Evidence Mapping of Social Sciences Methods Used in Fragile and Conflict-Affected Settings (FCAS) (2015–2025)

Report prepared by
Lucas Sempé
Suvarna Pande
Cem Yavuz
Etienne Lwamba
Ailbhe N. Finnerty Mutlu
Zahra Premji

December 2025

Global Research and Technology Development
Funded by UK Government
RCC Research Commissioning Centre
3ie International Initiative for Impact Evaluation
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 the report

This report, *Evidence Mapping of Social Sciences Methods Used in Fragile and Conflict-Affected Settings (FCAS) (2015–2025)*, examines how researchers design and implement methodological approaches in contexts marked by instability, insecurity, and institutional fragility. Drawing on 5,327 studies identified through a comprehensive mapping process, it analyses patterns in data collection, research design, analytical techniques, sectoral focus, and geographic distribution, revealing both dominant practices and critical gaps in the evidence base. Findings from this work will inform future research commissioning, highlight opportunities for methodological innovation, and support more context-appropriate, rigorous, and policy-relevant evidence generation in FCAS environments.

Review process

This report was reviewed by Professor Ulrike Theuerkauf (UEA), Anil Thota (International Initiative for Impact Evaluation (3ie)), Tess Morris (Foreign, Commonwealth and Development Office (FCDO)), Ella Spencer (International Initiative for Impact Evaluation (3ie)/Research Commissioning Centre (RCC)) and Professor Mansoob Murshed (Institute of Social Studies (ISS)).

Funding

This project is funded by Global Research and Technology Development (GRTD), the research portfolio of the UK government's Foreign, Commonwealth and Development Office (FCDO). It is managed through its Research Commissioning Centre (RCC) led by International Initiative for Impact Evaluation (3ie) and University of Birmingham. 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, Lucas Sempé, at lsempe@3ieimpact.org

Suggested citation: Sempé, L., Pande, S., Yavuz, C., Lwamba, E., Finnerty Mutlu, A. N., and Premji, Z. 2025. *Evidence Mapping of Social Sciences Methods Used in Fragile and Conflict-Affected Settings (FCAS) (2015–2025).*

Evidence Mapping of social sciences methods used in fragile and conflict-affected settings (FCAS) (2015-2025)

Lucas Sempé[1,✉], Suvarna Pande[1], Cem Yavuz[1], Etienne Lwamba[1], Ailbhe N. Finnerty Mutlu[2], and Zahra Premji[3]

2025-12-07

Contents

[1] International Initiative for Impact Evaluation
[2] EPPI Centre, Social Science Research Unit, UCL Institute of Education, University College London
[3] Independent information specialist

Correspondence: Lucas Sempé <lsempe@3ieimpact.org>

List of figures and tables

1. Executive Summary

1.1 Background and Objectives

This evidence mapping exercise analysed research methodologies employed in Fragile and Conflict Affected States (FCAS) from 2015-2025 aiming to understand how methodological approaches are used to understand social sciences in fragile contexts. As 2024 marked the fourth most violent year since the Cold War's end, with 61 state-based conflicts across 36 countries, this study examined how researchers adapt methodological frameworks to contexts characterised by institutional weakness, ongoing violence, and social fragmentation.This study uses the World Bank's FY24 Fragility and Conflict Situations (FCS) classification (Bank (2023)), including all countries identified as in conflict, plus Libya, and excluding Ukraine. In this framework, conflict denotes severe insecurity driven by politically motivated violence from state or non-state actors, either between armed groups or targeting civilians.

The study addressed five core research questions examining empirical approaches, validity and rigour mechanisms, research prioritisation, cross-cutting issues integration, and methodological adaptations specific to FCAS environments. Using a comprehensive evidence mapping methodology, we analysed the methodological landscape to identify patterns, gaps, and opportunities for enhancing research quality and policy relevance.

1.2 Methodology

The evidence mapping employed a multi-stage methodology combining traditional systematic review approaches with innovative AI-assisted screening methods. We conducted comprehensive searches across eight academic databases and four grey literature sources, identifying 265,011 initial records. Following deduplication and temporal filtering (post-2015), 96,424 records advanced to screening using iterative machine learning classifiers achieving 89% recall.

Geographic analysis utilised spatial intersection with Uppsala Conflict Data Program events to identify studies in conflict zones. Full-text retrieval employed dual strategies (Zotero and API-based approaches), successfully obtaining 23,779 documents. AI-assisted screening reduced human workload by one-third while maintaining 90%+ sensitivity. Final analysis included 5,327 studies meeting all inclusion criteria, with structured data extraction using standardised protocols across 23 variables encompassing publication metadata, geographic scope, methodological approaches, sectoral classification, and conflict exposure analysis.

1.3 Key Findings

This evidence mapping identifies significant trends and structural imbalances in how social science research is conducted in fragile and conflict-affected settings (FCAS) between 2015 and 2025. The analysis of 5,327 studies reveals that quantitative designs dominate the methodological landscape, accounting for roughly three-quarters of all analyses, while mixed methods approaches—though less prevalent—show higher levels of contextual validity and analytical depth. Experimental and quasi-experimental designs remain rare, comprising less than two per cent of studies, reflecting both ethical and logistical barriers to their implementation in conflict-affected environments.

Geographically, research is heavily concentrated in countries classified as having nationwide FCAS exposure, particularly across the Sahel and Horn of Africa. However, as discussed in Note 1, this concentration largely reflects the methodological structure of the geographic inclusion criteria rather than intrinsic differences in research productivity or conflict intensity. Despite this artifact, clear geographic and sectoral asymmetries remain: health and social protection dominate the evidence base (representing nearly 60 per cent of all studies), while governance, justice, and infrastructure research remain markedly under-represented.

The methodological analysis demonstrates a predominance of cross-sectional and observational studies designed for short-term, descriptive objectives. Primary data collection remains the norm, often through structured surveys, although studies increasingly integrate administrative, remote-sensing, and digital data sources to mitigate access and security limitations. Median sample sizes remain modest—97 for primary data studies and 33 for secondary data analyses—suggesting a research environment defined by adaptive, resource-constrained fieldwork rather than large-scale statistical generalisation.

Results suggest that variation in sample size across studies reflects a balance between methodological ambition and fieldwork feasibility rather than clear typological distinctions in data source or design. The predominance of modest sample sizes among primary data studies points to a research environment shaped by adaptation and resource constraints. Study scale, therefore, emerges less as a marker of rigour than as an artefact of circumstance. These findings highlight the importance of interpreting study scale as a contingent outcome of context-specific challenges rather than a proxy for analytical strength or evidential quality. To conflate size with quality risks misjudging the credibility of research that, while limited in scope, may offer deeply contextualised and policy-relevant insights into complex and volatile settings.

Causal inference methods are applied in 48 per cent of studies, though most rely on non-experimental identification strategies such as regression models, matching, or panel designs. Advanced analytical techniques—machine learning, Bayesian methods, and spatial analysis—remain marginal (<6 per cent), indicating persistent capacity and infrastructure constraints. Nevertheless, mixed data-source designs show growing methodological diversity, often integrating qualitative inquiry to explore causal mechanisms, contextual dynamics, and social meaning.

Across sectors, methodological pathways display strong disciplinary imprints. Health research remains anchored in quantitative-survey paradigms, while social protection and education studies exhibit greater methodological diversity, combining qualitative and participatory approaches with econometric analysis. Sub-sectoral concentration further narrows the thematic landscape: over two-thirds of health research focuses on public health and infectious disease, and social protection studies cluster around cash transfers and livelihood interventions, leaving governance, and institutional resilience comparatively neglected.

Research activity remains uneven across regions and sectors. Most studies are concentrated in countries where research access is feasible and in sectors that align with major funding streams, such as health and social protection. In contrast, governance-related research—including work on public administration, justice systems, and state capacity—represents less than five per cent of the total evidence base. This scarcity is notable given the central role that governance structures play in shaping how fragility evolves. The absence of robust governance evidence limits understanding of how institutional performance, accountability, and local authority systems contribute to recovery or deterioration in FCAS contexts.

Taken together, these findings depict a rapidly expanding but uneven field of research. FCAS scholarship is increasingly methodologically plural and technologically adaptive, yet still constrained by donor-driven priorities, security limitations, and epistemic inequities. The evidence base remains strongest where research is most feasible rather than where fragility is most acute. These patterns underscore the need for methodological reform, stronger local research capacity, and more balanced investment across regions, sectors, and methodological traditions to ensure that future evidence generation aligns with the realities of conflict-affected development. Clarifying the distinction between fragility and conflict, refining causal attribution terminology, and ensuring transparent coding validation together enhance the interpretive robustness of this evidence map.

2. Introduction

Fragile and Conflict-Affected States (FCAS) continue to bear a disproportionate burden of global violence, with 2024 marking the fourth most violent year since the end of the Cold War. These regions have recorded the highest number of conflicts since 1946, with a total of 61 state-based conflicts across 36 countries. Globally, 2024 was the fourth most violent year since 1989, with approximately 129,000 battle-related deaths (Rustad, 2025) .

According to the PRIO report on Conflict Trends (Rustad, 2025), in terms of regional distribution, Africa has emerged as the epicentre of global conflict, with a dramatic escalation in state-based and non-state violence. Major conflicts include the Ethiopian Tigray war, Sudan civil war, Democratic Republic of Congo war, regional transnational terrorism in West Africa like in Mali and Niger. Asia also saw 17 instances of continuing state-based conflicts dominated by Myanmar post-coup warfare, Yemen's civil war and Afghanistan's Taliban governance conflicts. Rustad (2025) notes that in addition to the level, the intensity and multi-layered patterns of the violence enforce a disproportionate burden of complexity in understanding the nature of the conflict in these countries.

The study of research methodologies in FCAS has evolved considerably over the past two decades, moving from ad hoc approaches to more systematic frameworks for understanding and addressing the unique challenges these contexts present. The conceptual foundation for FCAS research emerged from recognition that traditional social science methodologies, developed primarily for stable institutional environments, require substantial adaptation when applied to contexts characterised by institutional weakness, ongoing violence, and social fragmentation (Mazurana et al., 2013; Wood, 2006).

Early definitional frameworks conceptualised fragility primarily through binary classifications distinguishing "failed" from "functioning" states (Stewart & Brown, 2009). However, contemporary approaches have adopted more nuanced, multidimensional conceptualisations that recognise fragility as existing along continuums rather than discrete categories. The World Bank's updated classification methodology, implemented in 2020, differentiates between countries experiencing high levels of institutional and social fragility and those affected by violent conflict, providing more precise analytical categories for research design (Bank, 2020). Similarly, the FCDO uses data on state stability from the United Nations and the World Bank, along with local knowledge from embassies, networks, and intelligence services, to define fragile states and issue conflict-affected travel advisories. The Fund for Peace's Fragile States Index employs a triangulated approach to measuring fragility, combining quantitative data, qualitative analysis, and expert validation through its proprietary Conflict Assessment System Tool (CAST) framework (Messner et al., 2017).

FCAS countries definitions by the World Bank

Fragility, Conflict, and Violence (FCV) contexts are classified by the World Bank Group under its Fragility and Conflict Situations (FCS) framework, which identifies countries most affected by institutional fragility or active conflict. In this framework, conflict refers to a state of severe insecurity arising from the use of lethal force by organised actors—such as government forces, non-state groups, or other irregular entities—motivated by political objectives. This violence may occur between opposing armed groups or take the form of one-sided attacks deliberately directed at civilians (Bank, 2023). For the purposes of this study, we adopt the World Bank's FY25 FCS classification and include all countries designated as being in conflict situations, as well as Libya, while excluding Ukraine from the analytical sample.

While fragility and conflict frequently overlap, they represent analytically distinct dimensions within the World Bank's Fragility, Conflict and Violence (FCV) framework. Fragility refers to chronic institutional weakness, limited state capacity, and vulnerability to shocks, whereas conflict denotes the active manifestation of organized violence among political or social groups. Some countries in the FCV list experience deep structural fragility with relatively low levels of violence (for example, Lebanon or Mozambique), while others endure sustained warfare despite more functional institutions (such as Sudan or Yemen). Recognising this distinction is essential because methodological constraints arise differently in each case: fragility complicates data reliability and institutional access, whereas conflict primarily limits fieldwork feasibility and ethical safety. Throughout this report, we use the term fragile and conflict-affected settings (FCAS) inclusively, while noting that methodological implications vary along these two intersecting dimensions.

This review aims to map the empirical approaches commonly used in FCAS research, providing insights into how researchers address validity and rigour in these settings. Additionally, we aim to provide sector-wise mapping of the most studied questions, their cross-cutting nature, equity dimensions, and adaptations from the standard methodology used in FCAS environments. This work builds upon a well-established research tradition at 3ie focused on fragile and conflict-affected settings (Djimeu, 2014; Gaarder & Annan, 2013; Lwamba et al., 2022; A. Sonnenfeld et al., 2021; H. B. Sonnenfeld Ada; Chirgwin, 2020; Thissen & Ansari, 2024).

This Evidence Map (EM) will move beyond "what works" to provide a landscape for "what methods work for which questions under what conditions" in fragile contexts, pillars of realist evaluations (Government Social Research, 2021). The resulting framework will provide a methodological toolkit that recognises the legitimacy of diverse approaches when appropriately matched to research questions and contextual constraints.

2.1 Evolution and methodological challenges of research methodological frameworks

Research in Fragile and Conflict-Affected Settings (FCAS) presents a fundamental paradox: contexts where evidence is most urgently needed for effective intervention are precisely those where traditional “gold standard” methodologies are often least feasible. A recent Evidence Gap Map, completed for the FCDO, highlighted these gaps due to the sparse availability of experimental and quasi-experimental research in these contexts {Ravat et al. (2025)}.

The methodological landscape for FCAS research has been shaped by recognition that conventional research approaches often prove inadequate or inappropriate when applied to contexts characterised by violence, displacement, and institutional breakdown. Jacobsen & Landau (2003) identified what they termed the “dual imperative” facing researchers in these settings: the need to produce academically rigorous research while simultaneously generating knowledge that can inform urgent policy decisions under conditions of extreme uncertainty and limited access.

There is growing recognition that no single research approach can adequately address the multifaceted nature of research questions arising in FCAS contexts. Early methodological discussions primarily focused on adapting existing quantitative approaches to address sampling and data quality challenges (Haer & Becher, 2012). However, the field has increasingly adopted mixed methods approaches that combine quantitative data collection with qualitative methodologies, which are better suited to capturing complex social dynamics and local perspectives (Cohen & Arieli, 2011).

The emergence of conflict-sensitive monitoring and evaluation frameworks and ethnographic peace research (Millar, 2018) represents a significant methodological development. The Department for International Development (2010) guidance framework emphasises the importance of conflict-sensitive monitoring and evaluation, noting that “all activities in situations of conflict and fragility should be monitored for inadvertent negative impacts.” This represents a shift from traditional research approaches toward methodologies that explicitly account for the potential of research activities themselves to influence conflict dynamics. BetterEvaluation (2024) identifies the evolution from early “Do No Harm” frameworks in the late 1990s through to contemporary approaches that integrate conflict sensitivity, participatory methods, and adaptive management principles. Millar (2018) emphasises the importance of long-term engagement, participant observation and contextual understanding in such settings. However, such approaches raise unique ethical and practical challenges, leading to solutions like “limited immersion” (Krause, 2021), which seek to maintain ethnographic sensibilities while acknowledging the constraints imposed by security concerns and ethical considerations.

Sampling and data quality issues present fundamental challenges in such scenarios. Traditional sampling methods become complicated when populations are displaced, sampling frames are outdated or nonexistent, and security concerns restrict access (Hoogeveen & Pape, 2020). Data quality concerns arise from multiple sources, including high non-response rates, systematic selection biases, measurement errors due to fear and mistrust, and the challenge of maintaining data collection protocols under rapidly changing conditions (Haer & Becher, 2012).

The distinct challenges in FCAS regions require sometimes standard social science methods. Woodward et al. (2017) conducted a comprehensive qualitative study identifying eight primary categories of challenges: sampling difficulties, data quality concerns, ethical complexities, security constraints, institutional capacity limitations, access restrictions, measurement challenges, and sustainability issues. These challenges are interconnected and compound one another, creating what they describe as a “cascade of methodological compromises” that researchers must navigate. Khan Mohmand et al. (2017) provide a systematic analysis of how these challenges manifest across different types of FCAS contexts. They demonstrate that methodological constraints vary significantly depending on the specific configuration of fragility and conflict within national and subnational settings. In contexts with high conflict but low fragility, researchers may face access restrictions but have relatively good existing data sources. In contrast, highly fragile contexts with low conflict may require more extensive primary data collection efforts but allow for longer-term engagement with communities.

Technological innovations have opened new possibilities for rigorous data collection in FCAS contexts while potentially reducing some traditional methodological constraints. Approaches to address these challenges include the use of third-party monitoring systems, remote sensing technologies (Bank, 2020), Computer-Assisted Telephone Interview technology (Hoogeveen & Pape, 2020), and mobile phone-based data collection methods that can operate under security constraints while maintaining data quality standards (Gibson et al., 2017). However, technological solutions also introduce new methodological challenges and potential biases. Similarly, remote sensing approaches, while valuable for certain types of analysis, cannot capture the subjective experiences and local meanings that are often central to social science research questions.

2.2 Gaps in Current Methodological Knowledge

Despite significant advances in FCAS research methodologies, substantial gaps remain in our understanding of how different approaches perform under various types of constraints and how methodological choices affect research outcomes. Systematic mapping of varying sampling strategies, data collection methods, and analytical approaches is rare, making it difficult for researchers to make informed methodological decisions based on empirical evidence of effectiveness.

Khan Mohmand et al. (2017) identify a critical gap in matching methodological approaches to specific research questions and contextual conditions. Their analysis suggests that research design should be driven by two primary criteria: the specific research question being addressed (the ‘why’ criteria) and the particular configuration of fragility and conflict in the study context (the ‘where’ criteria).

The field also lacks comprehensive frameworks for assessing the quality and reliability of research conducted under the compromised conditions typical of FCAS contexts. Traditional measures of research quality, developed for stable research environments, may be inappropriate for evaluating studies conducted under significant constraints. Indeed, methods designed for controlled or well-resourced settings may not be the best means of generating realistic or actionable insights in contexts marked by volatility and uncertainty. Yet, alternative quality assessment frameworks remain underdeveloped (Woodward et al., 2017). Khan Mohmand et al. (2017) note that different methods contribute to rigour in various ways—some through measurement precision, others through causal inference capabilities, and yet others through narrative coherence and systematic analysis of mechanisms.

Furthermore, there is limited systematic documentation of how methodological adaptations affect the generalisability and policy relevance of research findings. While scholars increasingly acknowledge the need for methodological flexibility in FCAS contexts, the implications of this flexibility for knowledge accumulation and evidence-based policymaking remain poorly understood. Cost-effectiveness analysis of different methodological approaches is particularly underdeveloped, with limited guidance available on the trade-offs between methodological rigour, ethical considerations, and resource constraints (Khan Mohmand et al., 2017).

The rapid evolution of both technological capabilities and conflict dynamics also means that methodological knowledge quickly becomes outdated. New forms of violence, changing patterns of displacement, and evolving information technologies create ongoing demands for methodological innovation that current knowledge production and dissemination systems struggle to meet. The emergence of hybrid methods, such as ethnographic studies that embed qualitative insights within quantitative instruments, demonstrates the potential for methodological innovation but also highlights the need for a systematic evaluation of these approaches across different contexts.

2.3 The need for evidence mapping

The identified gaps in FCAS research are not unexpected, as conducting rigorous research in these settings presents numerous obstacles, including security concerns, ethical considerations, rapidly changing environments, a shortage of local research capacity, and logistical challenges. Despite these constraints, a large body of research in these contexts employs methodological approaches beyond traditional experimental designs yet remains understudied or undervalued from a methodological perspective. However, the persistence of traditional methodological approaches may reflect not only practical constraints but also the limited pool of local research partners deemed to meet externally defined standards of rigour. These standards, rooted in the expectations of international research investors, tend to privilege conventional designs over adaptive or context-sensitive approaches, constraining the development of methods better suited to the realities of FCAS environments.

The complexity and diversity of methodological approaches employed in FCAS research, combined with the rapid evolution of both contexts and methods, create a compelling case for systematic evidence mapping to identify patterns, gaps, and priorities for future methodological development. Traditional narrative reviews, while valuable, sometimes cannot adequately capture the full scope of methodological innovation occurring across different disciplines, sectors, and geographic contexts.

As Mansilla et al. (2024) highlight that there is a “continuing risk of mismatch between decision-maker’s needs and the evidence that is made available to support decision-makers.” This disconnect underscores the importance of systematically understanding the methodological approaches that have been successfully employed in FCAS settings and how they can be more effectively aligned with specific research questions. Petticrew & Roberts (2003) argue for moving beyond rigid “hierarchies of evidence” toward “typologies of evidence,” recognising that “different research methods are more or less good at answering different kinds of research question.” This emphasis on methodological appropriateness is particularly relevant in FCAS contexts, where experimental designs may not always be feasible or appropriate.

White & Phillips (2012) further support this perspective, noting the value of small-n approaches for impact evaluation in situations where statistical tests of significance between treatment and comparison groups are not possible. They emphasise that robust causal inference can still be achieved through careful attention to the underlying mechanisms that connect interventions to outcomes, establishing causation “beyond reasonable doubt by collecting evidence to validate, invalidate, or revise the hypothesised explanations.” Traditional research organisations have shown cautious acceptance of the argument for small-n approaches, recognising that robust causal inference can emerge from detailed analysis of mechanisms linking interventions to outcomes. Methods such as realist evaluation, process tracing, and contribution analysis have gained credibility as legitimate tools for understanding causation in complex settings. However, despite this rhetorical openness, institutional preferences remain anchored in large-n, statistically driven designs—particularly randomised controlled trials—which continue to dominate funding and influence. Evidence from reviews of donor evaluation portfolios and methodological guidance (e.g. Bamberger et al. (2010), Stern et al. (2012), White (2013), Bamberger et al. (2016), Vaessen & Raimondo (2019)) suggests that, although many agencies endorse methodological pluralism in principle, experimental and quasi-experimental approaches remain the default benchmark for rigour. Qualitative and small-n designs are typically positioned as complementary rather than central, reflecting a persistent institutional preference for statistical proof over explanatory understanding.

Evidence mapping approach offers several advantages for understanding the current state of FCAS research methodology. First, it can systematically identify and categorise the full range of methodological approaches being used, revealing patterns that may not be apparent from individual studies or unsystematic reviews. Ideally, researchers should systematically align methodological choices with both the research questions and the contextual conditions. The current literature lacks a comprehensive mapping of how these matches are actually implemented across different studies (Khan Mohmand et al., 2017).

Second, evidence mapping can potentially help researchers identify methodological innovations that address specific types of constraints and research questions, providing practical guidance for those facing similar challenges. Khan Mohmand et al. (2017) identify nine distinct methodological approaches, from controlled comparisons and quasi-experiments to visual methods and digital data collection, each with specific strengths and limitations that make them more or less suitable for different research purposes and contexts. Systematic mapping could reveal how these methods are actually combined in practice and what effects they have in real-world applications.

Third, evidence mapping can identify gaps where methodological development has been limited, highlighting priorities for future research and development efforts. Methods should be combined differently depending on whether research aims to assess contextual conditions, provide descriptive details on social and political action, identify causal pathways, or support operational learning; however, systematic documentation of how such combinations work in practice remains limited (Khan Mohmand et al., 2017).

Finally, evidence mapping can contribute to the development of more systematic frameworks for methodological decision-making in FCAS contexts by documenting what approaches have been tried, under what conditions, and with what outcomes. This systematic documentation is essential for moving beyond ad hoc methodological adaptation toward more principled approaches to research design in challenging contexts. While Khan Mohmand et al. (2017) provide a theoretical foundation for such systematic decision-making, empirical validation through comprehensive evidence mapping of existing practice is needed to refine and operationalise these frameworks.

Building on these insights, this project aims to develop an Evidence Map focused on research methodologies used in FCAS contexts. According to the literature, qualitative approaches, scoping reviews, and mixed methods are commonly employed in these settings; each adapted to address the inherent complexities and ethical considerations of fragile contexts (Bertone et al., 2019; Campbell, 2017; Duggan & Bush, 2014; Woodward et al., 2017). This research methods EM will provide a comprehensive understanding of the methodological landscape in FCAS research by systematically identifying, classifying, and analysing the frequency and application of these various research methods.

2.4 Research questions

This Evidence Map (EM) addresses five core research questions that collectively examine the methodological landscape of research conducted in fragile and conflict-affected settings.

These questions collectively address the “what methods work for which questions under what conditions” framework essential for evidence-informed policy making in challenging development contexts.

