experiment, then select: Natural experiment in which exposure to treatment is random Regression Discontinuity Design (RDD) Difference-in-Differences (DID) / Fixed effects estimation Instrumental variable (IV) estimation Endogenous treatment-effects models (including endogenous switching regression, and other methods synonymous to the Heckman two step model) Statistical matching (includes PSM or statistical weighting) Interrupted time series (ITS) Synthetic controls If observational, then select: Time series (without interruption) Other
	Additional Methods	Select additional method if any. If none, select not applicable. [Open Answer]

Estimate Information	Analysis type for this effect size	Free text, what type of analysis was used (Regression, 2SLS, ANCOVA, etc.)
	Estimate Type	Type of data for this effect size: 1 = Continuous - means and SDs, 2 = Continuous - mean difference and SD, 3 = Dichotomous outcome - proportions, 4 = Regression data - dichotomous outcome, 5 = Regression data - continuous outcome
	Unit of analysis	What is the unit of analysis? UOA for this effect size: 1= Individual, 2= Household, 3= Group (e.g., community organisation), 4= Village, 5 = Other, 6 = Not clear
	Source	Note the page number, table number, column, and row you used to extract the data [Open Answer]
Treatment variable information	Treatment	Record the treatment variable as written in the model (e.g., the variable name the author uses, such as ("Teleconnection x Time") [Open Answer]
	Treatment type	Describe the types of treatment variable used: i) binary; ii) continuous; iii) categorical; iv) other
	Comparison	1=No intervention (service delivery as usual), 2=Other intervention, 3=Pipeline (waitlist) control (still service delivery as usual)
	Describe Comparison Group	Describe the comparison group [Open Answer]
	Subgroup	Is this analysis of a subgroup or estimating heterogeneous effects? 0=no, 1=yes
	Subgroup information	Describe the subgroup or variable interacted with the treatment variable if applicable (e.g., boys, girls). If no subgroup, select not applicable [Open Answer]
Outcome Information	Outcome description	Record the outcome for the corresponding effect size. Use this open answer field to enter, in the author's own words, a description of the outcome. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the outcome. To the extent possible, be sure to include numbers, units, population, and comparators. Include page numbers with every excerpt extracted.
	Outcome codes	Choose an outcome code for each corresponding effect size: i) Health; ii) Conflict and Violence; iii) Economics; iv) Migration; v) Food and nutrition security
	Outcome sub-group	Choose an outcome sub-group code for each corresponding effect size: ● Health Direct injuries and fatalities Disruption of health services Morbidity and mortality Cholera

- Enteric infectious diseases
- Malaria
- Zoonotic and vector-borne diseases
- Respiratory infections and ailments
- Mental health and psychological effects
- Heat stress
- Other communicable diseases

- Conflict and violence
  - Local conflict
  - Trans-border conflict
  - Domestic abuse/IPV
  - Crime
  - Extremism
  - Disputes
  - Militarized conflicts
- Economics
  - Total income and wealth
  - Aggregated production
  - Employment
  - Productivity
  - Trade
  - Consumption and expenditures
  - Prices
  - Investments
  - Economic supply chains
  - Inequalities
  - IT
  - Empowerment
- Migration
  - Internal/domestic
  - International/cross-border
  - Transhumance

		Economic/labour ● Food and nutrition security Food and nutrition security Malnutrition
	Outcome description	Record the outcome for the corresponding effect size. Use this open answer field to enter, in the author's own words, a description of the outcome. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the outcome. To the extent possible, be sure to include numbers, units, population, and comparators. Include page numbers with every excerpt extracted.
	Post-intervention or change from baseline?	0 = Post-intervention, 1 = Change from baseline
Estimate data	Slope coefficient	Trend estimate for time series designs
	Data points before	Number of data points before treatment (interrupted time series only)
	Data points after	Number of data points after treatment (interrupted time series only)
	Mean treatment	Outcome mean for the treatment group
	SD treatment	Outcome standard deviation for treatment group
	Mean Control	Outcome mean for the comparison group
	SD Control	Outcome standard deviation for control group
	Mean difference	Overall mean difference (treatment - control)
	SE difference	Standard error of the overall mean difference
	Tstat difference	t-statistic of mean difference
	p-value difference	p-value of mean difference
	Odds ratio	Odds ratio reported in the study
	SE odds ratio	Odds ratio standard error reported in the study
	Risk ratio	Risk ratio reported in study
	SE risk ratio	Risk ratio standard error
	Coeff reg	Report the regression coefficient of the treatment effect
	SE reg	Report the associated standard error of the regression coefficient.
	Tstat reg	Report the associated t statistic of the effect size (coefficient/SE)

