

Systematic Review of Evaluation Reports

Effects of Development Aid to South Sudan and Mali, 2008 - 2021

Methods

Christoph Zuercher (principal investigator) and Patrick Labelle (research librarian)

Table of Contents

Why this review	2
Research questions	2
Type of systematic review	2
Search strategy	3
Databases	3
Sample search string for database search	3
Bi-lateral donors	5
Multilateral and international organizations	5
Repositories of impact evaluations in international development	6
Major developmental NGOs	7
Google	8
Outreach	9
Inclusion criteria	9
Screening	11
PRISMA Diagram	11
Ex-post categorizing studies according to sectors	11
Final sample per country, sector and type of study	12
Data Extraction and Reporting	13
Using ATLAS.ti for in depths analysis	13

Why this review

The review intends to provide solid evidence about the impacts of aid to Mali and South Sudan between 2008 and 2021. Both these countries are among the most fragile countries in the world, and there is a need to better understand what works and what does not work in these contexts.

The review replicates the “Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008 – 2018. Summary Paper”.¹ It complements the existing study on Afghanistan and further contributes to generalizable insights about the effectiveness of aid interventions in fragile and conflict affected states.

Research questions

The main objective of this review is to identify all evaluation reports on aid to Mali and South Sudan (such as impact evaluations, performance audits, formative evaluations, bilateral country-level evaluations, systematic reviews, etc.) produced by a wide range of relevant bi- and multilateral actors, across all aid sectors. The identified studies are summarized with regard to the effects of aid across all sectors.

The following four research questions were specified in the Term of Reference:

- First, what type of interventions in what sectors were effective (or not)?
- Second, which evaluation methods proved adequate for the contexts of Mali and South Sudan?
- Third, what are lessons learned for future evaluations in fragile contexts?
- Fourth, what are overarching lessons about effectiveness of aid interventions in fragile and conflict affected contexts?

Type of systematic review

This systematic review is a convergent mixed-methods review², which includes both statistical/experimental and qualitative evaluations in the final synthesis.

While well executed statistical/experimental evaluations are often thought to produce more robust results than qualitative evaluations, good qualitative evaluations can also provide valuable insights. Furthermore, the majority of available evaluations is qualitative. By not taking into account solid qualitative evaluations, a lot of valuable information would be lost. Also, restricting the study to statistical/experimental studies could introduce bias since this type of studies tends to be much more prevalent in sectors such as health and nutrition, and less so in

¹ Zürcher, Christoph. 2020. Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008 – 2018. Summary Paper. BMZ, Berlin and Bonn. Available at <https://christophzuercher.weebly.com/papers--reports.html>

² On convergent mixed-methods review, see Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris J, Garside R, Hannes K, Pantoja T, Thomas J. Chapter 21: Qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

other sectors. For these reasons, it was decided to also include solid qualitative evaluation reports.

In our reporting of findings, the evidence from statistical evaluations is kept separate from the evidence from qualitative studies. This allows readers to see how the evidence from quantitative and qualitative studies ‘speaks to each other’.

Search strategy

We revised and expanded the search strategy used in Zürcher (2020).³ We searched in relevant databases, on the website of relevant multi- and bilateral donors, in relevant aid document repositories, and on the websites of relevant NGOs. We also searched google.com and, in order to capture documents in French, on google.fr.

Databases

The following databases were searched:

1. Academic Search Complete
2. AfricaBib.org
3. Cairn
4. EconLit
5. Érudit
6. GenderWatch
7. Global Health
8. International Pol. Sci. Abstracts
9. MEDLINE
10. PAIS Index
11. Pascal (up to 2015)
12. RePEc / IDEAS
13. Web of Science
14. Worldwide Pol. Sci. Abstracts

³ Zürcher, Christoph. 2020. Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008 – 2018. Summary Paper. BMZ, Berlin and Bonn. Available at <https://christophzuercher.weebly.com/papers--reports.html>

Sample search string for database search

Below we give the search string for PAIS on PROQUEST. The search strings for other platforms are available on request.