RQ1: Empirical approaches in FCAS research
What empirical approaches have commonly been used for research in fragile and conflict-affected settings (FCAS)? This question maps the distribution of methodological approaches across the research corpus, identifying dominant paradigms and emerging methodological innovations adapted to challenging operational contexts.

RQ2: Validity and methodological rigour
How do researchers address validity and methodological rigour when conducting studies in FCAS contexts? This examines the specific adaptations, compromises, and innovations researchers employ to maintain research quality under constraints including limited access, security concerns, and institutional instability.

RQ3: Research question prioritisation
What research questions are most commonly addressed by studies conducted in FCAS? This analysis identifies thematic priorities in FCAS research, revealing both concentration areas where substantial evidence exists and neglected domains requiring future attention.

RQ4: Cross-cutting issues integration
To what extent do research methodologies in FCAS contexts address cross-cutting issues such as gender equity and inclusion? This question assesses whether research designs systematically incorporate equity considerations or whether such dimensions remain peripheral to core methodological frameworks.

RQ5: Methodological adaptations
What adaptations to standard methodological approaches have researchers explicitly developed to overcome the unique challenges of FCAS environments? This examines documented innovations in sampling, data collection, analysis, and ethical protocols specifically designed for fragile contexts, providing practical guidance for future research design.

3. Data and Methodology

3.1 Methodology used in the mapping and synthesis process

The World Bank's 2025 Fragile and Conflict-affected Situations list provided the foundational framework for country selection (World Bank Group, 2025), complemented by FCDO travel advisories to distinguish between countries experiencing nationwide FCAS exposure versus countries with regionalised FCAS exposure to conflict and instability (GOV.UK, 2024). The final country selection included Afghanistan, Burkina Faso, Cameroon, Central African Republic, Democratic Republic of Congo, Ethiopia, Haiti, Iraq, Lebanon, Mali, Mozambique, Myanmar, Niger, Nigeria, Somalia, South Sudan, Sudan, Syrian Arab Republic, West Bank and Gaza (territory), and Yemen. Ukraine was excluded from the original World Bank list due to its unique geopolitical context, while Libya was added based on its persistent fragility indicators and extensive conflict exposure, ensuring the dataset captured the full spectrum of contemporary fragile and conflict-affected contexts relevant to development research.

> ⓘ Terminology Standardisation
>
> Throughout this report, we standardise the terminology describing levels of conflict exposure. "Countries with nationwide FCAS exposure" refers to contexts where the entirety of the national territory meets fragility or conflict criteria, while "countries with regionalised FCAS exposure" designates states where only specific subnational areas are affected. Earlier designations such as "fully conflict-affected" or "partially conflict-affected" have been replaced with these standardised terms to improve precision and alignment with the geographic screening framework.

This evidence mapping employed a multi-stage methodology to identify and analyse social science research in Fragile and Conflict-Affected States (FCAS), combining traditional evidence mapping approaches with innovative AI-assisted screening methods. The search strategy, conducted in May-July 2025, identified 265,011 records across eight academic databases and four grey literature sources. Following deduplication using both R scripts and EPPI-Reviewer (removing 133,447 duplicates) and temporal filtering to post-2015 studies, 96,424 records advanced to title and abstract screening. The screening process utilised an iterative machine learning approach, developing five successive classifiers through EPPI-Reviewer with progressive performance improvement from 40% to 89% recall, ultimately identifying 34,763 potentially relevant records through human-AI collaboration.

The methodology integrated geolocation analysis for location extraction and geocoding (Cambon et al., 2021), successfully identifying 24,278 studies located in countries with regionalised FCAS exposure through spatial intersection with the Uppsala Conflict Data Program events (Davies et al., 2025).

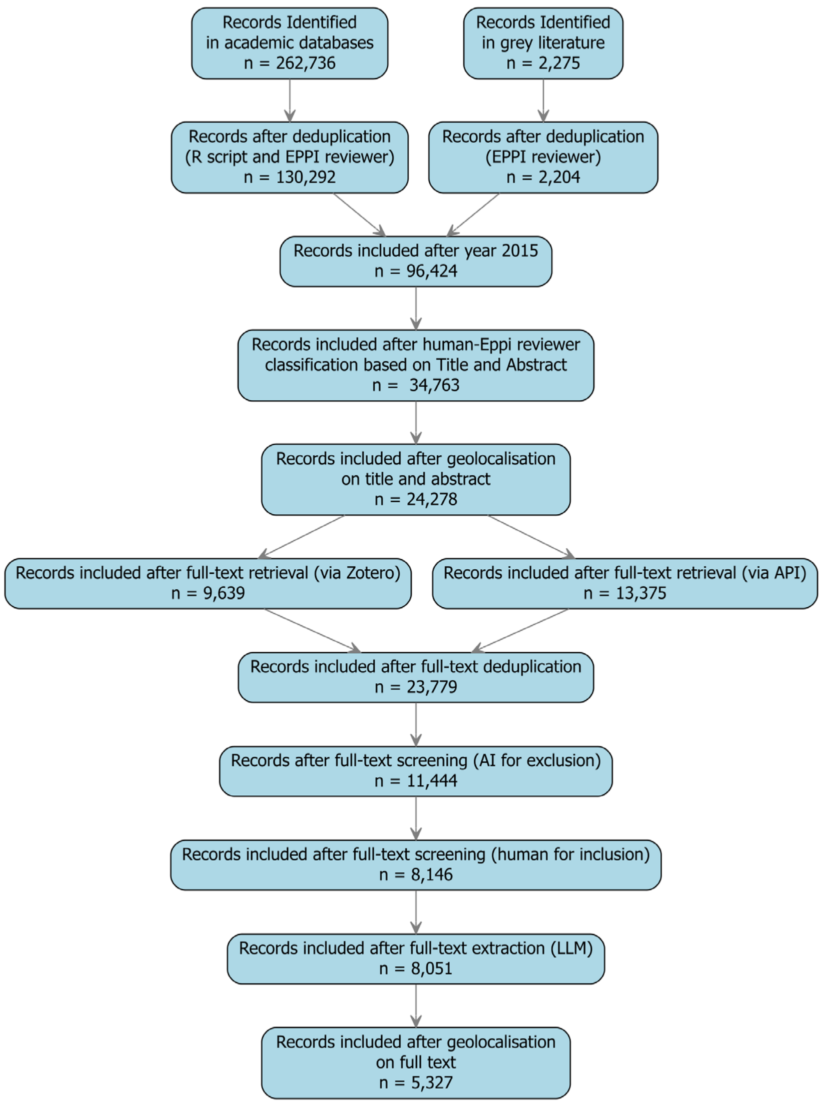
Full-text retrieval employed a dual strategy combining Zotero and API-based approaches, successfully obtaining 23,779 unique documents from multiple sources including Crossref, Semantic Scholar, and publisher APIs (Elsevier and Willey).

We implemented a two-stage screening approach to optimise the balance between efficiency and accuracy in the process In Stage 1, AI performed initial screening to identify and exclude obvious cases that clearly failed inclusion criteria (e.g., non-empirical studies, wrong geographic focus, non-health topics), achieving a 32% reduction in human workload while maintaining high specificity to avoid incorrectly excluding relevant studies. In Stage 2, human reviewers examined only the documents classified as “INCLUDE” by the AI, focusing their expertise on borderline cases and nuanced decisions where contextual judgment was required. This approach leveraged AI’s strength in rapidly processing large volumes and applying consistent rules while compensating for AI’s limitations in handling complex, ambiguous cases, ultimately achieving 90%+ sensitivity compared to 31% with AI-only screening, while reducing overall human screening burden by one-third.

A structured extraction process using ChatGPT-4.1 Mini successfully extracted standardised data from 8,146 studies to support comprehensive evidence mapping and analysis.

This reduced the corpus to records after full-text screening. A second geolocation verification based on full-text resulted in 5,327 final studies meeting all inclusion criteria.

Figure 1: Prisma diagram — flow of record selection and screening steps.



The screening and selection methodology, along with the final number of retained studies, is summarised in Figure 1.

3.2 Analytical Framework

The analytical framework integrates quantitative pattern identification with qualitative assessment of methodological innovation, enabling a comprehensive mapping of “what methods work for which questions under what conditions”—a perspective essential for evidence-informed policy making in fragile contexts.

We also employ network analysis to visualise and examine research collaboration patterns. By mapping co-authorship and institutional linkages, we reveal the structure and dynamics of scholarly networks, identifying key hubs and interdisciplinary connections within the research ecosystem. Together, these approaches offer a multidimensional understanding of the dataset.

3.2.1 Data Source Classification

The first dimension categorises studies according to their underlying data sources. This classification recognises that the type of data available fundamentally shapes both research opportunities and limitations. We distinguish four categories:

Primary Only studies are those that generate original data to address the research question. This includes field-based research, surveys, experiments, and ethnographic studies. Classification as “Primary Only” was based on explicit statements about original data collection and the absence of references to pre-existing datasets.

Secondary Only studies draw entirely on existing datasets, administrative records, or other previously collected information. Examples include analyses of demographic surveys, administrative databases, satellite imagery, and archival records. These were identified through clear references to secondary data sources and no evidence of new data collection.

Mixed Sources studies combine both primary and secondary data within a single research design. This category reflects the growing use of methodological triangulation—combining new and existing data to strengthen validity and expand analytical reach. Inclusion required evidence of both original data collection and use of existing datasets.

Not Specified studies lack sufficient methodological detail to determine data origin, highlighting concerns about transparency and reproducibility in reporting.

3.2.1.1 Primary Data Collection Approaches

Within studies using primary data, we identified a range of approaches that vary in technique, resource needs, and analytical potential. This framework recognises methodological diversity and emphasises the importance of selecting approaches that align with research questions and context.

Experimental approaches involve deliberate interventions or randomised designs such as field experiments and controlled trials. These methods enable strong causal inference through structured manipulation and treatment assignment.

Biomedical Measures include studies incorporating physiological or health-related data, such as biomarkers, anthropometry, or other laboratory-based measures requiring specialised equipment and protocols.

Spatial/GPS Methods refer to studies using geographic information systems, satellite imagery, or location-tracking to collect or analyse spatially explicit data.

Network Analysis includes approaches that map social, organisational, or institutional relationships, capturing the relational dimensions of data.

Participatory Methods encompass community-based and collaborative data collection strategies such as participatory action research. These approaches prioritise engagement and co-production of knowledge with participants.

Ethnographic methods rely on immersive fieldwork and sustained observation, capturing social dynamics and lived experience through long-term engagement.

Multi-technique approaches integrate multiple data collection methods—such as combining surveys, interviews, or document review—to generate complementary forms of evidence.

Structured Collection includes standardised instruments such as surveys and questionnaires, enabling systematic and replicable data gathering.

Semi-structured approaches include flexible interview guides or open-ended surveys that balance consistency with adaptability.

Basic Collection refers to data collection that is simple, unstructured, or insufficiently described, serving as a marker of limited methodological transparency.

3.2.1.2 Analytical Techniques

Our classification of analytical approaches captures the diversity of quantitative and qualitative techniques used across studies. Rather than implying a hierarchy, it highlights the varied analytical strategies researchers employ to address different kinds of questions and data structures.

Machine Learning covers algorithmic and computational approaches such as neural networks and predictive modelling, which enable pattern detection and classification beyond traditional statistical methods.

Causal Inference includes techniques explicitly designed to identify causal relationships, such as instrumental variables, regression discontinuity, difference-in-differences, and propensity score matching.

Structural Modeling refers to approaches like structural equation modelling or factor analysis that explore relationships among latent and observed variables.

Multilevel Modeling includes methods such as hierarchical linear models and mixed effects models that account for clustered or nested data structures.

Survival Analysis covers approaches for analysing time-to-event data, including hazard models and duration analysis.

Longitudinal Methods include panel data and time series analysis, addressing change and stability across repeated observations.

Bayesian Methods apply probabilistic modelling and inference grounded in prior distributions, offering a flexible framework for uncertainty quantification.

Network Analysis techniques model and interpret relational data structures using graph theory and related tools.

Spatial Analysis includes methods for analysing geographically referenced data, such as spatial statistics and GIS-based techniques.

Meta-Analysis synthesises quantitative findings across studies through systematic aggregation of effect sizes and heterogeneity assessment.

Advanced Regression encompasses specialised regression frameworks such as logistic, multinomial, and ordinal models.

Basic Regression includes standard linear regression and related statistical modelling.

Statistical Tests refer to conventional hypothesis tests such as t-tests, chi-square, and ANOVA.

Qualitative Analysis includes thematic and content analysis, grounded theory, and other systematic approaches to interpreting textual or visual data.

Descriptive Only refers to studies that summarise data through descriptive statistics without inferential or modelling components.

4. Results and Analysis

The dataset comprises **5327 academic papers and grey literature** systematically extracted using structured content analysis protocols. Each publication was coded according to **23 standardised variables** encompassing: (1) publication metadata including authors, year, and institutional affiliations, (2) geographic scope covering **263** unique study countries and specific regional locations, (3) methodological approaches including data collection methods, analysis types, and sample characteristics, (4) sectoral classification using World Bank taxonomy across **12** distinct sectors, (5) thematic alignment with UN Sustainable Development Goals, and (6) conflict exposure analysis through geospatial intersection with Uppsala Conflict Data Program events to classify studies by fragility context and conflict intensity levels.

4.1 Geographic Coverage and Conflict Classification

Considering the methodological dynamics outlined in Note 1, the geographic distribution of studies in our dataset reflects a pronounced concentration in countries classified as having nationwide FCAS exposure. A total of 4,769 studies (88.3%) were conducted in these settings, with Burkina Faso (1,098 studies), Afghanistan (697 studies), and Mali (496 studies) contributing most substantially. Together, these three countries account for 2,291 studies, or 42.4% of all research within FCAS contexts.

Countries with regionalised FCAS exposure contribute 634 studies (11.7%), led by Iraq (128 studies), Nigeria (121 studies), and Lebanon (102 studies). This distribution reflects the methodological filtering effect described earlier rather than inherent disparities in research capacity. On average, countries with nationwide FCAS exposure yield approximately 397 included studies each, compared with an average of 70 for those with regionalised exposure. The relative concentration of studies in nationwide FCAS contexts thus stems from the geographic screening criteria rather than differences in conflict severity or research productivity. Table 1 summarises the country-level distribution of included studies.

Table 1: Geographic concentration — studies cluster in nationwide FCAS settings, reflecting inclusion criteria and accessibility.

Country	Number of Included Studies	Total number of studies	Approx. Population	Public universities	Inclusion rate
Burkina Faso	1,098	10,472	23,550,000	3	10.49
Afghanistan	697	10,729	42,650,000	39	6.50
Mali	496	9,007	25,200,000	3	5.51
Sudan	470	14,037	51,700,000	36	3.35
Haiti	394	3,912	11,900,000	1	10.07
Somalia	373	4,857	19,700,000	2	7.68
Niger	352	15,134	27,900,000	4	2.33
Syria	323	13,010	25,600,000	10	2.48
South Sudan	294	3,213	12,200,000	8	9.15
Libya	119	5,212	7,500,000	12	2.28
Palestinian Territories	81	5,871	5,600,000	10	1.38
Central African Republic	72	1,728	5,300,000	1	4.17
Iraq	128	10,349	47,000,000	35	1.24
Nigeria	121	14,069	237,500,000	125	0.86
Lebanon	102	3,578	5,800,000	1	2.85
Ethiopia	81	11,557	135,500,000	45	0.70
Democratic Republic of the Congo	71	4,477	112,800,000	40	1.59
Cameroon	54	2,594	29,100,000	8	2.08
Chad	36	1,691	21,000,000	2	2.13
Mozambique	30	1,439	35,600,000	10	2.08
Myanmar	11	2,206	54,900,000	55	0.50

Note 1: Understanding Inclusion Rate Variation: A Methodological Artifact

The significant variation in inclusion rates across FCAS countries—ranging from 0.50% to 10.49%—is primarily a consequence of the differential geographic screening criteria applied, rather than reflecting differences in conflict intensity, research infrastructure quality, or research feasibility.

The Geographic Screening Effect

For countries with nationwide FCAS exposure, all studies conducted anywhere within the country were eligible for inclusion after passing relevance screening. This resulted in substantially higher inclusion rates (Burkina Faso: 10.49%, Afghanistan: 6.50%, Mali: 5.51%) because no additional geographic filter was applied. Critically, analysis of conflict exposure among included studies reveals that the vast majority of research in these countries occurs in relatively stable areas with zero recorded conflict deaths or events in proximity to study locations. Despite being classified as FCAS territories, most research sites in these countries are not located in active conflict zones—researchers naturally gravitate toward areas where fieldwork is safer and more feasible.

In contrast, countries with regionalised FCAS exposure required studies to be specifically located within designated conflict-affected regions to meet inclusion criteria. This additional geographic requirement created a dual filter: first, researchers in these countries tend to conduct studies in stable regions (as evidenced by the weak relationship between national university counts and FCAS-region studies), and second, the geographic screening process excluded all research conducted outside FCAS boundaries. The result is dramatically lower inclusion rates (Nigeria: 0.86%, Ethiopia: 0.70%, Iraq: 1.24%) despite these countries having substantial overall research productivity.

Research Infrastructure and the Geographic Mismatch

The relationship between research infrastructure and study output reveals the structural basis for these inclusion patterns. Among countries with more than three public universities, there is a moderate positive relationship between university count and studies initially found in databases fo rcountries with regionalised FCAS exposure ($R^2 = 0.493$), confirming that national research infrastructure does drive research productivity. However, this relationship effectively disappears after geographic screening is applied, demonstrating that research output occurs predominantly outside conflict-affected regions.

In countries with nationwide FCAS exposure, university counts show essentially no relationship with either studies found ($R^2 = 0.011$) or studies included ($R^2 = 0.064$). This weak relationship suggests that in contexts of pervasive fragility and conflict, factors beyond simple infrastructure counts—such as international research partnerships, humanitarian organisation presence, or donor funding mechanisms—may be more important determinants of research output. The visualisation is striking: countries with regionalised FCAS exposure with over 120 public universities yielded fewer than 130 included studies each, while countries with nationwide FCAS exposure with just 25-30 universities produced 400-600+ included studies. This pattern confirms that national university infrastructure in countries with regionalised FCAS exposure is geographically concentrated in stable regions, away from conflict-affected populations (see Figure 2 and Figure 3).

The Conflict Intensity Paradox

Paradoxically, analysis of conflict exposure reveals negative correlations between conflict intensity and inclusion rates (r = -0.43 for conflict zones intersected, r = -0.33 for deaths, r = -0.37 for events). This counterintuitive finding reflects the methodological structure: in countries with nationwide FCAS exposure, the absence of geographic filtering means included studies are predominantly from stable areas with minimal conflict exposure. In countries with regionalised FCAS exposure, the geographic filter ensures that only studies from designated conflict regions qualify—and these regions tend to be areas of higher conflict intensity where research is more challenging. The few studies that do emerge from countries with regionalised FCAS exposure (such as Lebanon with 304 average deaths, Cameroon with 443 deaths, and Iraq with 5,295 deaths per study location) show substantially higher conflict exposure than any studies from countries with nationwide FCAS exposure, yet represent only a small fraction of each country's total research output.

Implications

The inclusion rate variation therefore reflects a methodological reality rather than substantive differences in research capacity or conflict conditions. The geographic screening process successfully identified research conducted in conflict-affected settings, but the differential application of geographic criteria—necessary versus not necessary based on country-level FCAS classification—creates the observed pattern. The analysis of research infrastructure confirms that national capacity exists in countries with regionalised FCAS exposure but remains geographically misaligned with conflict-affected populations. The low inclusion rates from countries with regionalised FCAS exposure represent a genuine evidence gap: despite having research capacity and productivity at the national level, these countries conduct minimal research in their conflict-affected regions compared to the volume of research in their stable areas. This structural mismatch—where universities exist but not within FCAS regions—represents a critical barrier to generating evidence from fragile and conflict-affected settings.

Figure 2: Studies and university density — distribution of records by public university count.

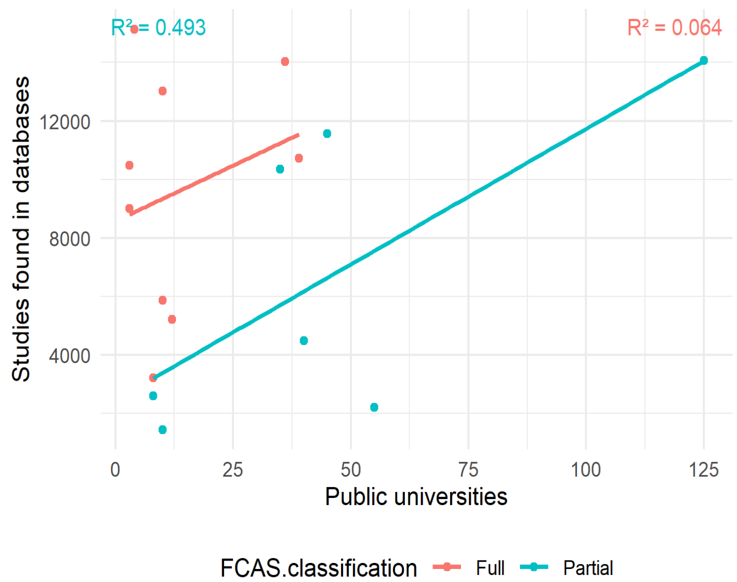
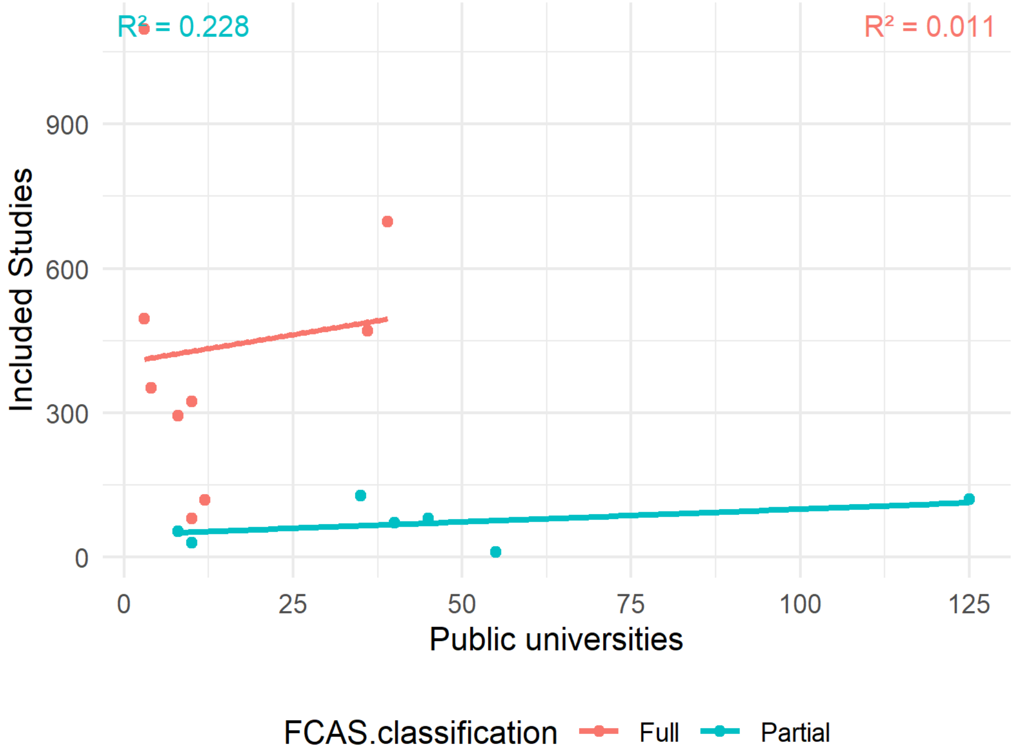


Figure 3: University density vs inclusion rate — relationship between research infrastructure and study inclusion in FCAS.



The spatial distribution of research locations, presented in Figure 4 and Figure 5, illustrates the uneven geographic pattern of research activity across fragile and conflict-affected settings. The global overview map (Figure 5) distinguishes between **countries with nationwide FCAS exposure** (shaded in red) and **countries with regionalised FCAS exposure** (highlighted in orange), underscoring the structural basis of the inclusion criteria described earlier. It shows that countries with nationwide FCAS exposure are concentrated across the Sahel belt—from Mauritania and Mali through Niger, Chad, Sudan, and Somalia—with additional representation in Afghanistan and Yemen. In contrast, countries with regionalised FCAS exposures, such as Nigeria, Cameroon, Ethiopia, Mozambique, Myanmar, and Lebanon, are geographically dispersed and characterised by conflict zones confined to specific subnational regions.

