ABSTRACT

This study investigates the relationship between real-world events and conflict onset, focusing on the quantity and pattern of these events over time. As the generation and collection of data has seen a significant increase in recent years, understanding the potential of real-world event data is crucial aiding decision-making during conflict. The initial analysis reveals a substantial increase in the number of such events leading up to conflict, despite periodic fluctuations over the years. Moreover, the research identifies a greater occurrence of these events in the days preceding conflict compared to regular days. Importantly, the study also establishes a significant positive relationship between the interaction of population and GDP and the frequency of events both prior to conflict and during regular days. These findings suggest that shifts in population and GDP have a direct impact on the volume of real-world events. By shedding light on these patterns, this research can aid in the development of more informed decision-making strategies during conflict situations.

Keywords: data, conflict, event data, decision-making.

TABLE OF CONTENTS

1. INTRODUCTION

The exponential increase in data production in the world has been a notorious phenomenon that took place over the last century. One of the main reasons why data has received significant attention is because it continuously proves to be valuable in improving decision-making processes in nearly every field. In the military realm, the value of data is clear through the production and use of intelligence - the collection and analysis of data that is used to guide decision-making. Intelligence efforts have been present for military purposes since biblical times (Sheldon, 2007) and today it is reflected in the work of several government agencies and individual groups that dedicate their work solely to providing their side with intelligence that can be helpful before, during, and after conflict.

Data are the records of information regarding real-world phenomena. In military use, for example, they can be recorded information about a party getting hostile towards another or about an opponent's equipment. In order for the data to be useful for military decision-makers, real-world information needs to be recorded, that is, become data, so that these data can be collected and processed. Once this process is done, the data need to be transmitted to the decision-makers, which involves the interconnectivity of agents in the field and the speed with which they can transmit this information to the relevant parties. While a lot of attention has been paid to the amount of data produced and collected, less attention has been paid to the amount of human-created events that take place in reality that eventually become valuable data for military decision-makers.

McChrystal (2015) shows that the high interconnectedness of individuals and almost instantaneous communication led to a significant change in the decision-making structure in the field of conflict. That is because these changes lead to more complexity, and the hierarchical military structure is not optimized for this level of complexity. The General noticed that his unit had far more resources than Al-Qaeda, but that it was not reflected in the field. The superiority in technological means and data was not giving the United States the advantage that should be expected. This all changed when he switched the decision-making structure of his unit to one of decentralized decision-making, where more agents had access to information and the freedom to act, taking into consideration the goals of the entire organization and working in a team of teams to achieve the broader organization's goals. This is evidence that the events that take place in the real world along with their sharing speed and reach do affect the military decision-making process.

One important part of data as it relates to conflict is the volume of events that take place prior to conflict. Advancements in technology allowed the development of more methods and equipment to collect data, which added to the higher interconnectedness of the world and the speed with which data can be transmitted, led to a richness of instant data available to the parties involved in a conflict. Within the overall data collected for conflict, an extremely important category is the data that refers to human-created events that lead up to the conflict. These are events that take place, for example, when General X makes a remark about State Y, and then State Y moves troops to its border with State Z, and so on. These are two events that take place, which due to technological advancements, are now fastly recorded, become data, and are made available almost instantly to a higher number of decision-makers. While the amount of data produced has been a topic of interest, little is known about the side of data that regards the actual events that take place prior to conflict onset.

The events that lead to a conflict are different from the broader spectrum of data, for they are naturally limited. The number of human-created events that take place is bounded by the number of individuals involved. That happens because there is just so much a human being can do in a day. Data in general is almost boundless, for it is found anywhere in the environment. For example, an intelligence agency can collect a massive amount of data about a country's military arsenal, with their specifications recorded to a fine level of granularity; but there is only a limited number of events generated by this country that can take place. Therefore, it is not necessarily the case that the number of events taking place prior to conflict actually matches the trend of the amount of data being created and collected.

The number of events that take place prior to conflict onset does not compose the entire value that can be taken from data for the purposes of conflict decision-making, but it certainly plays an important role. The increase in the amount of data affects the structure of decision-making, and so does the number of events that take place prior to conflict. When parties decide on whether to engage in military action, a bulk of the data that is used is composed of the events that are leading up to the escalation and onset. More events taking place in the real world means more events that decision-makers need to take into account. This affects the structure of military decision-making. If the structure with which military leaders make decisions is changing, the structure with which they make decisions should change accordingly. The aim of this research is to contribute to the understanding of data as it relates to conflict decision-making by exploring the following research question:

How has the number of events that take place prior to conflict onset changed over time?

It is hoped that by answering this question, some light will be shed on the mechanism of data usefulness insofar as it regards decision-making in the context of conflict onset. This will bring the events that lead up to conflict to the discussion table as an important factor to be taken into consideration.

The secondary goal of this research is to contribute to the quantitative literature in the field by creating a multi-purpose dataset that incorporates the number of events taking place in the world into a conflict dataset. In looking for data to answer the research question that is the focus of this paper, both datasets for conflict and number of events were found, but none that combined both sources of data. This is a reflection of the lack of investigation regarding the role that the number of events plays in conflict, which required the creation of a completely new dataset that carefully joined the two areas of interest. This new dataset was created to answer the research question here posed, but it will also allow further research to be undertaken.