<p>MAINSUBJECT.EXACT("Foreign Aid" or "Humanitarian Aid" or "Development Policy" or "Development Programs" or "Development Strategies" or "International Relief" or "Economic Assistance" or "Disaster Relief") OR MAINSUBJECT.EXACT.EXPLODE("Development")</p>
<p>TI,AB((foreign OR international* OR humanitar* OR development OR multilateral OR multi-lateral OR bilateral OR bi-lateral OR economic* OR government OR "public sector*" OR civic OR civil OR democra* OR election* OR electoral OR decentrali* OR communit* OR donor* OR budget* OR financ* OR savings OR tax OR trade OR "private sector*" OR conditional OR governance OR oversea* OR hospital* OR health* OR nutrition* OR sanitation OR water OR hygiene OR famine* OR hunger OR agricult* OR farm* OR livestock* OR irrigation OR flood OR drought* OR "natural resource*" OR mining OR food OR "human right*" OR empower* OR livelihood* OR education* OR vocational OR school* OR teacher* OR literacy) NEAR/3 (aid OR assistance OR intervention* OR cooperati* OR co-operati* OR project* OR program* OR policy OR policies OR support* OR fund* OR relations OR charit* OR relief OR subsid* OR contribution* OR grant* OR development* OR strateg* OR initiative*))</p>
<p>TI,AB((technical OR media* OR employment OR social OR gender OR women OR girls OR capacity OR stabilization OR disaster* OR emergenc* OR reconstruction OR infrastructure OR road* OR highway* OR rail* OR hydro* OR electric* OR energy OR land OR community OR energy OR military OR civil OR justice OR judicial OR law OR peace OR security OR police OR disarm* OR crisis OR conflict* OR "anti-corruption" OR "nongovernmental organi?ation*" OR ngo* OR "international organi?ation*" OR igo*) NEAR/3 (aid OR assistance OR intervention* OR cooperati* OR co-operati* OR project* OR program* OR policy OR policies OR support* OR fund* OR relations OR charit* OR relief OR subsid* OR contribution* OR grant* OR development* OR strateg* OR initiative*))</p>
<p>TI,AB(credit OR microcredit OR "micro-credit" OR "micro-finance" OR microfinance OR "micro loan*" OR microloan* OR "conditional cash transfer*" OR cct OR entrepreneur* OR "micro enterprise*" OR microenterprise* OR "small and medium enterprise*" OR sme OR smes)</p>
<p>TI,AB(mali OR malian* OR "south sudan*")</p>
<p>MAINSUBJECT.EXACT("Evaluation" OR "Program Evaluation" OR "Effects" OR "Efficiency" OR "Effectiveness" OR "Performance")</p>
<p>TI,AB(evaluat* OR impact* OR effect* OR overview* OR outcome* OR lesson* OR findings OR performance OR results OR monitoring OR efficien* OR assess* OR audit* OR review* OR analy*)</p>
<p>6 OR 7</p>
<p>1 AND 5 AND 8 (limited to 2008- from results page)</p>

2 AND 5 AND 8 (limited to 2008- from results page)
3 AND 5 AND 8 (limited to 2008- from results page)
4 AND 5 AND 8 (limited to 2008- from results page)
Results of lines 9-12 were selected and added to the "Selected Items" folder (duplicates from the four sets are removed automatically)

Bi-lateral donors

After the database search, searches were conducted on the website of bilateral donors. These websites provide a variety of user interfaces. When offered, we filtered for “evaluations” and then searched or filtered for Mali and South Sudan. In other cases, we began with searching for Mali or South Sudan, and then hand-searched for evaluation reports.

1. US / USAID (Development Experience Clearinghouse)
2. UK (Foreign, Commonwealth & Development Office, formerly DfID)
3. Canada (GAC)
4. Australia (DFAT)
5. New Zealand (MFAT)
6. Germany (KfW, GIZ and BMZ)
7. France (Agence française de développement AFD)
8. Italy (Italian Agency for Development Cooperation (AICS)
9. Sweden (SIDA)
10. Norway (NORAD)
11. Denmark (Danida)
12. Finland (Finnida)
13. Belgium (Enabel)
14. Netherlands (Ministry of Foreign Affairs)
15. Switzerland (DEZA)
16. Japan (JICA)
17. China (China International Development Cooperation Agency (CIDCA)

Multilateral and international organizations

Searches were also conducted on the website of multilateral and international organizations. These websites provide a variety of user interfaces. When offered, we filtered for “evaluations” and then searched or filtered for Mali and South Sudan. In other cases, we began with searching for Mali or South Sudan, and then hand-searched for evaluation reports.