The detailed country maps (Figure 4) highlight the spatial mismatch between research activity and conflict exposure. Blue dots indicate the geographic coordinates of study sites, while orange shading marks areas classified as conflict-affected. Across nearly all countries with regionalised FCAS exposure, the vast majority of studies are located outside the orange-shaded conflict regions, clustering instead around national capitals, university towns, and relatively stable administrative zones. For example, in Nigeria and Ethiopia—each with over a thousand studies—most research occurs in southern and central regions, far from high-conflict areas in the northeast and west. Similar patterns are visible in Cameroon, Iraq, and Lebanon, where research clusters coincide with areas of greater institutional presence and accessibility rather than conflict intensity.

Taken together, the maps provide visual confirmation of the **geographic screening effect** described in Note 1: inclusion patterns are driven not by differences in national research capacity or conflict severity, but by the interaction between geographic eligibility criteria and the spatial concentration of research infrastructure. The maps make visible the central methodological insight of this analysis—namely, that the geography of evidence in fragile contexts reflects where research can feasibly occur, rather than where the need for evidence may be greatest.

Figure 4: Countries with nationwide (dark orange) and regionalised (light orange, black borders) FCAS exposure

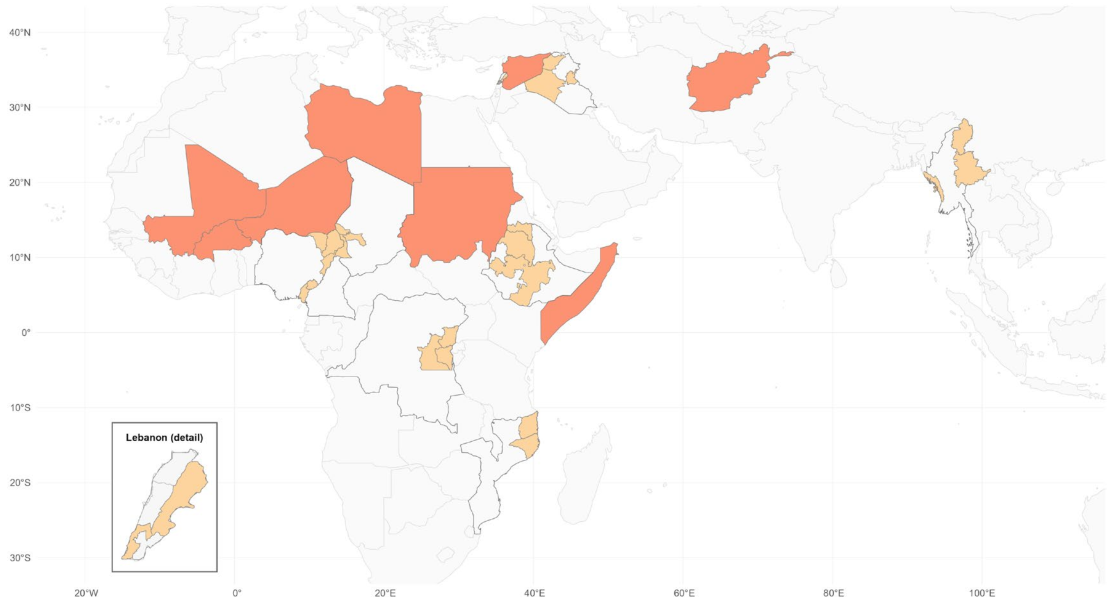
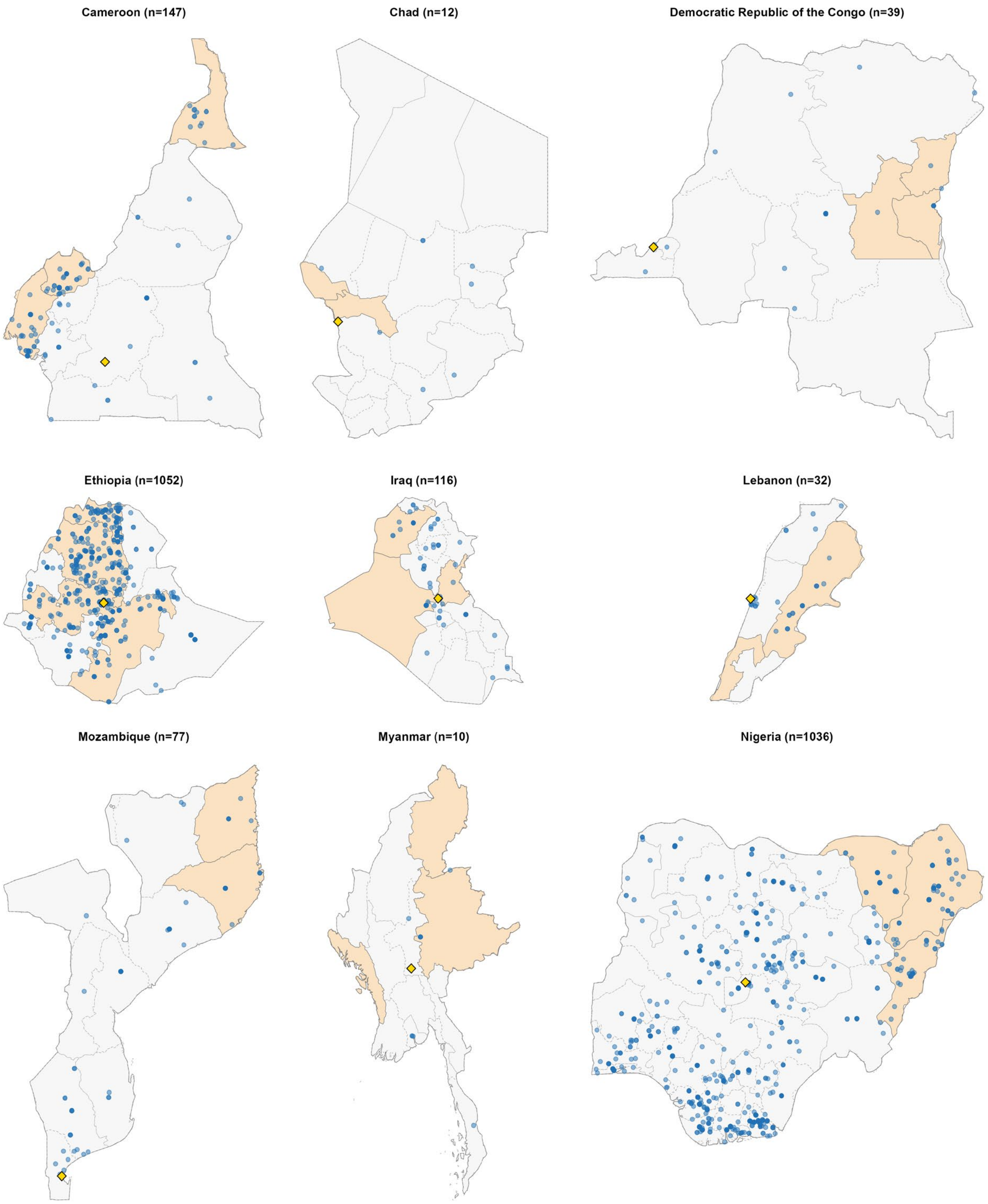


Figure 5: Geolocalisation of research conducted in countries with regionalised FCAS exposure



The studies analysed show that 81.7% of studies examine single country: nationwide conflict (representing the majority of cases). Cases involving conflict with stable regions account for 12.1% (a substantial proportion) (see Table 2).

Table 2: Conflict exposure distribution — most included studies are single-country analyses in nationwide FCAS settings.

Conflict Classification	Number of Studies	Percentage
Single Country: Nationwide Conflict	4351	81.7
Conflict with Stable Regions	646	12.1
Partial Territorial Conflict	131	2.5
Multi-Country: Mixed Conflict Statuses	126	2.4
Multi-Country: Nationwide Conflict	73	1.4

4.2 Sectoral Distribution

Table 3 shows a classification of studies utiling the World Bank Sectors taxonomy. The sectoral distribution shows an overwhelming focus on health (2078 studies, 39%), followed by social protection (1146 studies) and agriculture, fishing and forestry (987 studies). These top three sectors account for 79.1% of all studies, indicating strong research concentration. The remaining sectors each represent less than 6.4% of the total, including a small number of unclassified cases.

Table 3: Sectoral distribution — health, social protection, and agriculture dominate; governance and infrastructure are under-represented.

World Bank Sector	Number of Studies
Health	2078
Social Protection	1146
Agriculture, Fishing and Forestry	987
Education	340
Public Administration	223
Transportation	117
Water, Sanitation and Waste Management	116
Industry, Trade and Services	112
Financial Sector	109
Energy and Extractives	67
Information and Communication Technologies	29
Unclassified	3

4.3 Sustainable Development Goals (SDGs)

Similarly, we present in Table 4, a classification of studies according to SDG pillars. The analysis of SDG pillar distribution reveals a strong emphasis on people (3802 studies, 71.4%), which encompasses poverty, hunger, health, education, gender equality, and water/sanitation issues. Partnership follows with 710 studies (13.3%), focusing on institutions and global partnerships. Prosperity accounts for 568 studies (10.7%), covering energy, economic growth, innovation, and sustainable cities. The remaining pillars show limited representation: peace (129 studies, 2.4%) and planet (118 studies, 2.2%). The top two pillars account for 84.7% of all studies, highlighting the research community's focus on social development and governance frameworks over environmental sustainability themes.

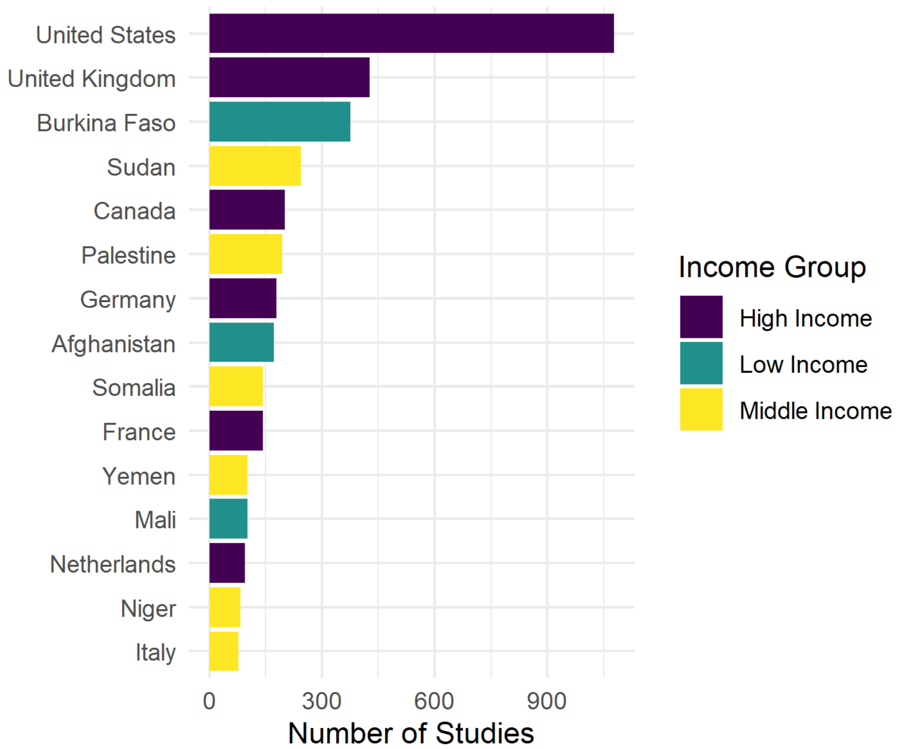
Table 4: SDG alignment — emphasis on 'People' pillars (health, education, equity); limited environment and peace coverage.

SDG Pillar	Number of Studies	SDG Range	Percentage
People	3802	SDGs 1-6	71.4
Partnership	710	SDGs 16-17	13.3
Prosperity	568	SDGs 7-11	10.7
Peace	129	SDGs 15-15	2.4
Planet	118	SDGs 12-14	2.2

4.4 Author Characteristics

The geographic distribution of research reveals significant concentration, with the top 15 countries producing 3609 studies (67.7% of the total). "Producing countries" here refer to the institutional affiliations of first authors, used as a proxy for the primary location of research production. United States leads with 1078 studies, followed by United Kingdom (427) and Burkina Faso (376) (see Figure 6). When examined by income classification, Middle Income countries account for 2331 studies (43.8%) across 125 nations. Notably, 97 countries contribute 20 or fewer studies each, highlighting the research concentration among a few highly productive nations.

Figure 6: Top author countries — top 15 first-author affiliations.



4.5 Temporal Dynamics and Research Intensity

The temporal analysis suggests a broadly stable development-research landscape across the past decade (see Figure 7). Output rose steadily through 2020—with 2,907 papers published in 2015–2020—before easing in the most recent period: 2,379 papers in 2021–

2025 (note that 2025 is incomplete and 2024 is the last full year). Rather than indicating a sudden collapse, the post-2020 softening should be read cautiously: it may reflect shifting publication lags, the tail-effects of the pandemic on research pipelines, or changes in indexing practices, rather than an abrupt decline in scholarly interest. The fitted linear regression trend (shaded 95% CI) therefore excludes incomplete years to avoid misinterpreting the current calendar-year snapshot as a substantive downward turn.

Notably, sample sizes have shown consistency across these periods, with mean sample sizes of 224 and 228 participants respectively, and median values remaining virtually unchanged at 70 and 68 participants. This stability suggests that established research methodologies and access patterns have proven resilient to external disruptions. See Table 5 for detailed breakdowns by period.

Figure 7: Publication trends — yearly evolution of research output.

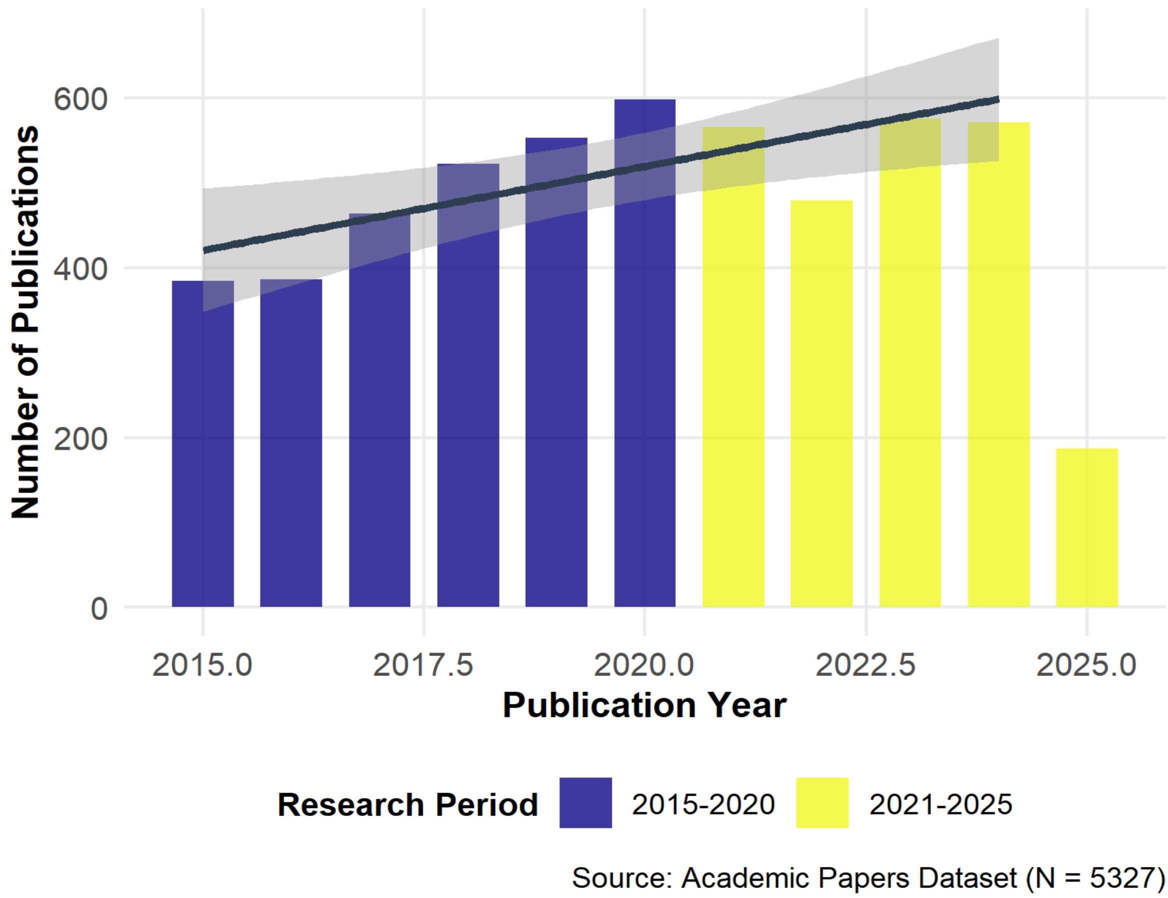


Table 5: Temporal comparison of research output and sample-size statistics - years 2015–2020 vs 2021–2025

Period	Papers	Percentage (%)	Mean Sample	Median Sample
2015-2020	2,907	54.6	223.8	70
2021-2025	2,379	44.7	228.2	68

Note: Sample sizes exclude missing values and extreme outliers

4.6 Methodological Overview

4.6.1 Research Design Landscape

The research design landscape demonstrates marked variation in methodological characteristics and prevalence across different approaches (see Figure 8). Cross-sectional and observational designs together comprise the majority of studies, reflecting their practicality and adaptability for descriptive and correlational research questions in fragile and conflict-affected contexts. Their predominance could suggest that researchers frequently adopt these designs because they are well-suited to limited data availability, short fieldwork windows, and logistical constraints. At the same time, the wide range of analytical strategies within these designs indicates that their application is far from uniform—some studies employ simple descriptive frameworks, while others integrate more complex modeling or mixed-methods interpretation.

Experimental designs, including both randomized and quasi-experimental studies, represent a smaller proportion of the total evidence base but offer distinct analytical possibilities for causal interpretation. Their relatively limited number (70 studies combined, less than 2% of the total) could suggest that resource demands, ethical considerations, and implementation barriers restrict the feasibility of experimental work in many FCAS contexts. Where conducted, such studies tend to concentrate in thematically bounded, intervention-oriented areas supported by external funding or long-term institutional partnerships.

Mixed methods designs are a notable feature of the landscape, comprising 671 studies—roughly 12% of the total—with over 60% integrating both qualitative and quantitative elements. This prevalence could indicate a methodological response to complex research environments where triangulation strengthens validity and enhances interpretive scope. Mixed approaches appear across design types, from observational to longitudinal and experimental, suggesting that integration of data sources and analytic paradigms is becoming a common strategy rather than an exceptional one.

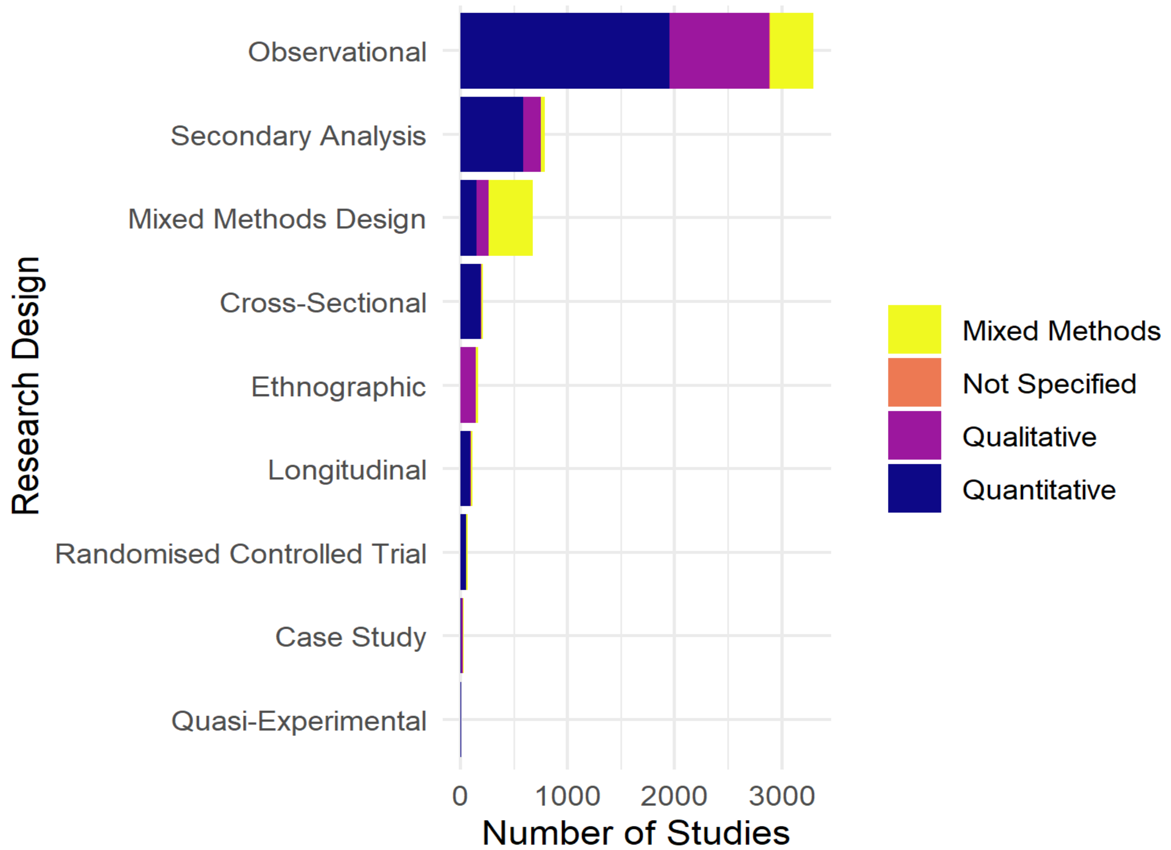
Ethnographic and case study designs show a distinctive profile, characterised by a stronger presence of qualitative and mixed methods paradigms (over 85% of ethnographic and 80% of case study research). This pattern could reflect the suitability of these approaches for generating contextualised insights and engaging directly with participants' lived experiences, particularly where quantitative measurement is constrained. The diversity within these categories—from brief field interactions to extended engagements—illustrates the flexibility of qualitative inquiry to adapt to field conditions while maintaining interpretive depth.

Longitudinal designs (113 studies) and secondary analyses (785 studies) further expand the methodological range. Longitudinal studies, the majority of which are quantitative, could indicate efforts to capture temporal dynamics or intervention effects where sustained data collection is feasible. Secondary analyses, dominated by quantitative approaches (74%), appear to leverage large-scale datasets such as the DHS or World Bank surveys for population-level insights. The relatively small share of mixed or qualitative secondary analyses may reflect both data accessibility and the structural limitations of available datasets rather than differences in analytical ambition.

Across the entire dataset, quantitative approaches dominate (approximately 70–75% of all analyses), with mixed methods and qualitative paradigms contributing complementary perspectives within specific designs. This distribution could suggest that quantitative approaches remain the most accessible and transferable across contexts, while mixed and qualitative methods are employed more selectively to explore mechanisms, meaning, and context.

Taken together, these findings depict a heterogeneous and adaptive methodological environment. Rather than reflecting a linear hierarchy of designs, the observed variation likely arises from pragmatic methodological choices shaped by feasibility, data availability, and research purpose. The diversity of approaches underscores that methodological form is best interpreted in relation to context—balancing analytical precision, field constraints, and the knowledge needs of fragile and conflict-affected settings.

Figure 8: Design types and paradigms — distribution of research designs and analytic approaches.



4.6.2 Sample Sizes

Sample sizes across included studies exhibit substantial heterogeneity, reflecting variation in research design, methodological orientation, and contextual constraints inherent to fragile and conflict-affected settings. Among studies employing primary data collection, sample sizes range from small qualitative investigations to large-scale quantitative surveys with up to 9,999 participants. The median sample size of 97 participants could suggest that most primary data studies are designed for focused, context-specific analysis rather than for broad statistical generalisation. This distribution may also indicate that research scale is frequently determined by operational feasibility and resource availability, with smaller studies representing adaptive responses to access, security, and logistical constraints.

Studies utilising secondary data—including administrative records, national demographic surveys, and existing databases—show a lower median analytical sample of 33, though maxima approach 10,000. This pattern could reflect both the selective use of subnational datasets and the inclusion of large national analyses, suggesting that secondary data studies encompass a broad spectrum of analytical scopes and data structures.

The large-scale survey category demonstrates similar variability (median = 65; interquartile range = 5.5–307; maximum = 5,117), which could indicate methodological differences in sampling frames, population coverage, and survey design. While some of these studies draw on nationally representative samples, others are limited to geographically or demographically defined populations, consistent with programmatic or intervention-focused objectives.

