Causes of Ceasefire Failure

A survival analysis, 1989-2017

Tora Sagård

Master thesis in Political Science

Department of political science

UNIVERSITY OF OSLO

Spring 2019

Word Count: 32468

Causes of Ceasefire Failure

A survival analysis, 1989-2017

Tora Sagård

Spring 2019

2019

Causes of Ceasefire Failure

Word Count: 32468

Tora Sagård

http://www.duo.uio.no

Print: Peace Research Institute Oslo

Abstract

Why do some ceasefires last several years before failure, while others collapse within days? Previous research on ceasefires have not addressed ceasefire durability directly, and those that have, have conflated ceasefires with peace. They have thus overlooked the fact that most ceasefires are preliminary, declared during a peace process. Similarly, the research has only focused on ceasefires that are agreed upon beforehand and have overlooked the fact that most ceasefires are unilaterally declared. To address these knowledge gaps, this thesis explores ceasefire durability by looking at all ceasefires related to a peace process. The thesis argues there are especially three conditions that can make some ceasefires fail faster than others: (1) a history of failed ceasefires, (2) the comprehensiveness of the ceasefire, and (3) the parties to the ceasefire. The thesis investigates whether these conditions affect ceasefire durability by using a new dataset from ETH-PRIO that comprises all ceasefires found in civil conflicts between 1989 and 2017.

Several survival analyses were conducted to test these conditions empirically. The results show that both the history of failed ceasefires and the comprehensiveness of ceasefire have an impact on ceasefire durability. There was no effect of multiple rebel groups on ceasefire duration. This suggests that ceasefires are more durable in conflicts that have not experienced prior ceasefire failure and that ceasefires are more durable if they are more comprehensive. These findings suggest that ceasefires are not always the best option when negotiation peace, and if adversaries decide to declare a ceasefire, they should include some mechanisms to ensure its durability.

Acknowledgements

First and foremost, I would like to thank my two wonderful supervisors, Siri Aas Rustad and Scott Gates– without their help, this thesis would not be possible. Their commitment to my thesis, their feedback, and all the hours they have put into reading my drafts have been extremely valuable and highly appreciated. I would also like to thank Siri Aas Rustad, Håvard Mokleiv Nygård and Håvard Strand for hiring me to be their research assistant on the Ceasefire Project. Working on the project alongside writing this thesis has given many important insights, both as I was able to use the dataset I am part of coding, but also because I have learned so much about ceasefires in the process. I also want to thank PRIO for taking me on as a master student. It has been very fruitful to sit in an environment full of knowledgeable, kind and helpful researchers. I would also like to thank Jonas Vesty and Christoph Dworschak for helping me out with R, the statistical software used in this thesis, and Julia Palik and Sverke Saxegaard for valuable input. To my friends and family, thank you for putting up with me the last weeks before hand-in, I know I can be a handful sometimes. All remaining errors are mine alone.

Tora Sagård

Oslo, 23.05.2019

As the data used for analysis is not open for the public yet, replication data for the analysis and the do-files from R can be provided upon request. Contact: torasagard@gmail.com

Table of Contents

List of Figures

List of Tables

Introduction

Why do some ceasefires last so much longer than others? In El Salvador in 1989, a ceasefire lasted only one day before an attack killed 127 people. Similarly, an Afghani ceasefire in 1994 lasted two days before 200 people were killed. Other ceasefires manage to last months or even years before they are violated. This was the case in the Central African Republic in 2008, where the ceasefire lasted more than four years before the conflicting parties decided to take up their arms (Clayton, Rustad, Nygård & Strand, 2019b) . There is a common feature of all the mentioned ceasefires: they have all been declared during a peace process. Indeed, ceasefires are found in nearly all peace processes and are frequently used by adversaries as a tool to show their commitment to finding a peaceful settlement. Nonetheless, as the first two examples show, many ceasefires fail almost immediately after they have been declared. These failed ceasefires, in turn, impact the peace process, yet we know little about the effects. The literature on ceasefires has neglected the effect of failed ceasefires on a peace process. Moreover, the few studies that do exist have solely focused on ceasefires that were meant to end the conflict (Fortna, 2004; Werner and Yuen, 2005; Cunningham, 2011). Most ceasefires are, however, preliminary. They are declared not to end the conflict, but rather to ensure a temporary interruption of fighting. This thesis expands the current literature on ceasefires, by investigating all ceasefires that are found in civil conflicts between 1989 and 2017. Gaining a more nuanced understanding of the factors that impact ceasefire durability, we can broaden our understanding of peace processes generally and conflict resolution specifically. Thus, the research question is; *what can explain why some ceasefires fail faster than others?*

The central theoretical argument of the thesis is that trust is a prerequisite for having durable ceasefires. To endure, a ceasefire requires the compliance of all parties to a conflict. Yet, any party may receive strategic advantages by attacking an adversary that has laid down its arms. Hence, without trust, ceasefires cannot be expected to hold. Following Clayton and Sticher (2018), ceasefires are seen as part of a larger bargaining process, where information is exchanged, and intentions are exposed (p. 3). Belligerents use this information to determine under what conditions they should uphold the ceasefire, or if there is more to gain from defection. This thesis argues there are three conditions that can impact the level of trust and thus the durability of ceasefires: (1) a history of previous ceasefires, (2) the comprehensiveness of ceasefires, and (3) the parties to the ceasefire.

This thesis contributes to research on ceasefires in three important ways. First, the thesis provides a comprehensive theoretical foundation with regards to ceasefires and their durability, focusing specifically on trust. The second contribution concerns the empirics: I use a new dataset, the ETH-PRIO Civil War Ceasefire Dataset. The dataset is the first to include all types of ceasefires found in civil conflicts as found in the Uppsala Conflict Data Program (UCDP) Armed Conflict Dataset between 1989-2017. The dataset is an ongoing project, expected to be finished in September 2019. As such, the analysis in this thesis is limited to the already coded countries. These countries are to be found in most of Africa, the Americas, Asia and some in Europe[1]. The dataset is originally an event dataset, but to answer the research question of this thesis, it was transformed into a survival dataset. The survival dataset will be published alongside the original dataset once finished. Not only is this the first study to use the new dataset, but this thesis is also the first to measure ceasefire failure. This is done because many ceasefires have unclear end-points. Thus, the thesis creates a framework in which future research can draw on when studying ceasefire failure. The third contribution concerns the findings: this thesis is the first study that systematically assesses ceasefire durability in civil conflict by studying all ceasefires related to peace processes. This includes preliminary ceasefires, as well as unilaterally declared ones, two aspects of ceasefires that have received scant scholarly attention. One cannot understand the structures that make ceasefires durable without investigating all types of ceasefires. Indeed, understanding these structures are important for both short-term humanitarian relief and the prospects for lasting peace. Although the dataset is still being coded, the results from this thesis will be the first of its kind and offer some explanations for why some ceasefires fail faster than others.

Using survival analysis, three hypotheses were tested. *The first hypothesis* stipulated that conflicts having experienced failed ceasefires in the past are expected to have shorter ceasefires in the present. The results from the empirical analysis lend support to this hypothesis. This means that ceasefire history matters. Ceasefires found in conflicts were ceasefire failure has occurred fail faster than ceasefires found in conflicts were no prior conflict has occurred. This is an interesting finding, suggesting that ceasefires are not always the best option for in a peace process for producing lasting peace. Indeed, if there is a history of failed ceasefires, mediators and adversaries should focus on building up trust, rather than pushing for a ceasefire. *The second hypothesis* theorized that more comprehensive ceasefires should last longer than less comprehensive ceasefires. The hypothesis was also supported by the empirical analysis. This suggests that ceasefires if expected to hold, should include some

[1] The list of included countries can be found in Table A1 in the Appendix.
[2] Both the Correlates of War (COW) and UCDP/PRIO Armed Conflict Dataset (ACD) have 1000

mechanisms to ensure its durability. This is in line with previous research, which has suggested that ceasefire agreements that include mechanisms reduce the chances of renewed warfare (Fortna 2003b, p. 365). *The third hypothesis* stipulated that ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts with only one rebel group. This hypothesis was not supported in the empirical analysis. This suggests that there is no difference in how fast ceasefires fail between multi-rebel conflicts and dyadic conflicts.

The chapter proceeds as follows: first, I define and clarify the main concepts used in the thesis, namely *civil conflict*, *ceasefire*, and *ceasefire failure*. Then, the theoretical framework of the thesis will be presented, before I highlight the relevance of the thesis and my contribution. Lastly, the structure of the thesis will be outlined.

1.1 Key concepts

1.1.1 Civil Conflict

This thesis investigates ceasefires that are found in civil conflicts. As such, one of the concepts that need to be elaborated before moving on is civil conflict. The conventional literature on civil war and conflict usually highlight four things that need to be true to count as a civil conflict: (1) the conflict must be violent; (2) violence must be targeted against the state; (3) the violence must be guided by some sort of organized opposition; and (4) the conflict takes place within one country (Cunningham, 2011, p. 24). Thus, protests, human rights violations or other forms of state repression do not constitute a civil conflict, although they can be politically important (ibid). Similarly, violent actions such as genocide, terrorism, counterterrorism or the targeting of civilians are not considered civil conflict by itself. The use of force to address the incompatibility must be systematic, implying that it is organized and sustained over an extended period (Bartusevičius & Gleditsch, 2019, p. 228).

Some countries experience several conflicts at the same time, and a fifth criterion is therefore introduced: (5) the conflict is fought over the same *set of issues* (Cunningham 2011, p. 24). The five criterions lead to the definition this thesis will use, which is the definition of armed conflict as stipulated by the Uppsala Conflict Data Project (UCDP hereafter). They define state-based 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 a calendar year” (Nils

Petter Gleditsch, Wallensteen, Eriksson, Stollenberg, & Strand, 2002; Pettersson & Eck, 2018). To be counted as a civil conflict then, 25 battle-deaths is needed within a calendar year. This separates it from war, where the normal threshold is usually set at 1000 battle-related deaths[2]. Contested incompatibility refers to an "incompatible difference of objective … a desire on the part of both contestants to obtain what is available to only one, or only in part" (Bartusevičius & Gleditsch, 2019 p. 288). As such, the conflict is fought over the same set of issues, as emphasized by Cunningham (2011), and these issues are usually indivisible.

Both interstate conflicts and civil conflicts are considered state-based conflicts, where the former involves two or more states, whereas the latter involves one state and one or more non-state groups (Vestby, 2018, p. 10). As the third criterion above recognizes, one of the parties of the conflict must be an organized opposition. Thus, interstate conflicts are not relevant in this regard, and the thesis will only focus on civil conflict. The thesis will use the terms armed conflict, intrastate conflict and civil conflict interchangeably, while also largely using the term "conflict" alone to refer to civil conflict.

1.1.2 Ceasefire

Ceasefires are an understudied area, and the lack of research is reflected in the fact that no commonly recognized definition of the concept exists (Åkebo, 2016). Smith (1995) argues that historically, there has been a "great deal of confusion surrounding the definition of ceasefire" (p. 265), and Åkebo (2016) have noted that how the concept of a ceasefire has been understood in research, has largely been determined by the aim of the study (p. 30). It is not just confusion surrounding the definition that has gained attention, however, but also how the term ceasefire differs from other related terms such as truce, armistice, cessation of hostilities, and suspension of arms (Smith, 1995, p. 265). Despite Smith (1995) writing this more than 20 years ago, attempts at distinguishing the terms persist today, and the terms have been utilized interchangeably because of their ambiguous meaning (Smith, 1995, p. 266; Karakus & Svensson, 2017, p. 3). In this thesis, the terms will be used interchangeably with ceasefire, and no emphasis will be paid on separating the different terms from one another.

This thesis will use a broad and generic definition proposed by ETH-PRIO Civil War Dataset on Ceasefires. This dataset defines a ceasefire to

[2] Both the Correlates of War (COW) and UCDP/PRIO Armed Conflict Dataset (ACD) have 1000 battle-death as the threshold for civil war.

be an explicitly declared intention, by at least one belligerent, to suspend hostilities from a specific point in time. This [definition] of a ceasefire captures the full range of security arrangements through which belligerents might agree to temporarily suspend and/or terminate hostilities (Clayton et al., 2019b).

This definition implies that a ceasefire can serve a wide range of functions and that their nature can be diverse. Indeed, the definition allows the thesis to consider ceasefires that are declared both during a conflict, and ceasefires that are declared to end a conflict. This means that ceasefires can be both preliminary and definite. Preliminary ceasefires are ceasefires declared before a political agreement is reached and are part of most peace processes. This is the most common type of ceasefire. Definite ceasefires, in contrast, apply if the ceasefire is declared with an intention to resolve the conflict (Clayton, Rustad, Nygård & Strand, 2019a) . This suggests that ceasefires can serve as a way of ending the conflict but does not necessarily do so, and are usually one element of a wider political, social and economic process (Karakus & Svensson, 2017, p. 3; Chounet-Cambas, 2016, p. 6). In addition, although most preliminary ceasefires are part of a peace process, this is not necessary, and their scope varies (Karakus & Svensson, 2017, p. 4). As such, preliminary ceasefires encompass all ceasefires except those that are declared to end the conflict.

The definition also implies that ceasefires can be declared unilaterally-, bilaterally- and multilaterally, and this distinguishes the definition from others, as most studies on ceasefires have tended to focus on agreements (i.e. bilateral or multilateral ceasefires)[3]. A unilateral ceasefire means that the ceasefire is declared by only one of the belligerents. If another actor in the conflict reciprocates the unilateral ceasefire, this is regarded as a separate unilateral ceasefire. All actors in the conflict can, therefore, be in a ceasefire together, but they are regarded as unilateral because they *did not agree beforehand* to cease hostilities. I argue it is important to include also unilateral ceasefires when attempting to catalogue the main reasons behind ceasefire durability. In the Philippines for example, 149 ceasefires have been declared between 1989-2017. Of these, 124 were unilaterally declared (Ryland et al., 2018, p. 2). It is reasonable to believe at least some of these have had an impact on the bargaining process and conflict dynamics and should, therefore, be included when investigating ceasefires.

The last issue to address is the scope of the ceasefire. Some are limited in scope, declared during a religious holiday or to ensure the safe passage of humanitarian aid. Their purpose is not related to a peace process, although they can still be declared during ongoing

[3] Åkebo 2016; Fortna 2004; Clayton and Sticher 2018; Karakus and Svensson 2017

negotiations. Others have a broad scope, declared as part of a peace process and with the intention to resolve the conflict (Chounet-Cambas, 2011, p. 6). This thesis will include ceasefires that are preliminary and definite, unilateral, bilateral and multilateral, but only include the ones that have a broad scope (i.e. ceasefires that are declared as part of a peace process).

1.1.3 Ceasefire failure

This thesis is interested in investigating the duration of ceasefires, and to do so, an account for how failure is understood is needed. There have previously been few attempts at quantitatively assess the durability of ceasefires and their failure, and those who have assessed it, have used ceasefire agreements as a proxy for peace. The existing research has therefore drawn much from the durability of peace literature, conflating a failed ceasefire as failed peace. Fortna (2004), for example, measures the duration of peace from the date of a ceasefire to the start of another war between the same two belligerents (p. 48). This is similar to the definition proposed by Call (2012), who defines failed peace as "the recurrence of internal armed conflict where a prior civil war is widely perceived to have ended" (p. 8). These ways of defining failed peace cannot be directly transferred to defining failed ceasefires in this thesis for two reasons. The first is that an initiation of a ceasefire does not necessarily translate into ending a war. This thesis defines a ceasefire "to be an explicitly declared intention, by at least one belligerent, to suspend hostilities from a specific point in time" (Clayton et al., 2019b). As such, the definitions of failed peace only consider definite agreements and ceasefires meant to end a conflict, whereas I also consider preliminary ceasefires. The second reason these approaches must be separated from ceasefire failure is that failure of peace does not occur until conflict or war resumes. To have a ceasefire means to lay down your arms, even just for a temporary time. A ceasefire thus often constitutes less of a commitment than other means (which are usually meant to end a conflict), as belligerents more easily can revoke their commitment to the ceasefire. As such, ceasefires are easier to make, but also easier to break. To measure a failed ceasefire the same way as failed peace can therefore be misleading, as they are conceptually different.

To define ceasefire failure then, I move away from the definitions on the failure of peace, and rather define it as a recurrence of organized violence in a conflict between one or more of the conflicting parties. Violence in this thesis is understood very narrowly and is defined as explicitly related to the number of people *killed* in battle (Murshed & Gates, 2005,

p. 129). Hence, a recurrence of violence occurs when at least one of the belligerents in the ceasefire opens fire which results in battle-death. Following UCDP, a “battle-related death is the use of armed force between warring parties in a conflict dyad” (Nils Petter Gleditsch, Wallensteen, Eriksson, Sollenberg, & Strand, 2002; Pettersson & Eck, 2018). The reason for simplifying violence down to battle-related death is because ceasefires in its general form mean to lay down your arms, and a battle-death is a simple measure to know that the hostilities have resumed. In chapter 4, I will elaborate on how ceasefire failure is measured.

1.2 Theoretical Framework

Ceasefires are often seen as a simple measure to end violence and are viewed as an important peace-building measure (Kolås, 2011, p. 781). As a result, they are frequently used to mitigate conflict. Yet, as Darby (2001) has pointed out “the word itself [ceasefire] acknowledges that there has been a truce rather than a surrender and that neither side has abandoned the option of returning to the use of force” (p. 8). Indeed, declaring a ceasefire means the belligerents lay down their main bargaining chip –their arms, and to trust that the other parties in a conflict will not take advantage of this can be foolish (Chounet-Cambas, 2011, p. 7). Consequently, maintaining a ceasefire requires commitment and cooperation, something that is easier said than done after several years of fighting.

Drawing insights from bargaining and cooperation theory, this thesis argues that when trust is high (or vice versa, insecurity is low), cooperation and committing to a ceasefire will be easier. Nonetheless, ceasefires only survive if all parties to the ceasefire uphold it, and the imminent threat that an adversary will renege on its agreement makes committing to it harder. This is especially true as ceasefires can be declared for tactical and strategic reasons. To create stable and long-lasting ceasefires thus require the belligerents to find the ceasefire a better option than continued warfare. As such, ceasefires follow a bargaining pattern, where belligerents continuously seek to investigate under what conditions the ceasefire is a better alternative than defecting from it. This thesis argues that there are three conditions that can make trust harder to maintain, and thus make some ceasefires fail faster than others: (1) the history of ceasefires, (2) the comprehensiveness of ceasefires, and (3) the parties to the ceasefire.

The thesis argues that the history of previous ceasefires can impact on the trust between belligerents, and thus impact on the duration of ceasefires. Kirschner (2015) claims that trust is shaped by both past- and current behaviors (pp. 24-25). As such, a history of many failed ceasefires in the past will probably influence the adversaries' ability to uphold ceasefires in the present. Indeed, if there is a history of violating the ceasefires, this might incentives defection, as an ambush can be a strategic advantage. *The first hypothesis* of the thesis is therefore that a history of failed ceasefires in the past will make ceasefires less durable in the present. It is not just the history of previous ceasefires that can impact on trust and thus ceasefire duration. The thesis argues that also the comprehensiveness of the ceasefire can have this effect. Ceasefires are rarely created equal, where some come as a result of intense and extensive negotiations, whereas others are rushed informal truces. The comprehensive ceasefires are probably, as a result, more stable, as the actors at several occasions have met and thus built up their trust. Comprehensive ceasefires often also include mechanisms such as a ceasefire committee or the deployment of third-party troops, which makes them costlier to defect from than less comprehensive ceasefires. *The second hypothesis* is thus that more comprehensive ceasefires reduce insecurity and should make them more durable. The last condition, the parties to the ceasefire, refers to how the presence of multiple rebel groups affects the duration of a ceasefire. It has been suggested that bargaining becomes especially hard when there are internal divisions within the opposition movement, as this exacerbates information and credibility problems (Cunningham, 2013, p. 660). More actors in a conflict should thus exacerbate trust issues, making it harder to maintain a ceasefire. *The third hypothesis* of the thesis posits that ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts containing only one rebel group.

1.3 Relevance and Contribution

This study is important from both an academic and policy perspective. Ceasefires are important to understand as they can help save lives by alleviating suffering in the short term but also help end conflicts in the long term. However, academic research offers little insight to enable informed decisions regarding the use of ceasefires, and present research has tended to focus on definite ceasefire agreements, disregarding the fact that most ceasefires are preliminary. If we can understand why some ceasefires fail faster than others - this can help policymakers and mediators understand whether a ceasefire is a correct tool to use or not. Indeed, it might shed new lights on when ceasefires should be declared or if they should be declared at all. It might show that conflicts become harder to resolve because of many failed ceasefires and that more comprehensive agreements are needed. To investigate in detail what causes some ceasefires to fail faster than others, this thesis can allow practitioners and researchers understand what approaches are most likely to be effective in creating durable ceasefires, but also what approaches are most likely to be effective in resolving conflicts (Cunningham, 2011, p. 184).

1.4 Structure of Thesis

The thesis will be structured as follows. Chapter 2 describes the existing literature on ceasefire duration, focusing primarily on conceptual- and research design issues found in the current literature, and how this thesis plans to fill these gaps. Chapter 3 lays out the theoretical foundations of the thesis, drawing insights from both bargaining- and cooperation theory. Based on the theory presented, three hypotheses are derived. Chapter 4 then turns to explain what datasets are used to answer the three hypotheses derived from chapter 3. It continues by explaining the structure of the original dataset, and how I have changed it to a survival dataset for this thesis. Lastly, the chapter explains how ceasefire failure is measured. Chapter 5 addresses the research design and methodology, where the choice of model and estimation technique, as well as some diagnostics of the chosen model, will be highlighted. The thesis will employ survival analysis and use the Cox model as the main estimation technique. It will then explain how the variables used in the analysis are operationalized. In chapter 6, the hypotheses are finally put to empirical test. The results give support to two of the three hypotheses: *ceasefire history*, and *comprehensiveness of ceasefire*. The last hypothesis, the *parties to the ceasefire*, had no significant effect on ceasefire durability. Finally, chapter 7 sums it all up, and concludes with the main findings of the thesis, while also highlighting some limitations of the thesis and what future research should focus on.

2 Literature Review

This literature review will start with outlining the overall current lack of research on the subject, trying to highlight some of the main research gaps found in the ceasefire literature. It will move on with presenting the relevant literature regarding my research question, focusing mainly on their limitations. As such, it will do so by first stratifying the research into two main camps: (1) conceptual issues, and (2) research design issues. Lastly, the main gaps this thesis tries to fill will be identified.