The remainder of this paper will address the research question through the use of the newly created dataset. It begins with a review of the relevant literature on the use of event data prior to conflict and the decision-making structure in the military, which leads to the outline of a theoretical framework and testable hypotheses regarding the trend of events prior to conflict onset. Then, the data collection process, creation of the new dataset and methods of analysis employed are explained in detail. Next, the results section contains the outcomes of the tests conducted. Finally, the paper discusses the conclusions from the findings and possible avenues for future research.

2. LITERATURE REVIEW

The literature review covers two sides of the relationship between event data and decision-making in conflict scenarios. First, it explores the use of event data prior to conflict through an overview of conflict early warning systems, which are extremely sophisticated and useful systems that profit from the use of event data prior to conflict. The purpose of this overview is to show how important event data are to conflict studies and practice, as well as to identify the gap that this research seeks to fill. Second, it explores the work of General McChrystal, who shows how the events related to conflict affect the decision-making structure of the military. In reviewing his work, it is possible to identify his arguments from in-field experience about the changing dynamics of data and the effect it has on the decision-making structure. It is also possible to identify the gap in his research that refers to the number of events that take place prior to conflict. Once the use of event data prior to conflict and the relevance to decision-making are elicited and the gap in the literature is identified, this section builds upon previous works to construct a theoretical framework that will be used to understand the effect of event data in the decision-making structure. Finally, the hypotheses that will be investigated to answer the research question are laid out.

2.1 The Use of Event Data Prior to Conflict

The use of events as data has been used extensively for forecasting, warning and preparing for possible armed and violent conflict in the short term. In fact, there have been considerable efforts to use events' data for immediate conflicts. This is reflected in the several early warning systems that have been developed around the world. These systems have benefitted from advancements in data collection and techniques of analysis to achieve a great level of sophistication (Cederman and Weidmann, 2017).

Early warning systems for conflict came to the discussion table with the Yom Kippur War in 1973 and the Falkland Islands in 1982 (Peck, 1998). The escalation in these conflicts was rather unanticipated and sparked a series of debates on the need for systems that could anticipate such events. For example, the Falkland Islands were not even on the map of the United Nations when the conflict began. However, the technology was not ready for systems that could track events all over the world, and reporting of events would lean mostly on the reporting of agents on the field. Then, the Rwandan Genocide of 1994 brought the topic of an integrated early warning system back to being a topic of interest.

As a result of the genocide and other armed conflicts that took place by the end of the Cold War, the first early warning systems for conflict started appearing. For example, Swisspeace launched FAST in 1998, which used event data to anticipate impending armed conflicts and political crises. FAST employs both qualitative and quantitative analysis. The data are collected by local information networks (LINs) in the countries covered by the project, and it is analyzed by experts. Then the "[a]utomated event-data analysis promotes timely evaluation of information provided by the LINs and is thus extremely important for FAST's early warning purpose" (FAST Early Warning Unit, 2002, p.3).

The advancement in technology of the 21st century has allowed for a shift from human data collection - as it used to be done by agents on the field - to automatic processes of data collection that extract event data from several sources, of which most notably are news articles. Today, early warning systems have become incredibly complex and are increasingly being employed to assist decision-makers.

In the United States, the most advanced of such systems is the Integrated Crisis Early Warning System (ICEWS). The system was created by the Defense Advanced Research Projects Agency (DARPA) and was further developed on a contract with Lockheed Martin Corporation, which oversees the project today (Meier 2010). The system has several parts to its process: iData collects the data, iTrace monitors it, iCast forecasts it and iSent shares it with other parties. The ICEWS data repository contains over 45 million stories worldwide. These are processed through deep and shallow parsing techniques to produce a set of events which are coded into 300 different types. For each event, a score of intensity from hostile to cooperative is attached, the main actor is identified (e.g. country, individual). Additionally, several other characteristics such as time index and other parties involved are included. The monitoring tool traces who is doing what to whom, when, where, and how. It monitors the entire globe through a series of text analysis techniques that automatically take in and process the news feeds. It then forecasts destabilizing events of interest for 167 countries in the world for the following 6-month period. The forecasts incorporate over 80 heterogeneous integrated models, leading to a weighted forecast that is superior to any of the individual forecasting methods. Finally, ICEWS shares data with intelligence analysts, including data sources, analytic trending, sentiment volumetrics, geographic mapping, network analysis and other useful visualizations.

In Africa, the most advanced early warning system is the ECOWAS Early Warning and Early Response Network (ECOWARN), which operates in West African countries. It is an observation and monitoring tool that is used for prevention and decision-making in the context of conflict (Sagna, 2009). The goal of this system is to improve the response to early warning recommendations, as well as to reduce the potential for violence that is triggered by national elections. ECOWARN is composed of sub-regional zone offices that assess the data collected. There is a clear focus on human security reflected in the analysis of political (e.g. human rights and democracy), economic (e.g. food shortages), social (e.g. unemployment), security (e.g. arms and civil-military relations), and environmental (e.g. drought and natural disasters) indicators daily (Ndinga-Muvumba and Lamin 2006). These are some examples of the 93 indicators taken into consideration to produce intelligence that is sent to leaders as an attempt to warn them of potential threats. The work is both quantitative and qualitative. The computer program collects and categorizes data from field reporting and news sources, and agents also enter and categorize data manually. Experts then produce reports and create contingency plans. According to Ndinga-Muvumba and Lamin, the system requires sustainable long-term funding to operate, as it counts on sophisticated technology and several agents in offices over different countries.