Taken together, these results could suggest that variation in sample size reflects a balance between methodological ambition and fieldwork feasibility rather than clear typological distinctions across data source types. The predominance of modest sample sizes among primary data studies implies a research environment characterised by adaptive, resource-constrained field designs. These findings underscore the need to interpret study scale as a contingent methodological outcome shaped by context-specific constraints rather than as an indicator of analytical strength or evidential quality.

See Table 6 for detailed descriptive statistics.

Table 6: Sample size summary — primary studies have larger median samples (~97) than secondary analyses (~33).

Sample Size Descriptive Statistics

Data source type	Number of studies	Median sample	First quartile	Third quartile	Maximum sample
Large-Scale Survey	131	65	5.5	307.00	5117
Primary Data	3352	97	20.0	319.25	9999
Secondary Data	1216	33	10.0	150.75	9884

4.6.3 Data Collection Analysis

4.6.3.1 Primary vs Secondary Data Utilisation

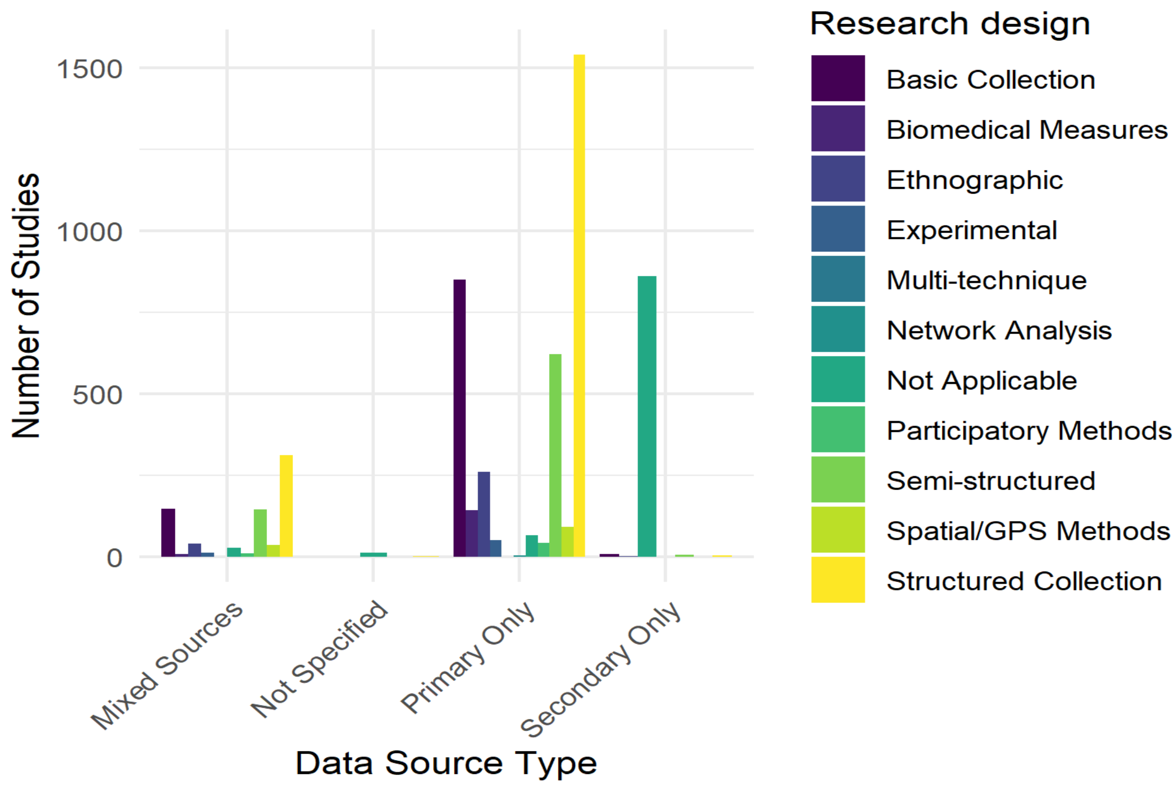
The analysis of data collection approaches reveals marked variation in methodological diversity across different data source types (Figure 9). Primary data collection dominates the research landscape, accounting for most studies, with structured collection methods being the most common (1,541 studies). This emphasis on primary data reflects the particular information needs of conflict-affected settings, where secondary data can be unreliable or unavailable.

Studies that draw on mixed data sources show a wide range of approaches, combining structured (312 studies) and semi-structured methods (146 studies) with more specialised techniques such as spatial/GPS mapping (37 studies) and ethnographic work (41 studies).

Secondary-only studies tend to rely on existing datasets or administrative records, with 861 classified as “Not Applicable” for primary collection type, highlighting their observational nature.

The notable presence of biomedical measures within primary data studies (144 cases) signals an increasing integration of health-related assessments into broader development research. Overall, this methodological variety reflects how researchers adapt their designs to challenging environments and to the interdisciplinary demands of studying complex development issues.

Figure 9: Data source types — collection modalities and research designs



4.6.3.2 Secondary Data use

The analysis of secondary data use shows clear variation in how different data sources are applied across studies (Table 7). Many secondary sources are used in relatively simple ways, with most studies employing descriptive or basic analytical approaches. “Not Applicable” sources make up the largest share (520 studies), and only a small proportion involve more in-depth analysis (1.7%), suggesting potential to extend how existing datasets are used rather than evidence of misuse.

Widely recognised datasets such as the Demographic and Health Surveys (DHS), World Bank surveys, and national surveys appear to be used primarily for foundational analysis rather than more complex modelling. Across 164 DHS studies, 159 World Bank studies, and 69 national survey studies, none applied advanced analytical designs. This may reflect barriers such as limited access to disaggregated data, resource constraints, or gaps in analytical capacity, rather than shortcomings of the data themselves.

Specialised datasets—including administrative records, conflict databases, remote sensing, and environmental monitoring—show somewhat broader but still underdeveloped use. Administrative data and conflict datasets are more frequently applied in structured analyses (2% and 1.5% respectively), but technical sources such as remote sensing and surveillance data remain underused, pointing to a disconnect between data availability and practical capacity to work with them.

The large and varied category of “Other Secondary” sources (467 studies, 0.6% with extended analysis) underscores both creativity and constraint in how researchers access and apply data. Overall, these patterns highlight a need for stronger support in secondary data analysis—through training, data access, and methodological collaboration—rather than a lack of quality in the data themselves.

Table 7: Secondary data sources — common datasets (DHS, World Bank, administrative, remote sensing, conflict data) and usage patterns.

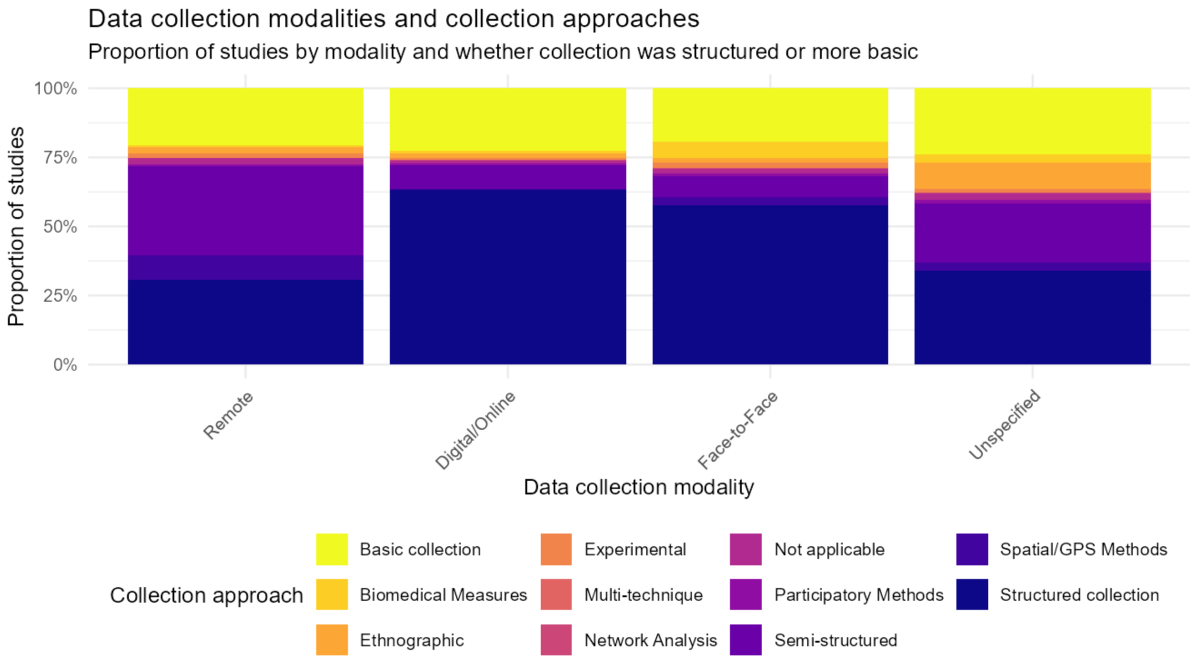
secondary_data_type	Total	High Quality %	Basic	Medium	Very High	High
Not Applicable	520	1.7	478	33	8	1
Other Secondary	467	0.6	461	3	2	1
DHS	164	0.0	164	0	0	0
World Bank Surveys	159	0.0	158	1	0	0
National Surveys	69	0.0	69	0	0	0
Conflict Data	65	1.5	62	2	1	0
Administrative Data	50	2.0	47	2	1	0
Remote Sensing	50	0.0	50	0	0	0
Surveillance Systems	50	0.0	50	0	0	0
Environmental Data	35	0.0	32	3	0	0

4.6.4 Data Collection Modality Trends

The analysis of data collection modalities reveals how technological and logistical choices shape the character and robustness of research conducted in constrained environments (see Figure 10).Digital and online modalities are often paired with more structured collection approaches and with study designs that report mixed methods, reflecting the workflows those methods support. Face-to-face approaches range from open-ended, exploratory fieldwork to standardised household surveys, underscoring their continuing importance for capturing local detail where digital reach is limited. Remote and telephone methods produce mixed outcomes, including both quick, light-touch surveys and careful multi-wave designs adapted to access limitations. Studies that combine modalities or report mixed methods tend to show clearer documentation of procedures and triangulation across sources, suggesting that methodological integration improves confidence in findings.

Importantly, these patterns show that modality and collection approach are pragmatic responses to field conditions: digital tools can increase efficiency but risk excluding people without access; in-person work improves reach and contextual grounding but can be costly or risky. Rather than treating sample size or technical features as stand-alone markers of value, we interpret collection choices as contingent adaptations to ethical, logistical and security trade-offs; assessments of study contributions should be rooted in those contextual realities.

Figure 10: Data collection modalities and design robustness.



4.6.5 Statistical Analysis approaches

Patterns of statistical and analytical practice vary considerably across research designs (Figure 11). Cross-sectional studies most often use descriptive statistics (61.4%), reflecting their role in mapping conditions and establishing baselines in conflict-affected contexts. Yet these studies also draw on a range of extended analyses, including spatial techniques (13.5%) and multilevel models (3.9%), showing that many researchers are combining descriptive and relational approaches to explore complex dynamics.

Experimental and quasi-experimental studies follow expected patterns, with randomised controlled trials (84.1%) and quasi-experimental designs (83.7%) primarily using causal inference methods. This concentration reflects the analytical requirements of testing interventions and estimating programme effects in real-world settings.

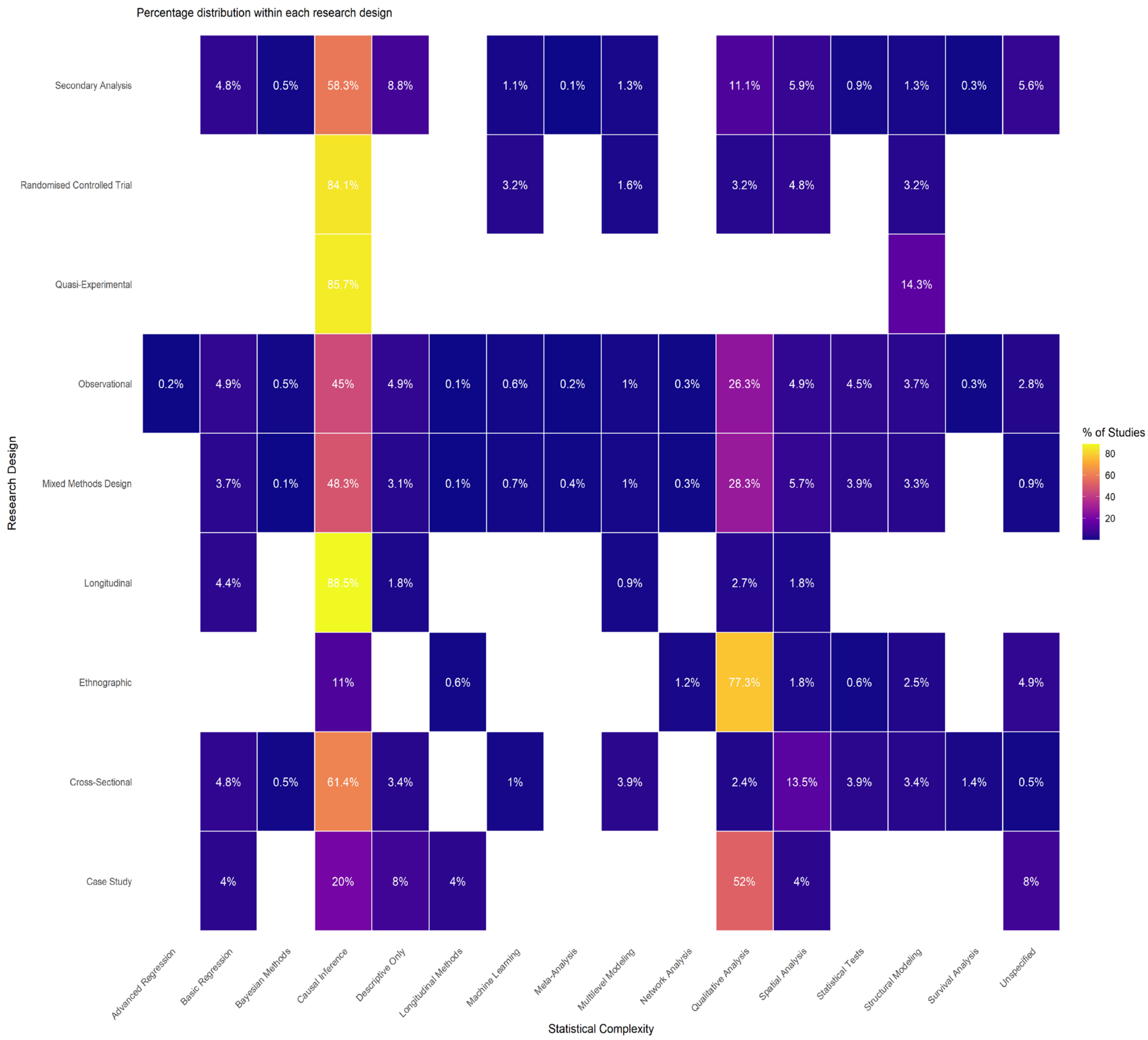
Secondary analyses show more varied practice. Most rely on descriptive approaches (58.3%), while a smaller share use qualitative analysis (11.1%) or spatial and multilevel methods (5.9% and 1.3%, respectively). This pattern points to the different ways existing datasets are being used, often shaped by data structure, access, and analytic capacity rather than by design choice alone.

Ethnographic studies remain largely qualitative (77.3%), consistent with their interpretive aims, though a minority (11%) also employ causal or mixed analytical strategies, reflecting creative integration across methods. Mixed methods designs display the broadest spread—descriptive (48.3%), qualitative (28.3%), and spatial (5.7%)—illustrating how researchers blend analytical forms to connect quantitative and contextual insight.

The range of statistical and analytical approaches reflects the diversity of research design choices. The analysis highlights how different analytical strategies are matched to the types of data available and the questions being addressed, showing that methodological variation arises from practical and contextual adaptation rather than levels of complexity.

Although advanced analytical methods—machine learning, Bayesian modelling, and spatial analysis—account for less than six per cent of the corpus, their prevalence shows a modest upward trajectory after 2020. This growth likely reflects the diffusion of open-source computational tools, wider data availability, and the entry of interdisciplinary teams combining social science and data-science expertise. However, the persistence of a low baseline highlights the continued need for capacity-building and methodological training to ensure that innovation in analytical techniques is accompanied by corresponding advances in transparency and interpretability.

Figure 11: Statistical methods by design — heatmap of approaches across research designs.



4.6.6 Causal Inference Methods Analysis

Approximately half of the studies in the corpus attempt some form of causal modelling, but the strength of identification strategies varies considerably. Most rely on non-experimental techniques—standard regression models, fixed-effects panel estimators, or matching approaches—to approximate causal relationships under substantial data and design constraints. Only a small fraction employ explicit counterfactual designs such as randomized controlled trials or quasi-experimental frameworks (difference-in-differences, regression discontinuity, or instrumental variables). In this sense, the share of “causal inference” studies reflects methodological aspiration rather than confirmed causal validity. These findings underscore both the ambition of researchers to engage causal questions and the structural limitations that inhibit the application of robust counterfactual designs in FCAS environments (see Figure 12).

Randomised Controlled Trials (RCTs) continue to anchor causal identification, drawing on randomisation as a direct means of estimating intervention effects. Quasi-experimental designs also feature prominently, using approaches such as matching, regression discontinuity, and natural experiments to address causal questions where randomisation is not feasible.

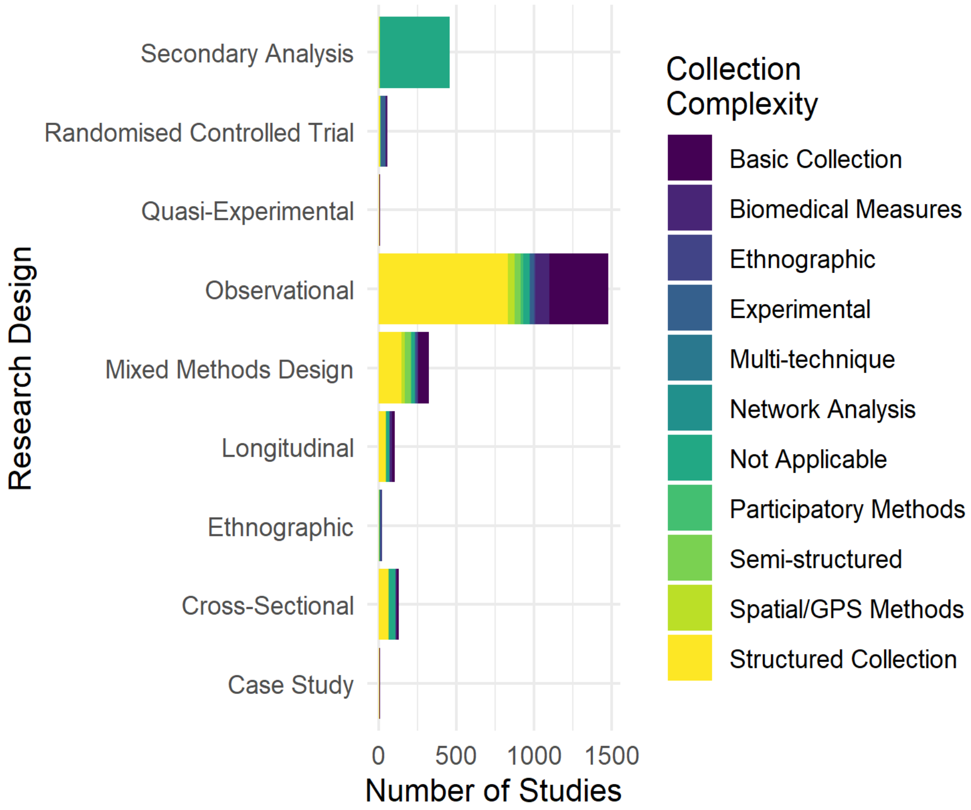
The notable presence of causal inference within observational studies reflects the growing use of creative identification strategies and modelling approaches that extend beyond purely descriptive analysis. These applications show how researchers adapt causal tools to real-world data conditions, often building on policy or programmatic variation rather than experimental design.

Cross-sectional and longitudinal studies apply causal reasoning in different ways. Cross-sectional studies frequently use instrumental variables and other single-period techniques, while longitudinal research often employs difference-in-differences and related panel data methods that use change over time for causal identification. These patterns highlight how temporal and contextual structure shape analytical opportunities.

In mixed methods studies, causal inference is often combined with qualitative inquiry to explore not just whether an effect occurs but how and why. This integration supports richer explanations of causal processes and strengthens the interpretive value of findings.

Variation in data collection approaches across causal inference studies—ranging from structured field data to biomedical and spatial data—demonstrates how researchers align analytical aims with data realities. Rather than reflecting a single model of causal analysis, the evidence points to a plural and adaptive set of practices shaped by disciplinary traditions, data availability, and the practical challenges of working in development and conflict-affected settings.

Figure 12: Causal methods — usage patterns and research contexts.



4.6.7 Methodological pathways

The flow of methods across stages of research—from data source selection, through research design, to analytical technique—shows recognisable patterns in how studies are constructed and how different methodological elements connect (Figure 13). These pathways illustrate both consistent design logics and areas where researchers adapt methods to suit data realities and research goals.

Primary data sources most often link to cross-sectional and observational designs that rely on descriptive or regression-based analyses. This configuration reflects the practicality and accessibility of direct data collection, especially in settings where time, security, or resource constraints shape what is feasible. The prevalence of these pathways underscores how primary research tends to prioritise responsiveness and contextual understanding.

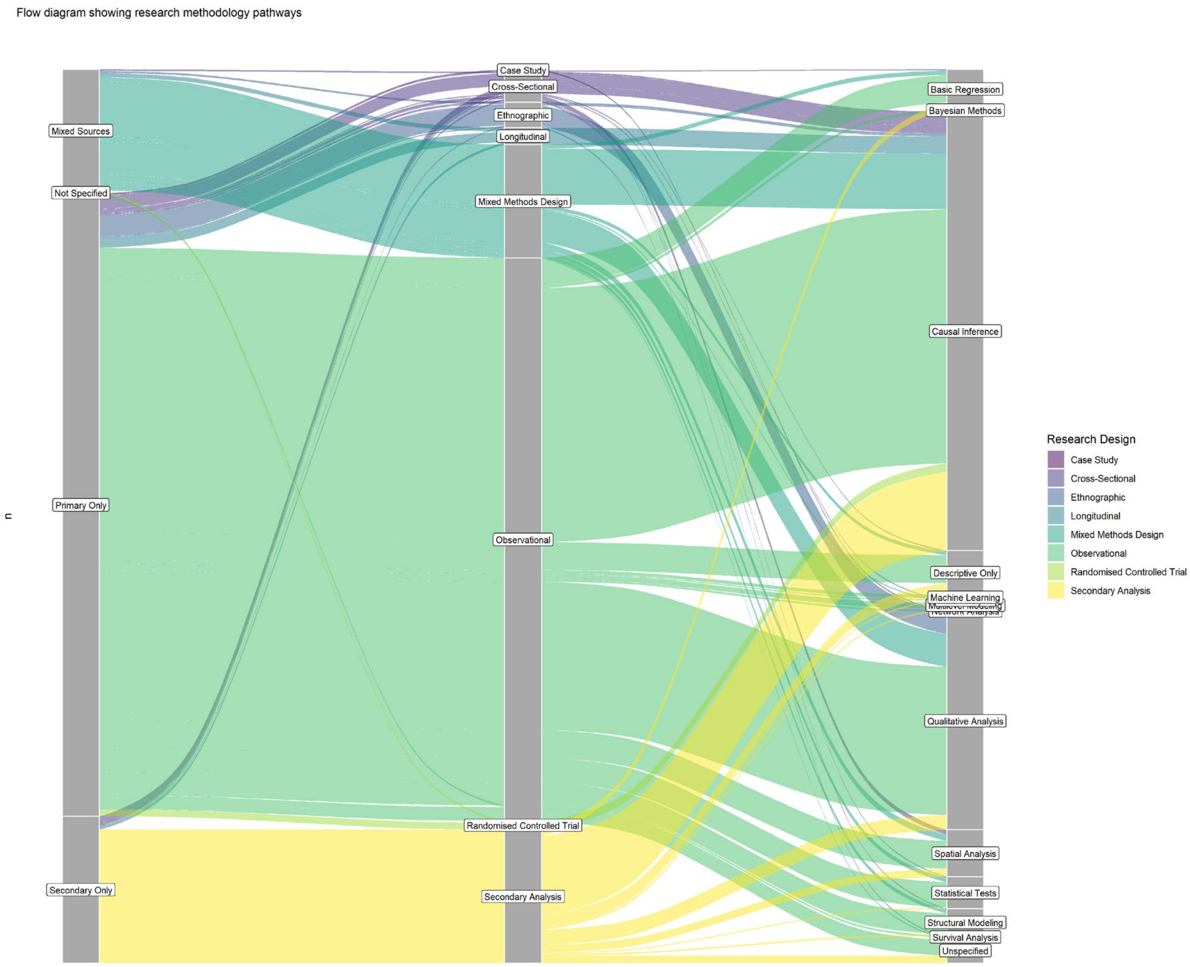
Secondary data sources show greater variation in analytical application, frequently connecting to secondary analysis designs and a wider range of regression and specialised techniques. This diversity highlights the potential of large-scale datasets for deeper statistical exploration, while also pointing to the need for technical support and analytical resources to make full use of existing data infrastructures.