CI_LB reg	Report the associated Lower bound of the 95% Confidence interval of the effect size. If CI is reported for a different confidence level, indicate that in the notes section.
CI_UP reg	Report the associated Upper bound of the 95% Confidence interval of the effect size. If CI is reported for a different confidence level, indicate that in the notes section.
P value exact	Exact p value if given, if not, record as written in the manuscript (e.g., $p < .001$, or $p > .05$)
Clusters treatment	Number of clusters - treatment group
Clusters control	Number of clusters - control group
Clusters total	Number of clusters - total sample
N treatment	Sample size - treatment group
N control	Sample size - control group
N total	Sample size - total sample
periods (1 if cross sectional)	Record how many time-period there are in the evaluation (e.g., cross section is 1, panel data with 3 measurements is 3)
Does the sample size need to be corrected?	Often in panel data, models will report number of observations rather than number of participants. In this column you will indicate 1="Yes" if the sample size needs to be divided by the number of periods, and 0="No" if either it is cross-sectional data, or if the authors have already divided the number of observations by the number of panel assessments and thus no correction is necessary.

Appendix 6: Risk of Bias Assessment Tool for Quasi-Experimental Designs

Code	**Coder**	**General**	**General**	**General**
Question		Time taken to complete assessment	Study first author	Outcome
Coding		Minutes	Open answer	Open answer
Criteria				
Decision-rules				
Response	Core Team			
Response	**Core Team**			
General	**General**	**General**	**General**	**1: Selection bias - Assessment**
Study design: What type of study design is used?	Methods used for analysis: Which methods are used to control for selection bias and confounding?	Ethical clearance	Study registration	1 - Mechanism of assignment: was the allocation or identification mechanism able to control for selection bias?
1= Natural experiment: randomised or as-if randomised 2= Natural experiment: regression discontinuity (RD) 3= CBA (non-randomised assignment with treatment and contemporaneous comparison group, baseline and end line data collection) – individual repeated measurement 4= CBA pseudo panel (repeated measurement for groups but different individuals)	1 = Statistical matching (PSM, CEM, covariate matching) 2 = Difference in differences (DID) estimation methods 3 = IV-regression (2-stage least squares or bivariate probit) 4 = Heckman selection model 5 = Fixed effects regression 6 = Covariate adjusted estimation 7 = Propensity weighted regression	Open answer	Open answer	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear

5= Interrupted time series (with or without contemporaneous control group) 6= Panel data, but no baseline (pre-test) 7 = Comparison group with end line data only	8 = Comparison of means 9 = Other (please state)			
	-	Provide any details of ethical research clearances granted. Report unclear if this information is not available.	Provide any details of study registration, including registry IDs, etc.	

1: Selection bias - Justification	1: Selection bias - Justification	2: Confounding - Assessment	2: Confounding - Justification	2: Confounding - Justification
For regression discontinuity designs	For assignment based non-randomised programme placement and self-selection (studies using a matching strategy or regression analysis, excluding IV)	2 - Group equivalence: was the method of analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding?	For regression discontinuity design	For non-randomised trials using difference-in-differences methods of analysis
Open answer	Open answer	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	Open answer	Open answer
a) Allocation is made based on a pre-determined discontinuity on a continuous variable (regression discontinuity design) and blinded to participants or; b) if not blinded, individuals reasonably cannot affect the assignment variable in response to knowledge of the participation decision rule; c) and the sample size immediately at both sides of the cut-off point is sufficiently large to equate groups on average.	a) Participants and non-participants are either matched based on all relevant characteristics explaining participation and outcomes, or; b) all relevant characteristics are accounted for.** c) and the data set used contains relevant variable that are measured in a relevant way (i.e. they were not collected for a different purpose initially and therefore are good proxy for some characteristics). **Accounting for and matching on all relevant characteristics is usually only feasible when the programme allocation rule is known and there are no errors of targeting. It is unlikely that		a) 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; b) 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; c) Significant differences in covariates of the individuals have been controlled in multivariate analysis; and for cluster-assignment, authors control for external cluster-level factors that might	a) The authors use a difference-in-differences (or fixed effects) multivariate estimation method; b) the authors control for a comprehensive set of individual time-varying characteristics, and for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme**; c) 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 (for example, by examining correlation with determinants of outcomes, in both treatment and comparison groups);