In Europe, the EU Conflict Early Warning System (EWS) has been developed to help decision-makers manage risk factors and prioritize resources (EU Conflict Early Warning System, 2014). The system employs both quantitative and qualitative measures. On the quantitative side, it scans for situations that are deteriorating or at high risk around the globe, combining them into an index created by the European Commission Joint Research Center. The qualitative input is then given by EU staff and expert analysts from member countries. The countries that ranked high for being at risk and could benefit from EU aid in analysis and action are identified. Then, EU staff in the field and headquarters analyze the selection, setting goals that serve as preparation for early preventive or peacebuilding actions. Finally, the actions are analyzed and fed into the following cycle of assessment and analysis. The system has produced positive results, receiving praise from senior decision-makers, officials, delegations, and even the Member States themselves.

ICEWS, ECOWARN and EWS are some examples of sophisticated conflict early warning systems, but several others with different degrees of complexity exist. These systems employ different methodologies and technology for data collection and analysis, they operate in different continents and countries, and they share intelligence with different classes of decision-makers for slightly different purposes. The one thing they share in common is that they all employ event data as the main source of information. This shows the power and importance of event data for conflict escalation and onset. However, what these systems don’t take into consideration is the number of events that take place prior to conflict. Their focus is on the content of the events, as their main goal is to foresee and therefore prepare recommendations prior to conflict onset. The number of events and its effect on decision-makers is still an unexplored topic.

2.2 Data and the Military Decision-Making Structure

General McChrystal provides a groundbreaking work in the literature on the relationship between the changing structure of data dissemination in conflict and the structure of military organizations (McChrystal, 2015). During his command of the JSOC (Joint Special Operations Command), McChrystal noticed that the incredibly superior technology and data of the United States in comparison to Al-Qaeda was not reflected in the battlefield. He argues that even though the world has much more data and advanced analytical methods today, it is also less predictable. He attributes this to the high interconnectedness of individuals and the almost instantaneous communication. According to him, this leads to complexity during conflict, which defeats the traditional hierarchical military structure.

McChrystal's work provides a basis for the framework employed in this research. At the basis of the decision-making structure in the military is the actual phenomena in the world - the events that take place in reality. These are the focus of this research. These events, when recorded, become data. These data can then be collected by agencies, groups and individuals to be turned into intelligence. On this side of the process, the level of interconnectedness and the speed of connectivity play a significant role in getting these data into the decision-making individuals. As McChrystal argues, in our current environment of high interconnectedness and instant connectivity, a hierarchical structure is not optimal, for there is so much data flowing that until it reaches the individuals working in the field, new data is already available and the old data is not as relevant anymore.

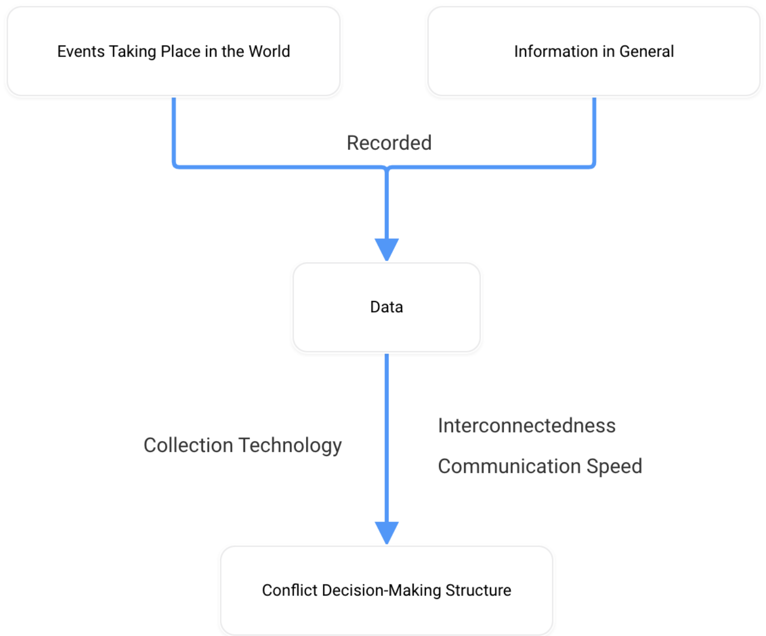
McChrystal's highlight of the new dynamics of interconnectedness and communication as a factor that highly affects the decision-making structure in the military is a significant contribution. In fact, it doesn't only apply to the military, and several companies have changed their structure to the one recommended by the General[1]. It is important to note that behind these changes in interconnectedness and communication are the actual events taking place in the world. Interconnectedness and communication are part of the channeling mechanism that disseminates these data, and the events partially compose the actual information that is captured as data. However, there is a gap in understanding how the trend of these events has taken place over time.

[1] McChrystal recommended shifting the organizational structure from a vertical hierarchy to a horizontal cooperation. Instead of having individuals at the top of the hierarchy getting access to the data and making decisions, he proposed to empower individuals on the field, giving them access to the data and allowing them to make decisions based on their experience and judgment. The rationale is that in the past, data would flow more slowly and the higher experience of leaders at the top of the hierarchy would make a significant difference in the decisions made. However, with so much information flowing, by the time that the data reaches the individual at the top of the hierarchy, the decision is no longer useful for the agent in the field, as new data has become available and the old data is no longer relevant.