Mixed data sources demonstrate the most varied methodological trajectories, often linking to mixed methods designs that combine quantitative and qualitative approaches. These studies commonly employ causal inference, multilevel modelling, or structural analysis, showing how integration across data types can enable more layered explanations and strengthen interpretation.

The flow analysis also indicates where certain pathways are less developed. For example, many primary data studies could incorporate a wider range of analytical approaches, while some secondary analyses might benefit from qualitative or participatory components to contextualise results. These patterns suggest areas for shared learning and methodological exchange across research traditions.

Experimental studies predictably converge toward causal inference methods, but the mapping also points to growing interest in approaches such as machine learning for treatment heterogeneity and structural modelling for understanding underlying mechanisms. These developments reflect an expanding toolkit rather than a fixed hierarchy of methods, showing how researchers adapt analytical choices to evolving questions and data possibilities.

Figure 13: Research flow — pathway from data sources to analytic diversity



4.6.8 Methodological pathways by Sector

The analysis of methodological pathways by research sector demonstrates clear disciplinary clustering alongside varying degrees of methodological diversification across domains (see Figure 14). Health sector studies exhibit a strong adherence to quantitative paradigms, typically employing survey-based data collection combined with statistical or econometric analysis. This pathway remains the dominant configuration within the evidence base, reflecting both the institutional maturity of global health research and the widespread availability of standardised instruments suitable for comparative and large-scale analysis. While this methodological coherence contributes to internal validity and replicability, it also indicates a limited adaptation to the complex causal structures and contextual variability characteristic of fragile and conflict-affected settings.

By contrast, research in the social protection and education sectors demonstrates comparatively higher methodological heterogeneity. Social protection studies frequently integrate quantitative and qualitative components, combining administrative or survey data with participatory and ethnographic approaches to capture behavioural, institutional, and community-level dynamics. Education research, although smaller in volume, shows the most balanced methodological profile, often adopting mixed-methods and case study designs that account for local context, institutional capacity, and cultural specificity. The distribution illustrated in Figure 14 suggests that while disciplinary conventions continue to shape methodological choices, there is a gradual but discernible convergence towards pluralistic and adaptive research frameworks across sectors. This evolution signals a broader shift from rigid disciplinary orthodoxy toward context-sensitive methodological pluralism within FCAS research.

Taken together, these sectoral patterns indicate that the geography of methodological practice remains uneven, with certain sectors—and their dominant research pathways—concentrated in specific regions and institutional networks. The following section (Section 4.6.9) examines how these disciplinary imbalances translate into geographic-sectoral coverage gaps that shape the overall distribution and policy relevance of FCAS evidence.

Figure 14: Methodological pathways — mapping sectors to data collection approaches.

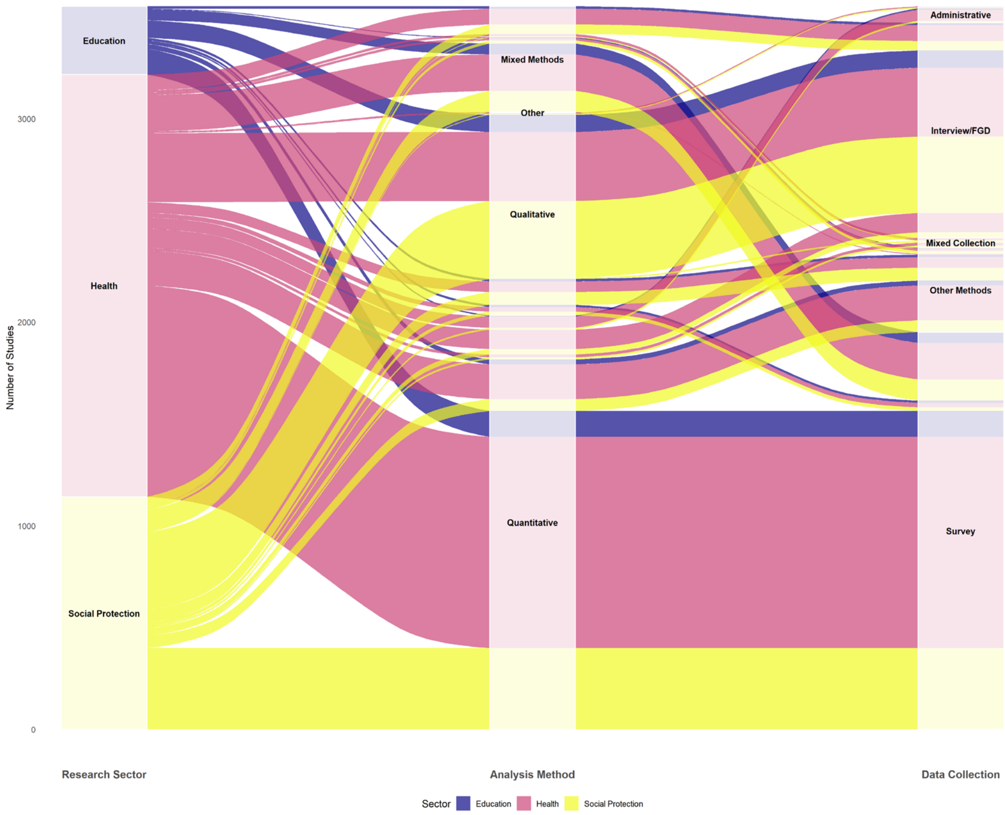


Table 8: Cross-sector methods comparison — health favors structured surveys; social protection and agriculture use more mixed-methods; governance leans qualitative.

Sector	Distinct Pathways	Total Studies	Concentration (%)
Health	14	2074	50.0
Social Protection	14	1143	35.0
Education	9	335	37.9

Note: Concentration shows percentage using most common pathway

4.6.9 Geographic-Sectoral Coverage Gaps

The analysis of geographic–sectoral distribution reveals significant asymmetries in research coverage across fragile and conflict-affected settings (see Figure 15). At first glance, certain countries—such as Burkina Faso, Mali, and Afghanistan—appear to dominate the evidence base across several major sectors, particularly health and social protection. However, as noted in Note 1, this pattern largely reflects a methodological artifact arising from differential geographic screening criteria applied during inclusion rather than intrinsic variations in research intensity, capacity, or conflict exposure. In countries classified as having nationwide FCAS exposure, all studies meeting thematic relevance were included, whereas for countries with regionalised FCAS exposure, inclusion was restricted to studies specifically located within conflict-affected subnational regions. This structural difference accounts for the apparent clustering of research in a limited set of contexts.

When adjusted for this methodological effect, Figure 15 highlights a more substantive issue: the uneven intersection between sectoral focus and conflict geography. Health and social protection studies predominate across most regions, while sectors such as governance, justice, and infrastructure remain comparatively under-represented. Approximately one-third of potential country–sector combinations show no identifiable research presence, indicating systemic gaps in evidence generation. These absences are not evenly distributed but tend to coincide with areas of highest conflict intensity and weakest institutional access—precisely the contexts where robust evidence is most critical yet most difficult to obtain.

Overall, the observed geographic–sectoral imbalances underscore the importance of interpreting research concentration patterns within their methodological and operational context. Strengthening evidence coverage will depend not only on expanding research to under-studied sectors but also on designing inclusion frameworks that mitigate the structural biases identified in Note 1. The following section (Section 4.6.10) explores how these disparities manifest at the sub-sectoral level, identifying patterns of thematic concentration that further constrain cross-context comparability and policy transferability.

Figure 15: Spatial breakdown by sector — geographic clustering of people-focused sectors vs sparse governance/infrastructure coverage.

Table 9: Geographic coverage by sector — spatial distribution of study sites across sectors.

Sector	Countries	Total Studies	Mean per Country	Maximum
Health	52	2751	52.9	507
Social Protection	52	1645	31.6	193
Agriculture, Fishing and Forestry	51	1249	24.5	342
Education	41	404	9.9	85
Public Administration	38	295	7.8	58
Water, Sanitation and Waste Management	34	162	4.8	34
Industry, Trade and Services	37	154	4.2	21
Transportation	26	130	5.0	21
Financial Sector	24	125	5.2	28
Energy and Extractives	28	85	3.0	19
Information and Communication Technologies	12	33	2.8	11

4.6.10 Sub-sectoral Research Concentration

The sub-sectoral analysis demonstrates a high degree of concentration within a narrow range of research topics across major sectors, as summarised in Table 10 and visualised in Figure 16. Within the health sector, five sub-areas—public health systems, maternal and child health, infectious disease control, nutrition, and reproductive health—account for nearly 74 per cent of all health-related studies. In contrast, themes such as mental health, health-system governance, and non-communicable diseases collectively represent less than 10 per cent, revealing the persistent dominance of short-term, intervention-oriented research over system-level or structural inquiry.

Social protection research exhibits a comparable pattern, with livelihood recovery, cash transfers, and community-based resilience initiatives constituting over two-thirds (68 per cent) of all outputs. Far fewer studies address the institutional and policy dimensions of social protection—particularly fiscal sustainability, governance mechanisms, and social accountability frameworks—despite their centrality to long-term stability in FCAS contexts. In education, concentration is somewhat lower but still marked: access and enrolment studies account for 42 per cent of publications, while teacher training, curriculum reform, and digital learning together comprise less than 20 per cent.

As shown in Figure 16, this narrowing of thematic focus produces a steep gradient across sub-sectors, with research intensity declining rapidly beyond the most operationally tractable domains. Such patterns mirror donor funding priorities and data availability rather than proportional policy importance. The resulting evidence landscape privileges measurable, short-cycle interventions at the expense of system-level or governance-oriented questions, limiting both the generalisability and transformative potential of the FCAS research corpus.

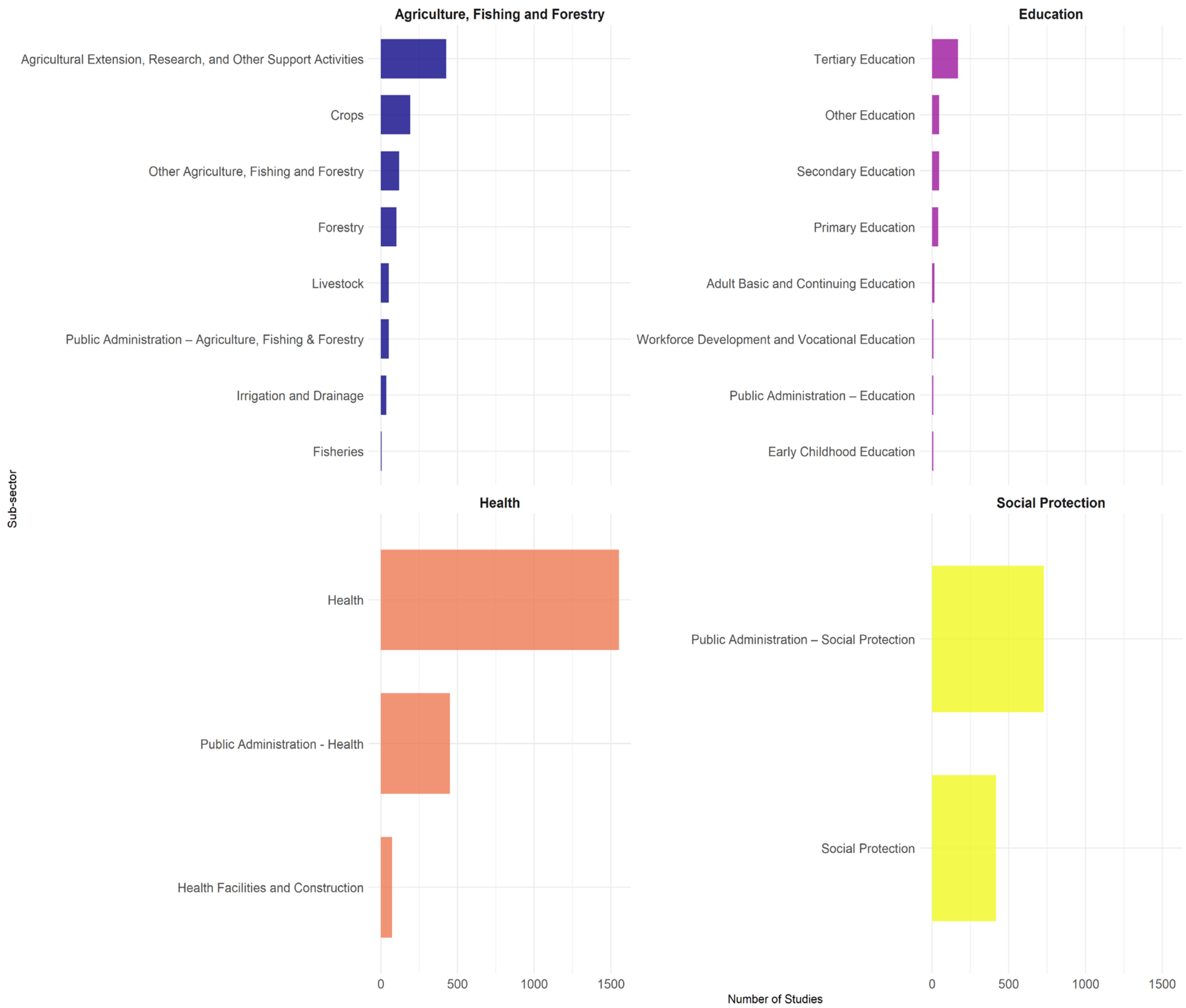
These findings underscore the need for deliberate diversification of sub-sectoral research portfolios and for funding strategies that incentivise studies addressing institutional reform, long-term resilience, and cross-sectoral linkages.

Table 10: Within-sector concentration — a few sub-topics account for a disproportionate share of studies.

Sector	Sub-sector	Studies	Mean Sample	Median Sample
Health	Health	1,554	271	66
Social Protection	Public Administration – Social Protection	729	183	48
Health	Public Administration - Health	450	191	50
Agriculture, Fishing and Forestry	Agricultural Extension, Research, and Other Support Activities	426	264	118
Social Protection	Social Protection	417	193	46
Agriculture, Fishing and Forestry	Crops	193	230	120
Education	Tertiary Education	169	172	88
Agriculture,	Other Agriculture,	120	206	99

Fishing and Forestry	Fishing and Forestry			
Agriculture, Fishing and Forestry	Forestry	102	183	140
Health	Health Facilities and Construction	74	240	78
Agriculture, Fishing and Forestry	Livestock	53	223	118
Agriculture, Fishing and Forestry	Public Administration – Agriculture, Fishing & Forestry	52	170	71
Education	Other Education	46	127	30
Education	Secondary Education	45	167	80
Education	Primary Education	40	170	60

Figure 16: Sub-sector distribution — within-sector topical breakdown across major research areas.



4.7 Determinants of research design in FCAS contexts

The regression analysis reveals fundamental structural determinants of research design in FCAS contexts (see Table 11). The coefficient of 1.011 indicates that quantitative studies have sample sizes approximately 175% larger than qualitative studies (exp(1.011) = 2.75 times larger).

Contrary to expectations, high-income country authorship shows minimal effect on sample sizes (-0.9%), suggesting that **resource advantages may not translate directly into larger studies within FCAS operational constraints.** This finding challenges assumptions about North-South research capacity differentials and may reflect collaborative arrangements that level resource playing fields or selection effects where only feasible studies proceed.

The model explains 14.4% of sample size variance, leaving 85.6% unexplained—**indicating that context-specific factors, local partnerships, and field-level adaptations substantially shape research possibilities beyond observable institutional characteristics.** This suggests significant scope for innovative methodological approaches tailored to specific FCAS contexts.

Table 11: Regression results — predictors of log(sample size) with robust SEs (design, sector, conflict exposure).

Predictor	Coefficient	Std. Error	p-value
Intercept	-0.115	26.962	0.997
Quantitative Method	1.011	0.042	< 0.001
High-Income Author	-0.009	0.043	0.837
Health Sector	-0.001	0.041	0.986
Recent Study (≥2020)	-0.008	0.078	0.916
Publication Year	0.002	0.013	0.872

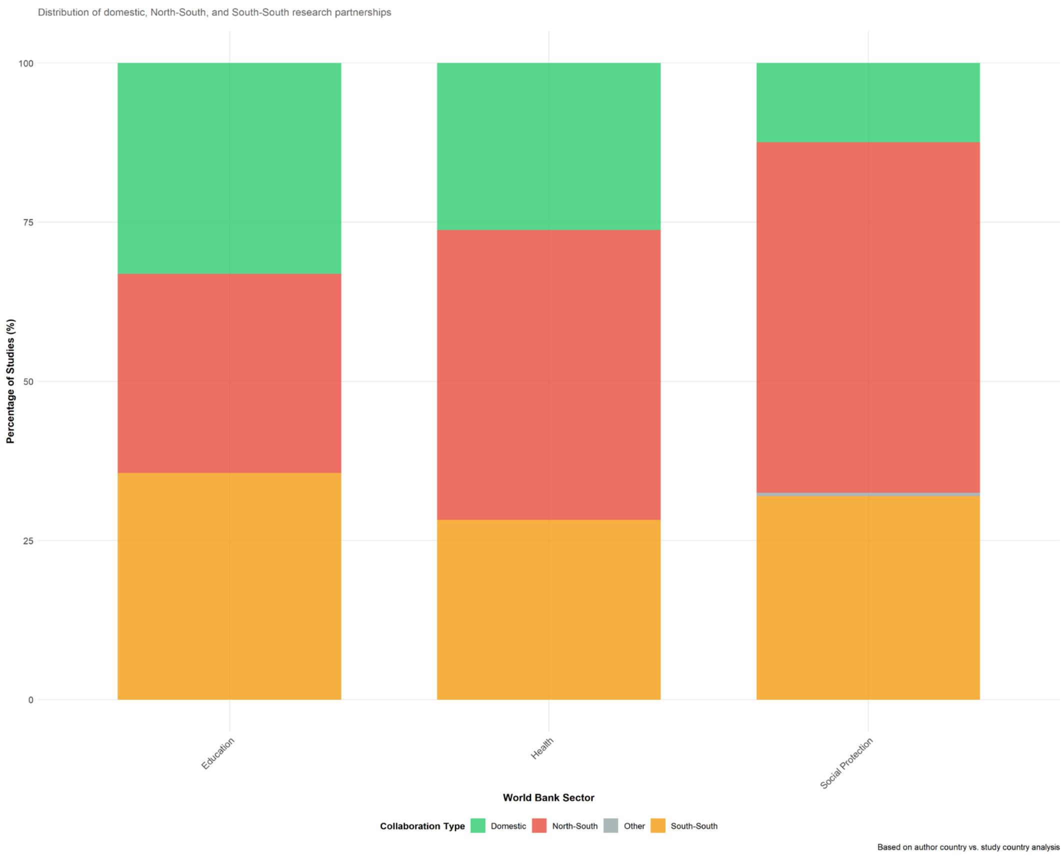
Note: Dependent variable: log(sample size); N = 3732 studies

Research collaboration patterns expose the political economy of development knowledge production (see Figure 17). **North-South partnerships account for 137, 1352, 1040 studies across sectors, with Health showing the strongest concentration (45.5%), reflecting the global health architecture where funding flows from DAC countries through multilateral institutions to FCAS implementation sites.** This pattern institutionalises dependency relationships where research agendas, methodological frameworks, and analytical priorities originate in high-income countries despite implementation in FCAS contexts.

Agriculture's higher South-South collaboration (%) suggests **emerging horizontal knowledge networks around shared agro-ecological challenges and indigenous farming systems,** potentially offering models for more equitable research partnerships. Education's high domestic research proportion (33.1%) reflects both the context-specificity of educational systems and possible **limited international investment in FCAS education research, contributing to evidence gaps in comparative educational policy analysis.**

The uneven distribution of methodological approaches across sectors found in the sample may reflect differences in access, priorities, and operational feasibility rather than a uniform pattern of methodological development. Expanding approaches used in underrepresented sectors, such as governance or justice, could strengthen the overall diversity and relevance of the evidence base.

Figure 17: Collaboration networks — research collaboration patterns across sectors.



5. Limitations

The scope and interpretation of this evidence mapping are subject to several methodological and structural limitations. First, the search strategy (see Section 9.0.1) primarily operationalised the concept of violent conflict through the term *"conflict,"* combined with proximity operators and related qualifiers such as *"violence," "group conflict,"* and *"interethnic conflict."* While this framing ensured conceptual coherence with development and peacebuilding literatures, it excluded other terminologies widely used in political science and security studies, including *"war," "insurgency," "terrorism,"* and *"civil unrest."* As demonstrated by sensitivity testing, this definitional boundary reduced retrieval coverage and likely limited the inclusion of studies employing alternative disciplinary taxonomies of organised violence.

Second, the geographic distribution of included studies reflects a methodological artifact arising from the differential application of geographic screening criteria, as discussed in Note 1. Countries classified as having nationwide FCAS exposure were included in full, whereas those with regionalised exposure required explicit geographic linkage to conflict-affected zones. This produced an apparent over-representation of certain contexts—particularly in the Sahel—while under-representing others with comparable levels of fragility. The resulting imbalances should therefore be interpreted as products of methodological design rather than as indicators of actual research intensity or capacity.

Third, limitations in sampling and data availability remain intrinsic to research conducted in fragile and conflict-affected settings. Populations in displacement, inaccessible regions, or high-risk zones are systematically excluded from many empirical studies, contributing to underrepresentation of the most affected communities. Such constraints also extend to the evidence corpus itself, as the availability of published studies is shaped by where research can be safely and feasibly conducted. Consequently, while the map captures dominant methodological patterns, it cannot fully account for unseen evidence or informal research activities occurring beyond formal publication channels.

Fourth, despite the inclusion of major grey literature sources, the corpus is primarily English-language and biased toward outputs indexed in international repositories. Important regional studies and local policy documents—particularly those published in Arabic, French, or national institutional archives—are likely underrepresented. This linguistic and indexing limitation constrains the inclusivity of the evidence base and may skew findings toward anglophone methodological traditions.

Finally, the use of automated tools for screening and extraction, while significantly improving efficiency and recall, introduces additional uncertainties. Machine-learning classifiers and large language models can misclassify documents or inconsistently interpret methodological details, particularly where study metadata are incomplete or poorly structured. Although quality assurance procedures mitigated these risks, a degree of classification error remains inherent. Furthermore, the evidence base reflects the state of research as of mid-2025 and may not capture subsequent methodological innovations or shifts in conflict dynamics. Future iterations can mitigate these uncertainties through periodic model retraining on manually validated datasets and clearer documentation of classification criteria. Combining automated screening with targeted expert review at critical stages of extraction would further enhance accuracy and interpretive consistency.

Research in FCAS settings raises unique ethical challenges that extend beyond standard institutional review procedures. Limited state authority, high mobility, and persistent insecurity complicate informed consent and data protection. Researchers must also consider the potential re-identification of vulnerable individuals through digital or spatial data and the risk that research activities themselves can alter local power relations. The growing use of remote data-collection technologies introduces additional concerns about surveillance, data sovereignty, and participant comprehension. These ethical tensions do not merely constrain methodology; they shape the kind of knowledge that can be safely and responsibly produced. Recognising and documenting such dilemmas is therefore integral to transparency and should be viewed as a core component of methodological rigour in FCAS research.

Taken together, these limitations highlight the interpretive boundaries of this mapping exercise. They do not undermine the validity of the patterns identified but rather situate them within the operational, linguistic, and structural constraints characteristic of research in fragile and conflict-affected contexts. Future iterations should address these issues through expanded multilingual coverage, refined inclusion logic, and continued methodological experimentation combining human expertise with transparent AI-assisted processes.

6. Discussion and Conclusions

This evidence mapping shows clear progress in the scope and diversity of research conducted in fragile and conflict-affected settings between 2015 and 2025. The volume of studies has increased, and researchers have developed a range of practical approaches for working under difficult conditions. At the same time, the analysis highlights continuing imbalances that limit how well current evidence reflects the realities of fragility and conflict.