2.1 Research Gaps in the Ceasefire Literature

Research on ceasefires remains scant, despite a growing agreement over the important role they play. Indeed, ceasefires are found in nearly all civil conflicts; yet their presence in the literature is far from proportional to their presence in war. Smith (1995) contended over two decades ago the importance of understanding the processes that take place when trying to end a conflict, and that a ceasefire constitutes a necessary part. He stressed that the realization that all conflicts must end at some point "seems to be lost on political leaders generally and on war theorists in particular" (p. 6). He claimed that there existed very few major works on war termination, but that ceasefires seemed to be of even less concern. This realization still holds today and is evident in the lack of research on the subject. Höglund (2004) for example, stressed that little is known about the "factors that influence the sequencing of peace processes regarding ceasefires" (p. 24) and that the relationship between ceasefires, negotiations, and peace agreements in peace processes have both empirically and theoretically received limited attention (ibid). Kolås (2011) has argued that ceasefires are viewed primarily "as a means to an end" (p. 781) and have received little scholarly attention relative to other parts of a peace process. Similarly, Åkebo (2016) claimed that although ceasefire agreements are often mentioned in the literature, they are seldom the centre of attention (p. 19). Winokur (2018) argued that the conflict resolution literature has largely overlooked the study of ceasefires, and to the extent it has been studied, particularly ceasefire duration has received little attention (p. 7). This is a significant oversight, he further reasons, as "most major violent conflicts involve calls for and agreements on ceasefires, and that these ceasefires are intimately linked to the broader processes of waging and resolving war" (p. 7).

It is evident that many scholars agree that ceasefires have received far less attention than deserved in the literature, and this view has persisted more than two decades. Despite this acknowledgement, small amounts of research have been done. There thus exist many research gaps when it comes to ceasefires, and some of this blame can probably be attributed to a previous lack of data, especially evident with regards to quantitative studies. Indeed, much of the existing literature on ceasefires are qualitative in nature and based on case studies[4]. The remainder of this chapter will try to catalogue the main gaps and limitations in the existing literature in relation to my research question: *what can explain why some ceasefires fail faster than others?*

2.1.1 Conceptual Issues

As described in the previous section, there is a general research and knowledge gap about ceasefires. Moving beyond this acknowledgement and turning the focus towards relevant literature to this thesis, one quickly realizes that there are some issues regarding the existing literature, both in terms of conceptual- and research design. The aim of this thesis is to investigate why some ceasefires fail faster than others, and a good starting point then is to look to the literature on ceasefire duration. Yet, as previously mentioned by Winokur (2018), ceasefire duration has received little attention in the scholarly debate. Consequently, few works exist on the subject, and those that exist have usually conflated ceasefire with peace.

Fortna (2004) has written one of the most well-known books on ceasefires and their duration. In her book *Peace Time – Cease-Fire Agreements and the Durability of Peace*, Fortna sets to understand the relationship between ceasefire agreements and the duration of peace after war. She investigates why ceasefires sometimes fail and sometimes hold, and what belligerents and the international community can do to make them more likely to last. Her results suggest that mechanisms such as demilitarized zones, peacekeeping operations and external guarantees to enforce an agreement are the most important drivers to enhance the durability of peace (p. 211). Werner and Yuen (2005) use Fortna's data to investigate what makes belligerents already in a ceasefire maintain their commitments. Their results indicate that it is not only the structure of agreement that matter for ensuring post-conflict peace, as Fortna emphasized. Although they agree with Fortna that mechanisms to enforce an agreement can facilitate cooperation, this is not enough if the belligerent themselves prefer war to peace on the current terms of agreement (pp. 261-264). Rather, their results indicate that "unnatural ceasefire that come about as a consequence of third-party pressure are significantly more likely to fail" (p. 261).

[4] Two exceptions are Virginia Page Fortna (2004) "Peace Time: Cease-Fire Agreements and the Durability of Peace", and Werner and Yuen (2005) "Making and keeping peace".

The abovementioned literature suggests that a common conceptualization has been to assume that ceasefires are always declared to end a conflict. The evidence from the ETH-PRIO Civil War Dataset on Ceasefires (CWCD) reveals that most ceasefires declared in conflicts are preliminary. Although they are preliminary, many of the ceasefires are still related to a peace process. As such, it might be misleading to draw too much from the existing work if one is interested in the duration of ceasefires, as this thesis is. When investigating the duration of ceasefires, it is important to include all ceasefires that can have an impact on the broader peace process, not just a few definite ceasefires. This is because, which will further be elaborated upon in the theory section, ceasefires should be considered as part of a larger process and not isolated events. As such, all ceasefires that are related to a peace process can impact on the prospects for peace.

A similar problem that arises when investigating the ceasefire literature is that much of it embeds ceasefire agreements as being part of general peace negotiations. Although correctly assuming that ceasefires are an integral part of a peace process and that it is a prerequisite in the war-to-peace transition, the literature has tended to assume that ceasefires are always agreed upon beforehand as a result of negotiations. This is consequential, as the empirical evidence from the ETH-PRIO CWCD reveals that a significant number of all ceasefires are declared unilaterally with an intention of promoting the peace process or negotiations. By excluding these unilateral ceasefires, one might miss important information relating to the conflict dynamics and conflict termination. Åkebo (2016), for instance, aims to understand the war-to-peace transition by analyzing ceasefire agreements in relation to peace processes, focusing on the Aceh and Sri Lanka peace process specifically. She considers ceasefires as a tool in peace processes to change patterns of behavior, attitudes and the relationship between the actors, and finds that ceasefires can have "significant negative impact on the broader dynamics of peace processes" (p. 5). She argues that ceasefire agreements reached during the early phases of a conflict can impact on the resolution of conflict, both if the ceasefire holds and if it fails (ibid, p. 176). These findings are indeed very interesting, but unfortunately, she only considers bilateral ceasefires in her analysis. This means that she excludes all unilaterally declared ceasefires, despite them being related to a peace process. As she only considers bilateral ceasefire agreements, she overlooks the impact unilateral ceasefires can have on the behavior, attitudes and relationship between the actors.

Likewise, Smith (1995) investigates the transition from war to peace and the unclear road that follows. He seeks to understand what makes wars end, and claims ceasefires are the most obvious sign of this. Indeed, he argues that no war ends without one, and it thus becomes important to understand them. The aim of his study is to "catalogue the most common barriers to successful cease-fires in international and civil wars" (p. 3). He concludes that there are several obstacles that get in the way of a ceasefire agreement and argues that these obstacles will apply beyond the cases he studies. The most common of these obstacles are adversaries' unwillingness to consider a ceasefire (ibid, p. 257). As such, Smith focuses more on ceasefire onset; what make belligerents agree to ceasefires, rather than on the success or failure of already existing ceasefires. Smith's study, therefore, does not try to answer how a ceasefire agreement might relate to the broader peace process, as Åkebo's study does.

Karakus and Svensson (2017) assess the effect of local ceasefires agreements in Syria and explores why some of them are respected whereas others are violated. They find that the two main explanations the literature previously have emphasized – the quality of agreements and external third-party intervention – largely fail to explain the variation in ceasefire success (p. 1). They argue that in conflicts where the actors are well defined, it might be hard for mediators and monitors to correctly assess the situation (p. 6), especially if the conflict has multiple rebel groups. Informal and domestic approaches should, therefore, be emphasized over formal and external approaches (ibid). This is a very interesting finding; especially since the third hypothesis of this thesis is that the number of rebel groups will impact on ceasefire duration. Nonetheless, Karakus and Svensson (2017) assess only ceasefire agreements and thus exclude unilateral ceasefires.

The literature so far suggests that Åkebo (2016), Smith (1995), and Karakus and Svensson (2017) consider only bilateral ceasefires, i.e. ceasefires declared as a result of negotiations. As such, the literature investigated until now, including the literature on ceasefire duration, excludes important observations in their analysis. Whereas Åkebo (2016), Smith (1995) and Karakus and Svensson (2017) consider both preliminary and definite bilateral ceasefires, Fortna (2004) and Werner and Yuen (2005) only consider definite ceasefires. Consequently, the literature examined so far excludes unilateral ceasefires. Karakus and Svensson (2017) say that they exclude unilateral ceasefires because "it puts us in line with most of the previous research on ceasefire agreements […], but also because of the practical problems of identifying these types of unilateral offers in a systematic and comprehensive matter" (p. 7). With the new ETH-PRIO CWDC, one finally has a dataset that contains these unilateral ceasefires. It is problematic that this has not been available in previous research, especially if one has been interested in understanding the impact of ceasefires on conflict resolution and conflict management. As Karakus and Svensson (2017) have noted "unilateral ceasefires are interesting to study as part of signaling within the overall dynamics of peace processes" (p. 7). It is important to create a better understanding of what makes ceasefires last and what makes them fail. Yet, it then becomes important to include all relevant observations that can influence on the duration. In the analysis of the duration of ceasefires in this thesis, then, all ceasefires that are related to a peace process will be included.

2.1.2 Research Design Issues

It is not only conceptual issues that exist in the current literature on ceasefires. There are also some research design issues that need to be addressed. Firstly, very few quantitative studies exist on ceasefires, and especially on ceasefire duration. Fortna (2004) and Werner and Yuen (2005) are the exceptions, but they conflate ceasefires with peace and thus assume that ceasefires are always a result of negotiations and are meant to end the conflict. Consequently, most studies on ceasefires are qualitative. This is not so surprising, as ceasefires are very important to understand, not just in general, but also in specific contexts and conflicts. In addition, there has been a previous lack of quantitative data on the subject, which has probably attributed to the lack of quantitative studies.

There is, to my knowledge, only one qualitative article that addresses ceasefires and their duration. Winokur (2018) studies the relationship between ethnic civil war and the lengths of time belligerents are willing to cooperate while still disagreeing (p. 3). He emphasizes three mechanisms that will have an impact on ceasefire duration: territorial satisfaction, the relative balance of power, and actor coherence. Ceasefire durability, then, should be higher when belligerents are satisfied with territorial holdings, are not in a mutually hurting stalemate, and are highly cohesive (p. 3-4). Winokur (2018) test the variables by using John Stuart Mill's method of difference, comparing two ceasefire agreements signed during the Bosnia-Herzegovina civil war. His results suggest that both territorial satisfaction and the relative balance of power have an impact on ceasefire duration. The results also indicate that actor cohesion does not have a direct causal effect (p. 5).

Winokur's (2018) results suggest that actor cohesion does not affect the duration of ceasefires. This finding is very interesting as it is the opposite of the third hypothesis of this thesis. Yet despite his results, he focuses only on one specific conflict and on two specific ceasefire agreements. Drawing generalizing conclusions from his results can, therefore, lead to wrongful assessments of the dynamics of a conflict. In addition, his focus is on ethnic civil war, and not on civil wars in general, which can also impact on the results. To investigate the conditions that impact ceasefire duration using a quantitative method, then, allows us to draw more general conclusions with regard to ceasefire failure.

There is only one qualitative work that assesses ceasefire duration. However, there are other qualitative papers that are interested in the effects ceasefires might have on conflict resolution. Kolås (2011), investigating in-depth the Naga militancy in India, for instance, argues that the ceasefires in the Naga conflict has made resolving the conflict more difficult, as the ceasefires has disrupted the internal cohesion of the armed actors, as well as contributed to the way the armed groups have operated (p. 790-791). This has divided some actors while empowered others, leading to a prolonging of the conflict. This is a very interesting finding, as it suggests that ceasefires can impact on actor fragmentation. The study, however, is only of the Naga militancy, and investigating this more large-scale to see if the results also hold for other conflicts is of interest.

The literature reviewed thus far is important for understanding the relationship between ceasefires and their interaction with conflict termination. Although no general propositions can be made, the literature agrees that ceasefires can potentially have significant impacts on the outcome of peace processes and conflict termination, as well as on conflict dynamics. A greater understanding of the role of ceasefires in these processes should, therefore, be of high value. The literature on how ceasefires can be used as a tool to end conflict does not, however, directly address why some ceasefires collapse, and why there is such variation in their duration.

There are some more research design issues that need to be addressed. In addition to conflating ceasefires with peace, Fortna (2004) and Werner and Yuen (2005) only studies ceasefire agreements *between* states and do not consider ceasefires civil war. Today, almost all conflicts are *within* states. Yet, Fortna (2004) argues that some measures are likely to have the same effect in both inter- and intrastate conflict, for example third-party guarantees (p. 215). Nevertheless, she also stresses that there are important differences between the two types of conflicts and that some measures might need modification. Indeed, she argues that the most important difference between civil war and interstate war is that in the former, belligerents cannot leave “the fundamental political issues unsettled” (ibid).

Another important gap that appears in both Fortna (2004) and Smith (1995) is that their empirical focus is mainly on conflicts and ceasefires that primarily occurred during the cold war. Since the end of the cold war, peacekeeping operations have become increasingly more common, especially in civil conflict. Indeed, the ending of the cold war opened an opportunity for states and organizations to increase their engagement in conflict resolution, and this has made peace operations become more common in the contemporary world. Thus, there is a new international environment that needs to be considered.

The empirical focus of this thesis will be on ceasefires declared in civil conflicts and drawing too much from Fortna (2004) and Werner and Yuen (2005) can, therefore, be misleading. As Fortna (2004) stipulates herself "there are different stakes and dimensions in civil conflicts compared to in interstate conflict, and this calls for further research" (Åkebo, 2016, p. 21). Similarly, this thesis will investigate ceasefires found in civil conflicts between 1989-2017, which is quite a different international community than what Fortna (2004) and Smith (1995) investigated. This is also something that needs to be considered before drawing too much from their results.

2.2 Filling the Gaps

There are several gaps in the literature, but unfortunately, this thesis cannot try to address all of them. The focus is thus on ceasefires *that are already declared* and will not try to answer why some conflicts see more ceasefires than others, or what explains why some groups declare ceasefires while others do not. These are interesting questions that should be addressed at later stages but will not receive any attention in this thesis.

The literature on ceasefire duration offers some insights into why some ceasefires fail faster than others. Yet, the literature does not agree on what the most important measures for a lasting and durable ceasefire are. Karakus and Svensson (2017) point to contextual differences in conflicts that need to be addressed when making ceasefire agreements, whereas Fortna (2004) mostly emphasize mechanisms that can enhance an agreement. Winokur (2018) and Werner and Yuen (2005) rather focus on structural factors present in the conflict itself to explain the durability of ceasefires. What is evident is that a more comprehensive study of ceasefire durability is needed, and with the new dataset from ETH-PRIO, this becomes possible. Hence, the goal of this thesis is to fill some of this gap, by trying to find answers to why some ceasefires fail faster than others. It will do so by looking at the effect a history of failed ceasefires in the past have on ceasefire duration, the effect the comprehensiveness of the ceasefire has on its duration, and the effect multiple rebel groups in a conflict has on ceasefire duration. It will expand the current literature, by focusing not only on definite or bilateral ceasefires but also on preliminary and unilateral ceasefires. It will also focus on civil conflicts after the cold war (1989-2017), hopefully capturing the global community as it is today, and how it has been the last three decades. This way, it might help mediators, belligerents and policy-makers understand not only why ceasefires fail, but also whether a ceasefire is the best option in a given conflict. Hopefully, this again can help creating and drafting stabile long-lasting ceasefires.

3 Theoretical Foundations

This chapter presents the theoretical framework of the thesis and proposes some arguments to explain why some ceasefires fail faster than others. As touched upon previously, ceasefires rely on trust and cooperation, and only manage to survive if the parties to the ceasefire find them beneficial. Yet, during a conflict, belligerents can have many incentives to violate an existing truce. Following Clayton and Sticher (2018), I argue that ceasefires are part of a larger bargaining process, where information is exchanged, and intentions are exposed (p. 3). Belligerents use this information of their opponents to determine whether they should uphold the ceasefire, or if there is more to gain from defection. As such, ceasefires follow a bargaining pattern, as the actors seek to investigate under what conditions a ceasefire is a better option than continued warfare. I argue these conditions are driven by the trust between the belligerents, as ceasefires are built on mutual dependence. Indeed, maintaining a ceasefire requires commitment and cooperation. This thesis tries to catalogue these conditions and argues that there are three main reasons that can explain why some ceasefires fail faster than others: (1) the history of failed ceasefires in the conflict, (2) the comprehensiveness of the ceasefire, and (3) the parties to a ceasefire. The arguments build on cooperation- and bargaining theory, where commitment- and cooperation issues, as well as information asymmetries, will be emphasized.

The chapter is structured as follows; it starts with explaining why trust is a vital component to ensure ceasefire durability. Indeed, trust is important for all three of the conditions stipulated. The chapter continues with going more deeply into the different mechanisms that can explain ceasefire duration, and hypotheses are derived. Finally, a summary of the hypotheses is presented.

3.1 Trust and Ceasefire Duration

A ceasefire can be a simple tool used to halt the fighting and show a willingness to resolve the conflict. In general, they can enable the conflicting parties to display their intentions and they are usually regarded as an integral step on the path to peace (Chounet-Cambas, 2011, p. 7; Kolås, 2011, p. 781). Accordingly, if perceived successful, a ceasefire can help build trust between the belligerents and can make it easier to cooperate and commit to agreements. If perceived unsuccessful, however, a ceasefire can deteriorate the existing relationship, making cooperation and commitment harder. As such, ceasefires are part of a larger bargaining process, where information is exchanged, and intentions are exposed (Clayton & Sticher, 2018, p. 3).

When belligerents declare a ceasefire, it is often regarded as a positive step towards ending the violent conflict. Nonetheless, ceasefires can also be declared for purely strategic and tactical reasons. A ceasefire can be declared although the conflicting parties prefer war to peace, as it can enable rebel groups to rearm, regroup and remobilize (Chounet-Cambas, 2011, p. 7-8). Similarly, the combatants may gain breathing space to recover from battle and revive their morale (Clayton & Sticher, 2018, p. 7).

The difficulty that arises for adversaries is to separate ceasefires that are declared with benign intentions from the ceasefires declared with malign intentions. Most ceasefires declared will be presented as having good purposes, and to know the real reasons behind the ceasefire can be problematic for the other conflicting parties. This is not to say that adversaries declaring a ceasefire with bad intent will always violate a ceasefire. They can only use the pause in fighting to gain strength if the ceasefire holds (Clayton & Sticher, 2018, p. 7). If the opposing parties to the conflict believe the ceasefire is declared with bad intent, however, they will have incentives to break the ceasefire. Otherwise, it might allow the devious actor to improve its military capabilities and gain a stronger foothold. This can result in the emergence of spoilers, defined by Stephen J. Stedman (1997) as "leaders and parties who believe that peace emerging from negotiations threatens their power, worldview, and interests, and use violence to undermine attempts to achieve it" (p. 5). Stedman argues that spoilers arise as all involved parties to a conflict rarely see peace at beneficial, and if they do, they rarely see it at the same time. Spoilers, therefore, disapprove of peace and will take active steps to undermine any potential settlement (Stedman, 1997, p. 7-8; Cunningham, 2011, p. 16).

Low levels of trust between the belligerents, then, can make the adversaries spoil a ceasefire, especially if the costs of defection are low. Certainly, uncertainty whether an adversary will renege on a ceasefire can make committing to it harder, especially when the combatants start to worry that their opponent might take advantage of them. This is what makes ceasefires problematic. As Darby (2001) has pointed out "the word itself acknowledges that there has been a truce rather than a surrender and that neither side has abandoned the option of returning to the use of force" (p. 8). This insecurity will surely be compounded in conflicts where the actors have a history of failed ceasefires in the past, when the ceasefire is informal or when the conflict contains several rebel groups fighting the state and each other simultaneously, as the number of potential spoilers rises. Consequently, because of trust- and commitment issues, ceasefires can fail although the parties in the ceasefire prefer peace to war. What matters for a successful ceasefire, then, is based on the will to cooperate, where trust between the combatants become a vital component.

There are several factors that can influence on the trust between belligerents, and thus influence the duration of ceasefires. In what follows, I will elaborate on the factors I deem most responsible for facilitating ceasefire failure, namely ceasefire history, the comprehensiveness of the ceasefires, and lastly, the number of actors active in the conflict.

3.1.1 Ceasefire History

Some conflicts drag on for decades without being resolved and see several ceasefires during this time. The ongoing civil conflict in the Mindanao region in the Philippines, for example, experienced 26 ceasefires between 1989 and 2017 that were related to the peace process. Out of these ceasefires, more than 70% failed[5]. It seems unlikely that these ceasefires have been independent of each other, especially since actors in conflicts draw on several sources to collect information, including past and current behaviors (Kirschner, 2015, p. 23). A history of violent conflict will probably fuel mistrust, as the actor's reputations are formed and changed through the course of the conflict depending on the actions. Indeed, events on the ground can alter the preferences the belligerents have regarding the outcome of the conflict, and combatants draw from these previous experiences when choosing their tactics on the battlefield (Rudloff & Findley, 2016, p. 22). Many failed ceasefires in the past, then, will likely influence this. As Fortna (2012) has pointed out "the more failed attempts at peace [ceasefires] in the past, the harder it is to work toward a negotiated peace in the future". As such, information the actors gain from previous interactions can exacerbate commitment problems, as the adversaries probably are afraid that the ceasefire is either declared with malign intentions or that they are just biding their time until they can resume fighting (ibid). This may increase the chances of spoiling, as the adversaries realize they can gain more from defecting. Indeed, if the actors to a conflict have experienced ceasefire failure in the past, this can increase the incentive to violate the ceasefire, as the parties to the ceasefire might assume the other part will violate the ceasefire and thus strike first to get an upper hand.

When a ceasefire fails, it is usually a result of renewed violence. As Lim and Lee (2015) argue, when a ceasefire is broken, "it means that one of the parties has taken renewed

---

[5] 19 out of 26 ceasefires in the dataset related to the peace process were recorded as failures, when using the strictest threshold for ceasefire failure at 1 battle-death.

military action against its old enemy" (p. 491). Although this is not always true, violence is usually a catalyst for ceasefire failure. Scholars have reached the not so surprising conclusion that violence often has a negative impact on peace negotiations. When violence flares up, negotiations have been postponed, stalled or cancelled (Höglund, 2004, p. 10). This is because the use of violence by one of the belligerents is often seen as a breach of faith, and as such, it increases fears and mistrust between the combatants (ibid, p. 31). Thus, a failed ceasefire can lead to a crisis in the peace process, which is fueled by the mistrust between belligerents. Ceasefires often constitute a vital component of a peace process, and this process is often dynamic and "based on the exchange of concessions or compromise offers with the adversary" (Mitchell, 1981, p. 198). The fact that the process is dynamic, signals that the path to peace is not linear, but rather goes backwards and forward at different speeds depending on the information available to the adversaries (Darby, 2001, p. 11). Thus, if a conflict has had many failed or successful ceasefires in the past, this can have an influence on the adversaries' abilities to trust each other and cooperate, and this again will impact on the prospects for peace. In this view, a ceasefire can build trust if perceived successful by all the adversaries. Nevertheless, a ceasefire can also destroy trust and exacerbate commitment problems if perceived unsuccessful. In other words, I expect past ceasefire failure to impact on future ceasefire duration. This leads to my first hypothesis:

H1: *Conflicts having experienced failed ceasefires in the past will have more short-lived ceasefires in the present.*

3.1.2 Comprehensiveness of Ceasefire

It is not just the history of ceasefires in a conflict that can impact on the duration of ceasefires. This thesis argues that also the nature of the ceasefire can impact on ceasefire duration. No ceasefires are created equal, and their comprehensiveness can differ greatly. Some ceasefires are written agreements, crafted after many years of formal peace negotiations. These ceasefires often include mechanisms for ensuring the implementation and enforcement of the ceasefire, such as ceasefire committees, external mediators and third-party troops. Other ceasefires are declared informally, perhaps as an effort to show sincerity towards upcoming negotiations. It is reasonable to expect this difference in comprehensiveness to impact on the duration of ceasefires.