2.3 Theoretical Framework

Data are extremely useful in the context of conflict. It has been shown that one important part of these data is event data. Event data refers to those events that take place in the world, the actions of individuals that affect the escalation and onset of the conflict. In today's world, more sophisticated technology to collect data added to the increased interconnectedness and instantaneous communication has provided military decision-makers with a massive amount of data to analyze and make decisions upon. The increased amount of data that reach decision-makers destabilizes the decision-making structure from the traditional hierarchies. These data that reach decision makers can be general data or events data. An increase in events data can be caused both due to an increase in methods to record it, the higher interconnectedness and instantaneous communication - but it can also be because there are more events taking place in the world. Thus, the framework developed in this section states that data can be divided into two parts: general data and event data. These compose data, which are then channeled to the decision-making structure. The channeling can be affected by more advanced methods of data collection, higher interconnectedness and instantaneous communication. A different amount of data flowing into decision-making structures then affects this structure, and it requires new organizational models to optimize the new pattern of data flow. An illustration of this framework is demonstrated in Figure 1.

Figure 1. An Illustration of the Theoretical Framework



2.4 Hypotheses

At the root of the intelligence that reaches decision-makers are the actual events taking place in the world. So far, there is no research that quantitatively explores the number of events taking place in the context of conflict. This is what this paper aims to do. The relevance lies in that, by better understanding the amount of information that takes place before conflict, it is possible to better optimize decision-making structures. The first hypothesis is based on the increased capability of recording events and the higher number of individuals involved in the military machinery, which is assumed to produce more events. Additionally, the onset of conflict is naturally associated with a chain of events that take place. Ceteris paribus, there should be more events in total prior to conflict onset, as there is more action than if there were no conflict escalation leading up to onset. Therefore, the first hypothesis is formulated as follows:

H1a: The average number of events prior to conflict has increased significantly over the years.

H1b: The average number of events prior to conflict onset is higher than the average number of overall events.

The first hypothesis explores the number of events that take place prior to conflict onset over time. However, it is equally important to know what factors are associated with this change in the number of events. As mentioned, the number of events that take place prior to conflict onset is expected to be a product of the amount of individuals involved in the conflict. Therefore, the number of events should depend on both population and GDP. The absolute population and absolute GDP might have an effect on the number of events, but the interaction between both is expected to be highly significant. That is because a higher population without the resources to make the events known does not increase the number of events that are recorded, and a high GDP without the population to produce these events also does not increase the number of events recorded. However, a higher population together with a higher GDP is expected to significantly increase the number of events recorded. Therefore, the second hypothesis states:

H2: The number of events prior to conflict and the overall number of events are positively associated with the interaction between population and GDP.

3. METHODS

This section explores the concepts relevant to this research by clearly defining them, describing how they were operationalized, and how the respective variables from distinct sources were merged together into a final dataset upon which subsequent analysis relies. Then, this section discusses the plan of analysis for the two hypotheses that this research seeks to test. Finally, it discusses the limitations of the methods employed.

3.1. Concept Definitions and Operationalization

The relevant concepts explored in this section are event data, armed conflict, and conflict onset. The following subsections provide a clear definition for each concept and explain the chosen way to operationalize the concept in a dataset variable, showing how the data were collected into a single final dataset.

Event Data

An event is a physical activity that takes place over the world. It takes the form of subject-action in time and space, in which a party engages in physical activity at some point in time at some geographical location. For example, information regarding a country's military arsenal is not an event - although it is data - but an agency finding information regarding a country's military arsenal is. It involves the subject (the agency) engaging in an action (finding information) and it necessarily involves some point in time and a geographical location. Data can be, therefore, subsetted to include the category of event data, which refers to those events that take place in the real world and are recorded to become data.

The GDELT project database constitutes an ideal source of data to operationalize events. The project monitors broadcast, print, and web news around the world, reaching almost every corner of every country. It processes data in over 100 languages and is able to track the point in time and the geographical location of each event. The database tracks events from 1979 to the present, updating every 15 minutes. The languages are translated into English, and it has the capacity to translate 98.4% of the data that is not in English. To exemplify one news source that is turned into events, the database "[...] takes a sentence like "*The United States criticized Russia yesterday for deploying its troops in Crimea, in which a recent clash with its soldiers left 10 civilians injured*" and transforms this blurb of unstructured text into three structured database entries, recording *US CRITICIZES RUSSIA*, *RUSSIA TROOP-DEPLOY UKRAINE (CRIMEA)*, and *RUSSIA MATERIAL-CONFLICT CIVILIANS (CRIMEA)*" (The GDELT Project, 2022). In this way, it records three events and identifies the point in time and geographical location.

For this research, the data was retrieved from a BigQuery search that automatically aggregated the number of events per day per country. The choice to aggregate it at the daily and country levels was made so that the data could be later merged with the conflicts dataset, as a day-country level was the finest level of granularity to which both datasets could be brought together. Similarly, in order to match the conflict data, the time span chosen was 1989-2020, which is the years that the conflict dataset covers entirely. In this way, the data collected represented the sum of daily events for each country in the world for every day from 1989 to 2020.