Research activity remains uneven across regions and sectors. Most studies are concentrated in countries where research access is feasible and in sectors that align with major funding streams, such as health and social protection. Areas such as governance, justice, and infrastructure receive less attention, even though they are central to understanding how fragility persists or recedes. This pattern reflects both operational constraints and the way funding priorities shape the research agenda.

Methodologically, the field is adaptive but not yet balanced. Quantitative and cross-sectional designs continue to dominate, largely because they are easier to implement and compare across settings. Mixed-methods and qualitative designs, though less common, are increasingly used to capture context and meaning, and they often yield deeper insight into the mechanisms shaping outcomes. Rather than a single preferred model, the mapping suggests a range of approaches that can be effective when matched appropriately to research questions and constraints.

The geography of authorship and institutional participation also remains uneven. Much of the research in FCAS contexts is led by institutions based outside those settings, with local researchers and organisations playing smaller roles. This imbalance limits opportunities for building lasting research capacity and for ensuring that local perspectives shape both study design and interpretation.

While quantitative and mixed-methods studies dominate the FCAS research landscape, qualitative and ethnographic traditions remain essential to understanding the social dynamics underlying fragile contexts. In many settings where statistical sampling is infeasible or ethically problematic, these approaches provide the only viable path to capturing local meanings, institutional trust, and lived experience. Ethnographic and participatory work also serve a corrective function: they surface informal governance mechanisms, social coping systems, and cultural interpretations of conflict that quantitative models often abstract away. Far from being secondary, qualitative inquiry constitutes a primary mode of methodological adaptation in FCAS research, enabling deeper contextualisation and helping bridge the gap between empirical measurement and social understanding. Incorporating such approaches more explicitly within future evidence syntheses would yield a more balanced methodological ecosystem.

Clarifying the distinction between fragility and conflict, refining causal attribution terminology, and ensuring transparent coding validation together enhance the interpretive robustness of this evidence map.

Overall, the evidence base has grown in size and technical quality but still shows structural gaps. These include limited cross-sector learning, weak data sharing between institutions, and a lack of comparative work across countries. Strengthening future research will require more consistent investment in local capacity, better integration of different types of evidence, and closer alignment between research priorities and the information needs of decision-makers working within fragile contexts.

7. Recommendations for Research Funders and Institutions

The findings from this evidence mapping point to the need for a more deliberate and coordinated approach to strengthening research practice in fragile and conflict-affected settings. The recommendations below emphasise depth and sustainability rather than expansion, recognising that progress depends on improving the quality, inclusivity, and usefulness of the evidence produced.

7.1 Strengthen Design Transparency and Contextual Fit

Future research should make its design choices more explicit and better aligned with the realities of the settings in which it operates. Clear documentation of sampling logic, data sources, and analytical assumptions is essential for assessing quality and comparability across studies. Funders can support this by requiring structured design statements and by providing space in reporting templates for reflection on how methods were adapted to local conditions. Improving transparency will not only enhance reproducibility but also help identify when methodological compromises are necessary and how they affect validity.

7.2 Prioritise Governance and Institutional Research

Governance emerged as the least developed area within the FCAS evidence base, despite its importance to understanding resilience, service delivery, and the functioning of public institutions. Funders should explicitly support studies that examine how governance operates under stress—covering topics such as administrative capacity, justice provision, local accountability, and political inclusion. Investments in this area should link empirical research with policy and implementation partners, ensuring findings are applicable to national planning and reform processes.

7.3 Invest in Local Capacity and Collaborative Infrastructure

A key finding of this mapping is the limited participation of researchers and institutions based in FCAS contexts. Sustainable improvement in research quality depends on long-term investment in local capacity—training, institutional infrastructure, and data systems—rather than one-off project partnerships. Funders should prioritise collaborative models that give local institutions leadership roles in study design, data ownership, and interpretation. Shared data repositories, cross-country networks, and mentorship programmes can promote cumulative learning while ensuring that locally generated evidence informs regional and global policy debates.

7.4 Promote Methodological Diversity and Data Integration

The evidence base remains dominated by single-method studies, with limited integration across data types and analytical approaches. Funders can help broaden this landscape by supporting research designs that combine quantitative, qualitative, and administrative data in a coherent framework. This does not mean promoting complexity for its own sake, but ensuring that methods are selected for their appropriateness to the research question.

Encouraging cross-sectoral studies and integrated data infrastructures will also facilitate learning across domains such as health, governance, and livelihoods, where issues of fragility often intersect.

7.5 Use of large-scale secondary data

Invest in strengthening the analytical capacity and technical support needed to fully leverage existing large-scale datasets. Secondary data sources offer significant potential for deeper statistical and comparative analysis, but realising this value requires targeted investment in data management skills, specialised analytical training, and access to appropriate software and infrastructure. Enhancing these capacities would enable researchers—particularly those working in resource-constrained or FCAS contexts—to extract more meaningful insights from existing data and reduce duplication of primary data collection efforts.

Appendix A

Table 12: Evidence mapping workflow — multi-stage identification and screening process.

Stage	Description	Methods/Tools
1. Search Strategy	Development and execution of comprehensive search strategy across multiple databases	Boolean searches, database-specific syntax
2. EPPI Human Piloting	Human calibration to define inclusion/exclusion criteria (300 abstracts)	EPPI-Reviewer, human consensus
3. EPPI Classifier Model	Machine learning classifier to identify potentially relevant studies	EPPI-Reviewer classifier
4. Geolocation of Included	Geographic location extraction for included studies	LLM-based location extraction
5. Full Text Retrieval	DOI scraping and PDF retrieval through multiple sources (Crossref, Semantic Scholar, etc.)	Zotero/API retrieval, PDF matching
6. Title/Abstract Parsing	Automated parsing of titles and abstracts from PDFs	Python pdf_parser_multy.py
7. AI-Human Screening	Two-stage screening with ChatGPT first review and human verification	GPT-4 + human review
8. Full Text Parsing	Full text parsing of final included studies	Python pdf_parser_multy.py
9. FTS Extraction	Data extraction from full texts using AI and manual methods	EPPI/ChatGPT + R scripts
10. FTS Accuracy Checks	Quality checks on extracted data	Human validation samples
11. Final Geolocation	Final geographic verification of study locations	ArcGIS geocoding

This exercise employed a comprehensive, multi-stage methodology to identify, screen, and analyse literature on health systems in Fragile and Conflict-Affected States (FCAS). The approach combined automated AI-assisted methods with human expert review to ensure both efficiency and accuracy across 265,011 initial records.

The process successfully reduced 265,011 initial records to 5,327 final included studies through:

1. **Comprehensive searching** across academic and grey literature
2. **Deduplication** using both automated and manual methods
3. **Progressive screening** combining human expertise with AI assistance
4. **Geographic validation** ensuring FCAS setting relevance
5. **Dual full-text retrieval** maximising document access
6. **Two-stage AI screening** with human verification
7. **Enhanced geolocation** for conflict zone identification

This methodology demonstrates the successful integration of traditional evidence gap map approaches with modern AI-assisted tools, achieving both efficiency and methodological rigor in evidence synthesis for FCAS research.

Search Strategy and Initial Identification

Academic Database Search

The academic search was conducted on May 25, 2025, across eight databases that were selected to provide comprehensive coverage of health systems literature in fragile and conflict-affected settings. The databases included Gender Studies and Africa-Wide through EBSCO, RePEc & Greenfile via EBSCO Discovery Service, Scopus through Elsevier, the International Bibliography of Social Sciences via ProQuest, and CAB Abstracts, Global Health, and EconLit through Ovid interfaces (See Table 13).

Table 13: Search strategy and yield — records retrieved from academic databases (May 2025).

Database Name	Interface	Number of results
Gender Studies	EBSCO	963
Africa-Wide	EBSCO	19,617
RePeC & Greenfile	EBSCO Discovery Service	14,389
Scopus	Elsevier	73,925
International Bibliography of Social Sciences	ProQuest	26,292
CAB Abstracts	Ovid	77,505
Global Health	Ovid	46,483
EconLit	Ovid	3,562
Total		**2,62,736**

Total academic records identified: 262,736

The search strategy was developed in consultation with information specialist Zahra Premji and incorporated multiple components to ensure comprehensive coverage. Country terms included a comprehensive list of FCAS countries and their various demonyms, recognising that studies might refer to populations using different terminology (for example, “Afghan,” “Afghans,” or “Afghani” when discussing Afghanistan). Fragility indicators captured both explicit mentions of state fragility through terms like “fragil* N4 (state OR states)” and contextual indicators such as “humanitarian” settings. Given that several target countries have only a “countries with regionalised FCAS exposure” designation, the search included sub-national specificity with region-specific terms for these countries. Due to the extremely high number of initial hits, the temporal scope was restricted to studies from 2010 onwards rather than the originally planned 2000 start date, following discussions about resource constraints and scope management.

Grey Literature Search

Grey literature searches were conducted between June 26 and July 9, 2025, using a pragmatic approach that balanced comprehensiveness with available resources. Due to time and resource constraints, the team restricted grey literature searching to those sources where 3ie had existing web scraping capabilities already established. This decision was made recognising that grey literature searching can be extremely labor-intensive when conducted manually, as publications on non-academic repositories often lack standardised metadata fields and require manual extraction of key bibliographic information. The broad inclusion criteria of this review would have made comprehensive manual grey literature searching prohibitively time-consuming, so the team focused on four sources where automated extraction was feasible.

Total grey literature records identified: 2,275

Deduplication Process

The deduplication process was implemented as a two-stage approach to ensure thorough removal of duplicate records while maintaining efficiency. For academic records, an initial deduplication was performed using R scripts, which successfully identified and removed 108,086 duplicate records from the original 262,736. The remaining 154,650 records were then uploaded into EPPI-Reviewer for further processing. EPPI-Reviewer's built-in deduplication algorithms identified an additional 24,358 records as duplicates, resulting in 130,292 unique academic records advancing to the screening stage. This two-stage approach was necessary because different deduplication algorithms may identify different types of duplicates, and the combination of R-based and EPPI-based approaches provided more comprehensive duplicate removal than either method alone.

For grey literature records, all 2,275 records were processed directly in EPPI-Reviewer, where 703 were identified as duplicates. This left 2,204 unique grey literature records for further screening. The combined total of unique records after all deduplication procedures was 132,496.

Year Restriction Filter

Following the initial deduplication, the research team made a pragmatic decision to implement a year restriction filter. Despite the comprehensive search strategy, the volume of records requiring screening remained challenging given available resources. After careful consideration of the research objectives and timeline constraints, the team decided to exclude all studies published before 2015. This decision removed 33,868 records from consideration, leaving 96,424 records for title and abstract screening. The 2015 cutoff was selected to ensure that the most recent and relevant literature would be captured while making the screening workload manageable, and importantly, this temporal restriction did not compromise the core objectives of the evidence mapping exercise.

Title and Abstract Screening

Human Calibration and Training

The title and abstract screening process began with an extensive human calibration phase designed to ensure consistent application of inclusion and exclusion criteria across all reviewers. Initially, all four coders collaboratively screened the same batch of 50 studies, discussing and reconciling their decisions to establish a shared understanding of the screening criteria. Following this initial calibration, the coders were divided into pairs to continue the training process through additional batches (See Table 14).

Table 14: Inter-rater agreement — reviewer concordance during calibration.

Batch	Coder_1	Coder_2	Agreement
1	Cem	Etienne	82%
1	Suvarna	Lucas	78%
2	Cem	Lucas	78%
2	Suvarna	Etienne	94%

Two additional training batches of 50 records each were screened in pairs until agreement rates between all pairs reached the target threshold of 75%. The training phase was considered complete when consistent decision-making was demonstrated across all reviewer pairs. Following successful completion of the training phase, each coder independently screened 100 records to provide the initial dataset for machine learning classifier development. This systematic approach to human calibration ensured that the subsequent machine learning models would be trained on high-quality, consistently coded data.

Machine Learning Classifier Development

The research team implemented an iterative approach to machine learning classifier development using EPPI-Reviewer's built-in capabilities. This process involved building five successive classifiers, with each iteration incorporating additional manually screened records to improve performance. The first classifier was developed based on the initial 646 manually screened records from the training phase, achieving reasonable accuracy (0.800) but with relatively low recall (0.400), indicating that while the model made few false positive predictions, it was missing many relevant studies (See Table 15).

Table 15: Classifier performance — metrics across iterative machine-learning classifiers.

Classifier	Records	Accuracy	AUC	Precision	Recall
#1	646	0.800	0.853	0.888	0.400
#2	2,343	0.782	0.855	0.827	0.814
#3	1,259	0.775	0.844	0.823	0.838
#4	1,572	0.839	0.898	0.897	0.887
#5	4,790	0.852	0.917	0.898	0.891

Each subsequent classifier iteration incorporated feedback from human screening of records identified by the previous classifier. The progressive improvement in recall from 0.400 in Classifier #1 to 0.891 in Classifier #5 demonstrates the effectiveness of this iterative approach. The final classifier (#5) achieved excellent performance across all metrics, with high accuracy (0.852), strong discriminative ability (AUC = 0.917), and balanced precision (0.898) and recall (0.891). This iterative development process was essential for creating a reliable automated screening tool capable of handling the large volume of records while maintaining high sensitivity for relevant studies.

Final Classifier Distribution

Table 16 shows the results of the final classifier chosen. Using Classifier #5, the remaining 90,930 records were distributed as follows:

Table 16: Final classifier score distribution — record probabilities from classifier #5.

Range	Count
0-9%	15,360
10-19%	27,616
20-29%	18,347
30-39%	10,455
40-49%	6,040
50-59%	4,345
60-69%	3,354
70-79%	2,613
80-89%	2,035
90-99%	765

Decision threshold: Records scoring ≥30% were included for full-text screening

- **Excluded**: 61,323 records (scoring <30%)
- **Advanced to next stage**: 29,607 records

Records included after human-EPPI classification: 34,763

Geographic Location Screening

LLM-Based Location Extraction

Due to existence of countries with regionalised FCAS exposure, GPT-4 was used to extract specific location information:

Countries with regional FCAS designation: Cameroon, Chad, DRC, Ethiopia, Iraq, Lebanon, Mozambique, Myanmar, Nigeria

Geolocation Methodology

Table 17: Geolocation methods — conflict intersection and spatial analysis approach.

Geolocation and Conflict Analysis Methodology

Processing Step	Methods	Results
Location Extraction	LLM-based extraction Country-specific regex Multi-country classificatio	|33,686 items processed 2,468 multi-country studies
Geocoding	ArcGIS World Geocoding Batch processing 85.2% deduplication	|11,515 successfully geocoded 98% within

	efficiency	expected boun
Conflict Zone Mapping	Uppsala Conflict Data 25km grid cells Intensity classification	|6,165 items in conflict zones Risk categories assigne
Exposure Classification	Point-in-polygon analysis Casualty-based levels Temporal context	|26 items flagged for manual review (0.2%)

Records included after geolocation on title and abstract: 24,218

Full-Text Retrieval

Dual Retrieval Strategy

Table 18: Full-text retrieval — strategies and retrieval outcomes.

Method	Process	Results
Zotero Retrieval	DOI scraping from Crossref Zotero library management	|9,839 records retrieved
API Retrieval	Multi-source API calls Semantic Scholar, OpenAlex Wiley, Elsevie	|13,375 records retrieve
Combined Processing	PDF matching and deduplication Compare_pdfs_suvarna_lucas.R	|23,779 unique records

Records included after full-text deduplication: 23,779

AI-Assisted Full-Text Screening

Two-Stage Screening Process

Stage 1: AI High-Sensitivity Screening (GPT-4)

- **Method**: ChatGPT used as first reviewer for exclusion
- **Performance**: 67.5% accuracy against human gold standard
- **Approach**: High sensitivity to minimise false exclusions
- **Results**: AI excluded ~27% of abstracts automatically

Table 19: AI screening validation — performance vs human gold standard (n=300).

Metric	Value
Overall Accuracy	67.5% (63.2-71.6%)
Sensitivity (Exclusions)	40.3% (32.1-48.9%)
Specificity (Inclusions)	78.6% (74.1-82.7%)
Cohen's κ	0.19 (poor agreement)

Stage 2: Strategic Human Review

- **Scope**: Human verification of AI-included records
- **Protocol**: Structured CSV-based review system
- **Focus**: Final inclusion determination with documented reasoning

Records after full-text screening (AI for exclusion): 11,444

Records included after full-text screening (human for inclusion): 8,146

Structured Data Extraction for Evidence Mapping

AI-Assisted Data Extraction Using ChatGPT-4.1 Mini

Following the completion of human verification screening, the research team implemented a comprehensive structured data extraction process to support the broader evidence mapping objectives. While the traditional mapping pathway continued with enhanced geolocation verification for final study inclusion, a parallel extraction process was conducted to capture standardised information from the larger pool of relevant studies identified through the screening process. The structured extraction utilised ChatGPT-4.1 Mini to systematically extract key data elements from 8,138 studies that had successfully passed the initial screening phases. This AI-assisted approach enabled comprehensive data capture across a large corpus of studies while maintaining consistency in extraction protocols. The extraction framework was designed to capture essential study characteristics including geographic location, study design, population characteristics, health system components addressed, intervention types, and key findings relevant to FCAS contexts. The decision to use ChatGPT-4.1 Mini for this extraction phase reflected both the scale of the task and the need for standardised data capture across diverse study types and reporting formats. This approach allowed the research team to build a comprehensive evidence base for mapping purposes while simultaneously conducting the more intensive final inclusion process for studies meeting the strictest inclusion criteria. The extracted data from these 8,138 studies formed the foundation for the evidence mapping analysis, providing a rich dataset for identifying research gaps, geographic distributions of evidence, and thematic patterns in health systems research conducted in fragile and conflict-affected settings.

Final Geolocation

Enhanced Geographic Verification

- **Conflict zone verification**: Cross-referenced with Uppsala Conflict Data
- **Administrative boundary validation**: Ensured geographic accuracy
- **Manual review**: Edge cases requiring expert judgment

Records included after geolocation on full text: 5,327

Appendix A.1: Search Strategy Example

The systematic search strategy employed multiple databases using structured Boolean queries. Table 20 demonstrates the specific search string construction used in the Gender Studies Database, showing the combination of key terms, Boolean operators, and field restrictions that ensured comprehensive coverage while maintaining precision.

Gender Studies Database (EBSCO) Search String:

Table 20: Example search string — gender studies database query sample.

#	Query	Results
S1	TI ((("Afghanistan" OR "burkina\ faso" OR "burkina\ fasso" OR "cameroon" OR "cameron" OR "cameroun" OR "central\ african\ republic" OR "ubangi\ shari" OR "chad" OR "democratic\ republic\ of\ the\ congo" OR "democratic\ republic\ congo" OR "congo" OR "zaire" OR "ethiopia" OR "haiti" OR "iraq" OR "lebanon" OR "lebanese\ republic" OR "mali" OR "mozambique" OR "portuguese\ east\ africa" OR "myanmar" OR "burma" OR "niger" OR "nigeria" OR "somalia" OR "south\ sudan" OR "sudan" OR "syria" OR "syrian\ arab\ republic" OR "west\ bank" OR "gaza" OR "palestine" OR "yemen" OR "libya" OR "libyan\ arab\ jamahiriya" OR "afghan" OR "afghans" OR "afghani*" OR "burkinabe" OR "burkinese" OR "cameroonian*" OR "central\ African" OR "central\ africans" OR "Chadian*" OR "congolese" OR "Ethiopian*" OR "Haitian*" OR "iraqian*" OR "iraqi*" OR "lebanese" OR "malian*" OR "mozambican*" OR "burmese" OR "myanma*" OR "nigerien*" OR "nigerian*" OR "somali*" OR "somalian*" OR "south\ sudanese" OR "sudanese" OR "syrian*" OR "palestinian*" OR "yemeni*" OR "Yemenite*" OR "yemenese" OR "Libyan*") AND (("fragil*" N4 ("state" OR "states")) OR "humanitarian" OR (("conflict" OR "conflicts") N3 ("affect*" OR "experienc*" OR "site" OR "sites" OR "zone*" OR "areas")) OR (("conflict" OR "conflicts" OR "violence") N3 ("group" OR "groups" OR "intergroup" OR "ethnic" OR "interethnic"))))) OR\ AB ((("Afghanistan" OR "burkina\ faso" OR "burkina\ fasso" OR "cameroon" OR "cameron" OR "cameroun" OR "central\ african\ republic" OR "ubangi\ shari" OR "chad" OR "democratic\ republic\ of\ the\ congo" OR "democratic\ republic\ congo" OR "congo" OR "zaire" OR "ethiopia" OR "haiti" OR "iraq" OR "lebanon" OR "lebanese\ republic" OR "mali" OR "mozambique" OR "portuguese\ east\ africa" OR "myanmar" OR "burma" OR "niger" OR "nigeria" OR "somalia" OR "south\ sudan" OR "sudan" OR "syria" OR "syrian\ arab\ republic" OR "west\ bank" OR "gaza" OR "palestine" OR "yemen" OR "libya" OR "libyan\ arab\ jamahiriya" OR "afghan" OR "afghans" OR "afghani*" OR "burkinabe" OR "burkinese" OR "cameroonian*" OR "central\ African" OR "central\ africans" OR "Chadian*" OR "congolese" OR "Ethiopian*" OR "Haitian*" OR "iraqian*" OR "iraqi*" OR "lebanese" OR "malian*" OR "mozambican*" OR "burmese" OR "myanma*" OR "nigerien*" OR "nigerian*" OR "somali*" OR "somalian*" OR "south\ sudanese" OR "sudanese" OR "syrian*" OR "palestinian*" OR "yemeni*" OR "Yemenite*" OR "yemenese" OR "Libyan*") AND (("fragil*" N4 ("state" OR "states")) OR "humanitarian" OR (("conflict" OR "conflicts") N3 ("affect*" OR "experienc*" OR "site" OR "sites" OR "zone*" OR "areas")) OR (("conflict" OR "conflicts" OR "violence") N3 ("group" OR "groups" OR "intergroup" OR "ethnic" OR "interethnic"))))) OR\ SU ((("Afghanistan" OR "burkina\ faso" OR "burkina\ fasso" OR "cameroon" OR "cameron" OR "cameroun" OR "central\ african\ republic" OR "ubangi\ shari" OR "chad" OR "democratic\	731

republic\ of\ the\ congo" OR "democratic\ republic\ congo" OR "congo" OR "zaire" OR "ethiopia" OR "haiti" OR "iraq" OR "lebanon" OR "lebanese\ republic" OR "mali" OR "mozambique" OR "portuguese\ east\ africa" OR "myanmar" OR "burma" OR "niger" OR "nigeria" OR "somalia" OR "south\ sudan" OR "sudan" OR "syria" OR "syrian\ arab\ republic" OR "west\ bank" OR "gaza" OR "palestine" OR "yemen" OR "libya" OR "libyan\ arab\ jamahiriya" OR "afghan" OR "afghans" OR "afghani*" OR "burkinabe" OR "burkinese" OR "cameroonian*" OR "central\ African" OR "central\ africans" OR "Chadian*" OR "congolese" OR "Ethiopian*" OR "Haitian*" OR "iraqian*" OR "iraqi*" OR "lebanese" OR "malian*" OR "mozambican*" OR "burmese" OR "myanma*" OR "nigerien*" OR "nigerian*" OR "somali*" OR "somalian*" OR "south\ sudanese" OR "sudanese" OR "syrian*" OR "palestinian*" OR "yemeni*" OR "Yemenite*" OR "yemenese" OR "Libyan*") AND (("fragil*" N4 ("state" OR "states")) OR "humanitarian" OR (("conflict" OR "conflicts") N3 ("affect*" OR "experienc*" OR "site" OR "sites" OR "zone*" OR "areas")) OR (("conflict" OR "conflicts" OR "violence") N3 ("group" OR "groups" OR "intergroup" OR "ethnic" OR "interethnic")))))