1. African Development Bank (AfDB)
2. African Union
3. European Bank for Reconstruction and Development (EBRD)
4. European Investment Bank
5. European Bank for Reconstruction and Development (EBRD)
6. European Investment Bank
7. European Union

8. International Bank for Reconstruction and Development (IBRD; part of the World Bank Group)
9. International Fund for Agricultural Development (IFAD)
10. International Monetary Fund (IMF)
11. International Organization for Migration (IOM)
12. UNMAS, United Nations Mine Action Service
13. United Nations (UN)
14. United Nations Children's Fund (UNICEF)
15. United Nations Conference on Trade and Development (UNCTAD)
16. United Nations Development Programme (UNDP), Evaluation Resource Center
17. United Nations Environment Programme (UNEP)
18. United Nations High Commissioner for Refugees (UNHCR)
19. United Nations Industrial Development Organization (UNIDO)
20. World Bank Group (esp. World Bank e-library: <https://elibrary.worldbank.org/>), CAS Completion Report Review, Country Performance Portfolio Review, IEG Evaluations, Impact Evaluation
21. World Food Programme (WFP), Evaluation Library
22. World Health Organization (WHO)
23. UN Women, GATE System
24. OECD DEREc

Repositories of impact evaluations in international development

Searches were also conducted in repositories of impact evaluations in international development. These websites provide a variety of user interfaces. When offered, we filtered for "evaluations" and then searched or filtered for Mali and South Sudan. In other cases, we began with searching for Mali or South Sudan, and then hand-searched for evaluation reports.

1. 3ie RIDIE (Registry for International Development Impact Evaluations)
2. 3ie Development Evidence Portal
3. AgEcon
4. AGRIS
5. BREAD
6. Center for Effective Global Action (CEGA)
7. CGIAR: Consultative Group on International Agricultural Research
8. DEval
9. GEF (Global Environmental Facility)
10. Global Facility for Disaster Reduction and Recovery:
11. ICNL Research Centre
12. IFPRI
13. Independent Development Evaluation, AfDB
14. J-Poverty Action Lab (J-PAL)
15. Millennium Challenge Cooperation
16. RePEC IDEAS

Major developmental NGOs⁴

Searches were also conducted on websites of selected NGOs. These websites provide a variety of user interfaces. When offered, we filtered for “evaluations” and then searched or filtered for Mali and South Sudan. In other cases, we began with searching for Mali or South Sudan, and then hand-searched for evaluation reports.

1. ACTED
2. ActionAid
3. Aga Khan Development Network
4. CARE International: <http://www.careevaluations.org/>
5. Catholic Relief Services: <https://www.crs.org/our-work-overseas/research-publications>
6. Danish Refugee Council
7. IRC: <https://www.rescue.org/reports-and-resources>
8. Médecins sans frontières
9. Mercy Corps: <https://www.mercycorps.org/research>
10. Oxfam International: <https://policy-practice.oxfam.org.uk/publications>
11. Plan International
12. Samuel Hall (evaluations): <http://samuelhall.org/category/publications/>
13. Save the Children
14. Welthungerhilfe
15. World Vision: <http://www.wvi.org/resources>
16. HALO Trust
17. Oxfam Novib
18. Save the Children NL

In addition, we will also screened studies provided by the Ministry of Foreign Affairs of the Netherlands from the following organizations:

1. CARE NL
2. World Vision Netherlands
3. Spark
4. IKV Pax Christi
5. VNG International
6. ZOA
7. AWEPA
8. The Global Partnership for the Prevention of Armed Conflict
9. SOMO
10. The Centre for conflict resolution
11. Search for Common Ground
12. Cordaid
13. Saferworld
14. Hivos
15. Netherlands Institute for Multiparty Democracy (NIMD)
16. ICCO
17. Agency for Cooperation and research in Development – Community (ACORD)
18. War Child
19. TEAR
20. Dorcas
21. Netherlands Red Cross