Armed Conflict

The definition of armed conflict is not universally agreed upon. Definitions disagree on whether a government needs to be one of the actors in the dyad and on a minimum number of casualties or another threshold so that it can constitute a conflict. There is a fine line between a full-blown conflict and a riot that went wrong, but it is not evident where to draw the line. For operationalization, this research employs the UCDP definition of armed conflict as: "[...] a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25-battle-related deaths in one calendar year" (UCDP, 2022).

The data for armed conflict was retrieved from the UCDP website. The dataset chosen was the Georeferenced Event Dataset (GED)[2] which contains events of violence that took place as part of an armed conflict. Consistent with the above-mentioned definition, for a conflict to be included in the dataset, it needs to meet the 25-battle-related deaths in a year criteria, and one of the parties needs to be the government of a state. This dataset is disaggregated at the event level, in which every event that is part of the conflict is included. With all the events for a conflict, it was possible to detect the day of conflict onset.

Conflict Onset

Conflict onset is the period of time in which an armed conflict begins. The purpose of identifying a period of conflict onset for this research is to differentiate the global events that take place during conflict onset from those that do not. Therefore, the operationalization of conflict onset took into consideration the 7-day period prior to the first conflict-related event for each unique conflict, inclusive. It might be argued that conflict onset should be the day of conflict onset, the day in which the conflict officially started. Since this research investigates the number of events prior to conflict, taking into consideration only the day of conflict onset would not be representative enough of the events that took place during the period of conflict onset. The 7-day period, which represents a week prior to the conflict, is long enough to detect changes in events that were leading up to the onset, yet it is short enough so that it does not include many events from before the conflict onset. For this reason, the 7-day period was chosen so that it incorporates the heavy escalation period together with the day of onset.

The UCDP dataset contains a unique conflict identifier variable for each conflict related event. The variable conflict onset was created by detecting the first date of each conflict-related event for each unique conflict as the final day of onset and subtracted 7 days from that date to mark the beginning of the period.

[2] The global version 22.1 was used.

Final Datasets

Two datasets were created from two different sources. The first contains the sum of daily global events for each country for each day from 1989 to 2020, and it was retrieved from the GDELT project. The second included all unique conflicts from 1989 to 2020 with the marked date of conflict onset, which represents the end of the period of conflict onset, and the previous 7-day date, which marks the beginning of the conflict onset period. The two datasets were combined to create two datasets that were used for analysis, which differ mainly in their aggregation level and additional country characteristics detail.

The first final dataset is at the day-country level. It was put together by merging GDELT dataset containing the countries, dates at the daily level, and number of events on that day, with the aggregated version of the UCDP dataset, containing countries, dates at the daily level, and a dummy variable for whether a specific day in a specific country marked a period of conflict onset or not. In this way, the first final dataset contained the variable date, which shows what date of the year it was at the daily level, the variable country, which indicates the country, the variable events, which shows how many events took place in that specific country on that specific day, and the variable conflict onset, which indicates whether that day for that country was a period of conflict onset. Table 1 shows the summary of variables and data sources for the first final dataset.

Table 1. Summary of Variables and Data Sources for Dataset 1.

Variable	Description	Source
Date	The date of the observation from 1989-2020. It takes the format day-month-year, with a minimum value of 01-01-1989 and a maximum value of 31-12-2020.	GDELT
Country	The country of the observation. It includes all the countries in the world.	GDELT
Number of Events	The number of events that took place on the respective date and country.	GDELT
Conflict Onset	A dummy indicating whether it is a day of conflict onset (7-day period) for the respective date and country. It takes value 1 if it is a day of conflict onset and value 0 otherwise.	UCDP GED

The second final dataset was created from the first final dataset. First, it was aggregated to the year-country level, creating the variable events representing the average daily events in that country in that year and another variable conflict onset events containing the average daily events for the days of conflict onset only. Then, it was merged with population and GDP per country per year from the World Bank. Finally, the region of the countries was added to the main dataset. The second final dataset is shown in Table 2.

Table 2. Summary of Variables and Data Sources for Dataset 2.

Variable	Description	Source
Year	The year of the observation from 1989-2020.	GDELT
Country	The country of the observation. It includes all countries in the world.	GDELT
Events	The average number of daily events for the respective year and country.	GDELT
Conflict Onset Events	The average number of daily events in the week prior to conflict onset for the respective year and country.	GDELT and UCDP GED
Population	The population for the respective year and country.	World Bank
GDP	The GDP in US dollars for the respective year and country.	World Bank
Region	The region of the respective country.	UCDP GED

3.2 Analysis Plan

The two datasets were employed to test for the hypotheses. The first dataset was employed to test for the first hypothesis and the second dataset was employed to test for the second hypothesis. The first hypothesis was divided into two parts. H1a is concerned with the number of events prior to conflict over time, hypothesizing that the average number of events prior to conflict has increased significantly over the years. In order to test this hypothesis, a line plot will be created to visualize the average number of daily events over the years from 1989 to 2020. Then, Pearson's correlation coefficient will be calculated to determine whether there is a linear relationship between the average number of events and the year. If there is a linear relationship, positive or negative, a simple linear regression model will be employed to determine whether this relationship is statistically significant.