Defecting from a ceasefire that is the result of extensive negotiations will probably have higher costs than defecting from a ceasefire that is the result of a spontaneous decision to halt fighting. As Clayton and Sticher (2018) have argued: "by accepting specific agreement provisions, such as monitoring and verification, conflict parties can tie their hands and send a costly signal that they no longer see war as the preferred long-term option" (p. 3). Indeed, comprehensive ceasefires, as mentioned, usually require extensive negotiations that often take place over many years. During these negotiations, trust between the parties is built, which leads to them agree on a ceasefire. The parties are dependent on trust for the ceasefire to survive, but it is also this trust that got the ceasefire in place. One can, therefore, expect the trust between belligerents to be stronger when they have managed to agree on specific mechanisms and declare a ceasefire, and the ceasefire is thus expected to last longer. In fact, literature has suggested that ceasefire agreements that include mechanisms reduce the chances of renewed warfare (Fortna, 2003b, p. 365). Although ceasefires are dependent on trust, they can also be driven by reciprocity and mutual deterrence – it is the fear of a counterattack that makes them stable. Ceasefires should thus include measures that reduce the incentives to fight. Otherwise, the ceasefires might incentivize the complete opposite, as defecting from an informal truce can gain the combatants a strategic upper hand. Fortna (2003b) have argued that

> for reciprocity and deterrence to work, several things must be true: the cost of reinitiating conflict must outweigh the incentives to attack; it must be easy to distinguish compliance from noncompliance; both sides must be reassured about each other's intentions, especially if there is a military advantage to striking first […] (p. 342).

This implies that ceasefires need to be created in such a way that makes maintaining them a better option than defecting from them. However, after several years of fighting, trust between the combatants is probably either extremely weak or absent and to assume that the actors will adhere to a ceasefire just because they said so might be foolish. To include measures to overcome this insecurity and ascertain the parties to the ceasefire that defection is a costly option, then, should make the ceasefire more durable than if such measures were not included. This leads to the second hypothesis:

H2: *More comprehensive ceasefires should last longer than less comprehensive ceasefires*

3.1.3 Parties to a Ceasefire

How does the presence of multiple rebel groups affect the duration of ceasefires? The theory presented thus far suggests that previous ceasefires failures found in a conflict, and the comprehensiveness of these ceasefires can impact on their duration. The arguments propose that when trust is high (or vice versa, insecurity is low), cooperation and committing to a ceasefire will be easier, which leads to longer-lasting ceasefires. In addition to these causes of ceasefire failure, it has also been suggested that bargaining becomes especially hard when there are internal divisions within the opposition movement, as this exacerbates information and credibility problems (Cunningham, 2013, p. 660). Indeed, Cunningham (2011) argues that conflicts involving multiple rebel groups have “fundamentally different dynamics from two-party ones, because the presence of additional combatants changes the incentives that groups have to negotiate and/or to fight” (p. 14). As such, combatants in fractionalized conflicts will face greater barriers resolving war than their dyadic counterpart, much due to the “dual contest” they face: “the contest in the pursuit of the common good for the group as a whole *and* a contest over private advantages with other factions in the movement” (Bakke, Cunningham, & Seymour, 2012, p. 266). Thus, rebel groups in a divided movement will not only fight the government, but they will also fight each other to reap as many benefits as possible. Fjelde and Nilsson (2012) argue that conflict between rebel groups erupt as they want to “secure material resources and political leverage that will help in the conflict against the government” (p. 605). So even though rebels within the same movement represent the same population, they still have different leadership and the possibility to act independently, leading them to often have different claims about what the population they represent wants (Cunningham, 2013, p. 663). This provides the state with multiple, competing views concerning the goals of the divided movement, which makes finding a settlement all parties prefer to war much harder.

A ceasefire declared in a conflict consisting of one coherent rebel movement and a government would face failure if either the state or rebels decide to violate, or spoil, the ceasefire. In contrast, ceasefires declared in a conflict with a divided opposition movement will not only face potential violence from the participating actors, but also from excluded ones. Hence, conflicts with multiple rebel groups can experience both inside and outside spoiling. Inside spoiling occurs when an adversary that is part of the ceasefire fails to uphold it, whereas outside spoiling occurs when parties that are excluded from the ceasefire decides to attack the parties in the ceasefire to make it fail (Stedman, 1997, p. 8).

Although a conflict involves several groups, this does not necessarily mean that all relevant actors are part of a ceasefire. Indeed, ceasefires can be declared both unilaterally-, bilaterally-, and multilaterally. When the opposition movement is highly divided, the chances that all actors are participating in the ceasefire seem unlikely, and even if they are, the chance that all have declared them with good intentions seems doubtful. Certainly, if a ceasefire is declared with all relevant actors, some might find it beneficial to violate the truce to gain an advantage over its opponents. After all, the reasons behind declaring a ceasefire can be ambiguous, and the more actors participating in the ceasefire, the more uncertainty and information asymmetry will persist regarding the intention. Consequently, inside spoiling can occur, either because one group thought other groups agreed to the ceasefire with devious intentions, because the adversaries might come to believe their own interests will be compromised in the future by the ceasefire, or they might see an unanticipated advantage to gain at the enemies' expense, and as a result chooses to renew hostilities (Fortna, 2008, p. 83). The more rebels participating in the ceasefire, then, the higher the likelihood that one or more belligerents believe they can gain from attacking its enemies. As a result, a ceasefire agreement will probably be unstable in fragmented civil conflicts, as mistrust and fear make the adversaries more quickly renew violence. Consequently, although all rebel groups in the rebel movement are included in the ceasefire, all must believe that the ceasefire is a better option than continue to fight. Since none of the actors in the ceasefire can be certain whether the other actors are planning an attack, this suspicion towards the other actors will make all sides extremely vary (Fortna, 2008, p. 84).

Although we have some instances of ceasefires containing all actors in a conflict, in most cases, the ceasefires declared only include some. When this is the case, a ceasefire can fail because of both inside and outside spoiling. Ceasefires can be broken because some groups are excluded from the ceasefire, and this outside spoiling happens when belligerents, who are not part of the ceasefire, use violence towards the involved parts in the ceasefire. This might force or incentivize the adversaries in the ceasefire to renew violence. Bilateral ceasefires, for example, often exclude some of the rebel groups in the conflict, making them only “partial” ceasefires (Cunningham, 2011, p. 185). This means that a truce between some groups is implemented while fighting continues between others. This is a highly unstable situation, as there is ongoing violence in the conflict simultaneously with the ceasefire. Certainly, if a bilateral ceasefire between one of the rebel groups and the government is declared, other rebels within the same movement might act as spoilers to increase their power position vis-à-vis the other rebel groups in the movement.

Ceasefires have become a common tool used by mediators and belligerents during negotiations to advance the peace process. In conflicts where there is an internal division within the opposition movement, a peace process might cause malcontent for some of the rebels, either because they are excluded, or because they are not satisfied with how the process unfolds. As a matter of fact, some adversaries might gain a greater influence over politics by resisting settlement (Cunningham, 2013, p. 664). As such, a ceasefire might be spoiled not only because some groups are displeased with the ceasefire, but also because they are displeased with the peace process in general. This is in line with the literature on movement fragmentation, which suggests that the risk of fragmentation is greater during negotiations (Fjelde & Nilsson, 2018, p. 552). Malcontent with ongoing negotiations then, do not just fuel ceasefire failure, but also fuel the rise of new rebel groups.

Until now, the focus has been on bilateral-, and multilateral ceasefires, that is, ceasefires that are agreed-upon before initiation by the adversaries. How do unilateral ceasefires fit into this? Stedman (1997) argued that spoiling could only occur when "at least two warring parties have committed themselves publicly to a pact or a peace agreement" (p. 7). This phrasing omits spoiling to occur during unilateral ceasefires. I argue that when it comes to ceasefires, spoiling can occur at all types since they are all part of a bargaining process and can thus influence a peace process. Unilateral ceasefires also require many elements of trust, as they are declared with a promise not to attack its enemy conditioned on the enemy not attacking them. Hence, also unilateral ceasefires are dependent on the adversaries' promise to their word and commit to the ceasefire. As such, unilateral ceasefires, although only one party has committed themselves publicly to the ceasefire, are still dependent on two or more warring factions to survive. Accordingly, even unilateral ceasefires can send a costly signal, and as Karakus and Svensson (2017) have argued "if designed in a way that makes them costly for the one that makes such a commitment [unilateral ceasefire], they can (at least theoretically) reveal the intentions of the initiator" (p. 7). In addition, many unilateral ceasefires are reciprocated, meaning that several warring factions have committed themselves to a ceasefire, albeit to different unilateral ones. The thesis, therefore, argues that spoiling can occur also in unilateral ceasefires, and if it occurs, trust is probably weakened as a consequence.

In sum, I argue that in a civil conflict, a fragmented rebel movement can make ceasefires fail faster compared to dyadic conflicts because of exacerbated trust issues. These trust issues usually stem from uncertainty and insecurity. When trust is low, committing to a ceasefire becomes harder, especially when there are several groups that can spoil the ceasefire. Similarly, when all relevant groups are part of a ceasefire, uncertainty whether someone will renege on the agreement becomes more probable. Fragmented rebel movements, then, will have ceasefires with shorter duration compared to conflicts that are dyadic. Based on the arguments presented above, the third and last hypothesis is:

H3: *Ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts containing only one rebel group.*

3.2 Summary of Hypotheses

Table 3.1 Summary of Hypotheses

Indicator	**Hypothesis**
History of Ceasefire	H1: *Conflicts having experienced failed ceasefires in the past will have more short-lived ceasefires in the present.*
Ceasefire Comprehensiveness	H2: *More comprehensive ceasefires should last longer than less comprehensive ceasefires*
Parties to the Ceasefire	H3: *Ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts containing only one rebel group.*

4 Data

In this chapter, the data used to test the hypotheses are presented. The chapter starts with introducing the main dataset that will be used to test the three hypotheses, namely the ETH-PRIO Civil War Dataset on Ceasefires. Then, the structure of the original dataset, and how the dataset has been changed in order to use it in survival analysis, is discussed. Lastly, the chapter explains how ceasefire failure is measured.

4.1 Dataset

The dataset I base my analysis on is the ETH-PRIO CWCD[6]. The aim of the dataset is to cover all ceasefires found in civil conflict between 1989 and 2017. The dataset is still being coded, meaning that the thesis will use the available data as of 22.04.2019[7]. As the dataset is not yet finished, there are several conflict countries that are excluded from the analysis. Most countries in Africa, Asia and the Americas are included in the analysis, whereas the Middle East and Europe[8] are still in progress. This could potentially lead to a selection bias in the analysis, meaning that the sample used is not representative of the population it is meant to analyze. If selection bias is present in the data, this can lead to inaccurate conclusions. I argue that the CWCD, while not completed, represent a broad range of countries that captures the different characteristics found in civil conflict. As it includes Sub-Saharan Africa, North Africa, the Americas and parts of Asia and Europe, the sample used for the analysis has countries ranging from low- to high-income countries. Hence, I argue selection bias is not an outspoken problem in the data.

Nonetheless, there are some countries that can be consequential not including, as they could potentially drive the results. Israel for example is regarded as a high-income country by the World Bank but is also a country that has been in conflict since the 1940s (World Bank Group). Based on the news articles involving both Israel and the word ceasefire[9], it seems likely that Israel has experienced many ceasefires. Syria and Yemen are other countries that should have been included. Both countries are ravaged by high-intensity war and represent

[6] The dataset is created as collaboration between ETH Zurich and PRIO, led by Govinda Clayton and Siri Aas Rustad. The coding of the dataset has been going on for three years and is expected to be finished in September 2019.
[7] List over the included countries can be found in table A1 in the Appendix.
[8] Azerbaijan, Georgia and United Kingdom are included from Europe
[9] Factiva have approximately 132600 news articles that mention both Israel and the word ceasefire. For comparison, India has approximately 61 800. Out of these articles, 164 ceasefires were found.

large humanitarian crises that have had several ceasefires through the conflicts[10]. As will be discussed further in section 4.3, ceasefire failure is measured using the UCDP Georeferenced Event Dataset (GED) (Sundberg & Melander, 2013). Syria is not included in their data, as the data collection did not yield consistency and clarity as other GED data (Croicu & Sundberg, 2017). In addition, large parts of the conflict in Yemen between the government and Houthi movement (Ansarallah) was excluded from the GED data because of a lack of stated incompatibility. As such the GED data only have data for the years 2014-2015, although the conflict erupted in 2004 (ibid). Because of the lack of data in other datasets, including Syria and Yemen could bias the results, as ceasefire failure would not be possible to capture in these conflicts. Hence, the only country that potentially could drive the results is Israel, and this country should be included in future research. What is worth mentioning is that UCDP is working on releasing the data of Syria, so for future research on ceasefire durability, this country should also be included (ibid).

Although the data is not finished by the time of this thesis submission, a great advantage is that the thesis is the first allowed to employ the dataset. This means that although the sample is not completed, the thesis will still be the first to test ceasefire durability quantitatively, using a dataset that also includes preliminary ceasefires. This allows me to shed new lights on the causes of ceasefire failure, expanding the existing literature. Another advantage is that I have been one of the main coders of the dataset at PRIO the past year, which make me know the dataset extremely well.

In the literature review, I stated that few attempts have previously been made to quantitatively assess ceasefires, and those that have tried, have used data that does not necessarily capture all aspects of ceasefires (for example Fortna, 2004; Cunningham, 2011; Werner & Yuen, 2005; Åkebo, 2016). These have either assessed ceasefires found in interstate wars, only focused on definite ceasefires meant to end a war, focused on ceasefires that are agreed-upon beforehand, or a mixture of these. The new dataset is a comprehensive compilation of all ceasefires, and thus include measures that have not been present in prior datasets[11]. More specifically, it includes all types of ceasefires found in civil conflicts, both preliminary and definite, unilateral-, bilateral-, and multilateral ceasefires, as well as including ceasefires with both a limited and a broad scope. Hence, the new dataset opens for investigating a plethora of questions regarding ceasefires and their impact on peace processes

[10] Karakus and Svensson (2017) have investigated ceasefires in Syria, whereas the ongoing coding at PRIO has found several ceasefires in Yemen.
[11] Does not include ceasefires found in interstate war

and conflict termination, questions that were previously unattainable to answer quantitatively. Smith (1995) argued more than two decades ago that theorists and researchers have tended to use terms related to ceasefires interchangeably. As the confusion persists today, the new dataset has adopted a broad definition on ceasefires, meaning that the dataset defines a ceasefire

> to be an explicitly declared intention, by at least one belligerent, to suspend hostilities from a specific point in time. This broad definition of a ceasefire captures the full range of security arrangements through which belligerents might agree to temporarily suspend and/or terminate hostilities (Clayton et al., 2019b).

The use of a broad definition makes it a very flexible dataset, allowing the researchers themselves to decide their area of focus without being hampered by non-available data. This thesis will make use of the broad definition of a ceasefire while excluding ceasefires that are not related to a peace process. The reason for this is that I am interested in the causes of ceasefire failure and the effect this might have on peace processes and conflict termination.

The CWCD is based on the UCDP/PRIO Armed Conflict Dataset (ACD), meaning that ceasefires are coded for the same countries and conflicts as found in the ACD. This makes the ceasefire dataset compatible with most of UCDP's datasets. A disadvantage with the ACD is that it only includes conflict-years where the conflict has reached a threshold of 25 battle-related death (BRD) during a calendar year. As such, in the ACD, a conflict can be active in 2001 and 2003 (i.e. reached 25 BRD), but not be recorded as an active conflict in 2002. This can be misleading, as it does not necessarily mean that the conflict was peaceful in the intermittent lull, just that fewer than 25 people were killed (Gates & Strand, 2004, p.12). The CWCD has taken this into consideration, by including ceasefires that are found up to three years before and after the first (or last) active conflict year in the country according to the UCDP (Clayton et al., 2019b). In addition, if there are indications that the conflict was active at a low-intensity level far beyond the stipulated start or end date in the ACD, the ceasefires should be included (ibid). As such, the ceasefire dataset includes also these intermittent lulls, which the ACD excludes. It includes these years in the coding to avoid unreliable reporting regarding how the conflict lapses, and to avoid having incomplete data.

4.2 Structure of Data

The CWCD is structured as an event dataset, where the unit of observation is a ceasefire declaration. The original dataset has 2157 observations[12], where each observation represents an actor-ceasefire. As such, a bilateral ceasefire between a government and a rebel group constitute two lines of observation. Each ceasefire is coded down to specific dates to stipulate when the ceasefire came into effect. If the ceasefire has a stipulated ending, this is also coded. As such, the dataset includes both successful and failed ceasefires. The question driving this thesis is what explains why some ceasefires fail faster than others, and the thesis is thus interested in investigating the ceasefires that have failed (or still have the possibility of failure, i.e. right-censored observations). To investigate this, survival analysis is used. Survival analysis, which will be discussed in chapter 5, is used when the aim is to explore how long something will survive until it fails. Failure, or the possibility of failure, is therefore a requisite when using this method, and successful ceasefires are (unfortunately) discarded from the analysis.

Another requirement when using survival analysis is to use a time-to-event dataset. I therefore needed to transform the original CWCD into duration data. In survival analysis, the event is usually regarded as the failure, and the time is the duration it takes before failure occurs. In my analysis then, I needed to create a duration variable and a ceasefire failure variable. The new format of the dataset will be published along with the original CWCD once finished. In survival analysis, unless you have interval data, the dependent variable is made up of two variables: duration and failure. In this thesis, the dependent variable more specifically is ceasefire length and ceasefire failure. The ceasefire length records the duration (in days) the ceasefire is in effect. This means that a ceasefire with a length of 30 lasted 1 month (or 30 days) until it failed. The ceasefire failure variable documents whether the ceasefire ends or not, and it thus records whether failure takes place. The variable is binary, and gives a 1 to ceasefires that have failed, and a 0 to ceasefires that has not yet failed by the end of the dataset (i.e. censored observations). After taking out the successful ceasefires, filtering out ceasefires that are not related to a peace process, and taking out ceasefires that have missing values on dyad-, and conflict id, there are 579 observations left[13].

[12] As of 22.04.2019

[13] Using the 1 BRD threshold.

4.3 Measuring Ceasefire Failure

Many ceasefires have an unclear ending, and consequently, it is both important and necessary to discuss what is considered to constitute a failure in this thesis. In chapter 1, ceasefire failure was defined as a "recurrence of organized violence in a conflict between one or more of the conflicting parties", where violence is understood to explicitly relate to the number of people killed in battle (Murshed & Gates, 2005, p. 129). One problem that arises is to determine how to measure ceasefire failure, i.e. how much violence is needed before the ceasefire fails because of a recurrence of violence? As so many ceasefires have unclear endings, this thesis will supplement the existing endings in the ceasefire dataset by creating my own additional de facto endings based on battle-related deaths. Using a dataset on Georeferenced Events from UCDP, it will match the duration of ceasefires with battle-related deaths. Hence, a ceasefire fails after a certain number of deaths have occurred between the belligerents that are part of the ceasefire, or if the ceasefire has failed according to the coding. As such, all ceasefires that ended because the stipulated time frame was over and did not reach the battle-death threshold before the ceasefire ended will be treated as "successful" and are not considered in the analysis.

There are especially three difficulties that arise when measuring ceasefire failure; (1) whose deaths should be counted, (2) how fast ceasefire failure can occur after onset, and (3) how many deaths are needed for failure. The first difficulty, whose deaths should be counted, relates to questions regarding what deaths in the conflict should be considered when measuring failure. For example, should deaths from both the government and the rebel group count, or is it enough that the BRD happens at either side? And what about civilian casualties? As was argued in the theory chapter, a ceasefire means in its most general form to suspend hostilities from a specific point in time by at least one belligerent (Clayton et al., 2019b). As such, attacking of civilians violates the terms of the ceasefire, and is therefore included in the measure of ceasefire failure. In fact, literature has suggested that civilians tend to be targeted in civil war (Gates & Strand, 2004, p. 3). To exclude killings of civilians can therefore lead to wrongful assessments of ceasefire duration, or even worse, treat some ceasefires as successful. Consequently, civilian casualties are included in the BRD threshold. A more difficult question arises as to how unilateral ceasefires fit into this. For bilateral and multilateral ceasefires, it makes sense that battle deaths from either the government or rebel side should constitute ceasefire failure, as the ceasefire is an agreement. Can failure be measured similarly for unilateral ceasefires? If the rebel group unilaterally declares a ceasefire, and the government attacks the rebels, has it failed? As discussed in the theory chapter, spoiling can occur in all types of ceasefires, also the unilateral ones, as unilateral ceasefires are declared with a promise not to attack its enemy as long as the enemy does not attack them. As such, they are also dependent on the adversaries' promise to their word and commit to the ceasefire, and although only one party have committed themselves publicly to the ceasefire, they are still dependent on two or more warring factions to survive. Because of this, I argue that the BRD can occur at either side of the conflict dyad, as the ceasefire de facto ends when fighting resumes. Some might argue that this is misleading, as unilateral ceasefires in many instances are conceptually different from bilateral- and multilateral ceasefires. Yet, I argue this way of determining whose deaths should be counted makes sense in terms of the theory presented, as a resumption of violence will probably represent a breach of trust between the warring parties.

The second difficulty relates to how fast ceasefire failure can happen after a ceasefire comes into effect. Can the ceasefire be violated the same day as it takes effect? Ideally, it would be preferable to look at BRD right after the ceasefire takes effect, but as we have measured ceasefires in days, and not a specific time of the day, one cannot know at what time of the day the ceasefire took effect. For example, several ceasefires are declared to take effect during the middle of the day, like one in Afghanistan in August 1992 that was scheduled to begin at noon. Others are declared to take effect at midnight, like one in Sri Lanka in January 1991 (Clayton et al., 2019b). As it is impossible to determine when the ceasefire has come into effect, it might bias the result if ceasefire failure is measured on the same day the ceasefire takes effect. This is especially true as adversaries sometimes fight until the ceasefire starts. The most famous example of this is from World War I, when a soldier was killed one minute before the ceasefire, which was meant to end the war, took effect (Hayes-Fisher, 2008). If one decides to measure ceasefire failure the same day the ceasefire comes into effect, and the threshold is met, it is impossible to know whether the attacked happened before or after it started. Consequently, ceasefire failure is measured from the day after the ceasefire is meant to take effect (i.e. at day +1).