⁴ Based on an expanded list by Morton, B., 2013. An overview of international NGOs in development cooperation. United Nations Development Program, case study 7

22. Conflict Dynamics
23. Stichting Red een Kind
24. LDSPS
25. Norwegian Church Aid
26. Woord en Daad
27. Mercy Corps Europe
28. Norwegian Refugee Council (NRC)
29. Helvetas
30. ACTED
31. Aidsfonds
32. Amref
33. Both Ends
34. COC
35. Fair Wear
36. FCAM
37. Free Press
38. GPPAC
39. IUCN
40. Mama cash
41. Milieudefensie
42. Plan International (and Plan International Nederland)
43. Rode Kruis (Red Cross)
44. Rutgers
45. SNV
46. Solidaridad
47. UTZ
48. PACJA
49. Wereld Natuur Fonds (WWF)
50. The Hunger Project
51. CMC – Mensen met een Missie
52. Wilde Ganzen/Ikon
53. Fair Wear Foundation
54. CREA
55. Just Associates
56. Kvinna till Kvinna
57. SIHA Network
58. Womankind Worldwide
59. World YMCA
60. Wemos

Google

We used the following five search strings for google.com and google.fr:

1. South Sudan AND evaluation
2. Mali AND evaluation
3. intitle:South Sudan AND (evaluation OR impact OR effect OR overview OR outcome OR lesson OR findings OR performance OR results OR monitoring OR assessment OR audit OR review OR analysis).
4. intitle:Mali AND (evaluation OR impact OR effect OR overview OR outcome OR lesson OR findings OR performance OR results OR monitoring OR assessment OR audit OR review OR analysis)
5. intitle:Mali ET (évaluation OU impact OU effet OU résultat OU leçon OU performance OU résultats OU audit OU analyse OU rapport)

We screened all search result pages until there were no more relevant hits, typically between 7 and 10 pages.

Outreach

We contacted the evaluation departments of all above listed bilateral donors and major NGOs, as well as the World Bank's country offices with a request to provide us with evaluations on aid to Mali and South Sudan.

Inclusion criteria

Studies were included when they met the following criteria:

1. Published in English between 2008 and 2021, and published in French between 2008 and 2021 (only for Mali).
2. Provide an assessment of the outcomes and impacts of projects, programs, multi-sectoral programs, and country-level assistance. Studies which only report outputs were not included.
3. Studies were included when they meet one of two thresholds for methodological quality:
 - a. They are what we call '**rigorous impact evaluations**', with a logically or statistically measured value for the counterfactual. Typically, rigorous impact evaluations use one of the following research designs:
 - Randomized controlled trials (RCTs)
 - Regression discontinuity designs
 - Natural experiments
 - Non-randomized studies with pre-intervention and post-intervention outcomes data in treatment and comparisons groups
 - Difference-in-difference designs
 - interrupted time series
 - Non-randomized studies with control for observable confounding, including various matching design

- Regression designs, including repeated cross-sectional regressions
- b. We also included studies which are mostly qualitative, but still meet a certain quality threshold. We call this type of study ‘**good enough evaluation**’. While rigorous evaluations are *likely* to capture the impacts of an intervention, it is *possible* that good enough evaluations capture the impacts. We used the following four necessary criteria for “good enough” evaluations.⁵
- The study must explicitly intend to assess outcomes and/or impacts of one or several specific interventions
 - The study must contain adequate primary data (typically quantified measures of outcomes, and/or data from interviews)
 - The study must demonstrate that it is plausible that the data is suited to attribute observed outcomes to the interventions
 - For interventions with a complex causal chain, the assumed theory of change / causal mechanisms / interventions logic is mentioned

These criteria were treated as ‘necessary’. Only studies which met all four criteria were included. We recognize that these criteria are, like all existing appraisal tools for non-experimental studies, open to interpretation. In order to minimize the effects of individual bias when assessing the quality of a study, the research team repeatedly applied these criteria to random subsamples of studies. Once all researchers had given their assessments, we discussed the reasoning for our assessments. By repeating this process, we worked toward a shared understanding of how to apply the criteria, resulting in more consistency when the criteria were applied by members of the research team. For the actual screening process, two researchers had to independently arrive at the same decision. The principal investigator acted as tie-breaker.