The second part of hypothesis one is concerned with the comparison between the average number of events prior to conflict onset and the overall number of events. H1b states that the average number of events prior to conflict onset is higher than the average number of overall events. In order to test this hypothesis, a t-test will be conducted comparing the means of daily events for days that fall in the period of conflict onset and the means of daily events for all days. It is expected that the mean of daily events for days that fall in the period of conflict onset is statistically higher than the mean of daily events for all days.

The second hypothesis explores the effect of the interaction between population and GDP on the number of events. It states that the number of events prior to conflict and the overall number of events is positively associated with the interaction between population and GDP. In order to test this hypothesis, a generalized linear random effects model was chosen. Due to the nature of the panel data, the generalized linear model family was chosen, and the random effects model was deemed appropriate as it allows for the differences between countries and years to be modeled as a random effect. There were two variations of the model, which differed only on the dependent variable. The first sought to explore the relationship between the number of events prior to conflict and the interaction between population and GDP. In this model, the number of events prior to conflict was the dependent variable. The second variation sought to explore the relationship between the number of events for all days and the interaction between population and GDP. The number of events for all days was the dependent variable in this model. For both models, the region of the country was also added for control. The data manipulation and analysis were conducted using R and the visualizations were produced using Eisengard AI.

3.3 Limitations

The main limitation of this research is the fact that the number of events recorded over time has increased partially due to an increase in the capabilities of data collection. The first hypothesis is time-dependent, as it seeks to analyze the trend of the number of events prior to conflict over time. The problem is that the number of events might change due to a change in the collection of data that was capacitated by a change in collection technology, not because of an actual change in the number of events. However, it is expected that events will not increase year after year in a predictable manner as the increase of data collection. The amount of data produced and collected increases yearly in a steady line. It is not expected that the same will happen with the number of events in the world, as they are expected to vary significantly between years - even though in an upward trend in 2020 when compared to 1989. If this is confirmed, then it is evidence that the real effect of the change in events taking place in the world is captured in the data, and not only the increase in data collected.

4. RESULTS

The first hypothesis is divided into two formulations. Hypothesis 1a states that the average number of events prior to conflict onset has increased over the years. Every year from 1989-2020 had at least 10 conflict onsets (2005) to a maximum of 68 (1989). Regarding the events at the country level, the minimum average daily number of events during the week prior to conflict onset is 1 (Nepal 1989; Niger 1990; Eritrea 1993) and the maximum is 8532.9 (Libya 2011). To test H1a, the analysis must be conducted at the global level. The minimum average daily number of events during the week prior to conflict onset is 22.3 (1992) and the maximum is 1518.2 (2015).

The number of events between 1989 and 1998 remained steady. Starting in 1999, the events increased significantly, falling in 2000 to reach a peak in 2003. It is by this time that General McChrystal identified the phenomenon of high interconnectedness at instantaneous speed, resulting in intelligence being quickly disseminated to more individuals. The data show that there are in fact more events taking place at this time. However, the number of events fell in 2004, and remained steady for a couple of years, increasing significantly in 2007 to a peak in 2009, when it fell again and increased in 2011. It fell again in 2012 and 2013 and then increased in 2014. The average number of daily events reached its peak in 2015 at more than 50 times the levels prior to 1999. It fell from 2016 to 2019, rising one last time in 2020 to a similar level to that found in 2012. It is clear that the average number of daily events has increased significantly in the past 3 decades, and it has seen a lot of volatility during this period. Additionally, the trend shows a significant variation both upwards and downwards from year to year, which is evidence that the effect captured cannot be solely attributed to an increased capacity of data collection. Figure 2 shows this trend.

Figure 2. Average Daily Number of Conflict Onset Events from 1989-2020.

The Pearson's correlation coefficient between the average number of daily events during the onset period and year is 0.44, demonstrating that there is a moderate positive association between the average number of daily events during conflict onset and year. To test this relationship statistically, a simple linear regression model was used to regress the average number of daily events during conflict onset on year. The coefficient of year is 36.2, which shows that on average, a one-year increase leads to an increase in the average number of daily events during conflict onset by 36.2 events. The coefficient is statistically significant with a t-statistic of 13.44 and a p-level < 0.0001. Therefore, it is possible to reject the null hypothesis that there is no relationship between the year and the average number of daily events during conflict onset.

To test H1b, a t-test was conducted. The t-test compared the means of daily events on the week prior to onset and all weeks, including all countries in the world in the measure of daily events for all weeks. The average number of events in the week prior to conflict onset is 336.8 daily events, and the average number of events including all days of the year for all countries is 226.6. The t-statistic is 3.06 and the p-value is less than 0.01. Thus, it is possible to reject the null hypothesis, concluding that there is evidence that the average daily events during the week of conflict onset for countries that faced conflict were higher than the average daily events for all days and all countries. This leads to the conclusion that there are more events prior to conflict onset than in general.

The second hypothesis states that the interaction between population and GDP is positively associated with the number of events prior to conflict. To test this hypothesis, a generalized linear random effects model was employed. The model took into consideration the year and the country, and regressed the log of daily average events prior to conflict on the log of population, the log of GDP, the interaction of the log of population and log of GDP, and region. The results show a significant negative relationship between the log of population and the log of GDP and the average number of events prior to conflict, and a significant positive relationship between the interaction of logged population and logged GDP and the average number of daily events prior to conflict. There were 5 regions: Africa, the Americas, Asia, Europe, and the Middle East. Relative to Africa, the Americas, Asia, and Europe were significantly lower, and the Middle East was not statistically different. The results are shown in Table 3.