	studies not based on randomisation or regression discontinuity can score “YES” on this criterion. There are different ways in which covariates can be taken into account. Differences across groups in observable characteristics can be taken into account as covariates in the framework of a regression analysis or can be assessed by testing equality of means between groups. Differences in unobservable characteristics can be taken into account through the use of instrumental variables (see also question 1.d) or proxy variables in the framework of a regression analysis, or using a fixed effects or difference-in-differences model if the only characteristics which are unobserved are time-invariant		confound the impact of the programme.	**Knowing allocation rules for the programme – or even whether the non-participants were individuals that refused to participate in the programme, as opposed to individuals that were not given the opportunity to participate in the programme – can help in the assessment of whether the covariates accounted for in the regression capture all the relevant characteristics that explain differences between treatment and comparison
Score “Yes” if criteria a), b), c) are all satisfied Score "Probably Yes" if there are minor differences in between both sides of the cut-off point but authors convincingly argue that the differences are unlikely to affect the outcome, OR individuals are not blinded	Score “Yes” if a) or b) and c) are satisfied Score "Probably yes" if a) or b) are addressed for but there is some doubt related to c), OR authors combined statistical matching and difference-in-difference to cope with unobservable differences, OR they only did statistical		Score "Yes, if criterion a), b), c) and d) are addressed. Score "Probably yes" if b) is not addressed but c) is addressed and differences in means are not large. Score “Unclear” if insufficient details are provided on controls; or if	Score "Yes, if a, b, c, d (if relevant) are addressed and baseline imbalances between groups were relatively low OR the method was combined by a statistical matching. Score "Probably yes" if all possible variables are controlled for and the

and there are low risk of them affecting the assignment but the authors do not mention it. Score “Unclear” if it is unclear whether participants can affect it in response to knowledge of the allocation mechanism. Score "Probably No" if there are differences between individuals on both sides of the cut-off point, and there are doubts that the differences are due to individuals altering the assignment OR the participants are blinded but there is evidence that the decisions that determined the discontinuity is based on differences between the two groups or differences in time. Score “No” if the sample size is not sufficient OR there is evidence that participants altered the assignment variable prior to assignment. If the research has serious concerns with the validity of the assignment process or the group equivalence	matching and there were clear rules for selection into the program (no self-selection). 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 "Probably no" if only a statistical matching was done and there was self-selection into the program. Score “No” if relevant characteristics are omitted from the analysis.		insufficient details are provided on cluster controls. Score "Probably no" if b) is not addressed (absence of a difference test or balance table) and there are doubt regarding the continuity on both sides of the cut-off point (a). Score “No” otherwise.	selection into the program was done according to clear rules, but baseline imbalances between groups were very large. Score “Unclear” if insufficient details are provided; or if insufficient details are provided on cluster controls. Score "Probably no" if some time-varying characteristics are not controlled for and the program was self-selected by the intervention groups. Score “No” if any of the criterion is not addressed.

completely fails, we recommend assessing risk of bias of the study using the relevant questions for the appropriate methods of analysis (cross-sectional regressions, difference-in-difference, etc.) rather than the RDDs questions.				

2: Confounding - Justification	**2: Confounding - Justification**	**3: Performance bias - Assessment**	**3: Performance bias - Justification**	**4: Spillovers, crossovers and contamination - Assessment**
For statistical matching studies including propensity scores (PSM) and covariate matching** **Matching strategies are sometimes complemented with difference-in-difference regression estimation methods. This combination approach is superior since it only uses in the estimation the common support region of the sample size, reducing the likelihood of existence of time-variant unobservable differences across groups affecting outcome of interest and removing biases arising from time-invariant unobservable characteristics.	For regression-based studies using cross sectional data (excluding IV)	3 - Performance bias: was the process of being observed free from motivation bias?	Performance bias - Justification	4 - Spillovers, crossovers and contamination: was the study adequately protected against spillovers, crossovers and contamination?
Open answer	Open answer	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	Open answer	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear
a) 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	a) The study controls for relevant confounders that may be correlated with both participation and explain outcomes (for example, demographic and socio-economic factors at	a) For data collected in the context of a particular intervention trial (randomised or non-randomised assignment), the authors state explicitly that the process of monitoring the	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages).	a) There were no implementation issues that might have led the control participants to receive the treatment (implementer's mistake). b) The intervention is