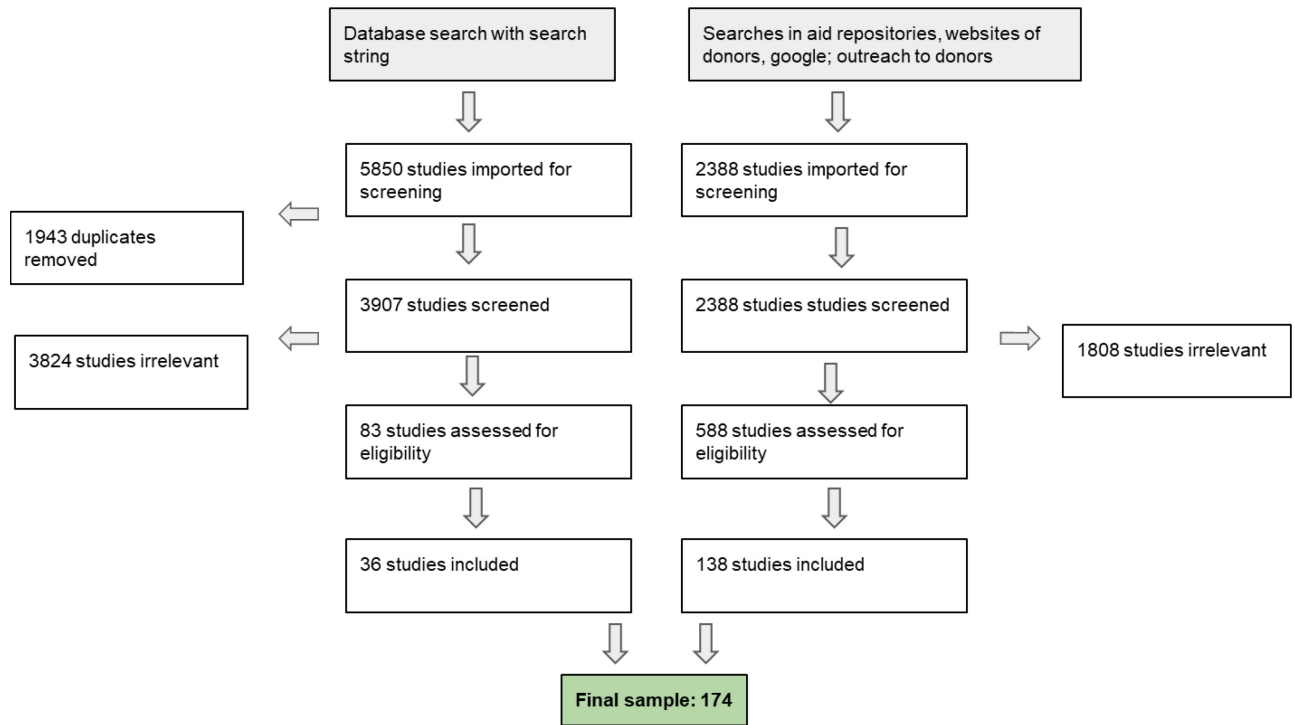
4. Also included were **country-level evaluations of bi-or multilateral donors**. Such studies were included by default and did not have to meet a threshold for methodological quality. Such evaluations are typically a lengthy and highly condensed synthesis of many program and project evaluations, and it is in our view not possible to develop a valid quality threshold. Nevertheless, these studies contain a wealth of aggregated information which we did not want to leave out of our analysis.