S2	TI ((("Cameroon" AND ("Bakassi\ Peninsula" OR "Central\ African\ Republic\ border" OR "Chad\ border" OR "Nigeria\ border" OR "Far-North" OR "North-West\ Region" OR "South-West\ Region" OR "limbe")) OR ("Congo" AND ("Kinshasa" OR "N'djili" OR "Ndjili" OR "Kimbanseke" OR "Nsele\ commune" OR "Menkao" OR "Kenge" OR "Mai-Ndombe" OR "Kasa\ Vubu" OR "Triumphal\ Road" OR "Barumbu" OR "Lingwala" OR "Central\ African\ Republic\ border" OR "Haut-Uele" OR "Ituri" OR "South\ Sudan\ border" OR "North\ Kivu" OR "Goma" OR "south\ Kivu" OR "Bukavu" OR "Maniema" OR "Tanganyika" OR "Haut-Lomami" OR "Kwamouth" OR "Bandundu" OR "Kasai")) OR ("Ethiopia" AND ("tigray" OR "amhara" OR "oromia" OR "Afar--Somali\ Border" OR "Benishangul-Gumuz" OR "gambella")) OR ("Haiti" AND ("Cite\ Soleil" OR "Port-au-Prince" OR "delmas" OR "Croix-des-Bouquets" OR "La\ Saline" OR "tabarre" OR "Petion-Ville")) OR ("Iraq" AND ("Anbar" OR "Ramadi\ City" OR "Basra" OR "Diyala" OR "Kirkuk" OR "Ninawa" OR "Salah\ al-Din" OR "Sadr\ City" OR "Baghdad" OR ("border*" AND ("Iran" OR "Syria" OR "Saudi\ Arabia" OR "Kuwait")))) OR ("Lebanon" AND ("southern" OR "aitaroun" OR "tyre" OR "Beqaa\ valley" OR "eastern" OR "Baalbek-Hermel" OR "Ain\ Ebel" OR "Tariq\ el\ Jdideh" OR "Bir\ Hassan" OR "Ghobeiry" OR "Chiayah" OR "Rizkallah\ Semaan\ road" OR "Old\ Saida\ road" OR "haret\ Hraik" OR "Burj\ Al\ Barajneh" OR "Mraije" OR "Laylake")) OR ("Myanmar" AND ("Chin\ State" OR "Kachin" OR "Kayah" OR "Kayin" OR "Mon\ State" OR "Rakhine" OR "Sagaing" OR "Magway" OR "Tanintharyi" OR "Shan\ State\ North" OR "North\ Mandalay" OR "Mandalay\ City" OR "Pyin\ Oo\ Lwin" OR "Yangon-Mandalay\ Expressway" OR "Bago")) OR ("Mozambique" AND ("Cabo\ Delgado" OR "Niassa" OR "Nampula")) OR ("Niger" AND ("Tillaberi" OR "Tahoua" OR "Diffa" OR "Maradi")) OR ("Nigeria" AND ("Borno" OR "Yobe" OR "Adamawa" OR "Gombe" OR "Kaduna" OR "Katsina" OR	10725

"Zamfara" OR "riverine\ area*" OR "Delta" OR "Bayelsa" OR "Akwa\ Ibom" OR "Cross\ River\ state*")) OR ("Chad" AND ("Lake\ Chad" OR "Lac\ Province*" OR "Eastern\ Chad" OR "Ouaddai" OR "Sila" OR "Wadi\ Fira" OR "Ennedi\ Est" OR "Southern\ Chad" OR "Logone\ Occidental" OR "Moyen-Chari" OR "N'Djamena" OR "NDjamena" OR "Central\ Chad")) OR ("Congo" AND "pool") OR ("Libya" AND ("Tripoli" OR "Eastern\ Libya" OR "Benghazi" OR "Southern\ Libya" OR "Fezzan" OR "Sirte" OR "Misrata")) OR ("Afghanistan" OR "Burkina\ Faso" OR "Central\ African\ Republic" OR "Haiti" OR "Libya" OR "Mali" OR "Niger" OR "Somalia" OR "south\ sudan" OR "Sudan" OR "Syria" OR "Syrian\ Arab\ Republic" OR "west\ bank" OR "gaza" OR "yemen"))) OR\ AB ((("Cameroon" AND ("Bakassi\ Peninsula" OR "Central\ African\ Republic\ border" OR "Chad\ border" OR "Nigeria\ border" OR "Far-North" OR "North-West\ Region" OR "South-West\ Region" OR "limbe")) OR ("Congo" AND ("Kinshasa" OR "N'djili" OR "Ndjili" OR "Kimbanseke" OR "Nsele\ commune" OR "Menkao" OR "Kenge" OR "Mai-Ndombe" OR "Kasa\ Vubu" OR "Triumphal\ Road" OR "Barumbu" OR "Lingwala" OR "Central\ African\ Republic\ border" OR "Haut-Uele" OR "Ituri" OR "South\ Sudan\ border" OR "North\ Kivu" OR "Goma" OR "south\ Kivu" OR "Bukavu" OR "Maniema" OR "Tanganyika" OR "Haut-Lomami" OR "Kwamouth" OR "Bandundu" OR "Kasai")) OR ("Ethiopia" AND ("tigray" OR "amhara" OR "oromia" OR "Afar--Somali\ Border" OR "Benishangul-Gumuz" OR "gambella")) OR ("Haiti" AND ("Cite\ Soleil" OR "Port-au-Prince" OR "delmas" OR "Croix-des-Bouquets" OR "La\ Saline" OR "tabarre" OR "Petion-Ville")) OR ("Iraq" AND ("Anbar" OR "Ramadi\ City" OR "Basra" OR "Diyala" OR "Kirkuk" OR "Ninawa" OR "Salah\ al-Din" OR "Sadr\ City" OR "Baghdad" OR ("border*" AND ("Iran" OR "Syria" OR "Saudi\ Arabia" OR "Kuwait")))) OR ("Lebanon" AND ("southern" OR "aitaroun" OR "tyre" OR "Beqaa\ valley" OR "eastern" OR "Baalbek-Hermel" OR "Ain\ Ebel" OR "Tariq\ el\ Jdideh" OR "Bir\ Hassan" OR "Ghobeiry" OR "Chiayah" OR "Rizkallah\ Semaan\ road" OR "Old\ Saida\ road" OR "haret\ Hraik" OR "Burj\ Al\ Barajneh" OR "Mraije" OR "Laylake")) OR ("Myanmar" AND ("Chin\ State" OR "Kachin" OR "Kayah" OR "Kayin" OR "Mon\ State" OR "Rakhine" OR "Sagaing" OR "Magway" OR "Tanintharyi" OR "Shan\ State\ North" OR "North\ Mandalay" OR "Mandalay\ City" OR "Pyin\ Oo\ Lwin" OR "Yangon-Mandalay\ Expressway" OR "Bago")) OR ("Mozambique" AND ("Cabo\ Delgado" OR "Niassa" OR "Nampula")) OR ("Niger" AND ("Tillaberi" OR "Tahoua" OR "Diffa" OR "Maradi")) OR ("Nigeria" AND ("Borno" OR "Yobe" OR "Adamawa" OR "Gombe" OR "Kaduna" OR "Katsina" OR "Zamfara" OR "riverine\ area*" OR "Delta" OR "Bayelsa" OR "Akwa\ Ibom" OR "Cross\ River\ state*")) OR ("Chad" AND ("Lake\ Chad" OR "Lac\ Province*" OR "Eastern\ Chad" OR "Ouaddai" OR "Sila" OR "Wadi\ Fira" OR "Ennedi\ Est" OR "Southern\ Chad" OR "Logone\ Occidental" OR "Moyen-Chari" OR "N'Djamena" OR "NDjamena" OR "Central\ Chad")) OR ("Congo" AND "pool") OR ("Libya" AND ("Tripoli" OR "Eastern\ Libya" OR "Benghazi" OR "Southern\ Libya" OR "Fezzan" OR "Sirte" OR "Misrata")) OR ("Afghanistan" OR

	"Burkina\ Faso" OR "Central\ African\ Republic" OR "Haiti" OR "Libya" OR "Mali" OR "Niger" OR "Somalia" OR "south\ sudan" OR "Sudan" OR "Syria" OR "Syrian\ Arab\ Republic" OR "west\ bank" OR "gaza" OR "yemen")))	
S3	(S1\ OR\ S2)	11089
S4	(S1\ OR\ S2)	963
NA	Limiters - Publication\ Date: 20100101-20251231	NA
NA	Narrow\ by\ Language: - dari	NA
NA	Narrow\ by\ Language: - pashto	NA
NA	Narrow\ by\ Language: - portuguese	NA
NA	Narrow\ by\ Language: - arabic	NA
NA	Narrow\ by\ Language: - french	NA
NA	Narrow\ by\ Language: - english	NA

Appendix A.2: Conflict Zone Classification

To systematically categorise conflict severity across study regions, we developed a four-tier intensity classification framework. Table 21 presents the death thresholds and geographical coverage criteria that determined conflict zone classifications, with higher intensity levels receiving priority attention in the analysis due to their greater impact on research feasibility and participant safety.

Table 21: Conflict intensity thresholds — grid-cell rules for spatial classification (low→very high).

Intensity_Level	Death_Threshold	Grid_Cells	Priority
Low	5-24	Multiple	Standard
Medium	25-99	Multiple	Moderate
High	100-999	Limited	High
Very High	1000+	Few	Critical

Appendix A.3: Target Countries and Regions

Our analysis focused on countries and regions identified through the World Bank’s Fragile, Conflict and Violence (FCV) framework. The selection includes both complete FCAS territory countries where conflict affects the entire nation, and countries with regional FCAS designations where conflict is concentrated in specific provinces or states. This geographical scope ensures comprehensive coverage of contexts where conflict exposure significantly impacts research environments and population access.

Complete FCAS Territory Countries:

- **Afghanistan**
- **Burkina Faso**
- **Central African Republic**
- **Haiti**
- **Libya**
- **Mali**
- **Niger**
- **Somalia**
- **South Sudan**
- **Sudan**
- **Syria**
- **Palestinian Territories (West Bank and Gaza)**

Countries with Regional FCAS Designation:

- **Cameroon**: Far-North, North-West, South-West Regions
- **Chad**: Lake Chad, Lac Province, Eastern Chad
- **DRC**: North/South Kivu, Ituri, Tanganyika, Maniema
- **Ethiopia**: Tigray, Amhara, Oromia regions
- **Iraq**: Anbar, Diyala, Ninawa provinces
- **Lebanon**: Beqaa Valley, Southern Lebanon
- **Mozambique**: Cabo Delgado, Nampula provinces
- **Myanmar**: Rakhine, Kachin, Shan States
- **Nigeria**: Borno, Yobe, Adamawa states

Appendix A.4: Spatial Analysis Methodology

The spatial analysis employed a grid-based approach to systematically identify and classify conflict zones across all target countries. Using 25km × 25km cells, we established minimum thresholds for conflict designation and applied the intensity classification system detailed in Table 21 to ensure consistent measurement across different geographical contexts.

Grid-Based Conflict Zoning:

- 25km × 25km cells covering all target countries
- Minimum threshold: ≥5 events AND ≥10 deaths per zone
- Intensity classification:
  - Low: 5-24 deaths
  - Medium: 25-99
  - High: 100-999
  - Very High: 1000+

Appendix A.5: Exposure Classification Logic

Individual studies were classified according to their proximity to and intersection with identified conflict zones. Table 22 outlines the five-level exposure framework, where studies conducted in capital cities or very high intensity zones received the highest exposure classification due to elevated security risks and potential impacts on data collection quality.

Table 22: Study exposure schema — five-level proximity/severity classification (No → Very high).

Level	Criteria
No Exposure	Outside all conflict zones
Low	1 zone ∩, Low intensity
Moderate	≥2 zones ∩ OR Medium intensity
High	High intensity zone
Very High	Very High intensity OR capital city zone

Appendix A.6 Technical Details during data analysis

Data Processing and Quality Control

Quality control procedures were implemented throughout the data processing pipeline to ensure reliability and completeness. Table 23 summarizes the data quality metrics, including variable completeness rates and the removal of extreme outliers, which informed subsequent analytical decisions and interpretation of results.

To ensure consistency across the 23 analytical variables used in the evidence map, we implemented a dual-phase quality control process. During calibration, two reviewers independently coded a random 10 per cent subset of studies, achieving inter-coder agreement above 0.88 (Cohen's κ). Discrepancies were discussed and resolved through consensus, after which a refined codebook was applied to the full dataset. Automated classification outputs from the AI-assisted system were benchmarked against this gold-standard subset, yielding precision and recall scores of 0.91 and 0.89 respectively. This process balanced efficiency with reliability and ensured that final classifications met acceptable standards for reproducibility in large-scale evidence mapping.

Table 23: Data completeness metrics — variable completeness, outlier handling, and record counts.

Data Processing and Quality Control Summary

Data Quality Metric	Count
total_records	5327
missing_publication_year	41
missing_authors	0
missing_sectors	0
missing_countries	0
valid_sample_sizes	4699
extreme_outliers_removed	0
Variable	Completeness (%)
record_id	100.0
authors	100.0
first_author_country	100.0

first_author_organization	100.0
research_year	100.0
topic_summary	100.0
world_bank_sector	100.0
world_bank_subsector	100.0
sdg	100.0
study_countries	100.0
study_regions	100.0
population	100.0
sample_size	100.0
data_collection_method	100.0
analysis_type	100.0
secondary_dataset	100.0
primary_data_techniques	100.0
data_analysis_methods	100.0
sdg_number	100.0
author_income_group	100.0
sample_category	100.0
data_source_type	100.0
publication_year	99.2
research_period	99.2
sample_numeric	88.2

Statistical Methods and Model Specifications

The analysis employed multiple statistical approaches to examine temporal trends, cross-sectoral patterns, and study quality variations. These models provided robust estimates while accounting for the heterogeneous nature of the research landscape in conflict-affected settings.

Temporal Analysis: Quadratic regression model: $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$

Cross-sectoral Analysis: Multiple regression: $\log(\text{Sample Size}) = \beta_0 + \beta_1 \text{Quantitative} + \beta_2 \text{High Income Author} + \beta_3 \text{Health Sector} + \beta_4 \text{Recent Study} + \beta_5 \text{Publication Year} + \epsilon$

Quality Index: Composite score: $Q = I_{\text{Large Sample}} + I_{\text{Mixed Methods}} + I_{\text{Multi-Country}} + I_{\text{Recent Data}}$

All statistical analyses employed robust standard errors and appropriate significance testing procedures.

Appendix B: Extended Results

Complete Sectoral Distribution

The World Bank sector classification reveals the breadth of research conducted in conflict-affected settings. Table 24 shows the complete distribution across all identified sectors, highlighting the concentration of studies in specific domains and the relative scarcity of research in others, which has implications for evidence-based policy development in these contexts.

Table 24: Full sectoral breakdown — sub-sectors and marginal fields in the FCAS evidence base.

World Bank Sector	Studies	Percentage (%)	Cumulative (%)
Health	2076	38.97	38.97
Social Protection	1143	21.46	60.43
Agriculture	842	15.81	76.23
Education	340	6.38	82.62
Environment and Natural Resources Management	133	2.50	85.11
Governance	130	2.44	87.55
Water	104	1.95	89.51
Urban Development	83	1.56	91.06
Finance	78	1.46	92.53
Private Sector Development	74	1.39	93.92
Energy & Extractives	47	0.88	94.80
Public Administration	40	0.75	95.55
Financial Sector	30	0.56	96.11
Information and Communication Technology	23	0.43	96.55
Tourism	17	0.32	96.87
Transport	17	0.32	97.18
Public Sector Governance	16	0.30	97.48
Macroeconomics, Trade and Investment	13	0.24	97.73
Environment and Natural Resources	12	0.23	97.95
Infrastructure	12	0.23	98.18
Economic Policy	10	0.19	98.37
Trade and Competitiveness	9	0.17	98.54
Economic Policy & Debt	8	0.15	98.69
Energy and Extractives	6	0.11	98.80
Industry and Trade	6	0.11	98.91
Economic Policy and Debt	5	0.09	99.01
Disaster Risk Management	4	0.08	99.08
Economic Policy, Trade and Investment	4	0.08	99.16
Energy	4	0.08	99.23
Water, Sanitation and Waste Management	4	0.08	99.31

Extractives	3	0.06	99.36
Labor Markets	3	0.06	99.42
Water, Sanitation and Hygiene (WASH)	3	0.06	99.47
Conflict, Security and Justice	2	0.04	99.51
Culture and Tourism	2	0.04	99.55
Extractive Industries	2	0.04	99.59
Extractives and Mining	2	0.04	99.62
Information and Communication	2	0.04	99.66
Justice and Public Administration	2	0.04	99.70
Justice and Rule of Law	2	0.04	99.74
Security	2	0.04	99.77
Conflict Prevention and Post-Conflict Reconstruction	1	0.02	99.79
Economic Policy and Debt Management	1	0.02	99.81
Economic Policy and Management	1	0.02	99.83
Economic Policy and Planning	1	0.02	99.85
Extractive Industries and Mining	1	0.02	99.87
Justice and Law	1	0.02	99.89
Macroeconomics	1	0.02	99.91
Macroeconomics and Economic Management	1	0.02	99.92
Macroeconomics, Trade & Investment	1	0.02	99.94
Media and Information	1	0.02	99.96
Urban, Resilience and Land	1	0.02	99.98
Water Supply, Sanitation and Waste Management	1	0.02	100.00

Geographic Distribution Analysis

The geographic concentration of research activity varies significantly across conflict-affected regions. Table 25 presents the top 25 study countries, revealing both expected patterns based on conflict prominence and surprising gaps that may indicate access barriers or research capacity constraints in certain high-priority contexts.

Table 25: Country rankings — number of included studies with cumulative percentages.

Study Country	Studies	Percentage (%)	Cumulative (%)
Burkina Faso	1098	14.26	14.26
Afghanistan	697	9.05	23.31
Mali	496	6.44	29.75
Sudan	469	6.09	35.84
Palestine	441	5.73	41.57
Haiti	394	5.12	46.69
Niger	352	4.57	51.26
Somalia	336	4.36	55.62
Syria	322	4.18	59.81
South Sudan	293	3.81	63.61

Yemen	292	3.79	67.40
Iraq	127	1.65	69.05
Nigeria	122	1.58	70.64
Libya	119	1.55	72.18
Lebanon	102	1.32	73.51
Kenya	95	1.23	74.74
Ethiopia	81	1.05	75.79
Central African Republic	72	0.94	76.73
Ghana	66	0.86	77.58
Uganda	59	0.77	78.35
Cameroon	54	0.70	79.05
Senegal	51	0.66	79.71
Israel	48	0.62	80.34
Tanzania	46	0.60	80.94
Democratic Republic of Congo	44	0.57	81.51

Appendix C: Scripts

```
import pandas as pd
import requests
import json
import time
import os
import hashlib
import pickle
from datetime import datetime
import numpy as np
import re
from concurrent.futures import ThreadPoolExecutor, as_completed
import threading
from threading import Lock


def build_prompt(record_id, text_block):
    """Build the extraction prompt with expert instructions and structured JSON output"""
    timestamp = int(np.floor(np.datetime64('now').astype(int) / 1e9))

    prompt = f"""
DOCUMENT ID: {record_id}
TIMESTAMP: {timestamp}

You are an expert academic evidence synthesis researcher with extensive experience in
systematic reviews, meta-analyses, and research methodology. You specialize in extracting
structured information from academic papers across multiple disciplines including
development economics, public policy, health, education, and social sciences.

Your task is to carefully read and analyze the following academic paper and extract key
information with precision and scholarly rigor. Pay particular attention to methodological
details, data sources, and policy relevance.

EXTRACTION INSTRUCTIONS:

1. Extract all authors with the following format: Last name, First name. Use a semicolon to
separate authors names.
   Example: "Smith, John; Garcia, Maria; Johnson, Sarah"

2. Extract the year of publication of the paper.
   Example: "2023"

3. Extract the country affiliation for the first author only. If not available, add "Not Specified".
   Example: "United States" or "Not Specified"
```

4. Extract the organisational affiliation for the first author only. If not available, add "Not Specified".
   Example: "Harvard University" or "World Bank" or "Not Specified"

5. Extract the year that data collection took place. If collection took place across multiple years, please specify the time frame in the following format: First year - Last year.
   Example: "2020" or "2018-2021" or "Not Specified"

6. Provide a three sentence summary of the research paper. Make sure to include the general focus of the study as well as its setting. Structure the summary in the following format: First sentence - general description of topic including setting of the study. Second sentence - general summary of the methods used including whether it is primary or secondary data and the sample size. Third sentence - note whether the study addresses FCAS in anyway, if it does, briefly mention how.
   Example: "This study examines the impact of microfinance programs on women's empowerment in rural Bangladesh. The research uses primary data collected through surveys with 1,200 women participants and employs logistic regression analysis. The study addresses FCAS relevance by examining vulnerable populations in areas prone to climate-related conflicts."

7. Extract the most relevant World Bank sector.
   Example: "Social Protection" or "Education" or "Health" or "Agriculture"

8. Extract the most relevant World Bank sub-sector.
   Example: "Primary Education" or "Rural Health" or "Social Safety Nets"

9. Extract the UN Sustainable Development Goal (SDG) which is most relevant to this study. Return the data by specifying the SDG number, followed by a two sentence justification.
   Example: "SDG 4: Quality Education. This study directly examines educational access and learning outcomes in primary schools. The research contributes to understanding how to ensure inclusive and equitable quality education for all children."

10. Extract the country/countries where the study was conducted. Use semicolons to separate multiple countries.
   Example: "Bangladesh" or "Kenya; Tanzania; Uganda" or "Multi-country study across Sub-Saharan Africa"

11. Extract the regions inside the country/countries (province, state, district, municipality, or other local administrative divisions). Use semicolons to separate multiple regions.
   Example: "Dhaka Division; Chittagong Division" or "Nairobi County; Kisumu County" or "Northern Province" or "Not specified"

12. Extract the main population of the research paper. If there is no main population, but data was collected at a certain level, specify the level, e.g., households.
   Example: "Women entrepreneurs" or "School children aged 6-12" or "Households" or "Healthcare workers"

13. Extract the sample size.
Example: "1,500 participants" or "450 households" or "Not specified"

14. Extract whether the research uses primary, secondary or both primary and secondary data. If the paper uses secondary data, name the secondary dataset used. Provide the text to justify this decision. If no clear name is available, provide the text which explains what this data is. If primary data is used, extract all primary data collection techniques used with page numbers to support this. Separate techniques with a semicolon and provide justification afterwards.
Example: "Primary data collected through face-to-face interviews (p. 15) and focus group discussions (p. 16)" or "Secondary data from World Bank Living Standards Measurement Study (LSMS) household surveys" or "Mixed: Primary surveys (p. 12) combined with administrative records from Ministry of Education"

15. Extract whether research paper uses quantitative, qualitative or both quantitative and qualitative techniques.
Example: "Quantitative" or "Qualitative" or "Mixed methods"

16. Extract the secondary dataset name if used, otherwise state "Not applicable".
Example: "Demographic and Health Surveys (DHS)" or "World Bank Enterprise Surveys" or "Not applicable"

17. Extract all primary data collection techniques used if primary data is collected, otherwise state "Not applicable". Separate techniques with a semicolon and provide justification with page numbers afterwards.
Example: "Structured interviews; Focus group discussions; Participant observation - as described on pages 23-25" or "Not applicable"

18. List all data analysis methods used in the paper separated by semicolons. Categorize them systematically and provide the text to justify this decision. Include:
- DESCRIPTIVE ANALYSIS: frequencies, means, medians, cross-tabulations, etc.
- STATISTICAL TESTS: t-tests, chi-square, ANOVA, non-parametric tests, etc.
- REGRESSION METHODS: OLS, logistic, multinomial, multilevel, fixed/random effects, etc.
- CAUSAL INFERENCE: IV, RDD, DID, matching, randomized experiments, etc.
- QUALITATIVE ANALYSIS: thematic analysis, content analysis, grounded theory, etc.
- ADVANCED METHODS: machine learning, structural equation modeling, meta-analysis, etc.