The last difficulty and the one deemed the most critical, relates to deciding how many battle deaths are "needed" before the ceasefire is said to have failed. As no one has attempted to use battle-deaths as a proxy for ceasefire failure, it is hard to both theoretically and empirically decide where the best cut-off point is. Gates and Strand (2004), when investigating the duration of civil wars, found that where the casualty threshold is set makes a difference (p. 27). In addition, they argued that it could lead to problems of selection bias (ibid, p. 5). Because of this, ceasefire failure is measured at different thresholds, namely at 1, 10 and 25 battle deaths.

The first cut-off point is a very strict measure of ceasefire failure and is chosen because as soon as there is a resumption of violence, one can argue that the break in fighting has de facto ended. Trust between belligerents will probably be weakened despite it being one only death – a promise has nonetheless been broken. Likewise, unilateral ceasefires can also weaken trust, as they are always declared towards some other actor. However, the one battle-death threshold might be a too strict measure. For example, a ceasefire declared by a rebel leader does not necessarily mean that the fighters on the ground have gotten the message right away, and as a result, "accidents" can happen before the ceasefire takes full effect. Although ceasefire failure is measured the day after initiation, accidents can still occur. A too strict measure on ceasefire failure may "fail" ceasefires where the actors are adhering to it. Empirical evidence also suggests that some actors believe that small violations are not enough to break a ceasefire, and having only a one battle-death threshold might fall into this category (Clayton et al., 2019b). To account for the fact that accidents and small violations can occur during a ceasefire, two additional thresholds are included.

The second measure, at 10 battle-deaths, is included to differentiate between high- and low-profile incidences of violence. Höglund (2004) argues that there are several characteristics that fall under the high-profile incidents of violence, one of them being mass casualty attack. A mass casualty attack is defined as "involving a large number of casualties in a single attack […] this means an attack in which ten or more people are killed" (p. 46). Differentiating between high- and low-profile incidents of violence is important as high-profile incidents more severely cause a crisis in the negotiation process (ibid, p. 47), and thus also more severely impact the trust and commitment issues the adversaries have towards each other.

The last measure is at 25 battle-deaths and is included following the UCDP definition of armed conflict, where a conflict must have at least 25 battle-related deaths per calendar year to count as an active conflict. Gleditsch et al (2002) argued that "25 deaths in a single year - is high enough for the violence to represent a politically significant event" (p. 617). As such, the 25 battle-death measure is included with hope of differentiating between comprehensive ceasefires meant to end the conflict (definite ceasefires)[14] and preliminary ceasefires. An important difference between these thresholds and UCDP's is that the threshold used in this thesis represents one event constituting 25 battle deaths (25 deaths in one single attack), whereas the UCDP has an accumulative measure. This applies also to the other battle-related thresholds chosen. Ideally, it would be preferable to measure ceasefire failure using an accumulative measure, as this would more correctly reflect the true aspects of the conflict. Unfortunately, after facing some computational problems, this is not something I am able to do. For future research, however, I intend to take this into account as well, as this will probably have interesting effects on the results.

[14] Following the previous literature, which has measured ceasefire failure in the same way as failure of peace (Fortna 2004; Werner and Yuen 2005).

5 Research Design and Methodology

The research design represents the overarching strategy for answering the research question. The thesis will use quantitative analysis, using a new dataset that encompasses ceasefires in almost all civil conflicts from 1989-2017. In the theory section, it was argued that because of trust and commitment issues, keeping a truce will be more difficult in situations where ceasefires have failed in the past, when the cost of defection are low (i.e. less comprehensive ceasefires), and when a greater number of rebel groups are present in the conflict, as possibilities of spoiling increase. The thesis thus posits three main arguments to explain why some ceasefires fail faster than others: (1) the history of ceasefires, (2) the comprehensiveness of ceasefires, as well as the (3) parties to the ceasefire. To investigate whether these arguments hold, survival analysis is used.

The chapter is structured as follows: it will begin by exploring the choice of statistical model the thesis will use, namely survival analysis. Second, it will highlight the estimation technique chosen and the reasons for this. Thirdly, important assumptions that survival models, and more specifically the Cox model assumes is highlighted, and how to deal with these. The second part of the chapter presents the operationalization of the dependent and independent variables before the control variables used in the analysis is presented. Lastly, the proportional hazard assumption is tested on the data.

5.1 Choice of Model: Survival Analysis

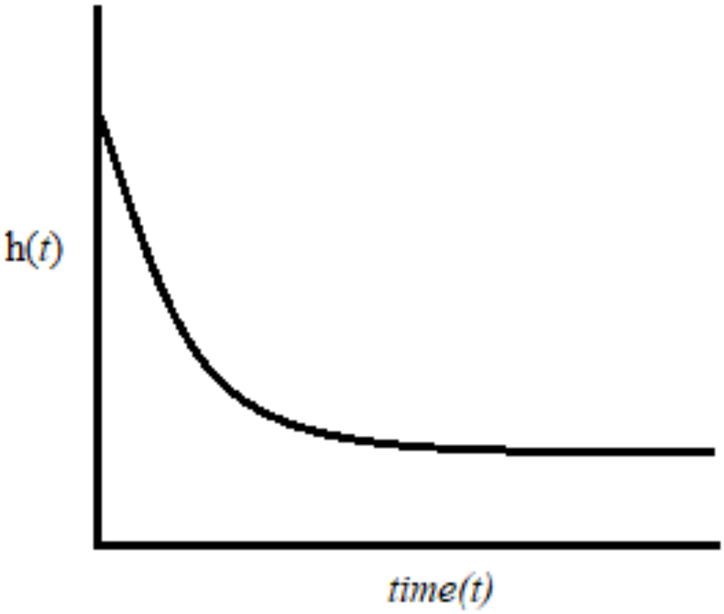
The aim of this thesis is to understand what makes some ceasefires fail faster than others. More specifically, it aims to understand (1) the effect previously failed ceasefire in a conflict can have on the length of time a ceasefire last, (2) whether the comprehensiveness of ceasefire matter for the ceasefire duration, and (3) the effects the parties to a ceasefire can have on duration of ceasefire. To test this, survival analysis is used. Survival analysis (also called event history analysis or duration models) is used when one is interested in studying events and their cause (Allison, 1984, p. 1). Duration models are therefore the best choice when the aim is to explore how long something will "survive" until it fails.

An event occurs when there is a relatively sharp transition from one state to another (Box-Steffensmeier & Jones, 2004, p. 1). Similarly, duration is understood as the length of time something lasts until an event. In this thesis, the "event" is ceasefire failure, while duration is the time a ceasefire survives until it fails. The dependent variable is ceasefire duration, which makes survival analysis the best-suited option. The choice to use survival analysis is further substantiated as the data can assume that the probability of experiencing some event changes over time (Cunningham, 2011, p.86). Hence, survival models allow the duration to be continuous. For example, with the duration of ceasefires, it is plausible to expect that ceasefires are much more likely to fail in their early days and that the probability of survival will increase the longer the ceasefire is holding. Ceasefires that fall apart within days are thus treated as less stable than ceasefires that last several months or even years. At some point in time, it is reasonable to expect that a ceasefire will generally continue to hold and that the probability of experiencing ceasefire failure will flatten out (Cunningham, 2011, p. 186). This is presented graphically in figure 5.1.

However, this does not mean that they always do so. Survival data does not assume that because some event has lasted for a given period, it will continue to last (Fortna, 2008, p.11). Although we know a ceasefire has lasted to date (the end of the dataset), we do not know how long it will last in the future (ibid). This is another advantage with event history models, as it can easily deal with censoring. Censoring occurs whenever an observation's full event history is unobserved. Right censoring is the most commonly observed in event data and is experienced because the study concludes prior to the termination of survival times (Box-Steffensmeier & Jones, 2004, p. 16). This is also the type of censoring that will be used in this thesis. Thus, ceasefires are censored if they are still in effect on 31.12.2017, as this is the end of the dataset. If censoring were not considered, ceasefires that last over this date would be treated as permanent.

Measuring duration has some difficulties, and a problem that occurs when working with ceasefires is that many have unclear endings. To know when a ceasefire fails (i.e. event occurs) can thus be troublesome in many instances. To cope with this, as discussed in section 4.3, different thresholds based on battle-related deaths are used to create ceasefire failure. Lastly, as the focus of this thesis is on the causes of ceasefire failure, the unit of analysis is on ceasefire duration (measured in days).

Figure 5.1 Hazard function for ceasefire failure



The hazard function h(t) showing the chances of ceasefire failure during a conflict

5.1.1 Choice of Estimation Technique

The choice of estimation technique is based on the assumptions made regarding the shape of the hazard function, where the hazard function models what periods have the lowest or highest chances of an event (Der & Everitt, 2008, p. 207). As figure 5.1 illustrates, it is reasonable to expect that ceasefires will have the highest chance of failure near the onset and that after a certain time; the hazard rate will flatten out. If a researcher has strong theoretical expectations that the risk of an event will either increase or decrease over time, one can specify the distribution function using a parametric model (Box-Steffensmeier & Jones, 2004, p. 21). If the shape of the hazard function is unclear, then using a parametric model can lead to misleading results that do not make any substantive sense (ibid).

The Weibull estimation technique is the most common among the parametric models in social sciences. The Weibull model is characterized by being monotonic, meaning that the baseline hazard can either increase, decrease or be flat with respect to time, yet only in one direction (Box-Steffensmeier & Jones, 2004, p. 25). As such, if one has a theoretical expectation that the hazard rate will be monotonically decreasing over time, as figure 5.1 illustrates, then a Weibull distribution can be the correct choice. Yet, if the researcher is not certain about the distribution, then choosing the parametric Weibull distribution can lead to deceptive results regarding the relationship between the duration time and the covariates (Box-Steffensmeier & Jones 2004, p. 21). As Larsen and Vaupel (1993) point out "in the analysis of duration data, if the functional form of the hazard has the wrong shape, even the best-fitting model may not fit the data well enough to be useful" (p. 96).

An alternative modelling strategy to the parametric model is the Cox Proportional Hazard model. The Cox model makes no assumptions regarding the underlying hazard function, thus leaving the distributional form of the duration times unspecified (Box-Steffensmeier & Jones 2004, p. 47). This means that it makes no assumptions about whether a ceasefire will be likely to last given that it has held thus far (Fortna, 2008, p. 11). Box-Steffensmeier and Jones (2004) argue that most

> research questions in social science should be chiefly concerned with getting the appropriate theoretical relationship 'right' and less concerned with the specific form of the duration dependency, which can be sensitive to the form of the posited model (p. 47).

They thus argue that specifying the distribution of the hazard rate is not necessarily so important if one does not already have strong theoretical expectations. Following this, the Cox Proportional Hazard Model will be the point of departure, leaving the hazard rate unspecified. However, as there are some theoretical expectations regarding the distribution of the hazard function, the model fit will be assessed.

5.1.2 Assessing Model Fit

This section will investigate whether the semi-parametric Cox model or the parametric Weibull model is best fitted to the data. A graphical Goodness-Of-Fit test, which is a comparison of the cumulative hazards function for a semi-parametric and parametric model, is used (Broström, na). The test indicates that the data does not follow a parametric distribution, indicating that the hazard rate should be left unspecified and that the semi-parametric Cox model should be used. The graph showing the distribution can be found in Figure A1, in the Appendix. The use of the Cox model is further substantiated after testing the models against each other using an Akaike Information Criterion (AIC hereafter). The AIC investigates the balance between good fit (high value of log-likelihood) and complexity, where complex models are more penalized than simple models. The method thus tries to find the model that both have few parameters but still fits the data well (Claeskens & Hjort, 2008, pp. 22-23). The interpretation of the AIC is that lower values indicate better fit, compared to models with a higher AIC. The AIC indicates that the Cox models have a better fit to the data than the Weibull models and the Cox model is therefore used in the main analysis. Golub (2008) have argued that the use of the AIC is "inherently limited because it provides only an informal means of discriminating between competing non-nested models" (p. 535). However, I argue that the AIC in combination with the graphical test is sufficient evidence for using a Cox model over the Weibull model, especially since the costs of imposing the wrong specification of the baseline hazard can be enormous (ibid, p. 536).

5.2 Diagnostics for the Cox Model

The Cox model makes several assumptions that are important to check, to make sure the data adequately fit the estimation technique chosen. If these assumptions are not tested, one runs the risk of having a biased model. There are especially three issues that should be considered when using survival analysis: (1) the proportional hazards assumption, (2) tied events, and (3) influential observations. Testing these assumptions and accounting for potential breaches is essential for ensuring the internal validity of the models.

5.2.1 Proportional Hazards

Both the Cox- and the Weibull model assume that the hazard rate is proportional (Box-Steffensmeier & Jones, 2004, p. 48). This means that both models assume that the hazard function of an observation follows the same pattern over time. When studying ceasefires, there are many theoretical reasons to expect that the effects of one or more predictor variables on the underlying hazard rate increases or decreases over time (Teachman & Hayward, 1993, p. 359). Consequently, the covariates may exhibit non-proportional hazards. Teachman and Hayward (1993), for instance, argued that "learning effects, shifts in life-course position, maturational changes and so on" can all cause this non-proportionality (p. 359). For example, one of the independent variables of this thesis is that earlier ceasefire failure found between warring factions will impact on future ceasefire duration. It is reasonable to expect that this effect is stronger for ceasefires that comes right after a failure, rather than a ceasefire that comes several years after the previous failure. If this is the case, the hazard rate of the covariate is non-proportional, something traditional event history analysis does not allow (Box-Steffensmeier, Reiter, & Zorn, 2003, p. 34).

Testing for whether the proportional hazard (PH) assumption is met is arguably the primary concern when fitting a model (Box-Steffensmeier & Jones, 2004, p. 131). It is important, as it can help verify that the researcher has correctly parameterized the model chosen (Cleves et al, 2012, p. 203). Indeed, misspecified proportional hazard models will overestimate the impact of variables whose associated hazards are increasing, while coefficient estimates for covariates in which the hazards are converging will be biased towards zero (Box-Steffensmeier & Jones 2004, p. 132).

Thus, regardless of whether you assume non-proportionality in your data or not, testing the assumption is necessary. Golub (2008) have argued that if non-proportionality is present in the data, the Cox model is superior to the Weibull model, as no adequate way for detecting and correcting the violations exists for parametric models (p. 537). If non-proportionality is present in the data, this provides another reason for choosing the Cox over the Weibull model. The proportional hazard assumption is investigated in the last section of this chapter, section 5.5, and will be further addressed also in chapter 6.

5.2.2 Influential Observations

When determining the adequacy of the fit of the model, it is also important to investigate whether any of the observations has a disproportionate influence on the estimated parameters (Cleves et al, 2012, p. 223). For example, when studying the duration of ceasefires, one wants to make sure that a single ceasefire, if removed from the data, will not make the relative hazard increase or decrease with a considerable amount (Box-Steffensmeier & Jones, 2004, p. 127). One way of examining whether influential observations are present in the model is through score residuals. Box-Steffensmeier and Jones (2004) say that when using score residuals, "we can plot the scaled change in a coefficient that would occur if observation *i* was dropped from the model" (p. 128). Hence, it tells whether the estimates would be affected if some of the observations were to be dropped. If an observation has large values of the score residual (large from 0), this indicates that it is influential, i.e. it has a high impact on the value of the coefficients. A plot showing the score residual of the different variables in the model suggest that influential observations are not a problem, as the DFbeta for all variables are smaller than -2 to 2 (Christophersen, 2013, p. 79). The plot can be seen in Figure A2 the Appendix. A plot using DFbetas to check for outliers also show that outliers are not a problem. This can be seen in Figure A3 in the Appendix.

5.2.3 Tied Data

The last issue that needs to be addressed is the possibility of the data being tied. Cases are tied when they have either identical event times or identical censored survival times (Golub, 2008, p. 539). Survival analysis assumes that the hazard function is continuous, which means that identical survival times are impossible. As a result, ties can be a serious problem (ibid).

When studying ceasefires, which are quite frequent events, it is plausible to expect that tied data will be present. This is especially true as the thesis studies all types of ceasefires, including unilateral ceasefires, which empirically have shown to often be reciprocated by other actors (hence having the same duration and failure time, while being separate observations). It is, therefore, reasonable to expect tied data to be present in the data. A plot showing whether the data is tied confirms this expectation: ties are indeed a problem[15]. At most, ten events have occurred concurrently. The problem when having tied data is that when two or more observations have identical event times, it is impossible to tell which of the events failed first. Consequently "it is not possible to discern precisely the composition of the risk set at the time of the failures" (Box-Steffensmeier & Jones, 2004, p. 54). For the Cox model, however, there are some simple methods that could be used to handle tied data. Two of these methods are the Breslow and Efron method. The Breslow method has shown to be adequate if the number of tied events is small at any given period. Given the data, where the number of tied events at the same time is large for several events, the approximation will be less precise. As an alternative, the Efron approximation can be used (Box-Steffensmeier & Jones, 2004, p. 55). A log likelihood and AIC test indicate that the Efron method for handling ties fits the data best and is the method that will be used in all the models. Lastly, as with data that exhibit non-proportionality, Golub (2008) argues that the ability the Cox model has of handling ties makes it superior to the parametric models (p. 539), which provides yet another reason for why the Cox model is a better choice for the data.

5.3 Operationalization of Variables

This section describes how the dependent variable *ceasefire duration*, and the three independent variables *earlier ceasefire failure; ceasefire comprehensiveness* and *number of actors in conflict* are operationalized.

5.3.1 Dependent Variable – Ceasefire Duration

This thesis is interested in studying events and their causes, and survival analysis is therefore used. Accordingly, the dependent variable is *ceasefire duration*, measuring the time between ceasefire onset and ceasefire failure. Ceasefire onset is the day the ceasefire is declared to take effect. The survival time of a ceasefire is determined by the occurrence of an event,

[15] The plot can be found in Figure A4 in the Appendix.

where ceasefire failure occurs when the battle-related death threshold is met, or when the ceasefire is announced to have failed.

Many ceasefires have an unclear ending, and as a result, battle-related deaths have been used to create a de facto ceasefire failure to supplement the already existing ceasefire failures in the data. No one has attempted to use battle-deaths as a proxy for ceasefire failure, and it thus becomes hard to both theoretically and empirically decide where the best cut-off point is. Gates and Strand (2004) have noted that the threshold chosen can lead to problems of selection bias (p. 4). Consequently, this thesis will investigate ceasefire failure where battle-deaths are set at different thresholds, namely where one-, 10-, and 25 are killed in one single event. The dataset also codes if the actors announce that the ceasefire has failed. These types of failures are included in instances where the BRD threshold is not met. This means that a ceasefire can either fail because it reaches the stipulated threshold, or because the actors to the ceasefire announce that it has failed.

To find battle-deaths, version 18.2 of the UCDP Georeferenced Event Dataset (GED) has been used. The dataset covers individual events of lethal violence by an organized actor (Croicu & Sundberg, 2017), and measures individual events in days. This is essential as ceasefire duration is also measured in days, meaning that when the ceasefire data is matched with GED, the ending can be stipulated on exact days. This is in line with Gates and Strand (2004), who note that measuring time in years lead to measurement problems (p. 13). The GED data also includes all events, regardless of whether the conflict is coded as active in the given year in the ACD. The ceasefire failure variable thus manages to capture failures also in conflicts with intermittent lulls. To create the variable, the GED and ceasefire data is matched. If a BRD at the given threshold occur after the start date of the ceasefire, the ceasefire is coded to have failed. Hence, the duration of the ceasefire starts with the initiation of the ceasefire and lapses on until the battle-death threshold is met. If a ceasefire starts but never reaches the threshold or is said to have failed, the ceasefire is censored on December 31, 2017.

5.3.2 Independent Variable – History of Ceasefire

The first independent variable *earlier ceasefire failure* is created using the ETH-PRIO Civil War Ceasefire Dataset. In the theory section, it was argued that conflicts having experienced failed ceasefires in the past would have more short-lived ceasefires in the present. To capture this effect, the variable *earlier ceasefire failure* is created. However, as no one has previously investigated the variation of ceasefire durability and the effect previous ceasefires might have on this, this is the first attempt to operationalize such a variable.

There are potentially many ways of operationalizing ceasefire failure. One way is to create a count variable that counts the number of times ceasefire failure has occurred, thus separating not only past failure from no past failure but also the degree to which past failure has occurred in the past. It might be plausible to expect it to be a difference between conflicts having experienced ceasefire failure once before, and conflicts having experienced it many times. Another way is to create a simple binary variable, capturing whether a conflict-dyad has experienced past ceasefires in the past or not. This is the way I have done it. Arguably, it might be more precise to use a count variable, as it is able to capture whether there is a variation within the ceasefires that have experienced failure or not. Yet, the angling of the hypothesis suggests that the binary variable is efficient for capturing the effect I am interested in. The wording of the hypothesis suggests that the interest is not in the variation within past failures, but rather if past failure overall has any impact on duration. For future research, however, expanding this variable can be interesting.

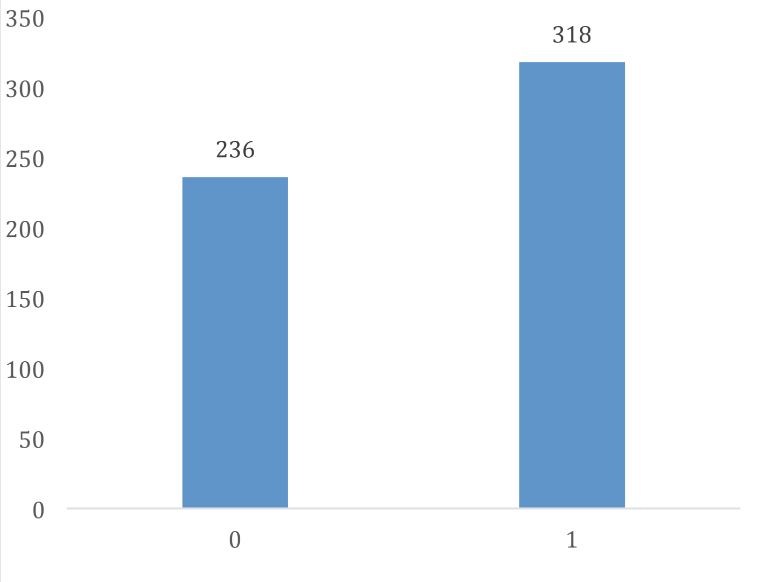
It is not just expanding the variable with relation to counting the previous failures that can be of interest. One might also want to include how long ago the failure occurred, or if there is a variation depending on the comprehensiveness of the ceasefire. As mentioned, the variable as it is used in this thesis only measures whether ceasefire failure has occurred or not. This means that the variable does not separate between failures that have occurred quite recently and failures that occurred many years ago. As argued in the methods chapter, it might be reasonable to expect that the effect of ceasefire failure is stronger for ceasefires that comes right after a failure, rather than a ceasefire that comes several years after previous failure. This is an effect that could be interesting to investigate in depth. A different variation that could be interesting to capture relates to the second independent variable, *ceasefire comprehensiveness*, and could measure whether the past failure was an informal truce, or if it was comprehensive. One might expect that the failure of a comprehensive ceasefire in the past will impact future ceasefires to a larger degree than failure of informal truces. Both the count variable, the time-since-event and comprehensiveness measures are all variations of the *earlier ceasefire failure* that could be included in the future when investigating the effect failed ceasefires in the past have on ceasefire duration. For now, a binary variable is used, as this simplistic way of investigating *earlier ceasefire failure* rightfully captures the hypothesis of interest.