relevant (for example, 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); and, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme b) in addition, for PSM Rosenbaum's test suggests the results are not sensitive to the existence of hidden bias; c) and, with the exception of Kernel matching, the means of the individual covariates are equated for treatment and comparison groups after matching; d) different matching methods including varying sample sizes yields the same results and authors take into account the use of control observations multiple times against the same treatment in their standard error calculation.	individual and community level) using multivariate methods with appropriate proxies for unobservable covariates, and, for cluster-assignment, authors control particularly for external cluster-level factors that might confound the impact of the programme; b) and a Hausman test with an appropriate instrument suggests there is no evidence of endogeneity**; c) and none of the covariate controls can be affected by participation; d) 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; **The Hausman test explores endogeneity in the framework of regression by comparing whether the OLS and the IV approaches yield significantly different estimations. However, it	intervention and outcome measurement is blinded, or argue convincingly why it is not likely that being monitored could affect the performance of participants in treatment and comparison groups in different ways (such as resulting in Hawthorne or John Henry effects). 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 from administrative records or in the context of a retrospective (ex post) evaluation.		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 not likely) or the potential effects of spill overs were measured (e.g. variation in the % of unit within a cluster receiving the treatment). c) There is no risk of contamination by external programs: the treatment and comparisons are isolated from other interventions which might explain changes in outcomes. d) There is nothing in the surveys that might have given the control participants an idea of what the other group might receive OR they did but there is no risk that this has changed their behaviours; AND the survey process did not reveal information to the control group that they did not have before (e.g. the study aims to measure increase in take up of a service or product that participants might not know

	plays a different role in the different methods of analysis. While in the OLS regression framework the Hausman test mainly explores endogeneity and therefore is related with the validity of the method, in IV approaches it explores whether the author has chosen the best available strategy for addressing causal attribution (since in the absence of endogeneity OLS yields more precise estimators) and therefore is more related with analysis reporting bias.			about) Authors might put something in place in the design of the study that allows to control for that survey effect (e.g. a pure control with no monitoring except baseline end line)
Score "Yes, if a, b, c, and d (if relevant) are addressed. Score "Probably yes" if the selection into the program was done according to clear rules, which are used for the matching but there are slight imbalances remaining after matching. Score "Unclear" if relevant variables are not included in the matching equation, or if matching is based on characteristics collected at end line; or if insufficient	Score "Yes, if a, b, c and d are addressed. Score "Probably yes" if all criteria are addressed but authors did not report the Hausman test (b). 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 "Probably no" if any	Score "Yes" if either criterion a) or b) are satisfied; Score "Probably yes" if the study is based on survey data collected during a trial and there is no obvious issue with the monitoring processes, but authors do not mention potential risks. Score "Unclear" if it is not clear whether the authors use an appropriate method to prevent Hawthorne and John Henry Effects (e.g. blinding		Score "Yes" if criterion a), b), c) and d) are satisfied; Score "Probably yes" if there is no obvious problem but there is no information reported on potential risks related to spill overs, contamination, or survey effects in the control group OR if there were issues with spillovers but they were controlled for or measured. Score "Unclear" if spillovers, crossovers, survey effects and/or contamination

details are provided on cluster controls. Score "Probably no" if the program was self-selected by the intervention groups or participants OR if the selection into the program was done according to clear rules but there is no baseline data available to match the participants or groups on. Score "No" if matching was done based on variables that are likely to be affected by the program or any other scenario that affect a), b) c) or d).	of the criterion other than b) is not addressed. Score "No" if none of the criterion are addressed.	of outcomes and, or enumerators, other methods to ensure consistent monitoring across groups). Hawthorne effects may result where participants know that they are being observed and John Henry Effects may result from participant knowledge of being compared. Score "Probably no" if there was imbalance in the frequency of monitoring in intervention groups, which might have influenced participants' behaviours. Score "No" if neither criterion a) or b) are satisfied;		are not addressed clearly. Score "Probably no" if any of the criterion a), b), c) or d) are not satisfied but the scale of the issue is not clear. Score "No" if any of the criterion a), b), c) or d) are not satisfied and happened at a large scale in the study.
4: Spillovers, crossovers and contamination - Justification	**5: Outcome measurement bias - Assessment**	**6: Reporting bias - Assessment**	**6: Reporting bias - Justification**	**7: Other bias - Assessment**
Spill-overs, crossovers and contamination - Justification	5 - Outcome measurement bias	6 - Selective analysis reporting: was the study free from selective analysis reporting?	Analysis reporting bias - Justification	7 - Other risks of bias: Is the study free from other sources of bias?
Open answer	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	Open answer	1= Yes, 4 = No
Justification for coding decision (Include a brief summary of	a) Outcome assessors are blinded, or the outcome measures are not likely to be	a) a pre-analysis plan is published, especially for prospective NRS but it	Justification for coding decision (Include a summary of	Score "Yes" if the reported results do not suggest any other sources of bias.