Table 3. Summary of Results for Conflict Onset Period Model

	Estimate	Standard Error	P-value
Intercept	30.01	10.15	< 0.001 **
Log Population	-2.61	0.6	< 0.0001 ***
Log GDP	-1.24	0.44	< 0.001 **
Log Population x Log GDP	0.12	0.03	< 0.0001 ***
The Americas	-1.45	0.38	< 0.0001 ***
Asia	-0.99	0.38	< 0.001 **
Europe	-1.69	0.41	< 0.0001 ***
Middle East	0.28	0.51	0.59

Significance levels indicated by: + p < .1, * p < .01, ** p< .001, *** p < .0001

Alternatively, when considering the number of daily events for all countries and all years, the log of population has a significant negative effect on the average number of daily events, but the log of GDP does not have a significant effect on it. The interaction of the log of population and the log of GDP has a positive significant effect, although a weaker one in magnitude than for the conflict onset events. Relative to Africa, the Americas, Asia, Europe and the Middle East all have a significantly lower average number of daily events. The results are shown in Table 4.

Table 4. Summary of Results for Conflict Onset Period Model

	Estimate	Standard Error	P-value
Intercept	9.19	3.64	0.011 **
Log Population	-1.89	0.22	< 0.0001 ***
Log GDP	-0.12	0.15	0.43
Log Population x Log GDP	0.075	0.009	< 0.0001 ***
The Americas	-1.91	0.21	< 0.0001 ***

Asia	-0.88	0.22	< 0.0001 ***
Europe	-1.98	0.22	< 0.0001 ***
Middle East	-0.86	0.29	< 0.001 **

Significance levels indicated by: + p < .1, * p < .01, ** p< .001, *** p < .0001

5. DISCUSSION AND CONCLUSION

The results of the tests are supportive of the hypothesis stated in this research. The number of events has been shown to be positively associated with years through a Pearson correlation coefficient and a simple linear regression model, which demonstrated a positive statistically significant relationship. This means that there is evidence that the number of events has been increasing over time, which is consistent with H1a. Additionally, it has been shown that the number of events does not increase in a consistent trend like data collection, leading us to conclude that there is in fact a mechanism leading to the number of events that take place in the real world and that the increase in events is not solely attributed to more events being captured due to the higher capabilities of data collection.

A t-test has shown that the mean number of events for conflict onset days is statistically significantly higher than the mean number of events for all days. This is evidence that conflict onset periods are associated with a higher number of events taking place in the world than average, which is consistent with H1b.

Generalized linear random effects models have been employed to test for the effect of the interaction between logged population and GDP on the number of events during conflict onset and on the number of events in general, respectively. The first model showed a negative statistically significant relationship between the log of population and the log of GDP and the log of the average number of daily events during conflict onset. This means that when population and GDP do not change together, the effect on the number of events for both variables is negative. However, the interaction between the log of population and the log of GDP was positive and statistically significant, which means that when both variables change together, their effect is positive on the number of events during conflict onset. When GDP increases, the effect of population on the number of events increases, and vice-versa. Therefore, GDP and population only impact the number of events positively when they go together. The effect of the interaction is consistent with hypothesis 2. Although the effect of population and GDP independently was not hypothesized, it is a significant finding that population and GDP actually decrease the number of events when they do not go together. There was also a regional factor in place. When compared to Africa, the Americas, Asia and Europe counted with less number of events during conflict onset. All were statistically significant. Alternatively, the Middle East counted with more events but was not statistically different from Africa.

The results were slightly different for the second model, which used events for all days as the dependent variable. In this model, population by itself was also significantly negatively associated with the number of events, but GDP was not. GDP by itself has no statistical relationship with the effect on the number of events. Consistent with hypothesis 2 and the first model, the interaction between the log of population and the log of GDP had a significant positive effect on the log of number of events. These three findings show that population has a significant negative effect on the number of events when by itself, GDP has a non-significant effect on the number of events when by itself, and population and GDP have a positive significant effect on the number of events when they change together. The regional effect in this model was also slightly different from the first model: all regions count with fewer events than Africa, and these results are all statistically significant.

This research has explored the trend in the number of events prior to conflict. It has found that the number of events prior to conflict has increased over time in a generally positive trend but with ebbs and flows in recent years and it has shown that the number of events during conflict onset is higher than average. It has also shown that the interaction between population and GDP has a significant positive effect on the number of events both during conflict and in general, as well as explored the effect of population and GDP individually and geographical region on the number of events. It has contributed to the literature by bringing the number of events to the discussion table as an important part of data for conflict that should be studied further. It has also contributed to the literature by creating the first dataset that includes variables for conflict and number of events, which allows researchers to get readily available data to test for additional hypotheses.