There is one potential problem with how the variable is created that needs to be addressed. As mentioned in the section describing the dataset, the CWCD is based on the UCDP/PRIO ACD. However, there is one large difference between these datasets that is consequential for the creation of this variable. The ACD records conflicts in the period 1946-2017, whereas the CWCD only codes ceasefires found in conflicts between 1989-2017. There are thus 43 years "missing" from the CWCD, and some of these conflicts have probably had a ceasefire in these 43 years. The variable, however, is unable to capture the effect of failed ceasefires before the start of the dataset, which means that the first ceasefire found in each conflict-dyad will be coded to not having experienced ceasefire failure in the past, although this is not something we know. Although this is a potential problem, because it is a systematic "error" i.e. the same applies to all the first cases, the problem is less troublesome. In addition, the variable underestimates the effect of earlier ceasefire, as it gives a "no failure" to all of the first failures in a conflict-dyad, rather than overestimating it. Finding an effect of earlier ceasefire failure, although the data underestimates the effect, therefore strengthens any potential results.

The variable used in this thesis is therefore a binary variable, given a 0 if no past ceasefire failure has occurred, and 1 otherwise. The variable is created through ordering and a loop function. The ordering is the first step, where the data is ordered according to the conflict the ceasefire is part of, then the date the ceasefire took effect, and lastly the dyad in which the ceasefire was part of. Correctly ordering the data is important when using a loop, otherwise, it will assign wrong values to the variable one tries to create. After the ordering, a loop is implemented. The loop runs through the data, assigning a value of 1 if the current value belongs to the sequence (i.e. if it meets the conditions stipulated). The conditions stipulate that past failure must occur within a conflict-dyad and that the previous ceasefire must have failed. After the loop is run, 318 ceasefires in the dataset have experienced failure in the past (1), compared to 236 that has not (0). Figure 5.2 shows the distribution in a histogram.

Generally, it is recommended to avoid the use of loops in R, as R supports vectorization (Theuwissen, 2015). This means that computing using loops will make the calculations slower than if ones use vectorization instead. However, the dataset does not have enough observations for this to become a problem, and this is therefore not considered problematic.

Figure 5.2: Ceasefire History



5.3.3 Independent Variable – Comprehensiveness of Ceasefire

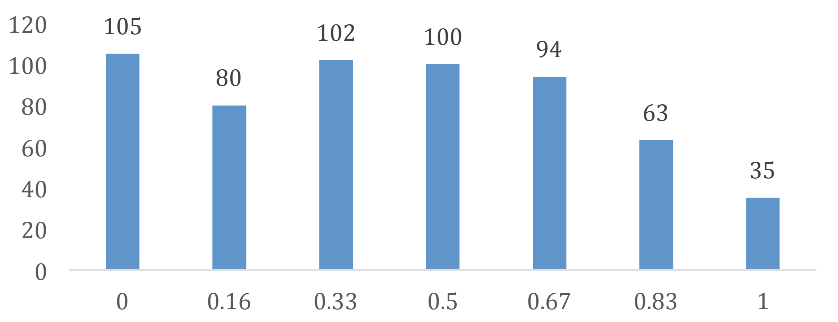
The second independent variable *ceasefire comprehensiveness* is also created using the ETH-PRIO CWCD. This variable tries to capture whether informal and comprehensive ceasefires differently affect the duration of the ceasefire.

As argued in the theory chapter, there is an expectation for comprehensive ceasefires to last longer than informal truces. This is in line with previous literature, which has suggested that "mechanisms implemented in the context of cease-fire agreements can help reduce the risk of another war" (Fortna 2003b, p. 365). What is not so clear-cut is what is meant by informal ceasefire and comprehensive ceasefire. Fortna (2003b), for example, have emphasized the importance of including mechanisms such as demilitarized zones, enforcement missions and the signing of formal agreements to have more long-lasting ceasefires (p. 363). Following Fortna (2003b), six variables from the CWCD seem relevant for measuring comprehensiveness. These are: *Written, Mediation, Implement, Enforcement* and *DDR*. The *written* variable captures whether the ceasefire declared was signed or not. Originally, the variable takes three categories: no written agreement (0), within the country hosting civil conflict (1), and on third party-territory (2). The two last categories (1 and 2) are merged, so the variable is a binary variable, measuring whether the agreement was signed or not. The *mediation* variable stipulates whether the ceasefire was mediated or not and is a binary variable. The *implement* variable is also binary and measures whether agreement on any mechanisms regarding the implementation of the ceasefire is included in the ceasefire. Such mechanisms can be provisions for monitoring, verification, separation of forces, or the establishment of a ceasefire committee (Clayton et al., 2019a). *Enforcement* captures whether any third-party troops are employed with the ceasefire. Originally, it has 4 different categories: no enforcement (0), enforcement by external state (1), enforcement by an international organization (2), and other (3). As the interest is whether enforcement measures were implemented or not, categories 1-3 are merged into one category, making it into a binary variable. *DDR* is the fifth and last variable from the CWCD and stands for disarmament, demobilization and reintegration. This variable capture whether there was any reference to DDR in the ceasefire, where it is enough that only one of the three mechanisms is present.

What becomes clear with some of these variables is that there can be broad variation within them, but that this variation is not captured. The variable *implement*, for instance, does not investigate what kind of implementation measure is present but rather gives an indication whether it was present or not. Some ceasefires probably include several implementation mechanisms, whereas others barely include 1. However, the variable treats these two cases as the same, assigning both ceasefires that mechanisms were agreed. The same goes for the variable *DDR*, as it only captures whether one or more was present or not. It thus fails to differentiate between ceasefires that mention all three aspects and ceasefires with reference to only one. One might argue that the variables cannot fully capture the comprehensiveness of the ceasefire. I agree to some extent. However, the variable *ceasefire comprehensiveness* is created using all five variables found in the CWCD, and this is a better option than looking at the variables in isolation. Indeed, it is plausible to expect that the most comprehensive ceasefires do not just include implementation mechanisms, but also have reference to enforcement or DDR, are written, and perhaps also mediated. As the mean value of all the variables are taken to create the new variable, I argue the variable capture the differences in comprehensiveness quite good.

In addition to the variables in the CWCD, one variable has manually been created for the thesis. The variable, labelled *bilateral_unilateral* captures whether the ceasefire was unilateral (0), bilateral (1) or multilateral (2). This variable is included as an additional measure of *ceasefire comprehensiveness.* This is done, as bilateral and multilateral ceasefires are harder to reach than unilateral ones, as one group can declare unilateral ceasefires alone. As with the other variables, the variable is merged into two categories, stating whether it was unilateral or not. The variable was then included when creating the ceasefire comprehensiveness. The complete *ceasefire comprehensiveness* variable thus includes the mean of six variables; *written, mediation, implementation, enforcement, DDR* and *bilateral_unilateral*, and is measured on a scale from 0 to 1, where 0 indicate no mechanisms, and 1 indicate all mechanisms. The distribution is shown in figure 5.3.

Figure 5.3: Comprehensiveness of Ceasefire



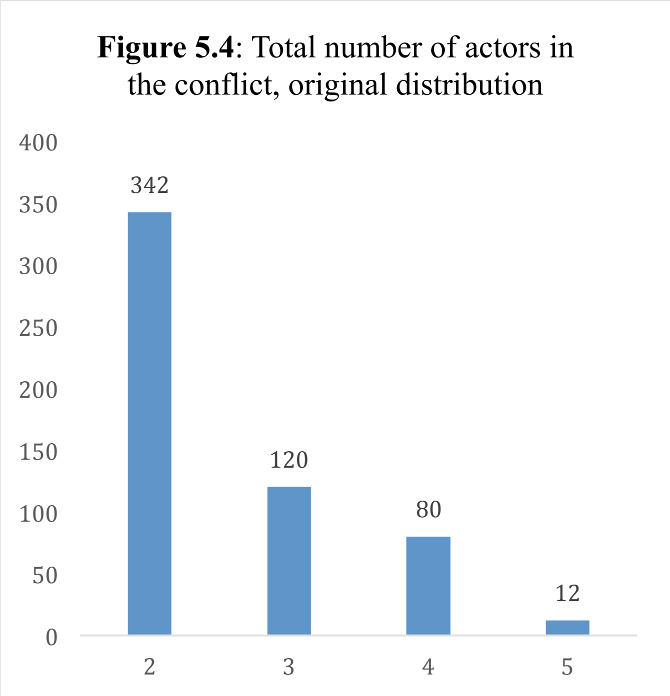
5.3.4 Independent Variable – Parties to the Ceasefire

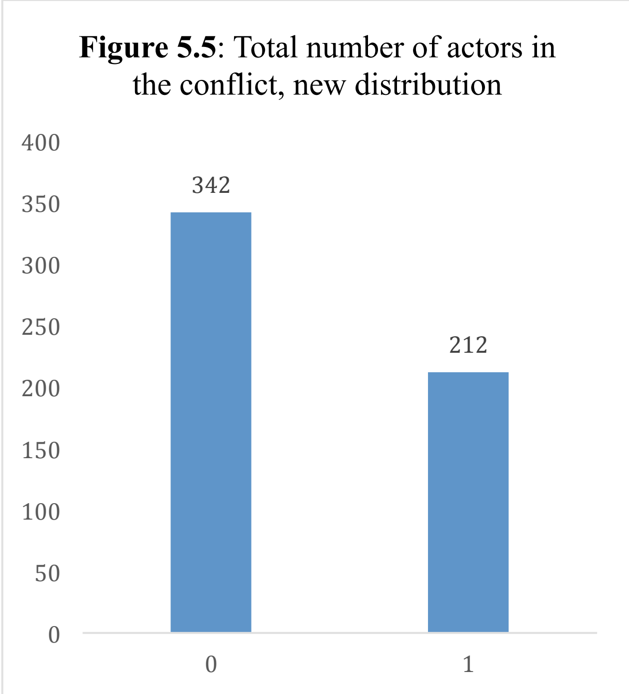
The last independent variable *number of actors* is created using version 18.2 of the UCDP/PRIO Armed Conflict Dataset (ACD). The third hypothesis of this thesis states that ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts containing only one rebel group. The hypothesis thus argues that it is the number of rebel groups in the conflict that impacts on ceasefire duration. As such, the variable is a simple count variable, where the number of participants in the conflict are grouped together. The ACD has information on the different actors that are active in a conflict in a given year. These actors are then disaggregated, grouped together with other actors found in the same conflict in the same year. All actors fighting in a conflict are, according to the UCDP definition of civil conflict, fighting over the same incompatibility. As such, actors fighting the same conflict are grouped together as a *rebel movement*. The government is included in the measure of the number of rebel groups in the conflict, so the smallest number of members in a conflict can be two participants: one rebel group and the government.

A difficulty that arises when creating the variable is that the CWCD codes ceasefires found in conflict-years where the ACD has not listed them as an active conflict (i.e. not reached the 25 BRD). As such, when matching the two datasets together, the variable receives a missing in situations where a ceasefire is found in a year that was not active. To deal with this issue, I have manually filled in the number of rebel groups in the conflict that year, giving it the same value as recorded in UCDP/PRIO ACD the previous year. If the conflict was not active the year before that either, a maximum of three years back in time is coded, following the coding from the CWCD.

There is a very right-skewed distribution regarding the actors to the conflicts the ceasefires occur in, seen in Figure 5.4. Figure 5.4 shows that most of the ceasefires in the dataset have consisted of the government and one rebel group only. Because there are relatively few ceasefires found in conflicts with 4 or more participants, the variable is made into a binary variable, separating conflicts including the government and one rebel group (given a 0), and conflicts with more than the government and one rebel group (1). Making it into a binary variable reduces the skewness, while also making the results a lot more intuitive to interpret, which is an advantage. The new distribution is shown in Figure 5.5.

This thesis measures fragmentation as the number of actors in a conflict. The literature on fragmentation, however, has stressed that opposition movements are more than just the sum of armed groups working towards a similar cause (Bakke, Cunningham & Seymour, 2012, p. 267). One can, therefore, argue that there is a substantial gap between the suggested measurements of fragmentation, and how it is done in this thesis. Yet, data on rebel movement fragmentation is still sparse, and to my knowledge, no comprehensive data on rebel groups exist for all civil conflicts between 1989-2017. The closest relevant dataset is the Ethnic Power Relations dataset from ETH, which identifies “all politically relevant ethnic groups and their access to state power in every country of the world from 1946 to 2017” (Vogt et al., 2015). However, the focus of the thesis is not on ethnic conflicts, but rather on rebel group- and conflict characteristics. Fragmentation has therefore been simplified to be the sum of armed groups in a conflict, which is also in line with hypothesis three of the thesis, namely that it is the number of rebel groups that influences ceasefire duration.





5.4 Control Variables

To establish if there is a relationship between the independent variables and dependent variable, it is important to control for factors that are likely to influence both. When choosing control variables, researchers often face a trade-off between bias and variance. Leaving out variables will usually introduce bias in the estimators, whereas having too few variables means having a smaller variance, as there are fewer unknown parameters to estimate (Claeskens & Hjort, 2008, p. 12). Ideally, a researcher should seek to find the proper balance between overfitting a model, meaning that too many control variables are included, and underfitting it, meaning that the researcher has too few parameters that do not manage to capture the signals correctly (p. 2). If a researcher has under fitted a model, omitted variable bias can be present, which means that some unmeasured factor could be causing the correlation between the dependent and independent variables. Finding the proper balance between parsimony and omitted variable bias, however, can be very challenging, especially when small amounts of research have been done on the subject. To ensure that relevant variables are included, a mix of theoretical expectations and statistics is employed.

Ceasefire duration is an understudied area, as emphasized in the literature chapter. Consequently, no broad consensus on important parameters to include when investigating ceasefire duration exists. I will therefore look to other related fields of study, such as the literature on the duration of peace, the literature on duration of war, and the literature on fragmentation, to find potential control variables to use in the analysis. These control variables are then statistically tested using an Akaike Information Criterion (AIC), to investigate whether the controls chosen brings anything new to the model. If not, they are discarded. This is done to hopefully find the proper balance between parsimony and omitted variable bias. The AIC investigates the balance between good fit (high value of log-likelihood) and complexity, where complex models are more penalized than simple models. The method thus tries to find the models that both have few parameters but still fit the data well (Claeskens & Hjort, 2008, p 22-23). The interpretation of the AIC is that lower values indicate parsimony, compared to models with a higher AIC.

The control variables presented below are all found from related fields of literature or are included out of own theoretical expectations. In addition, all the control variables included have shown that they give the model a lower AIC, and hence the models have a better fit when including them than excluding them[16]. A plot is run to make sure that the control variables are not correlated. The plot shows that correlation is not a problem. The plot can be found in figure A5 in the Appendix.

Conflict-level variables. Several characteristics of the conflict can affect the length of the ceasefire, the number of groups in the conflict, the history of the ceasefires and the comprehensiveness of the ceasefire. As such, the control variables *battle death* and *duration of conflict* are included. *Battle-death* in the conflict is included following the general agreement in the duration of peace literature arguing that civil wars having high-death tolls will be more likely to resume than less deadly conflicts (Fortna, 2003a; Hartzell, Hoddie & Rothchild, 2001). Likewise, Karakus and Svensson (2017) argued that ceasefires are likely to be less durable in difficult situations, underlining high-intensity fighting as one of these reasons (p. 9). High-death tolls can also lead to fragmentation. Mosinger (2018), for instance, emphasized that violence in a conflict can cause disagreements between groups in a rebel movement, causing the groups to fragment (pp. 71-72). I also argue that many BRD in a conflict can negatively influence the trust between the combatants, increasing the probability that a conflict-dyad has experienced a failed ceasefire in the past. Because of these arguments, a control variable to capture the number of deaths in a conflict is included. The variable is retrieved from Quality of Government Standard Dataset and measures the number

[16] The test is performed by comparing a model with only the dependent variable and independent variables, with a model where the control variables are included one at a time. This process is repeated for all the control variables.

of deaths in a conflict-dyad between warring factions, including civilian casualties. The variable is originally from UCDP and is measured from 1989-2017. This means that the conflicts that started before 1989 have only the battle-deaths recorded after 1989. This can bias the results. For example, Afghanistan has, according to the Armed Conflict Dataset, experienced war (more than 1000 people killed in a year) almost every year since 1978. This means that the variable only captures the battle deaths occurring after 1989, and thus have at least 10.000 deaths not included. Yet, to my knowledge, no variable exists that captures battle-death in conflict from 1946-1989, which is the start date of the ACD. Although some battle-deaths are consequently excluded from the analysis, I argue that this works against my theoretical argument rather than strengthens it. The theoretical expectation is that high death tolls in a conflict will lead to less durable ceasefires. The variable thus underestimates the effect, so if there is an effect of high death tolls this strengthens the argument. The variable is highly skewed, so to make it more normally distributed the variable is log transformed.

The second conflict-level variable is the *duration of conflict prior to the ceasefire*. Hartzell et al (2001) have contended that longer wars have a higher likelihood of stable negotiated settlement (p. 190). Karakus and Svensson (2017), however, have claimed that intractable conflicts can impact on ceasefire duration, as these conflicts are probably more difficult to solve (p. 9). Hence, there is disagreement in the literature over whether longer conflicts lead to stable agreements. The duration of conflict can also impact on the number of rebel groups. Mosinger (2018) have contemplated that maintaining unity amongst rebels becomes more difficult as the conflict drags on (pp. 71-72). This means that longer conflicts have a higher probability of experiencing fragmentation. When it comes to previous ceasefire in the conflict, it is reasonable to expect longer conflicts to have experienced more ceasefires than shorter conflicts and are thus more susceptible to having experienced a failed ceasefire in the past. Similarly, longer conflicts are probably more prone to have experienced negotiations, and thus are more capable in making comprehensive ceasefires. Yet, longer conflicts might also have fueled deeper mistrust between belligerents, making it harder to negotiate comprehensive ceasefires. A measure of conflict duration before the ceasefire is therefore included, to see whether longer conflicts lengthens or shortens the duration of ceasefires. The variable is created using the ACD and CWCD, measuring the difference in days between the start date of the conflict until the start-date of the ceasefire. This variable was also log transformed, as the distribution was skewed.

Country-level variables. In addition to including variables on the conflict-level, country-level variables are also included. A country can experience several civil conflicts simultaneously, and because of this, it is important to also include some country-level characteristics to capture whether there are any independent effects from the conflict-level variables. In addition, it is expected that the structural characteristics of a country can influence the adversaries' ability to trust and cooperate with each other. Two country-level variables are included: *Regime type* and *state resources*. The level of *democracy* is included as a control variable, as Derouen, Lea and Wallensteen (2009) have argued that the nature of government might play a role in the duration of an agreement, as higher democracy scores might lead to more successful peace negotiations, hence leading to higher endurance of agreements (p. 379). Similarly, I expect democracies to better provide assurance to the rebels that they will comply with a ceasefire than less democratic states. Violation of a ceasefire from the government side will be a more serious infringement compared to less democratic states, and democracies will thus be held more accountable by their population if the ceasefire fails. One might, therefore, expect that conflicts in democracies have not experienced failed ceasefires in the past. The variable is retrieved from the Quality of Government Standard dataset (Teorell et al., 2019). The variable is originally from the Freedom House and is a variable composed of several variables to measure the level of democracy. The variable includes political rights, civil liberties and the level of democracy. The variables are measured on a scale from 0-10, and these scores are averaged to create the level of democracy variable. The scale of the variable ranges from 0-10, where 0 is least democratic and 10 is most democratic. It has been suggested that this "averaged index performs better in terms of validity and reliability than its constituent parts" (Teorell et al., 2019, p. 301). The variable has been lagged 1 year, to deal with possible autocorrelation. Autocorrelation means that a time series is correlated with a delayed edition of itself.

State resources are also important to control for. There are several ways of measuring state resources, but this thesis follow Cunningham (2011), who uses GDP per capita as a proxy for state resources (p. 89). He argues that the measure can also be used as a proxy for state capacity and strength. Fearon and Laitin (2003) have argued that GDP per capita captures "a state's overall financial, administrative, police and military capabilities" (p. 80). A weak state will probably face larger difficulties fighting off the rebels, indeed, low state capacity favor insurgency (ibid). A weak state will thus face difficulties defeating the rebels, which can indicate that the power balance between the rebels and government are more symmetric than in places where the state has stronger state capacity. Both the rebels and government might realize that breaking a ceasefire is an easy and strategic way to get an advantage over its opponents, causing ceasefires in weak states to fail faster than ceasefires found in stronger states. This variable is also retrieved from the Quality of Government dataset (Teorell et al, 2019) but is originally from World Bank Group. It measures the gross domestic product divided by midyear population, in current US dollars. As the variable is highly skewed, it is log transformed. Both conflict-level variables are measured on a yearly basis.

Control variables not included. Several other control variables were considered to be included in the analysis. These were *type of conflict*, *ethnic fractionalization* and *quality of government*. Type of conflict was not included as it showed to not have any significant positive effect on the overall fit of the model, and thus did not decrease the AIC of the model. Ethnic fractionalization was originally included following the literature but was discarded as there were no good theoretical explanations for including it with regards to ceasefire failure. It was also discarded as the variable had many missing values, thus decreasing the number of observations in the main analysis. This was also the reason for not including quality of government, as it decreased the number of observations drastically when including it. After checking the control variables then, these control variables were not used in the analysis after all.

5.5 Checking the Proportional Hazard Assumption

As discussed in section 5.2.1, it is important to check the proportional hazard assumption when doing survival analysis. This section will investigate whether any of the variables in the data violates this assumption.