justification for rating, mentioning your response to all sub questions, cite relevant pages).	biased by their judgement. b) For self-reported outcomes: respondents in the intervention group are not more likely to have accurate answers due to recall bias; c) For self-reported outcomes: respondents do not have incentives to over/under report something related to their performance or actions, OR researchers put in place mechanisms to reduce the risk of reporting bias (researchers not strongly involved in the implementation of the program and it is clear that their answers to the survey will not affect what they receive in the future) OR authors have measured the risks of bias through falsification tests or measuring the effect on placebo outcomes in cases where there was a risk of reporting bias. d) Timing issue: the data collection period did not differ between intervention and comparison group; the baseline data is not likely to be affected by the beginning of the intervention or affects	should also be for retrospective studies b) authors use 'common' methods of estimation (i.e. credible analysis method to deal with attribution given the data available); c) There is no evidence that outcomes were selectively reported (e.g. results for all relevant outcomes in the methods section are reported in the results section); d) Requirements for specific methods of analysis: - For PSM and covariate matching: (a) Where over 10% of participants fail to be matched, sensitivity analysis is used to re-estimate results using different matching methods (Kernel Matching techniques); (b) For matching with replacement, no single observation in the control group is matched with a large number of observations in the treatment group. - For IV (including Heckman) models, (a) The authors test and report the results of a Hausman test for exogeneity ($p \leq 0.05$ is required to reject the null	justification for rating, mentioning your response to all sub questions, cite relevant pages).	Score "No" if other potential threats to validity are present, and note these here (e.g. coherence of results, survey instruments used are not reported)

	a small percentage of the study participants.	hypothesis of exogeneity); (b) the coefficient of the selectivity correction term (Rho) is significantly different from zero (P<0.05) (Heckman approach). - For studies using multivariate regression analysis, authors conduct appropriate specification tests (e.g. testing robustness of results to the inclusion of additional variables, or (very rare) reporting results of multicollinearity test etc).		
	Score “Yes” if criterion a), b), c) and d) are satisfied: Score "Probably yes" if there is a small risk related to any of a), b), c) or d) and there is no more information provided to justify the absence of bias OR if there was a high risk of bias but authors have either controlled it in their design or measured it with a placebo outcomes. Score “Unclear” if it there is a high risk related to any of a), b), c) or d) and there is no more information provided to justify the absence of bias.	Score “Yes” if a), b), c) and d) are satisfied OR if a) is not met and it is a retrospective NRS. Score "Probably Yes" if authors combined methods and reported relevant tests (d) only for one method OR if all the criteria are met except for a) and it is a prospective NRS Score "Unclear" if intended outcomes not specified in the paper OR if any of the requirements for d) are not reported. Score "Probably No" if b) is addressed, but authors did not present results for all outcomes announced in the		

	Score "Probably no" if there are high risk related to a), b), c) or d) and it is clear that authors were not able to control for this bias. Score “No” if there is evidence of bias.	method section OR did not meet requirement d) although reported. Score “No” if authors use uncommon or less rigorous estimation methods such as failure to conduct multivariate analysis for outcomes equations OR if some important outcomes are subsequently omitted from the results or the significance and magnitude of important outcomes was not assessed.		

7: Other bias - Justification	**8: External validity**
Other risks of bias - Justification	8 - External validity
Open answer	Open answer
Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages).	Open answer- what do authors say about external validity, if anything?

Appendix 7: Risk of Bias Assessment Tool for Observational Designs

1. Risk of Bias - Study design			
Was the analytical approach reasonable for the research question specific to the effect size extracted for this analysis?	Were tests supporting the identification strategy, selection of methods or results reported?	Was there adequate adjustment for confounding in the analyses?	Justification for the answers to the study design section.
1=Yes, 2=No, 3=Unclear, 4=N/A	1=Yes, 2=No, 3=Unclear, 4=N/A Indicate 'yes' for studies that report tests of assumptions, use goodness-of-fit measures for model selection, conduct robustness checks, perform sensitivity tests, assess prediction power, or provide other evidence supporting the study design.	1=Yes, 2=No, 3=Unclear, 4=N/A Indicate yes if authors identify and control from potential confounding factors.	Free text. Include page numbers

2. Risk of Bias - Data quality		
Were outcome measures objective and free from reporting bias?	Was the frequency and length of the data appropriate to answer the research question?	Justification for the answers to the data quality section.
1=Yes, 2=No, 3=Unclear, 4=N/A	1=Yes, 2=No, 3=Unclear, 4=N/A	Free text. Include page numbers

3. Risk of Bias Reporting		
Were conclusions consistent with the unit of analysis and reported results?	Did all reported results correspond to all intended analyses, avoiding "data dredging" and selective reporting?	Justification for the answers to the reporting section.
1=Yes, 2=No, 3=Unclear, 4=N/A	1=Yes, 2=No, 3=Unclear, 4=N/A	Free text. Include page numbers