While this research sought to investigate the trend in the number of events prior to conflict, the effect of the number of events on the actual conflict remains unexplored. As has been argued based on the work of General McChrystal, the amount of data affects the decision-making structure in the military - and event data is a significant part of data insofar as it relates to conflict. The literature would benefit from an exploration of the effect that the amount of data has on the decision-making structure in the context of conflict. Just as General McChrystal changed the decision-making structure of his unit, much more can be learned from exploring how the decision-making structures of sides from different conflicts have changed with the change in the trend of the number of events that take place. Additionally, more can be said about the relationship between the number of events, the type of decision-making structure and the outcome of the conflict. Because certain decision-making structures are optimized for a low number of events and can become destabilized when faced with a high number of events, this could affect the outcome of the conflict. Further, it is important to consider the number of events when predicting the course of conflict. This new metric can have a relationship with factors of the course of conflict, such as duration and casualties. It has been shown that the number of events prior to conflict is an important part of conflict data and that this data can be used with relevance in the context of conflict. It is hoped that further investigations will shed more light on the effect that real-world events have on conflict.

REFERENCES

Allen, P. T. "Predicting Levels of Conflict." Employee Relations 6(5) (1984): 25–28. https://doi.org/10.1108/eb055041.

Cederman, Lars-Erik, and Nils B. Weidmann. "Predicting Armed Conflict: Time to Adjust Our Expectations?" Science 355(6324) (2017): 474–76. https://doi.org/10.1126/science.aal4483.

Celiku, Bledi, and Aart Kraay. "Predicting Conflict," 2017. https://doi.org/10.1596/1813-9450-8075.

Davies, Shawn, Therése Pettersson, and Magnus Öberg. "Organized Violence 1989–2021 and Drone Warfare." Journal of Peace Research 59(4) (2022): 593–610. https://doi.org/10.1177/00223433221108428.

de Bruin, Sophie, Jannis Hoch, Nina von Uexkull, Halvard Buhaug, Jolle Demmers, Hans Visser, and Niko Wanders. "Projecting Long-Term Armed Conflict Risk: An Underappreciated Field of Inquiry?," 2022. https://doi.org/10.31223/x5hs5w.

Elfversson, Emma. "Cities and Armed Conflict: A Systematic Urban-Rural Coding of UCDP Conflict Events Data." Data in Brief 39 (2021): 107554. https://doi.org/10.1016/j.dib.2021.107554.

EU Conflict Early Warning System. European Union External Action, 2014.

"The GDELT Project." GDELT. Accessed May 12, 2022. https://www.gdeltproject.org/.

"GDP (Current US$)." Data. Accessed June 22, 2022. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.

Hegre, Håvard, and Nicholas Sambanis. "Sensitivity Analysis of Empirical Results on Civil War Onset." Journal of Conflict Resolution 50(4) (2006): 508–35. https://doi.org/10.1177/0022002706289303.

Hegre, Håvard, Håvard Mokleiv Nygård, and Peder Landsverk. "Can We Predict Armed Conflict? How the First 9 Years of Published Forecasts Stand up to Reality." International Studies Quarterly 65(3) (2021): 660–68. https://doi.org/10.1093/isq/sqaa094.

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal. "Predicting Armed Conflict, 2010-2050." International Studies Quarterly 57(2) (2012): 250–70. https://doi.org/10.1111/isqu.12007.

"ICEWS." Lockheed Martin, January 15, 2020. https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html.

Krummenacher, Heinz, Daniel Schwarz, and Matthias Siegfried. Information for Early Warning - The Use of Information Networks. Bern: FAST Early Warning Unit, 2002.

McChrystal, Stanley, Tantum Collins, David Silverman, Chris Fussell, Minami Yoshikawa, Chizuko Amacho, and Yoshihiko Takatori. Team of Teams. New York, NY: Penguin Publishing House, 2015.

Meier, Patrick. "DARPA's Crisis Early Warning and Decision Support System." Conflict Early Warning and Early Response (blog), March 10, 2010. https://earlywarning.wordpress.com/2010/03/20/early-warning-decision-support/.

O'Brien, Sean P. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." International Studies Review 12(1) (2010): 87–104. https://doi.org/10.1111/j.1468-2486.2009.00914.x.

Peck, Connie. Sustainable Peace: The Role of the UN and Regional Organizations in Preventing Conflict. Lanham, MD: Rowman & Littlefield, 1998.

"Population, Total." Data. Accessed June 22, 2022. https://data.worldbank.org/indicator/SP.POP.TOTL.

Reed, William. "A Unified Statistical Model of Conflict Onset and Escalation." American Journal of Political Science 44(1) (2000): 84. https://doi.org/10.2307/2669294.

Sagna, Augustin. The ECOWAS Early Warning and Response Network. Other. OECD, May 2009. https://www.oecd.org/swac/theecowasearlywarningandresponsenetwork.htm.

Sheldon, Rose Mary. Spies of the Bible: Espionage in Israel from the Exodus to the Bar Kokhba. London: Greenhill, 2007.

Sundberg, Ralph, and Erik Melander. "Introducing the UCDP Georeferenced Event Dataset." Journal of Peace Research 50(4) (2013): 523–32. https://doi.org/10.1177/0022343313484347.

"UCDP Definitions." Department of Peace and Conflict Research. Accessed July 22, 2022. https://www.pcr.uu.se/research/ucdp/definitions/.

USAID Reacting to Early Warning and Response Data in West Africa PROGRAM (REWARD). USAID, 2009.

Wang, Yu. “Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data: A Comment.” Political Analysis 27(1) (2018): 107–10. https://doi.org/10.1017/pan.2018.40.

Weidmann, Nils B., and Michael D. Ward. “Predicting Conflict in Space and Time.” Journal of Conflict Resolution 54(6) (2010): 883–901. https://doi.org/10.1177/0022002710371669.