One widely accepted way to test the proportionality assumption is to look at the Schoenfeld residuals after fitting a model (Box-Steffensmeier & Jones, 2004, p. 120). If the values returned are significant (p> 0.05), the variable is non-proportional. This method was used on the main model of the thesis, to check whether the covariates were proportional[17]. Two variables failed the test: the independent variable *earlier ceasefire failure* (0.035), and the control variable *log GDP/per capita* (0.0035). Moreover, the global test showed that the overall model did not meet the proportionality assumption (0.000329). Since the proportionality assumption was broken for one of the main explanatory variables, and the

[17] The plot showing the Schoenfeld residuals can be found in Figure A5 in the Appendix along with a test of the PH assumption, found in table A2. Table A2A show the variables after stratification has been done.

global test showed that non-proportionality was a problem for the overall model, this indicates that non-proportionality needs to be corrected for in the model. As such, the choice to use the Cox model over the Weibull is further substantiated, as the Cox model is superior to the Weibull model if the data is non-proportional (Golub, 2008, p. 537).

Therneau, Crowson and Atkinson (2019) argue that one of the simplest methods for correcting for non-proportionality is by using a step function, meaning that one splits the whole sample into subgroups, and uses different coefficients over the different time intervals (p. 17). With the example of the independent variable mentioned in section 5.2.1, for example, it was argued that it is reasonable to expect the change in time to be different from earlier periods after failure than later periods after failure. Using the step function, one can break the dataset into time-dependent parts, thus separating earlier periods after failure, from later periods after failure. The difficulty that arises is to determine what periods to break the data into?

A look at the data reveals that more than 50% of the ceasefires fail within the first 50 days after initiation. The cut-off point chosen when dividing the data into time-dependent parts is therefore at 50 days; the first 0-50 days, from 50-100 days, and greater than 100 days. The reason for choosing such a small cut-off relates to the fact that most failures occur quite early in the data. Indeed, after the first 100 days, only 75 more failures occur[18]. After stratifying the data into time-dependent parts, only the third period of the *log GDP/per capita* variable remains non-proportional, whereas all the periods of *earlier ceasefire failure* variable meet the PH assumption with a great margin. In addition, the global test indicates that non-proportionality is not a problem for the overall fit of the model. As *log GDP/per capita* only violates the proportionality assumption for one of the time periods, and the global test showed that the model in its entirety meets the assumption, the PH assumption is not regarded as a problem in the new model. Nonetheless, it has been suggested that the global test sometimes falsely indicate that the proportionality assumption is met (Box-Steffensmeier et al., 2003, p. 45; Golub, 2008, p. 537). Consequently, the PH assumption and the robustness of the models are addressed in more detail in chapter 6.

[18] 384 failures occurred in total.

6 Empirical Analysis

In this chapter, the analysis of the hypotheses is presented. In the theory section, I argued that because of trust and commitment issues, keeping a truce will be more difficult in situations where ceasefires have failed in the past, when the cost of defection are low (i.e. less comprehensive ceasefires), and when a greater number of rebel groups are present in the conflict, as possibilities of spoiling increase. The thesis thus posits three main arguments to explain why some ceasefires fail faster than others: (1) the history of ceasefires, (2) the comprehensiveness of ceasefires, as well as the (3) parties to a ceasefire. This section will investigate these claims using survival analysis, and more specifically the Cox model. The chapter is structured as follows. It starts with investigating the descriptive statistics, focusing on the three different hypotheses. It follows with the main empirical analysis, where the results are presented. The chapter then investigates the results when the ceasefire failure measure is changed to respectively 10- and 25 battle-deaths, before discussing the robustness of the models. Lastly, a summary of the findings is presented.

6.1 Descriptive Statistics

This section is divided into three parts, where the descriptive statistics for each of the hypotheses are introduced and discussed. This is done to see if there exist a connection between the dependent variable *ceasefire duration* and the three different explanatory variables. Descriptive statistics are used to describe the data in a meaningful way, making it easier to see whether there are some patterns in the data.

6.1.1 History of Ceasefires

The first hypothesis of the thesis posits that the history of ceasefires, i.e. whether a conflict-dyad has experienced ceasefire failure in the past, will impact on the duration of ceasefires. Table 6.1 shows the frequency of ceasefires that have experienced earlier failure and reveals that approximately 55% of all ceasefires declared in a conflict-dyad have experienced ceasefire failure in the past, whereas the rest have not. This number, as discussed in the operationalization of the variable, might be a bit misleading as the variable only records ceasefires occurring after 1989. The rate of ceasefires found in a conflict that have a history of failed ceasefires might, therefore, be larger than presented here.

Table 6.1: Frequency of Earlier Ceasefire Failure

Earlier Ceasefire Failure	N	%
0	260	44.9
1	319	55.1
Total	579	100.00

Table 6.2 expands on table 6.1 by including how many of the ceasefires experienced an event, i.e. ceasefire failure occurred[19]. It shows that approximately 93% of all ceasefires that had experienced failed ceasefires in the past failed, compared to 33% of the ceasefires that had not experienced earlier failure. In other words, almost a third more ceasefires failed in conflict-dyads where there was a history of ceasefire failure. This provides preliminary support for hypothesis one of this thesis. Yet, these numbers tell us nothing whether ceasefire failure occurred faster if the conflict-dyad has experienced failed ceasefire in the past. A great way to investigate the different survival times for the variable *earlier ceasefire failure* is by visualizing a Kaplan-Meier survival curve.

Table 6.2: Incident rate by Earlier Ceasefire Failure

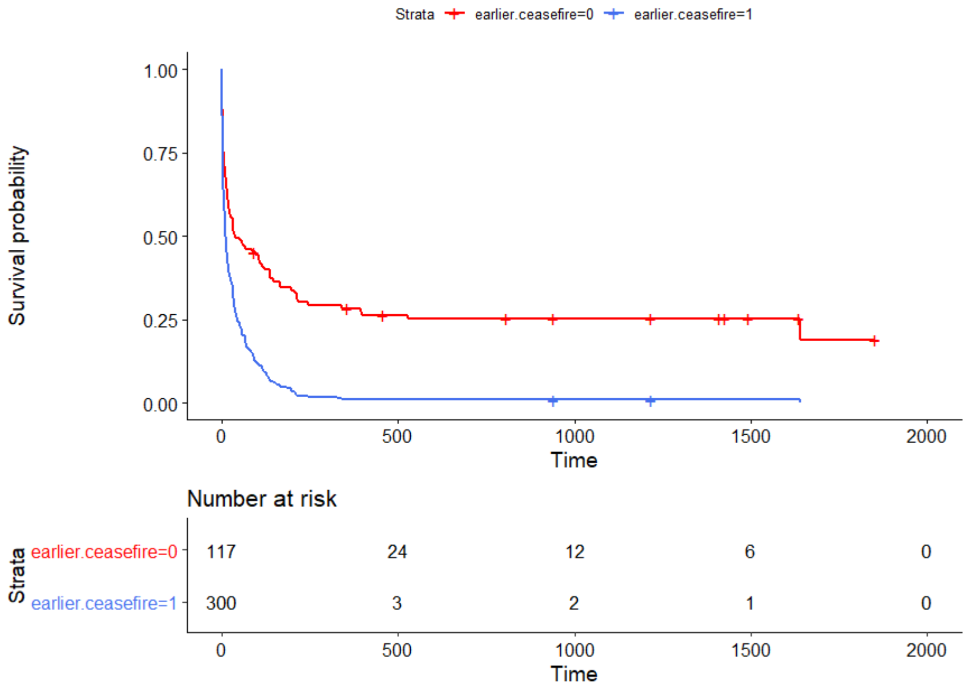
Earlier Ceasefire Failure	N	Events	%
0	260	87	33.46
1	319	297	93.10
Total	579	384	66.32

The Kaplan-Meier survival curve estimates the survival times of a ceasefire, meaning the difference in time from ceasefire onset until failure occur. If failure does not occur, the observation is right-censored. Figure 6.1 shows the difference in survival times between ceasefires that have not experienced failed ceasefire in the past (red line) and ceasefires that have (blue line). The figure has been limited to show only a duration of 2000 days. This is done to get a better impression of the failure of the ceasefires, as most of the ceasefires fail rather quickly. Indeed, the longest surviving ceasefire in the dataset lasted 1639 days[20].

[19] Failure measured using the strictest measure of 1 BRD.

[20] By only looking at duration of 2000 days, 146 observations that are right-censored are excluded from the graph. The original Kaplan-Meier plot with the whole duration period can be found in Figure A7 in the Appendix.

Figure 6.1: Kaplan-Meier Survival estimate of Earlier Ceasefire Failure



The figure shows that there is a difference in how fast a ceasefire fails depending on whether ceasefire failure has occurred in a conflict-dyad previously or not. The figure shows that the survival probability is extremely low and quite similar in the beginning, but after it reaches around 50%, the two lines separate. This indicates that ceasefires, in general, tend to fail fast regardless of its history right after initiation, but that after a certain time, ceasefires having experienced previous failure in the past fail faster than those who have not. In addition, the survival probability flattens out after about 500 days, meaning that ceasefires that have held for a long time generally continue to hold. The figure also shows that when ceasefires are declared in conflicts-dyads that previously have not had any failures (0), approximately 25% tend to survive beyond our coding period (i.e. they are right censored). For the ceasefires that have experienced a failed ceasefire in the past in the conflict-dyads, however, the survival rate is less than 10%. This is a profound difference, indicating that the history of ceasefires indeed impacts on their duration. To assess whether these preliminary findings are reliable, a log-rank test on the variable is run (Everitt & Hothorn, 2014, p 3). The log-rank test is "used to test the null hypothesis that there is no difference between the populations in the probability of an event […] at any time point" (Bland & Altman, 2004, p. 1073). Thus, if the p-value is significant ($p < 0.05$), it indicates that the results from the Kaplan-Meier plot are reliable and significant. The log-rank test for the variable *earlier ceasefire failure* indeed indicates that the survival times are different for both groups; hence, the p-value was significant. This brings further preliminary support for the first hypothesis, namely that ceasefire history impacts on ceasefire duration. The log-rank is presented in table A8 in the Appendix.

6.1.2 Comprehensiveness of Ceasefire

The second hypothesis of the thesis is that more comprehensive ceasefires should last longer than less comprehensive ceasefires. Table 6.3 shows the frequency of the comprehensiveness of the different ceasefires, ranging from 0 mechanisms to all mechanisms (6). It shows that the ceasefires are quite evenly distributed from 0-4 mechanisms, but that quite few ceasefires had 5 and 6 mechanisms. Table 6.4 expands table 6.3 by including how many of the ceasefires failed, i.e. event occurred. The table shows that the failure rate is highest for the ceasefires having no measures (90%), and then the failure rate steady decline until 4 measures are included (48%). From 0 measures to 4 measures, then, the survival rate almost doubles. Surprisingly, both at 5 and 6 measures, the failure rate is higher than when only 4 measures are included, but only slightly (59% for 5 measures, and 54% for all measures). This suggests that ceasefires with no measures at all are extremely fragile, and when several measures are included, they become more stable. However, it also suggests that after a certain point (i.e. 4 measures) they do not become any more stable, but rather proposes the opposite. This change, however, is marginal, and can be driven by the fact that so few ceasefires included more than 4 measures. These numbers provide preliminary evidence for my second hypothesis, namely that more comprehensive ceasefires are more stable than less comprehensive ceasefires.

Table 6.3: Frequency of Comprehensiveness of Ceasefire

Comprehensiveness Ceasefire	N	%
0	105	18.13
1	80	13.82
2	102	17.62
3	100	17.27
4	94	16.23
5	63	10.88
6	35	6.04
Total	579	100.00

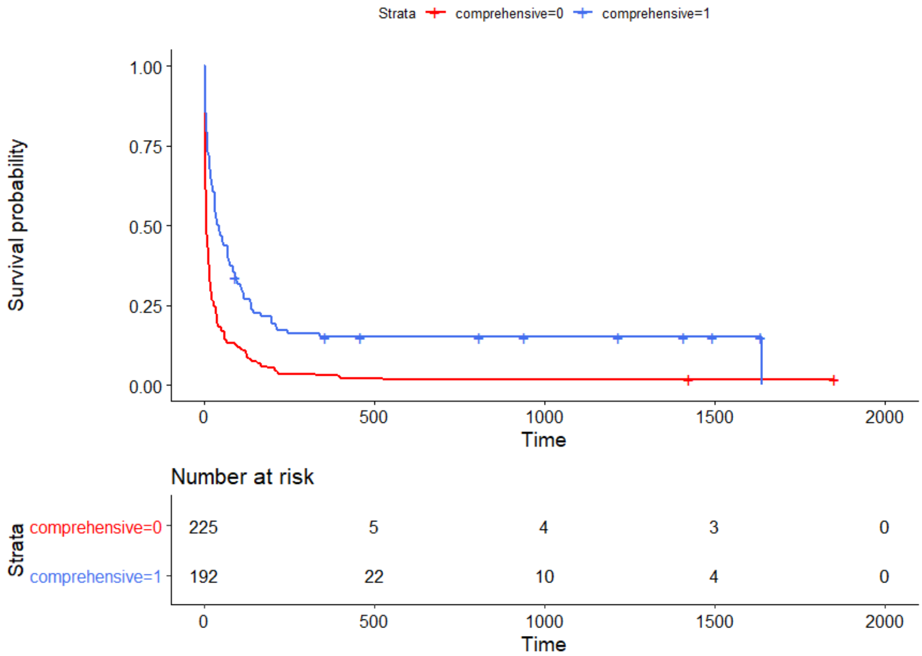
Table 6.4: Incident rate by Comprehensiveness of Ceasefire

Comprehensiveness Ceasefire	N	Events	%
0	105	95	90.48
1	80	58	72.5
2	102	68	66.67
3	100	62	62
4	94	46	47.92
5	63	37	58.73
6	35	19	54.29
Total	579	385	66.49

To investigate whether failure occurred faster for the less comprehensive ceasefires, a Kaplan-Meier survival curve is visualized. The Kaplan-Meier plot is seen in Figure 6.2. It shows the difference in survival times between comprehensive ceasefires (blue line) and less comprehensive ceasefires (red line). For visualization, the variable is made into a binary variable, where less comprehensive ceasefires have received a 0 (>0.5), and comprehensive ceasefires have received a 1 (<0.5). This plot is also limited to the first 2000 days after ceasefire initiation, as this allows for investigating more closely the effect of ceasefire comprehensiveness on survival rate[21]. The figure shows that there is indeed a difference in survival rate between the types of ceasefires. The lines start to separate after approximately 25% of the ceasefires have failed (survival probability 0.75). After that, one can see that less comprehensive ceasefires fail at a faster pace than the comprehensive ceasefires. The plot also suggests, as is similar to the Kaplan-Meier plot of *earlier ceasefire failure*, that the ceasefires in general tend to fail fast, but after it has survived approximately 400 days, the survival line flattens out. When assessing the reliability of the findings using a log-rank test, it suggests that the survival rate is significant. This means that the survival times for the categories are different, and that the result from the Kaplan-Meier is indeed reliable. This provides further preliminary support for hypothesis two. The log-rank test can be found in table A9 in the Appendix.

[21] The plot for the whole period can be found in Figure A8 in the Appendix.

Figure 6.2: Kaplan-Meier Survival estimate of Comprehensivness of Ceasefire



6.1.3 Parties to a Ceasefire

The last hypothesis of the thesis posits that ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts containing only one rebel group. Table 6.5 shows the frequency of ceasefires that are declared in conflicts with one rebel group (0) and more than one rebel group (1). The table indicates that approximately 38% of the ceasefires in the dataset are found in conflicts containing more than one rebel group fighting the state.

Table 6.5: Frequency of Number of Actors

Number of Actors	N	%
0	342	61.73
1	212	38.27
Total	554	100.00

Table 6.6 shows how many of the ceasefires that experienced failure. Surprisingly, 73% of the ceasefires found in conflicts containing only one rebel group failed, compared to only 63% of the ones that were found in conflicts with more than one group. This is an interesting finding, suggesting that ceasefires found in conflicts with only one rebel group aren't necessarily more stable than ceasefires found in multiparty civil war. This is contrary to the expectations, as I anticipated ceasefire failure to be more probable in multi-party civil conflicts.

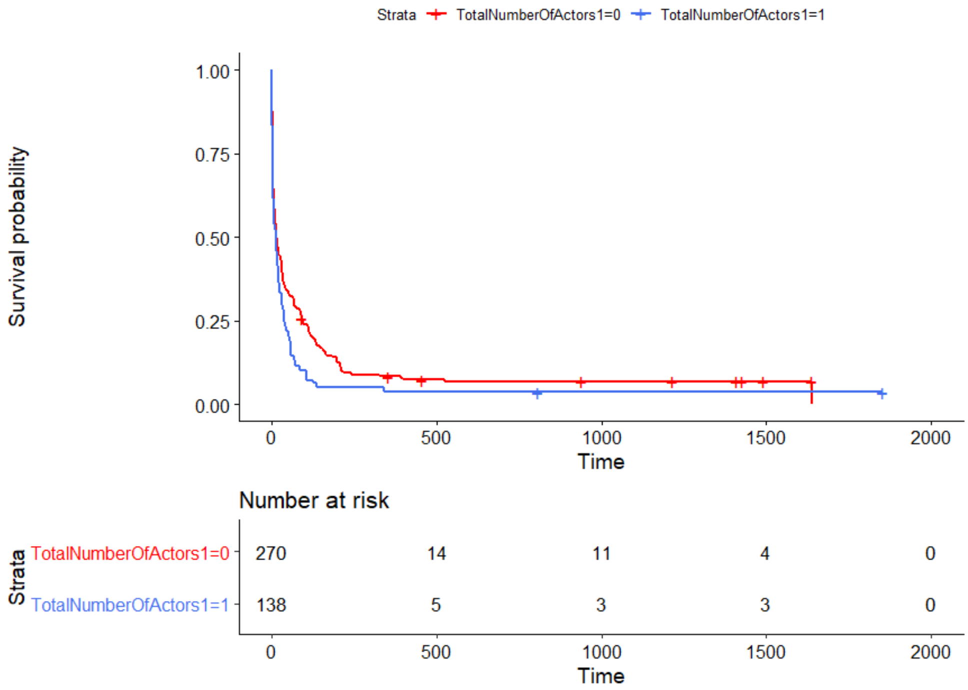
Table 6.6: Incident rate by Number of Actors

Number of Actors	N	Events	%
0	342	251	73.39
1	212	133	62.74
Total	554	384	69.31

To see whether there is a difference in survival times depending on the number of actors in the conflict, a Kaplan-Meier survival curve is estimated. Figure 6.3 illustrates the survival probability for the two outcomes of the variable, limiting the duration also here 2000 days[22]. The figure illustrates similar results to figure 6.2, but the results are less profound. Indeed, it seems like the survival probability is nearly identical for both types of ceasefires until approximately 65% have failed (at survival probability 0.35). This indicates that ceasefires, regardless of the number of rebel groups in the conflict, tend to fail extremely fast. Although the figure illustrates that after this point, conflicts containing more than one rebel group tend to fail faster than conflicts containing more than one rebel group, the difference is smaller than expected. Although the effect is small, the figure brings some preliminary support for the hypothesis, namely that ceasefires found in multiparty civil conflicts tend to fail faster than ceasefires found in conflicts containing only one group. To test whether the estimations from the Kaplan-Meier plot is reliable also for the variable *number of actors* I run a log-rank test. The results, which can be found in table A10 in the Appendix, indicate that the variable appears to have an insignificantly different survival curve, meaning that there is no difference between ceasefires found in dyadic conflicts compared to multi-party civil conflicts. This suggests that the Kaplan-Meier plot is not reliable and that the results are insignificant (Bland & Altman, 2004). As a result, it seems like the Kaplan-Meier Survival curve, after all, cannot bring preliminary support for the third hypothesis.

[22] The plot for the whole period can be found in Figure A9 in the Appendix.

Figure 6.3: Kaplan-Meier Survival estimate of Number of Actors in the conflict



6.2 Empirical Findings

Based on the tables and figures presented in section 6.1, there appears to be preliminary support for hypothesis one and two, whereas the support is less apparent for hypothesis three. Now, the different hypotheses will be statistically tested using survival analysis. The main parts of the statistical analysis will use the Cox Proportional Hazard model, but for robustness testing, the Weibull is also employed to compare the results. All tables report estimated hazard ratios. The interpretation of hazard ratios is quite straightforward. They are interpreted relative to a baseline of 1, where coefficients that have a hazard ratio greater than 1 indicate an increased risk of ceasefire failure, whereas coefficients with hazard ratios less than 1 indicate a decreased risk of ceasefire failure (Fortna, 2004, p. 85). For example, a binary variable with a hazard ratio of 2 would indicate a doubling of the hazard of renewed fighting (ibid). All models report robust standard errors clustered on country, and statistical significance for all models is indicated with an asterisk.

As discussed in chapter 4, finding the correct battle-death threshold is difficult both theoretically and empirically. As a result, ceasefire failure is measured at different thresholds, namely at 1-, 10- and 25 battle deaths. The main analyses will use the one battle-death threshold, seen in section 6.2.1. The next section will explore the data at the other thresholds and compare the results from the different models. Lastly, the robustness of the models is investigated, as they initially were unable to reject the null hypothesis of proportionality. Without addressing the problems existing in the data, presenting the data as it is can lead to biased results. Indeed, violations of the PH assumption can even render the results meaningless. To ensure the validity of the survival estimates presented, an investigation of the robustness of the main models is therefore necessary.

6.2.1 Survival Analysis – 1 Battle Death

Table 6.7 shows 4 different Cox models. The first three models show each of the independent variables and their interaction with the control variables, whereas the fourth model, the main model, has all variables included. Model 1 shows the first independent variable *earlier ceasefire failure*. As with the main model (model 4 in table 6.7), the variables *earlier ceasefire failure* and *log GDP/per capita* failed the assumption of proportionality and was stratified into three different periods. After the stratification, only the third period of *log GDP/per capita* showed signs of non-proportionality. This was not regarded as a problem since the global test showed that non-proportionality was not an issue for the overall model. Table A3 in the Appendix shows the proportionality assumption before the stratification, whereas table A3A shows the test after stratification has been done on model 1, table 6.7. Model 2 shows the second independent variable *ceasefire comprehensiveness* and its interaction with the control variables. This model was stratified on the *log GDP/per capita* control variable, as this variable showed signs of non-proportionality. After the stratification, all periods passed the test of proportionality. This can be found in table A4 and A4A in the Appendix. The third model includes the independent variable *number of actors*, and as with model 2, the *log GDP/per capita* variable was stratified into three separate time periods. After the stratification, also this model passed the proportionality assumption on all variables. This can be found in tables A5 and A5A in the Appendix.