Example: "Descriptive: means and frequencies for demographic variables; Regression: multilevel logistic regression to account for clustering at school level; Causal: difference-in-differences design exploiting policy variation - as described in methodology section pages 18-22"

CRITICAL FOR EVIDENCE SYNTHESIS: Pay special attention to:
- Effect size calculations and confidence intervals
- Methods for handling missing data

```
- Robustness checks and sensitivity analyses
- Multiple comparison corrections
- Any meta-analytical techniques if present

Use your expertise to make informed judgments when information is not explicitly stated. If
truly uncertain, use "Not specified" rather than guessing.

Respond with ONLY a valid JSON object. DO NOT include any explanatory text, markdown
formatting, or content outside the JSON structure:

{{
    "authors": "string",
    "publication_year": "string",
    "first_author_country": "string",
    "first_author_organization": "string",
    "research_year": "string",
    "topic_summary": "string",
    "world_bank_sector": "string",
    "world_bank_subsector": "string",
    "sdg": "string",
    "study_countries": "string",
    "study_regions": "string",
    "population": "string",
    "sample_size": "string",
    "data_collection_method": "string",
    "analysis_type": "string",
    "secondary_dataset": "string",
    "primary_data_techniques": "string",
    "data_analysis_methods": "string"
}}

ACADEMIC PAPER TEXT:
{text_block}
"""
    return prompt


def truncate_text(text, max_chars=50000):
    """Truncate text if too long (GPT-4.1 Mini has 1M token context)"""
    return text[:max_chars] if len(text) > max_chars else text


# Cache and state management
CACHE_FILE = "gpt41_mini_cache.pkl"
STATE_FILE = "extraction_state.json"
RESULTS_FILE = "gpt41_mini_partial_results.pkl"
```

```
# Thread-safe cache operations
cache_lock = Lock()
state_lock = Lock()

def load_cache():
    """Load response cache thread-safely"""
    with cache_lock:
        if os.path.exists(CACHE_FILE):
            with open(CACHE_FILE, 'rb') as f:
                return pickle.load(f)
        return {}

def save_cache(cache):
    """Save response cache thread-safely"""
    with cache_lock:
        with open(CACHE_FILE, 'wb') as f:
            pickle.dump(cache, f)

def load_state():
    """Load processing state for resuming"""
    with state_lock:
        if os.path.exists(STATE_FILE):
            with open(STATE_FILE, 'r') as f:
                return json.load(f)
        return {"completed_ids": [], "start_time": None, "total_papers": 0}

def save_state(state):
    """Save processing state for resuming"""
    with state_lock:
        with open(STATE_FILE, 'w') as f:
            json.dump(state, f)

def load_partial_results():
    """Load partial results for resuming"""
    if os.path.exists(RESULTS_FILE):
        with open(RESULTS_FILE, 'rb') as f:
            return pickle.load(f)
    return []

def save_partial_results(results):
    """Save partial results for resuming"""
    with open(RESULTS_FILE, 'wb') as f:
        pickle.dump(results, f)

def create_cache_key(text):
    """Create cache key from text"""
    text_hash = hashlib.md5(text.encode()).hexdigest()
```

```
    return f"gpt41_mini_{text_hash}"


def call_gpt41_mini(prompt, cache=None, max_retries=3):
    """Call GPT-4.1 Mini API with structured JSON output and retry logic"""
    if cache is None:
        cache = load_cache()

    cache_key = create_cache_key(prompt)
    if cache_key in cache:
        return cache[cache_key]

    headers = {
        "Content-Type": "application/json",
        "Authorization": f"Bearer {os.environ.get('OPENAI_API_KEY')}"
    }

    data = {
        "model": "gpt-4.1-mini-2025-04-14",
        "messages": [{"role": "user", "content": prompt}],
        "max_tokens": 4000,
        "temperature": 0.1,
        "response_format": {"type": "json_object"}
    }

    for attempt in range(max_retries):
        try:
            response = requests.post(
                "https://api.openai.com/v1/chat/completions",
                headers=headers,
                json=data,
                timeout=60
            )

            if response.status_code == 429:
                wait_time = min(60 * (2 ** attempt), 300)  # Exponential backoff, max 5 min
                print(f"Rate limit hit, waiting {wait_time} seconds...")
                time.sleep(wait_time)
                continue

            response.raise_for_status()
            result = response.json()["choices"][0]["message"]["content"]

            # Cache successful result
            cache[cache_key] = result
            save_cache(cache)
            return result
```

```
        except Exception as e:
            if attempt == max_retries - 1:
                print(f"GPT-4.1 Mini API error after {max_retries} attempts: {str(e)}")
                return None
            else:
                wait_time = 10 * (attempt + 1)
                print(f"Attempt {attempt + 1} failed, retrying in {wait_time}s: {str(e)}")
                time.sleep(wait_time)

    return None

def load_data_from_csv(file_path, limit=None):
    """Load data from CSV file"""
    try:
        # Check if file exists
        if not os.path.exists(file_path):
            print(f"❌ File not found: {file_path}")
            return None

        print(f"📥 Loading CSV file: {file_path}")

        # Load CSV file
        df = pd.read_csv(file_path, encoding='utf-8')

        # Apply limit if specified
        if limit and limit < len(df):
            df = df.head(limit)
            print(f"🔢 Limited to first {limit} rows")

        print(f"\n✅ Successfully loaded {len(df)} rows from CSV")
        print(f"📋 Columns ({len(df.columns)}): {list(df.columns)}")

        # Show basic info about the data
        print(f"\n📊 Data Overview:")
        print(f"  • Total rows: {len(df)}")
        print(f"  • Total columns: {len(df.columns)}")

        # Check for 'full_text' column
        if 'full_text' in df.columns:
            non_null_texts = df['full_text'].notna().sum()
            avg_text_length = df['full_text'].str.len().mean()
            print(f"  • 'full_text' column found: {non_null_texts} non-null entries")
            print(f"  • Average text length: {avg_text_length:,.0f} characters")
        else:
            print(f"  ⚠️  'full_text' column not found. Available columns:")
```

```
        for col in df.columns:
            print(f"    - {col}")

        # Show first few rows (preview) - truncate long text for readability
        print(f"\n👀 Data Preview (first 3 rows):")
        preview_df = df.head(3).copy()

        # Truncate text columns for preview
        for col in preview_df.columns:
            if preview_df[col].dtype == 'object':
                preview_df[col] = preview_df[col].astype(str).str[:100] + '...'

        print(preview_df.to_string())

        return df

    except UnicodeDecodeError:
        print("⚠  UTF-8 encoding failed, trying with 'latin-1' encoding...")
        try:
            df = pd.read_csv(file_path, encoding='latin-1')
            if limit and limit < len(df):
                df = df.head(limit)
            print(f"✅ Successfully loaded {len(df)} rows with latin-1 encoding")
            return df
        except Exception as e:
            print(f"❌ Error with latin-1 encoding: {str(e)}")
            return None
    except Exception as e:
        print(f"❌ Error loading CSV file: {str(e)}")
        return None


def parse_json_response(response_text):
    """Parse JSON response from GPT-4.1 Mini"""
    if not response_text or pd.isna(response_text):
        return None

    try:
        # Clean the response
        cleaned_response = response_text.strip()
        if cleaned_response.startswith('```json'):
            cleaned_response = cleaned_response[7:]
        if cleaned_response.endswith('```'):
            cleaned_response = cleaned_response[:-3]
        cleaned_response = cleaned_response.strip()
```

```
        # Parse JSON
        parsed = json.loads(cleaned_response)
        return parsed
    except json.JSONDecodeError as e:
        print(f"JSON parsing error: {e}")
        return None


def process_single_paper(paper_data, cache):
    """Process a single paper - designed for parallel execution"""
    record_id, full_text = paper_data

    try:
        truncated_text = truncate_text(full_text)
        prompt = build_prompt(record_id, truncated_text)

        start_time = time.time()
        response = call_gpt41_mini(prompt, cache)
        processing_time = time.time() - start_time

        if response:
            parsed_data = parse_json_response(response)
            if parsed_data:
                result = {
                    'record_id': record_id,
                    'processing_time': processing_time,
                    'original_text_length': len(full_text),
                    'truncated_text_length': len(truncated_text),
                    'extraction_successful': True,
                    'error': None,
                    **parsed_data
                }
                print(f"✅ Successfully processed paper {record_id}")
                return result
            else:
                result = {
                    'record_id': record_id,
                    'processing_time': processing_time,
                    'original_text_length': len(full_text),
                    'truncated_text_length': len(truncated_text),
                    'extraction_successful': False,
                    'error': 'JSON parsing failed'
                }
                print(f"❌ JSON parsing failed for paper {record_id}")
                return result
        else:
            result = {
```

```
            'record_id': record_id,
            'processing_time': processing_time,
            'original_text_length': len(full_text),
            'truncated_text_length': len(truncated_text),
            'extraction_successful': False,
            'error': 'API call failed'
        }
        print(f"❌ API call failed for paper {record_id}")
        return result

    except Exception as e:
        result = {
            'record_id': record_id,
            'processing_time': None,
            'original_text_length': len(full_text) if full_text else 0,
            'truncated_text_length': 0,
            'extraction_successful': False,
            'error': str(e)
        }
        print(f"❌ Error processing paper {record_id}: {e}")
        return result


def extract_papers_parallel(df, n_observations=None, max_workers=5, resume=True):
    """Extract information from papers using parallel processing with resume capability"""

    # Prepare data
    if n_observations:
        sample_data = df.head(n_observations).copy()
    else:
        sample_data = df.copy()

    sample_data['record_id'] = sample_data['record_id']

    # Load state and partial results for resuming
    state = load_state()
    partial_results = load_partial_results() if resume else []

    # Determine which papers to process
    if resume and state["completed_ids"]:
        completed_ids = set(state["completed_ids"])
        remaining_data = sample_data[~sample_data['record_id'].isin(completed_ids)]
        print(f"\n🔄 RESUMING: {len(completed_ids)} papers already completed")
        print(f"📋 Processing remaining {len(remaining_data)} papers")
    else:
        remaining_data = sample_data
```

```
    completed_ids = set()
    partial_results = []
    print(f"\n🚀 STARTING: Processing {len(remaining_data)} papers")

  if remaining_data.empty:
    print("✅ All papers already processed!")
    return pd.DataFrame(partial_results)

  # Update state
  state["total_papers"] = len(sample_data)
  if not state["start_time"]:
    state["start_time"] = datetime.now().isoformat()
  save_state(state)

  # Prepare data for parallel processing
  paper_data = [(row['record_id'], row['full_text']) for _, row in remaining_data.iterrows()]
  cache = load_cache()

  results = partial_results.copy()
  successful_extractions = len([r for r in partial_results if r.get('extraction_successful',
False)])

  print(f"⚡ Using {max_workers} parallel workers")
  print(f"📊 Current success rate: {successful_extractions}/{len(partial_results)} papers")

  # Process papers in parallel
  with ThreadPoolExecutor(max_workers=max_workers) as executor:
    # Submit all tasks
    future_to_paper = {
      executor.submit(process_single_paper, paper, cache): paper[0]
      for paper in paper_data
    }

    # Process completed tasks
    for future in as_completed(future_to_paper):
      paper_id = future_to_paper[future]

      try:
        result = future.result()
        results.append(result)

        # Update state
        state["completed_ids"].append(paper_id)
        save_state(state)

        # Save partial results every 10 papers
```

```
                if len(results) % 10 == 0:
                    save_partial_results(results)
                    success_count = len([r for r in results if r.get('extraction_successful', False)])
                    print(f"📈 Progress: {len(results)}/{state['total_papers']} papers, {success_count/len(results)*100:.1f}% success rate")

                if result.get('extraction_successful', False):
                    successful_extractions += 1

            except Exception as e:
                print(f"❌ Error processing paper {paper_id}: {e}")
                # Add error result
                results.append({
                    'record_id': paper_id,
                    'processing_time': None,
                    'extraction_successful': False,
                    'error': f'Future execution error: {str(e)}'
                })

    # Final save
    save_partial_results(results)

    return pd.DataFrame(results)


def create_clean_dataset(results_df):
    """Create a clean structured dataset from successful extractions"""
    successful_results = results_df[results_df['extraction_successful'] == True].copy()

    expected_columns = [
        'record_id', 'authors', 'publication_year', 'first_author_country',
        'first_author_organization', 'research_year', 'topic_summary',
        'world_bank_sector', 'world_bank_subsector', 'sdg', 'study_countries',
        'study_regions', 'population', 'sample_size', 'data_collection_method',
        'analysis_type', 'secondary_dataset', 'primary_data_techniques',
        'data_analysis_methods'
    ]

    # Ensure all expected columns exist
    for col in expected_columns:
        if col not in successful_results.columns:
            successful_results[col] = "Not extracted"

    clean_data = successful_results[expected_columns].copy()
    return clean_data
```

```
def generate_summary_report(results_df):
    """Generate comprehensive summary statistics and quality report"""
    total_papers = len(results_df)
    successful = len(results_df[results_df['extraction_successful'] == True])
    failed = total_papers - successful

    if total_papers > 0:
        avg_processing_time = results_df['processing_time'].mean()
        total_cost_estimate = calculate_cost_estimate(results_df)

        print("\n" + "="*60)
        print("🎯 ACADEMIC PAPER EXTRACTION SUMMARY REPORT")
        print("="*60)
        print(f"📊 Total papers processed: {total_papers}")
        print(f"✅ Successful extractions: {successful} ({successful/total_papers*100:.1f}%)")
        print(f"❌ Failed extractions: {failed} ({failed/total_papers*100:.1f}%)")
        print(f"⏱️  Average processing time: {avg_processing_time:.2f} seconds per paper")
        print(f"💰 Estimated total cost: ${total_cost_estimate:.2f}")

        if successful > 0:
            clean_data = create_clean_dataset(results_df)
            print(f"📋 Clean structured records: {len(clean_data)}")

        if failed > 0:
            print(f"\n❌ Failure Analysis:")
            error_counts = results_df[results_df['extraction_successful'] ==
False]['error'].value_counts()
            for error, count in error_counts.items():
                print(f"   • {error}: {count} papers")

        print(f"\n📁 Output Files:")
        print(f"   • gpt41_mini_raw_results.csv (all results)")
        print(f"   • gpt41_mini_structured_data.csv (clean data)")
        print(f"   • extraction_state.json (resume state)")

        # Performance metrics
        if successful > 0:
            successful_df = results_df[results_df['extraction_successful'] == True]
            avg_chars_processed = successful_df['truncated_text_length'].mean()
            print(f"\n📈 Performance Metrics:")
            print(f"   • Average characters per paper: {avg_chars_processed:,.0f}")
            print(f"   • Papers per minute: {60/avg_processing_time:.1f}")
            print(f"   • Estimated time for 10,000 papers: {(10000 *
avg_processing_time)/3600:.1f} hours")
```

```
def calculate_cost_estimate(results_df):
    """Calculate estimated API costs based on actual usage"""
    total_input_chars = (results_df['truncated_text_length'].sum() +
                         len(results_df) * 2000)  # prompt overhead
    total_output_chars = len(results_df) * 800  # estimated JSON response

    input_tokens = total_input_chars / 4
    output_tokens = total_output_chars / 4

    # GPT-4.1 Mini pricing
    input_cost = (input_tokens / 1000000) * 0.40
    output_cost = (output_tokens / 1000000) * 1.60

    return input_cost + output_cost


def cleanup_files():
    """Clean up temporary files after successful completion"""
    files_to_remove = [STATE_FILE, RESULTS_FILE]
    for file_path in files_to_remove:
        if os.path.exists(file_path):
            os.remove(file_path)
            print(f"🧹 Cleaned up {file_path}")


# Main execution
if __name__ == "__main__":
    # Configure API key
    os.environ["OPENAI_API_KEY"] = "your-api-key"

    # CSV file configuration - UPDATE THIS PATH
    csv_file_path = "C:/Users/LucasSempe/OneDrive - International Initiative for Impact
Evaluation/Desktop/llama_extract/human_final_review.csv"  # Update with your CSV file
path

    # Processing configuration
    PARALLEL_WORKERS = 3  # Adjust based on your OpenAI rate limits
    RESUME_MODE = True    # Set to False to start fresh

    try:
        print("🔄 LOADING DATA...")
        df = load_data_from_csv(csv_file_path, limit=10000)
        if df is None or 'full_text' not in df.columns:
            raise ValueError("Invalid data - missing 'full_text' column")
```

```
        print(f"\n🎯 EXPERT ACADEMIC EXTRACTION SYSTEM")
        print(f"⚡ Parallel workers: {PARALLEL_WORKERS}")
        print(f"🔄 Resume mode: {'ON' if RESUME_MODE else 'OFF'}")

        # Run extraction with parallelization and resume capability
        results = extract_papers_parallel(
            df,
            n_observations=10000,  # Start with 20 papers for testing
            max_workers=PARALLEL_WORKERS,
            resume=RESUME_MODE
        )

        # Create clean structured dataset
        clean_data = create_clean_dataset(results)

        # Save final results
        results.to_csv("gpt41_mini_raw_results.csv", index=False)
        clean_data.to_csv("gpt41_mini_structured_data.csv", index=False)

        # Generate comprehensive report
        generate_summary_report(results)

        # Show sample of results
        if len(clean_data) > 0:
            print(f"\n📋 SAMPLE OF EXTRACTED DATA:")
            print(clean_data[['record_id', 'authors', 'publication_year',
'world_bank_sector']].head())

        # Clean up temporary files on successful completion
        cleanup_files()

        print(f"\n🎉 EXTRACTION COMPLETED SUCCESSFULLY!")

    except KeyboardInterrupt:
        print(f"\n⏸️ EXTRACTION PAUSED - You can resume later by running the script
again")
    except Exception as e:
        print(f"\n❌ ERROR: {str(e)}")
        print(f"💡 You can resume processing by running the script again")
```

Appendix D: Included studies

Table 26: References included in the study

Complete Bibliography Information Total references: 5327 Complete bibliography with all 5327 references has been written to: complete_bibliography.txt *This file contains the full APA-formatted bibliography for all references included in the systematic review.*

References

Bamberger, M., Rao, V., & Woolcock, M. (2010). *Using mixed methods in monitoring and evaluation: Experiences from international development*. World Bank.

Bamberger, M., Vaessen, J., & Raimondo, E. (2016). *Dealing with complexity in development evaluation: A practical approach*. SAGE.

Bank, W. (2020). World Bank Group Strategy for Fragility, Conflict, and Violence 2020–2025. *Washington, DC: World Bank*.

Bank, W. (2023). *Classification of fragility and conflict situations for world bank group engagement (FY24)*. World Bank. https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/fcs-list

Bertone, M. P., Jowett, M., Dale, E., & Witter, S. (2019). Health financing in fragile and conflict-affected settings: What do we know, seven years on? *Social Science & Medicine*, *232*, 209–219. https://doi.org/10.1016/j.socscimed.2019.04.019

Cambon, J., Hernangómez, D., Belanger, C., & Possenriede, D. (2021). Tidygeocoder: An r package for geocoding. *Journal of Open Source Software*, *6*(65), 3544. https://doi.org/10.21105/joss.03544

Campbell, S. P. (2017). Ethics of research in conflict environments. *Journal of Global Security Studies*, *2*(1), 89–101. https://academic.oup.com/jogss/article-abstract/2/1/89/2959877

Cohen, N., & Arieli, T. (2011). Field research in conflict environments: Methodological challenges and snowball sampling. *Journal of Peace Research*, *48*(4), 423–435. https://doi.org/10.1177/0022343311405698

Davies, S., Pettersson, T., Sollenberg, M., & Öberg, M. (2025). Organized violence 1989–2024, and the challenges of identifying civilian victims. *Journal of Peace Research*, *62*(4).

Department for International Development. (2010). *Working effectively in conflict-affected and fragile situations-summary note*. Department for International Development.

Djimeu, E. W. (2014). The impact of social action funds on child health in a conflict affected country: Evidence from angola. *Social Science & Medicine*, *106*, 35–42.

Duggan, C., & Bush, K. (2014). The Ethical Tipping Points of Evaluators in Conflict Zones. *American Journal of Evaluation*, *35*(4), 485–506. https://doi.org/10.1177/1098214014535658

Gaarder, M., & Annan, J. (2013). Impact evaluation of conflict prevention and peacebuilding interventions. *World Bank Policy Research Working Paper*, *6496*.

Gibson, D. G., Pereira, A., Farrenkopf, B. A., Labrique, A. B., Pariyo, G. W., & Hyder, A. A. (2017). Mobile phone surveys for collecting population-level estimates in low-and middle-income countries: A literature review. *Journal of Medical Internet Research*, *19*(5), e139.

Government Social Research. (2021). *Brief introduction to realist evaluation*. https://assets.publishing.service.gov.uk/media/60f7fdf7d3bf7f56824cc634/Brief_introduction_to_realist_evaluation.pdf

GOV.UK. (2024). *Foreign travel advice*. GOV.UK. https://www.gov.uk/foreign-travel-advice

Haer, R., & Becher, I. (2012). A methodological note on quantitative field research in conflict zones: Get your hands dirty. *International Journal of Social Research Methodology*, *15*(1), 1–13. https://doi.org/10.1080/13645579.2011.597654

Hoogeveen, J., & Pape, U. (2020). *Data collection in fragile states: Innovations from africa and beyond*. Springer Nature. https://library.oapen.org/handle/20.500.12657/22858

Jacobsen, K., & Landau, L. B. (2003). The dual imperative in refugee research: Some methodological and ethical considerations in social science research on forced migration. *Disasters*, *27*(3), 185–206. https://doi.org/10.1111/1467-7717.00228

Khan Mohmand, S., Justino, P., Dowd, C., Scott-Villiers, P., Befani, B., Loureiro, M., & Shaw, J. (2017). *Innovative methods for research on social and political action in fragile and conflict-affected settings*. https://opendocs.ids.ac.uk/articles/report/Innovative_Methods_for_Research_on_Social_and_Political_Action_in_Fragile_and_Conflict-Affected_Settings/26480020

Krause, J. (2021). The ethics of ethnographic methods in conflict zones. *Journal of Peace Research*, *58*(3), 329–341. https://doi.org/10.1177/0022343320971021

Lwamba, E., Shisler, S., Ridlehoover, W., Kupfer, M., Tshabalala, N., Nduku, P., Langer, L., Grant, S., Sonnenfeld, A., Anda, D., et al. (2022). Strengthening women's empowerment and gender equality in fragile contexts towards peaceful and inclusive societies: A systematic review and meta-analysis. *Campbell Systematic Reviews*, *18*(1), e1214.

Mansilla, C., Sweetman, A., Guyatt, G., & Lavis, J. N. (2024). A taxonomy of demand-driven questions for use by evidence producers, intermediaries and decision-makers: Results from a cross-sectional survey. *Health Research Policy and Systems*, *22*(1), 78. https://doi.org/10.1186/s12961-024-01160-4

Mazurana, D., Jacobsen, K., & Gale, L. A. (2013). *Research methods in conflict settings: A view from below*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=hT8oAAAAQBAJ&oi=fnd&pg=PR7&dq=Research+methods+in+conflict+settings:+A+view+from+below.&ots=2HsllKIIB_&sig=E9rRThOTN4LdvT4sVF-KGYFGWmc

Messner, J. J., Haken, N., Taft, P., Blyth, H., Maglo, M., Murphy, C., Quinn, A., Brisard, L., Cuthbert, L., Fitzgerald, D., Effange, H. N., Fidler, C., Fiertz, C., Hanlon, M., Hoye, K., Kelly, A., Rosenberg, L., Shelton, C., Silverman, S., … Young, A. (2017). *Fragile states index*.

Millar, G. (2018). Engaging ethnographic peace research: Exploring an approach. *International Peacekeeping*, *25*(5), 597–609. https://doi.org/10.1080/13533312.2018.1521700

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology & Community Health*, *57*(7), 527–529. https://jech.bmj.com/content/57/7/527.short

Ravat, Z., Engelbert, M., & Bell, A. B. (2025). *Interventions to improve resilience and food security in the middle east, sahel, and horn of africa regions: A rapid evidence gap map* (p. 30). International Initiative for Impact Evaluation.

Rustad, S. A. (2025). *Conflict Trends: A Global Overview, 1946–2024 – Peace Research Institute Oslo (PRIO)*. https://www.prio.org/publications/14453

Sonnenfeld, A., Doherty, J., Berretta, M., Shisler, S., Snilstveit, B., Eyers, J., Castaman, K., Gupta, R., Anda Leon, M., Franich, A., et al. (2021). Strengthening intergroup social cohesion in fragile situations. *3ie Systematic Review*, *46*.

Sonnenfeld, H. B., Ada; Chirgwin. (2020). *Building peaceful societies: An evidence gap map*.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. Department for International Development (DFID).

Stewart, F., & Brown, G. (2009). *Fragile states*. https://ora.ox.ac.uk/objects/uuid:fb74a3ff-c2fd-4302-82af-c9fb6cf98821

Thissen, P., & Ansari, S. (2024). *Impact evaluation of the UN peacebuilding fund's peace and coexistence project in east darfur*.

Vaessen, J., & Raimondo, E. (2019). Theories of evaluation and the realities of evaluation practice in development. In V. Grybaite & N. Stame (Eds.), *Evaluation for the sustainable development goals: A new approach* (pp. 45–66). Palgrave Macmillan.

White, H. (2013). An introduction to the use of randomized control trials to evaluate development interventions. *Journal of Development Effectiveness*, *5*(1), 30–49.

White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework*. International Initiative for Impact Evaluation (3ie). https://doi.org/10.23846/WP0015

Wood, E. J. (2006). The ethical challenges of field research in conflict zones. *Qualitative Sociology*, *29*(3), 373–386. https://doi.org/10.1007/s11133-006-9027-8

Woodward, A., Sheahan, K., Martineau, T., & Sondorp, E. (2017). Health systems research in fragile and conflict affected states: A qualitative study of associated challenges. *Health Research Policy and Systems*, *15*(1), 44. https://doi.org/10.1186/s12961-017-0204-x

World Bank Group. (2025). *Fragile and conflict-affected situations (FCS) list: FY25* [Technical Report]. World Bank. https://thedocs.worldbank.org/en/doc/5c7e4e268baaafa6ef38d924be9279be-0090082025/original/FCSListFY26.pdf