⁵ These criteria were developed by the research team in discussion with the IOB. The objective was to develop threshold criteria for differentiating between ‘good enough’ evaluations and evaluations which are unlikely to produce reliable results. The aim was to define a threshold that is theoretically plausible and feasible (that is, relatively easy to apply to a very large number of studies). There is no shortage of appraisal tools for qualitative studies. A very useful overview which we consulted is Majid, Umair, and Meredith Vanstone. 2018. “Appraising Qualitative Research for Evidence Syntheses: A Compendium of Quality Appraisal Tools.” *Qualitative Health Research* 28 (13): 2115–31. A synthesis of the most often used criteria for assessing quality (based on 58 appraisal tools across various disciplines) is provided in Santiago-Delefosse, M., A. Gavin, C. Bruchez, P. Roux, and S.L. Stephen. 2016. “Quality of Qualitative Research in the Health Sciences: Analysis of the Common Criteria Present in 58 Assessment Guidelines by Expert Users.” *Social Science & Medicine* 148 (January): 142–51. Another useful tool is CASP - (Critical Appraisal Skills Program). CASP was developed and widely used for health science (see <https://casp-uk.net/>). Our criteria are based on a streamlined combination of the CASP criteria and the most commonly used criteria in social science appraisals tools as described in Santiago-Delfosse et al. (2014).

Screening

Once the search was completed, title/abstract screening took place. We used the screening software Covidence. Two researchers needed to independently agree or disagree on a study. Conflicting cases were resolved by a third researcher, usually by the principal investigator.

PRISMA Diagram



Ex-post categorizing studies according to sectors

The included studies were categorized according to aid sectors. The definition of these aid sectors was done in cooperation with the IOB team, taking into account both priority sectors of the Netherlands as well as DAC CRS purpose codes. Studies were sorted in 10 predefined aid sectors: women's rights; health; rural development and climate change; rule of law; stabilization; education; sustainable economic development; nutrition; humanitarian assistance and good governance.

These sectors were defined based on discussion with the commissioning agency IOB. When distributing studies, we typically followed the designation given by the studies themselves. This explains why some types of interventions can be found in more than one sector. For example, WASH intervention can be found in humanitarian aid, but also in health, or support for village savings associations can be found in rural development, humanitarian aid, and stabilization.

Final sample per country, sector and type of study

	Mali (2008 - 2020)		South Sudan (2008 - 2020)			
	Included Studies, total	(of which "rigorous")	Included Studies, total	(of which "rigorous")	Total per sector	(of which "rigorous")
Country level program evaluations	18	na	12	na	30	na
Women's Rights	9	4	4		13	4
Health and Nutrition	35	27	12	7	47	34
Rural development, climate	26	8	7	2	33	10
Rule of Law			1		1	0
Stabilization	4		12	1	16	1
Education	5	3	3		8	3
Sustainable economic development			3	1	3	1
Humanitarian assistance, refugees, migration (a)	3	1	6	2	9	3
Good governance	4		8		12	0
Afghanistan Reconstruction Trust Fund (b)					0	0

other					0	0
Total (per country)	104	43	69	13		
		Total Studies			174	
		(of which rigorous)			56	

Data Extraction and Reporting

Data extraction for evaluations was based on a predefined template. Extracted data included population, intervention(s), comparator, outcome, methods, and moderators. In addition, we also extracted data for sustainability, efficiency and for barriers (factors which were said to hinder better implementation).

The reporting structure follows the structure of the extracted data, that is, for each study we report on population, intervention(s), comparator, outcome, methods, and moderators, sustainability, efficiency and barriers.

In the reporting we kept apart findings from (1) rigorous impact evaluations (2) 'good enough evaluations' and (3) county level bilateral evaluations. Findings from country level evaluations are summarized in a separate report. Findings from experimental and qualitative project and program level evaluations are presented in one report, but original studies are referenced so that readers know whether a study was experimental or qualitative. This allows readers to see how the evidence from quantitative and qualitative studies 'speaks to each other', and to what degree the evidence from these different sources (with different degrees of robustness) are mutually supportive and thus increase confidence in the validity of the overall findings.

Using ATLAS.ti for in depths analysis

In order to provide an additional in depth-analysis of capacity building (marked as a priority by the IOB), we used qualitative data analysis software (ATLAS.ti) to search for and extract relevant information in the full texts of all studies. We ran a text search for paragraphs that contained 'capacity' and 'impact' or synonyms for impact. The identified paragraphs were coded as 'capacity building impact' and analyzed for information about the impacts of capacity building measures across all types of interventions.