Looking at model 1 and the independent variable, *earlier ceasefire failure*, the results suggest that the first (0-50 days) and second (50-100 days) time periods both are statistically significant and have a hazard ratio of more than 1. The third period also has a hazard ratio of more than one, but the variable is insignificant. This indicates that in general, there is an increasing risk of ceasefire failure when there has been a previous ceasefire, but this effect is different for the three time periods. Surprisingly, the effect is strongest for ceasefires found in the second period, suggesting that the chance of ceasefire failure when the conflict-dyad has experienced failure in the past is strongest between 50-100 days. This is interesting, as my initial expectation would be the opposite, namely that memory effects would lead to a higher failure rate closer to onset. For the first period, the hazard ratio is estimated to be 3.048, indicating a tripling of the hazard of ceasefire failure. For the second period, the effect is almost six times as large. In the theoretical framework, the argument was that conflicts having experienced failed ceasefires in the past would have more short-lived ceasefires in the present. Looking at model 4, where all the independent variables are included, the results are highly consistent. The hazard rate of *earlier ceasefire failure* for all three periods is smaller, but this difference is marginal. These results provide further support for the results from the descriptive statistics. In sum, there is strong evidence for hypothesis 1.

Looking at model 2 and the independent variable *ceasefire comprehensiveness*, the result is statistically significant at the 0.01% level and shows a hazard rate of less than 1. This means more comprehensive ceasefires decrease the risk of ceasefire failure. The variable is an index variable, meaning that for each increase in ceasefire comprehensiveness, the hazard is reduced by 0.281. This indicates that more comprehensive ceasefire does decrease the risk of ceasefire failure, and this effect is quite strong. Investigating the variable in model 4, the results are highly consistent across the models. The effect is significant at the 0.01% level, with a hazard rate marginally lower than in model 2. This gives strong support also for the second hypothesis, namely that more comprehensive ceasefires last longer than less comprehensive ceasefires. These findings also give supplementary support to the existing literature, which has suggested that ceasefire agreements that include mechanisms reduce the chances of renewed warfare (Fortna, 2003b, p. 365).

Looking at model 3 and the independent variable *number of actors*, the result is statistically significant at the 0.5% level and has a hazard rate of less than 1. This is interesting, as it indicates that ceasefires have a decreasing risk of failure when there are more rebel groups in the conflict. This is contrary to the theoretical expectation, which was that ceasefires declared in conflicts with multiple rebel groups will fail faster than ceasefires found in conflicts containing only one rebel group. The hazard rate is 0.629, meaning that when declared ceasefires go from having two actors in the conflict (one rebel group and the government), to more than two actors, the hazard of experiencing ceasefire failure is decreased by a third. Looking at the variable in model 4, the results are inconsistent. The variable is neither statistically significant, nor has a hazard rate far from 1. This indicates that the variable has no effect on the risk of ceasefire failure. This is in line with the results from the descriptive statistics, where the log-rank test suggested that the results from the Kaplan-Meier estimate were unreliable. The fact that the results from the two models are so inconsistent may be explained by one of the other independent variables, as the inclusion of the independent variables accounts for some of the variation. Running two additional models where each of the other independent variables are included in different models, the results suggest that the first independent variable *earlier ceasefire failure* is driving the results. These models can be seen in table A11 and A12 in the Appendix. This means that omitted variable bias was probably an issue in the third model, indicating that the effect of *number of actors* is not what explains why some ceasefires fail faster than others. Relying on the results from model 4 then, there is no support for the third hypothesis. The results from model 4 are contrary to Cunningham's (2011) findings, which found that conflicts involving more veto players lead to ceasefires that are less long-lasting (p. 188). However, Cunningham (2011) investigated only definite ceasefires, as well as focusing on veto players (i.e. rebel groups that are powerful enough to block agreement) (p. 31). This thesis, on the other side, investigated both definite and preliminary ceasefires, as well as looking at all groups present in conflict. As such, we investigated similar, yet different samples. Based on the results then, there is no support for the third hypothesis.

In terms of the control variables, they are highly consistent among all four models. In all four models, the variable *battle death* is significant at the 0.1% level and has a hazard rate of more than 1. This indicates that more battle deaths in a conflict make ceasefires less durable, compared to conflicts with fewer battle deaths. This is in line with the theoretical expectations and the previous literature, which has suggested that civil wars having high-death tolls will be more likely to resume than less deadly conflicts (Fortna, 2003b; Hartzell et al, 2001). In all four models, also the *duration of conflict* shows similar estimates and is significant. In model 1 and 4, the variable is significant at the 5% level, whereas in model 2 and 3, the significance is 0.1%. This difference in significance can probably be explained by the difference in stratification, as model 1 and 4 are stratified by two variables, whilst model 2 and 3 are only stratified by one variable. Nonetheless, the estimates are consistent across all four models. The results of the estimates are quite surprising, as the hazard ratio is smaller than one. This indicates that longer conflicts have a decreasing risk for ceasefire failure, compared to shorter conflicts. This gives some support to Hartzell et al (2001), who claimed that longer wars have a higher likelihood of stable negotiated settlement (p. 190). Although the results are statistically significant, the hazard rates for all three models are quite close to 1, suggesting that the overall effect of duration of conflict is not so strong. The variable *democracy* is neither statistically significant nor indistinguishable from 1. This applies to all four models. This suggests that democracy has no effect on ceasefire duration. The last control variable is *log GDP/per capita*. This variable was stratified in all four models, as it failed the PH test. Time group 1 has quite consistent estimates in all four models, with a hazard rate of less than one. This suggests that a higher GDP/per capita decrease the risk of ceasefire failure, at least near the onset (0-50 days). The results are statistically significant in three of the models, namely models 2-4. In period 2, models 1 and 4 have a hazard of less than 1, whereas models 2 and 3 have a hazard of more than one, indicating different risk of experiencing ceasefire failure. Nonetheless, the results are insignificant in all four models. Interestingly, the third period, for all four models, suggests that countries with higher GDP/per capita have an increased risk of ceasefire failure. Nevertheless, this difference might be explained, as the only surviving ceasefires after a certain time are the ones in countries with higher GDP per capita. The results are significant for models 2 and 3.

In the remainder of this chapter, the fourth model, which is the full model containing all the independent variables, will be the main model used for analysis. The next section compares the results from model four found in table 6.7, with the other battle-death thresholds at 10 and 25 BRD.

Table 6.7: Cox Model - Stratified

	Dependent variable:			
	(1)	(2)	(3)	(4)
Earlier Ceasefire Failure	3.068***			2.965***
	(0.197)			(0.199)
Ceasefire Comprehensiveness		0.281***		0.320***
		(0.257)		(0.245)
Number of Actors			0.629**	0.941
			(0.152)	(0.156)
Battle Death (ln)	1.319***	1.331***	1.352***	1.259***
	(0.055)	(0.061)	(0.058)	(0.056)
Duration of Conflict (ln)	0.895*	0.851***	0.862***	0.907*
	(0.050)	(0.043)	(0.045)	(0.049)
Democracy	0.996	1.018	1.026	1.001
	(0.028)	(0.029)	(0.029)	(0.028)
Strata: Earlier Ceasefire: tgroup 2	5.889**			5.692**
	(0.653)			(0.656)
Strata: Earlier Ceasefire: tgroup 3	2.305			2.079
	(0.449)			(0.451)
Strata: GDP/per capita (ln): tgroup 1	0.904	0.810*	0.842*	0.826*
	(0.106)	(0.103)	(0.100)	(0.113)
Strata: GDP/per capita (ln): tgroup 2	0.876	1.073	1.134	0.834
	(0.190)	(0.195)	(0.201)	(0.208)
Strata: GDP/per capita (ln): tgroup 3	1.369	1.896*	2.080*	1.248
	(0.163)	(0.219)	(0.236)	(0.167)
Observations	599	549	549	599
Log Likelihood	-1,208.376	-1,246.181	-1,256.008	-1,196.358
Wald Test	127.820***	85.950***	63.280***	157.580***
LR Test	159.237***	83.626***	63.972***	183.273***
Note:				*p<0.05; **p<0.01; ***p<0.001

6.2.2 Comparing the Thresholds

So far, the models analyzed have had the strictest battle-death threshold, at one battle death. What happens when the variables are analyzed at the other thresholds at 10 and 25 BRD? This section will investigate whether it makes a difference using different thresholds for ceasefire failure. The models investigated in this section are based on model 4 in table 6.7, meaning that all independent variables are included in the models. Before turning to the analysis, however, quick elaborations on the proportional hazard assumption of the models are needed.

The model with the 10 BRD follows almost the same pattern as the model with the 1 death threshold regarding violations of the proportional hazard assumption. Therefore, stratification is done on the same variables, namely the independent variable *earlier ceasefire failure* and the control variable *log GDP/per capita.* After the stratification, all variables pass the PH test with a good margin. In the model with the 25-battle death threshold, only the control variable *log GDP/per capita* fails to meet the PH assumption. In addition, the global test indicates that non-proportionality is present in the model. In this model then, only the *log GDP/per capita* is stratified[23]. After stratification, also this model shows that non-proportional hazard is no longer a problem. Nonetheless, since it has been suggested that the global test sometimes falsely indicate that the proportionality assumption is met, the robustness of the models will be discussed in more detail in the next section (Box-Steffensmeier, Reiter & Zorn, 2003, p. 45; Golub 2008, p. 537).

Table 6.8 shows the estimated parameters at the different BRD thresholds. Model 1 shows the original analysis, as found in model 4, table 6.7, where the one battle death threshold is used. Model 2 shows the 10 BRD threshold, whereas model 3 shows the 25 BRD threshold. What is striking, is that the hazard ratios are almost similar across all three models. The three models are to a large degree consistent with each other, both when it comes to statistical significance and the hazard rates. Model 3 deviates the most, which is not so surprising taken that the model is stratified differently from the other two models. Despite this, the variables give much of the same result, indicating that the models are robust. This also indicates that although stratification has been done differently for the three models, this does not seem to have large impact on the results, which gives support for stratifying the

[23] The test of the proportional hazard on the model with 10 BRD can be seen in table A6 in the Appendix. For the model with 25 BRD, this can be seen in table A7 in the Appendix.

variables. The largest inconsistency lies in the independent variable *earlier ceasefire failure*, where model 3 indicate a fourfold of the hazard of ceasefire failure for the whole period, whereas the other models give different hazards for the three periods. In both model 1 and 2, the first and third period of *earlier ceasefire failure* indicate that the hazard of experiencing failed ceasefires more than doubles if there is a history of failed ceasefires, regardless of the battle death threshold. It is only the first period that is statistically significant. For the second period, the hazard is more than five times as large for model 1, while it is only four times as large for model 2. Although there are some differences regarding the independent variable, all three models have the same direction, and close to the same significance.

In addition to the variable *earlier ceasefire failure* having somewhat different coefficients in the three models, also the *log GDP/per capita* show differing results regarding statistical significance. For all three models, the first period is statistically significant and has a decreasing hazard, i.e. countries with a higher GDP/per capita have a lower hazard of experiencing ceasefire failure. For all three models, the second period follows the same pattern as in period one, but the effect is insignificant. For the third period, however, there are some differences. All the coefficients have a hazard of more than 1, indicating increasing risk of ceasefire failure. This means that ceasefires that survive past the first 100 days will have a higher likelihood of failure than ceasefires found in countries with higher GDP/per capita. This effect is statistically significant for the third model. As was discussed in the section above, this can be due to the fact that the only surviving ceasefires after 100 days are the ones found in countries with higher GDP. The results from the three different models lend the same support to the hypotheses. All three models support hypothesis 1, which stated that conflicts having a history of failed ceasefires in the past have a higher ceasefire failure rate than ceasefires found in conflicts with no prior failures. They also lend support to hypothesis 2, that more comprehensive ceasefire are longer lasting than less comprehensive ceasefires. Lastly, all three models also suggest that the number of actors in a conflict has no impact on ceasefire failure.

Gates and Strand (2004), when investigating the duration of civil wars, found that where the casualty threshold is set makes a difference (p. 27). The results from the analysis suggest that this problem is not evident when studying ceasefire failure. The results are quite similar across all three models, and they give the same support to the hypotheses posited. Interestingly, the models suggest that as the ceasefire failure measure is increased (from 1 to

10 to 25 BRD), the model get a better fit to the data (i.e. lower value on the log likelihood). Running an AIC, it suggests the same; the 25-battle death threshold best captures the data[24].

Table 6.8: Cox Model – Stratified, Comparing the thresholds at 1, 10 and 25 BRD

	Dependent variable:		
	(1)	(2)	(3)
Earlier Ceasefire Failure	2.965***	2.666***	4.165***
	(0.199)	(0.228)	(0.228)
Ceasefire Comprehensiveness	0.320***	0.305***	0.425**
	(0.245)	(0.296)	(0.316)
Number of Actors	0.941	1.004	0.869
	(0.156)	(0.195)	(0.233)
Battle death (ln)	1.259***	1.371***	1.305**
	(0.056)	(0.082)	(0.091)
Duration of conflict (ln)	0.907**	0.866***	0.849**
	(0.049)	(0.063)	(0.062)
Democracy	1.001	1.042	1.031
	(0.028)	(0.035)	(0.040)
Strata: Earlier Ceasefire: tgroup 2	5.692**	4.344*	
	(0.656)	(0.682)	
Strata: Earlier Ceasefire: tgroup 3	2.079	2.495	
	(0.451)	(0.484)	
Strata: GDP/per capita (ln): tgroup 1	0.826*	0.659***	0.726*
	(0.113)	(0.151)	(0.173)
Strata: GDP/per capita (ln): tgroup 2	0.834	0.668	0.713
	(0.208)	(0.241)	(0.215)
Strata: GDP/per capita (ln): tgroup 3	1.248	1.395	1.588**
	(0.167)	(0.188)	(0.195)
Observations	599	580	558
Log Likelihood	-1,196.358	-758.763	-512.471
Wald Test	158.580***	134.340***	105.610***
LR Test	183.273***	153.151***	109.592***

Note: *p<0.05; **p<0.01; ***p<0.001

[24] AIC went from 2423 in the model with 1 BRD to 1049 with the model with 25 BRD.

6.2.3 Robustness of Models

Checking the robustness of the findings is important for ensuring its validity. In chapter 5, the Cox model was argued to be the superior choice of estimation technique based on the data. Nonetheless, to test the robustness of the estimates to alternative estimation methods, the hypotheses are also tested using the parametric Weibull model. This is done to further test the validity of the inferences, making sure that the estimates are not the result of model misspecification or non-proportionality bias.

As discussed in chapter 5, an important assumption that should be checked before presenting any results is the one regarding proportional hazards. The original Cox model, seen in model 1 table 6.9, failed the global test for proportionality. Mainly two variables were responsible for this; the independent variable *earlier ceasefire failure* and the control variable *log GDP/per capita* failed the test. Consequently, both variables were stratified, meaning that the variables were separated into different time periods. After stratifying, only the third period of the *log GDP/per capita* continued to violate the proportional hazard assumption, whereas the global test showed that non-proportionality should not be a problem. Nonetheless, Box-Steffensmeier et al (2003) argue that the global test sometimes falsely indicate that the proportionality assumption is met (p. 45). As such, this section will compare the results from the stratified Cox model found in model 4, table 6.7, with the original Cox model and a parametric Weibull model. This is done to see whether there are any major differences between the models.

In table 6.9, model 1 show the original Cox model before the proportionality assumption is addressed. Model 2 shows the stratified Cox, and the third model is a reproduction of model 1 using the Weibull distribution. In all the models, robust standard errors are clustered by country to address potential interdependence. Looking at model 3, it is clear that the estimated effects in the Weibull largely corresponds with the estimates from both the original Cox and the stratified Cox. All three models have similar hazard ratios, and their significance levels are also highly similar. This brings further support to H1 and H2, while also supporting the non-finding of H3. As with the models in table 6.8, the variable that is most inconsistent across the models is the independent variable *earlier ceasefire failure*. In model 1, the hazard of experiencing ceasefire failure quadruples when the conflict has experienced previous failed ceasefire in the past. In the Weibull model, the hazard is six times as large. For the stratified Cox model, the hazard of experiencing is three times as large, but this is only for the first period. The second period has almost a six doubling on the hazard, whereas the third period has a doubling on the hazard. The third period is not significant. The only other things that stand out appreciably, is that the variable *duration of conflict* is not statistically significant in the Weibull, and the *log GDP/per capita* is statistically significant in the first period of the stratified Cox model. Otherwise, the models are highly consistent. In the Weibull, the clustering on countries is statistically significant, but the coefficient is indistinguishable from 1, hence, there is no effect.

The last issue to address is the model fit. The log likelihood suggests that the stratified Cox model has the best model fit, as it has the lowest log-likelihood value. This is further supported when testing AIC, where the Cox stratified scores slightly better than the original Cox model. This substantiates the choice to stratify the Cox model. The fact that all three models give the same support to the three hypotheses suggest that the models are quite robust and that some inferences can be drawn. All models support hypothesis one, which posited that conflicts having experienced failed ceasefires in the past would have more short-lived ceasefires in the present. They also support hypothesis two, that more comprehensive ceasefires should last longer than less comprehensive ceasefires. Finally, all models reject hypothesis three, which stated that ceasefires declared in conflicts with multiple rebel groups would fail faster than ceasefires found in conflicts containing only one rebel group.

Table 6.9: Comparison between Cox, Stratified Cox and Weibull

	Dependent variable: ceasefire duration		
	Cox prop. Hazard	*Cox stratified*	*Weibull*
	(1)	(2)	(3)
Earlier Ceasefire Failure	4.108***	2.965***	6.124***
	(0.179)	(0.228)	(0.173)
Ceasefire Comprehensiveness	0.313***	0.320***	0.428***
	(0.325)	(0.296)	(0.235)
Number of Actors	0.946	0.941	0.895
	(0.243)	(0.195)	(0.158)
Battle Death log (ln)	1.248***	1.259***	1.354***
	(0.071)	(0.082)	(0.063)
Duration of conflict (ln)	0.893*	0.907*	0.914
	(0.056)	(0.063)	(0.048)
Democracy (t-1)	1.000	1.001	1.051
	(0.033)	(0.035)	(0.030)
GDP/per capita (ln)	0.893	0.826*	1.036
	(0.120)	(0.151)	(0.074)
Strata: Earlier Ceasefire: tgroup 2		5.692**	
		(0.682)	
Strata: Earlier Ceasefire: tgroup 3		2.079	
		(0.484)	
Strata: GDP/per capita (ln): tgroup 2		0.834	
		(0.241)	
Strata: GDP/per capita (ln): tgroup 3		1.248	
		(0.188)	
Cluster: country			0.999**
			(0.0004)
log(scale)			61,714.140***
			(2.056)
log(shape)			0.352***
			(0.048)
Observations	334	599	334
Log Likelihood	-1,204.635	-1,196.358	-1,314.888
Wald Test	149.230*** (df = 7)	157.580*** (df = 11)	
LR Test	166.718*** (df = 7)	183.273*** (df = 11)	
Note:			*p<0.05; **p<0.01; ***p<0.001

6.3 Findings

Based on the analyses and descriptive statistics presented above, there is strong support for hypothesis one and hypothesis two. There is no support for the third hypothesis in neither the descriptive statistics nor in the empirical analyses. The results are consistent among all models. The research question driving this thesis was *what can explain why some ceasefires fail faster than others*. The results from the analysis suggest that two factors can lead to faster ceasefire failure. The first is the history of ceasefires, where it was posited that conflicts having experienced failed ceasefires in the past would have more short-lived ceasefires in the present. This claim is supported by the empirical analysis. This is an interesting finding, as it suggests that ceasefires are not necessarily always a good idea. If a conflict has experienced a failed ceasefire in the past, then other options rather than a ceasefire might be preferable. Indeed, as Fortna (2012) has pointed out "the more failed attempts at peace [ceasefires] in the past, the harder it is to work toward a negotiated peace in the future". Although the analysis cannot directly address the role ceasefires have in prolonging a conflict, the results bring some preliminary evidence that ceasefire failure impacts on the conflict dynamics. As Smith (1995) has argued, ceasefires constitute a necessary part to end a conflict (p. 6). Since the history of failed ceasefires impact on its duration, it seems likely that conflicts become harder to resolve whenever there is a history of failed ceasefires. Certainly, violence has a negative impact on peace negotiations, and the use of violence is often regarded as a breach of faith between adversaries (Höglund, 2004, p. 10, 31). A failed ceasefire can, therefore, lead to a crisis in the negotiation process, which again can impact on the prospects for peace. This is consistent with Åkebo´s (2016) findings, which found that ceasefires could have "significant negative impact on the broader dynamics of peace processes" (p. 5). The findings of this thesis thus suggest that mediators and adversaries should not declare ceasefires just for the sake of it, but rather wait until trust has been restored or created. Otherwise, one might risk that trust is forsaken even more, leading to prolonged conflict. It is important to also highlight that although ceasefires can help end conflicts in the long term, they can also alleviate suffering in the short-term. Mediators should, therefore, consider the implications a failed ceasefire might have, compared to no ceasefire at all, keeping the civilians' best interest in mind.

The second factor that can lead to faster ceasefire failure is the comprehensiveness of ceasefires. The argument was that more comprehensive ceasefires should last longer than less comprehensive ceasefires. The results from the analysis, found in table 6.7, suggest they do. This finding brings supplementary evidence to the existing literature. Fortna (2004) for instance, found that mechanisms to enforce an agreement are the most important drivers to ensure ceasefire durability (p. 211). These mechanisms can be demilitarized zones, peacekeeping and external guarantees (ibid). Although she investigated interstate wars, this finding seems to be consistent also regarding civil conflict. The results from the analysis thus suggest that ceasefires, if expected to survive longer, should include some mechanisms to ensure its durability. More specifically, ceasefires that included four measures lasted the longest. Indeed, crafting coherent ceasefires that makes it costlier to defect is important for having longer-lasting ceasefires. As Fortna (2003a) has claimed "peace that is ushered in with a formal peace treaty may be more stable than an informal truce (p. 102). The findings of this thesis seem to agree with this. Indeed, ceasefires that include mechanisms, such as implementation efforts, enforcement missions or were mediated do last longer than ceasefires containing no such mechanisms. Nevertheless, Karakus and Svensson (2017) found that the quality of agreements largely fails to explain the variation in ceasefire success (p. 1). These results are the opposite of the finding of this thesis. Yet, Karakus and Svensson (2017) investigated local ceasefires in Syria. Syria is not included in my sample. When the CWCD is completed, and the UCDP GED includes Syria in their dataset, it will be interesting to see whether the results of this analysis changes.

The third theoretical argument was that ceasefires declared in conflicts with multiple rebel groups would fail faster than ceasefires found in conflicts containing only one rebel group. This argument was not supported by the analysis. This suggests there is no correlation between the number of rebel groups in the conflict, and how long ceasefires last. This is consistent with Winokur´s (2018) findings, as his results suggested that actor cohesion did not affect the duration of ceasefires. Although he only focused on one specific conflict and on two specific ceasefire agreements, the results from this thesis seem to support his findings. It is, however, contrary to Cunningham´s (2011) findings, which found that conflicts involving more veto players lead to ceasefires that are less long-lasting (p. 188). The results from this thesis suggest that there is no correlation between ceasefire duration and their failure when it is investigated using a dataset that includes both preliminary and definite ceasefires and unilateral ceasefires. It seems what matters for longer-lasting ceasefires, is not necessarily the conflict characteristics, but rather the characteristics of the ceasefire, where the history of previous ceasefire failure, and the comprehensiveness of the ceasefire, are the most important factors for explaining why some ceasefires last longer than others. Yet, the two control variables on the conflict-level, the *duration of conflict* and *battle death* were both statistically significant. The first control variable indicated that longer conflicts had a lower hazard of experiencing failure, but the effect was quite marginal. The second control variable indicated that conflicts having had more battle deaths increased the hazard of ceasefire failure. It is thus not only ceasefire characteristics that can explain why some ceasefires fail faster than others but also conflict characteristics. On the country-level, the *GDP/per capita* was significant in the first time period, suggesting that countries with higher levels of GDP had a lower hazard of ceasefire failure, compared to countries with lower levels of GDP. Lastly, the results suggest that where the BRD threshold is set, does not make any substantial difference. The model with the 25 battle deaths had a better fit than models with lower thresholds and is thus the preferable choice for future analysis. For future research, investigating the BRD at an accumulative measure is also of interest.

7 Conclusion

This thesis asks: What can explain why some ceasefires fail faster than others? Ceasefires are popular tool used in conflicts as they represent an easy solution to halt fighting. Indeed, they are found in nearly all civil conflicts and during all stages of a negotiation process. Yet, previous literature has offered little insight on ceasefires, and even less on ceasefire durability. Specifically, the current literature has tended to overlook the fact that most ceasefires are declared during negotiations, and rather focused on ceasefires that are meant to end war (Fortna, 2004; Werner & Yuen, 2005). This thesis has expanded the previous literature, by including all ceasefires that are declared during a negotiation process, thus including also preliminary ceasefires. This has allowed the thesis to investigate ceasefire durability in more detail, offering new insights to explain why some ceasefires fail faster than others.

This thesis has contributed to the research on ceasefire durability in three important ways. Firstly, the thesis has provided a comprehensive theoretical foundation with regards to ceasefires and their durability. The theory posited that trust between belligerents is important for maintaining a ceasefire, as ceasefires rely on mutual dependence – they only survive as long as all the parties to the ceasefire regard it as a better option than continued warfare. Yet, after several years of fighting, belligerents often have good reasons to mistrust each other, and to trust that the other parties to the ceasefire will uphold their word to the ceasefire can seem foolish. This is especially true as ceasefires can be declared for tactical and strategic purposes, meaning that there is always some insecurity whether someone is just using the ceasefire to gain an advantage. To succeed on a ceasefire thus hinges on the involved parties' ability to look past their disagreements, to compromise and cooperate. The thesis argued that there were especially three conditions that make trust between the belligerents harder, and thus make ceasefires fail faster; (1) the history of ceasefires in the conflict, (2) the comprehensiveness of the ceasefire, and (3) the parties to the ceasefire. From these conditions, three hypotheses were derived. The first posited that conflicts having experienced failed ceasefires in the past would have more short-lived ceasefires in the present. The second hypothesis theorized that more comprehensive ceasefires should last longer than less comprehensive ceasefires. The last and third hypothesis was that ceasefires declared in conflicts with multiple rebel groups would fail faster than ceasefires found in conflicts containing only one rebel group.

The second contribution to the research on ceasefires is the data; not only is this the first research to use the new ETH-PRIO Civil War Ceasefire Dataset, but this thesis has also transformed the data into a survival dataset to use in the empirical analysis. In addition, this thesis is also the first trying to measure ceasefire failure. As many ceasefires have an unclear ending, I created my own endings based on battle-related deaths. The argument was that as soon as there is a resumption of violence and people are killed, the ceasefire is said to be de facto over. The difficulty that arose was to decide where the best cut-off point is, especially since Gates and Strand (2004), when investigating the duration of civil wars, found that where the casualty threshold is set makes a difference (p. 27). Consequently, ceasefire failure was measured at three different thresholds, at 1-, 10-, and 25 battle deaths.

The last main contribution is the results and analysis. Using survival analysis, the thesis investigated whether the hypotheses influenced ceasefire durability. A Cox regression was used, and the main analysis used the 1 battle-death threshold. The results supported hypothesis one and two, while there was no support for the third hypothesis. This suggested that both the history of ceasefires, and the comprehensiveness of the ceasefire matter for ceasefire durability, whereas the parties to the ceasefires had no significant effect. When running the analysis with the other battle-related death thresholds and comparing the results, they showed to be highly consistent among the three models. This suggests that where the threshold is set, is not as consequential when studying the durability of ceasefires. Nonetheless, the thesis employed an incomplete dataset, and investigating the different thresholds with a full sample might lead to different results. Although the results from the thesis suggest that the threshold is not so important, it is important to continue investigating, preferable also with an accumulative measure.

The results from this thesis are highly relevant and constitute an important contribution to the literature. As ceasefires are so widely used in peace processes, understanding their impact is important. This is something previous literature to a large degree has overlooked. This thesis has shown that ceasefires, if declared after a failed ceasefire, might not be the best option. Indeed, if there is a history of failed ceasefires in the past, adversaries and mediators might want to find other ways to bring a halt to the fighting. Otherwise, it might hamper an existing peace process, leading to a prolonging of the conflict. If adversaries and mediators want to declare a ceasefire, they should make sure to craft coherent and comprehensive agreements that include mechanisms to reduce the incentives to defect. This is because the results of the empirical analysis suggested that ceasefires containing several mechanisms are more stable than ceasefires that do not include any mechanisms. To conclude, ceasefires are more stable in situations when there has not been a history of failed ceasefires, or when the ceasefires are more comprehensive. Mediators and adversaries should keep this in mind when negotiating ceasefires.

7.1 The need for Future Research

As with most scientific research, there are some limitations in the thesis that needs to be addressed. More specifically, there are five limitations that are highlighted. These are; (1) the operationalization of two of the independent variables, (2) that the dataset used is incomplete, (3) control variables and potential omitted variable bias, (4) that I do not use an accumulative battle-death threshold for measuring ceasefire failure, (5) and that I only look at failed ceasefires. I will now address these limitations and the need for future research.

The first issue relates to the operationalization of some of the variables, more specifically the independent variables *earlier ceasefire failure* and *number of actors*. In chapter 5, it was discussed that the independent variable *earlier ceasefire failure* is not able to capture variation. This means that the variable only manages to say whether a failure has occurred or not and does not address whether it has happened many times previously, or whether the ceasefire that failed was comprehensive or not. The theoretical argument was that trust between belligerents would be weakened if ceasefire failure has occurred in the past, but this trust will surely impact differently depending on the number of previous failures and the comprehensiveness of the ceasefire. This variation is something that could be of interest to study in more detail in the future, to get a deeper and better understanding of the effects previous failure has on ceasefire duration. Nonetheless, it was also argued that the operationalization was sufficient to answer the posited hypothesis. For future research, however, it might be interesting to look more deeply into this variation.

The second operationalization regards the third independent variable number of actors. The theoretical argument drew much of its insights from the literature on fragmentation, which emphasize that fragmentation is more than just the sum of armed groups in the conflict (Bakke, Cunningham & Seymour, 2012, p. 267). Despite this, the operationalization used only included the number of rebel groups. I argued that for the hypothesis, this was not consequential. For future research, however, it might be interesting to expand on this variable, by including measures on institutionalization and power distribution, as these are the measures Bakke, Cunningham and Seymour (2012) uses. One might find a different result than what was found in this thesis, which could bring support to Cunningham's (2011) findings. On the other side, one might find that the effect is insignificant, as was found in this thesis. Regardless of the results, investigating in more detail the effects of fragmentation on ceasefire duration can lead to more precise deductions regarding its impacts.

The second issue of the thesis is that it uses an incomplete dataset. This is not something I had the possibility to change, as the dataset is not scheduled to be finished before September 2019. As discussed in chapter 4, I argued that the dataset, while not completed, represented a broad range of countries that captures the different characteristics found in civil conflict. As such, it was not consequential for the analysis to use the incomplete sample. For future research, however, the whole dataset should be used, and it will be interesting to see whether the results stay the same, or if there are large differences.

The third limitation pertains to the control variables chosen and potential omitted variable bias. Ceasefire duration is an understudied area, as emphasized in the literature chapter. Consequently, no broad consensus on important parameters to include when investigating ceasefire duration exists. To decide on what control variables to use, I turned to other related fields, and the control variables chosen was a result of both theoretical expectation and statistical methods. Yet, there is always a possibility that the control variables chosen are not the correct ones, although statistical testing supported the decision to include them. In addition, there is always the possibility that some important variables are excluded from the analysis, and that omitted variable bias is present in the data. Although it was argued the control variables chosen was a good mixture of bias and variance, one cannot rule out the possibility that some other unexplored factor was driving the results. For future research, I would advice exploring other factors that can lead to ceasefire duration. This can both weaken and strengthen the results from the empirical analysis.

The fourth limitation of this thesis is that I was not able to measure ceasefire failure with an accumulative battle death threshold. This is something I was hoping to do, but unfortunately, I did not have the time or computational skills to make this happen. The analysis suggested that it did not make any substantial difference on the results which of the thresholds was used, but all of the thresholds measured events. For future research, it could be very interesting to see whether an accumulative measure changes any of the results. If no large change occurs, this gives further evidence to the results. If a change occurs, however, one needs to investigate in more depth the implications of the different thresholds.

The last limitation and probably the most consequential relates to the fact that I only studied ceasefires that have failed. The dataset contains many ceasefires that ended after the stipulated time frame, and these ceasefires were not investigated. As such, the analysis only contained ceasefires that were regarded as failures or ceasefires that still had the possibility of failing. The reason for this is that the estimation technique, survival analysis, is only able to capture events, where the event in this thesis was ceasefire failure. Consequently, the estimation technique is not sufficient if one is interested in investigating all ceasefires, not just the failures and the ceasefires that still have the possibility of failing. This is something that could be interesting to investigate in the future, but to do so; a different estimation technique is necessary.

To conclude, there are several limitations of this thesis, but most of them have been addressed throughout the thesis. As this is the first systematic study of ceasefire duration where all ceasefires related to a peace process are included, I argue that the thesis has to a large degree managed to fill some of the existing gaps. Nonetheless, there are still many aspects that should be given further attention, and these aspects have been highlighted in this section. With the new Civil War Ceasefire Dataset from ETH-PRIO, these questions can be answered in the future.

Bibliography

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*: Sage.

Bakke, K. M., Cunningham, K. G., & Seymour, L. J. M. (2012). A plague of initials: Fragmentation, cohesion, and infighting in civil wars. *Perspectives on Politics, 10*(2), 265-283.

Bartusevičius, H., & Gleditsch, K. S. (2019). A Two-Stage Approach to Civil Conflict: Contested Incompatibilities and Armed Violence. *International Organization, 73*(1), 225-248.

Bland, J. M., & Altman, D. G. (2004). The logrank test. *Bmj, 328*(7447), 1073.

Box-Steffensmeier, J. M., Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling: A guide for social scientists*: Cambridge University Press.

Box-Steffensmeier, J. M., Reiter, D., & Zorn, C. (2003). Nonproportional hazards and event history analysis in international relations. *Journal of Conflict Resolution, 47*(1), 33-53.

Broström, G. (na). check.dist. Graphical Goodness-Of-Fit Test. Retrieved 24.04.2019, from https://http://www.rdocumentation.org/packages/eha/versions/2.6.0/topics/check.dist

Call, C. T. (2012). *Why peace fails: the causes and prevention of civil war recurrence*: Georgetown University Press.

Chounet-Cambas, L. (2011). *Negotiating ceasefires: dilemmas & options for mediators*: HD Centre for Humanitarian Dialogue.

Christophersen, K.-A. (2013). *Introduksjon til statistisk analyse: regresjonsbaserte metoder og anvendelse: med oppgaver og oppgaveløsninger*: Gyldendal.

Claeskens, G., & Hjort, N. L. (2008). Model selection and model averaging: Cambridge University Press.

Clayton, G., Rustad, A. S., Nygård, M. H., & Strand, H. (2019a). *Ceasefire Project Codebook, V. 4.2: Unpublished.*

Clayton, G., Rustad, A. S., Nygård, M. H., & Strand, H. (2019b). *Civil War Ceasefire Dataset: Unpublished.*

Clayton, G., & Sticher, V. (2018). Ceasefire agreements and the dynamics of violence during civil conflict. *Unpublished.*

Cleves, M., Gould, W., Gould, W. W., Gutierrez, R., & Marchenko, Y. (2008). *An introduction to survival analysis using Stata*: Stata press.

Croicu, M., & Sundberg, R. (2017). UCDP GED codebook version 17.1. *Department of Peace and Conflict Research, Uppsala University*, 1-38.

Cunningham, D. E. (2011). *Barriers to peace in civil war*: Cambridge University Press.

Cunningham, K. G. (2013). Actor fragmentation and civil war bargaining: How internal divisions generate civil conflict. *American Journal of Political Science, 57*(3), 659-672.

Darby, J. (2001). *The effects of violence on peace processes*: United States Inst of Peace Pr.

Der, G., & Everitt, B. S. (2008). *A handbook of statistical analyses using SAS*: Chapman and Hall/CRC.

DeRouen Jr, K., Lea, J., & Wallensteen, P. (2009). The duration of civil war peace agreements. *Conflict Management and Peace Science, 26*(4), 367-387.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American political science review, 97*(1), 75-90.

Fjelde, H., & Nilsson, D. (2012). Rebels against rebels: Explaining violence between rebel groups. *Journal of Conflict Resolution, 56*(4), 604-628.

Fjelde, H., & Nilsson, D. (2018). The rise of rebel contenders: Barriers to entry and fragmentation in civil wars. *journal of Peace Research, 55*(5), 551-565.

Fortna, P. V. (2012). Do Ceasefires Ever Work? Why the tenuous Syrian truce could end up being a step back for peace. . *Foreign Policy* Retrieved from https://foreignpolicy.com/2012/10/27/do-ceasefires-ever-work/

Fortna, V. P. (2003a). Inside and out: Peacekeeping and the duration of peace after civil and interstate wars. *International Studies Review, 5*(4), 97-114.

Fortna, V. P. (2003b). Scraps of paper? Agreements and the durability of peace. *International Organization, 57*(2), 337-372.

Fortna, V. P. (2004). *Peace time: Cease-fire agreements and the durability of peace*: Princeton University Press.

Fortna, V. P. (2008). *Does peacekeeping work?: shaping belligerents' choices after civil war*: Princeton University Press.

Gates, S., & Strand, H. (2004). *Modeling the duration of civil wars: Measurement and estimation issues.* Paper presented at the presentation at the Joint Session of Workshops of the ECPR, Uppsala.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. *journal of Peace Research, 39*(5), 615-637.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Stollenberg, M., & Strand, H. (2002). Armed Conflict 1946-2001: A New Dataset. *journal of Peace Research, 39(5)*

Golub, J. (2008). Survival analysis *The Oxford handbook of political methodology*.

Hartzell, C., Hoddie, M., & Rothchild, D. (2001). Stabilizing the peace after civil war: An investigation of some key variables. *International Organization, 55*(1), 183-208.

Hothorn, T., & Everitt, B. S. (2014). *A handbook of statistical analyses using R*: Chapman and Hall/CRC.

Höglund, K. (2004). *Violence in the midst of peace negotiations: cases from Guatemala, Northern Ireland, South Africa and Sri Lanka.* Institutionen för freds-och konfliktforskning.

Karakus, D. C., & Svensson, I. (2017). Between the bombs: Exploring partial ceasefires in the syrian civil war, 2011–2017. *Terrorism and Political Violence*, 1-20.

Kirschner, S. (2015). *Trust and fear in civil wars: ending intrastate conflicts*: Lexington Books.

Kolås, Å. (2011). Naga militancy and violent politics in the shadow of ceasefire. *journal of Peace Research, 48*(6), 781-792.

Larsen, U., & Vaupel, J. W. (1993). Hutterite fecundability by age and parity: strategies for frailty modeling of event histories. *Demography, 30*(1), 81-102.

Lim, J., & Lee, G. (2015). Who Beats Plowshares into Swords? Determinants of Commitment to Ceasefire Agreements. *The Korean Journal of International Studies, 13*(3), 489-512.

Mitchell, C. R. (1981) *The Structure of International Conflict*: Springer.

Mosinger, E. S. (2018). Brothers or others in arms? Civilian constituencies and rebel fragmentation in civil war. *journal of Peace Research, 55*(1), 62-77.

Murshed, S. M., & Gates, S. (2005). Spatial–horizontal inequality and the Maoist insurgency in Nepal. *Review of development economics, 9*(1), 121-134.

Pettersson, T., & Eck, K. (2018). Organized violence, 1989-2017. *journal of Peace Research, 55(4)*.

Rudloff, P., & Findley, M. G. (2016). The downstream effects of combatant fragmentation on civil war recurrence. *journal of Peace Research, 53*(1), 19-32.

Ryland, R., Sagård, T., Landsverk, P., Nygård, M. H., Strand, H., Rustad, A. S., . . . Lander, D. (2018). The Effects of Ceasefires in Philippine Peace Process. PRIO: PRIO

Smith, J. D. D. (1995). *Stopping wars: Defining the obstacles to cease-fire*: Routledge.

Stedman, S. J. (1997). Spoiler problems in peace processes. *International security, 22*(2), 5-53.

Sundberg, R., & Melander, E. (2013). Introducing the UCDP georeferenced event dataset. *journal of Peace Research, 50*(4), 523-532.

Teachman, J. D., & Hayward, M. D. (1993). Interpreting hazard rate models. *Sociological methods & research, 21*(3), 340-371.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Alvarado, N. A., & Svensson, R. (2019). *The Quality of Government Standard Dataset, version Jan19.* . Retrieved from: http://qog.pol.gu.se

Therneau, T., Crowson, C., & Atkinson, E. (2019). Using Time Dependent Covariates and Time Dependent Coeffiscients in the Cox Model. Retrieved 11.03.2019, from https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

Theuwissen, M. (2015). How to write the first for loop in R. from https://http://www.r-bloggers.com/how-to-write-the-first-for-loop-in-r/

Vestby, J. (2018). *Climate, development, and conflict: Learning from the past and mapping uncertainties of the future.* (Doctoral Thesis), University of Oslo Retrieved from https://http://www.duo.uio.no/bitstream/handle/10852/61058/Vestby-PhD--2018.pdf?sequence=3&isAllowed=y

Vogt, M., Bormann, N.-C., Rüegger, S., Cederman, L.-E., Hunziker, P., & Girardin, L. (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution, 59*(7), 1327-1342.

Werner, S., & Yuen, A. (2005). Making and keeping peace. *International Organization, 59*(2), 261-292.

Winokur, J. (2018). *Before the Peace: Ceasefire Durability in Ethnic Civil Wars* (Honours), Connecticut College

World Bank Group, T. World Bank Country and Lending Groups Retrieved 24.04.2019, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519#High_income

Åkebo, M. (2016). *Ceasefire agreements and peace processes: A comparative study*: Routledge.

Appendix

Table A1: List of countries in the analysis

Afghanistan	Angola	Azerbaijan	Burundi	Cameroon
Central African Republic	Chad	Colombia	Comoros	Congo
Djibouti	Egypt	El Salvador	Ethiopia	Guatemala
Guinea-Bissau	India	Ivory Coast	Kenya	Lesotho
Liberia	Malaysia	Mali	Mexico	Morocco
Mozambique	Myanmar	Nicaragua	Niger	Nigeria
Pakistan	Philippines	Rwanda	Senegal	Sierra Leone
Somalia	Sri Lanka	Thailand	Uganda	United Kingdom
Haiti*	Trinidad and Tobago*	Panama*	Venezuela*	Peru*
Paraguay*	Uzbekistan*	China*	Laos*	Guinea*
Eritrea*	Algeria*	Mauritania*		

*Countries with asterisk have not had any ceasefires

Checking Assumptions of Cox regression

Figure A1: Graphical test showing the Goodness-Of-Fit of the data

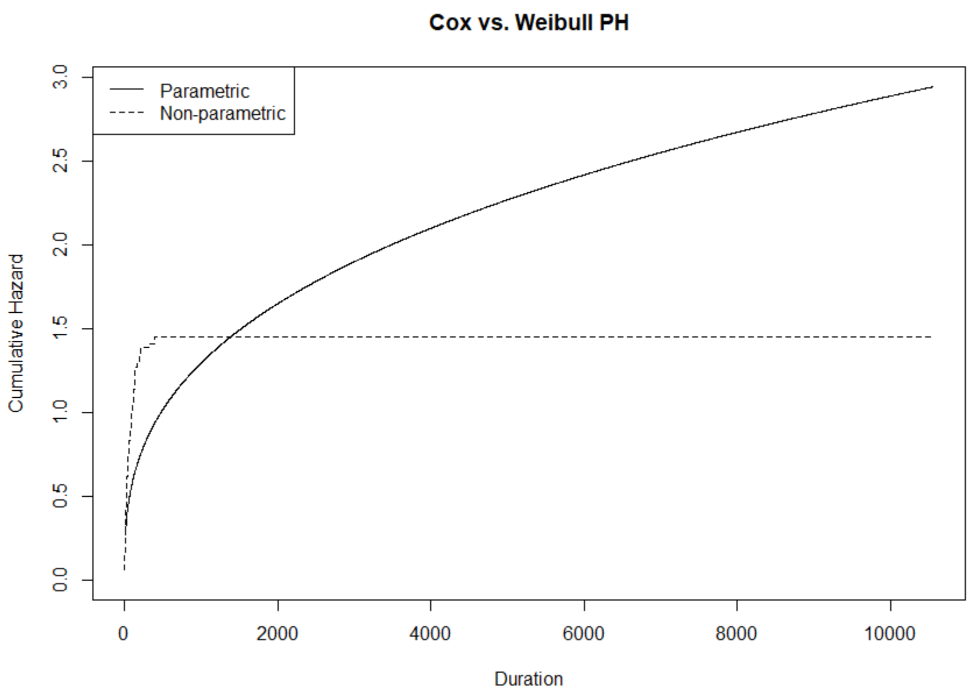


Figure A2: Influence Diagnostics using Score residuals and DFbetas

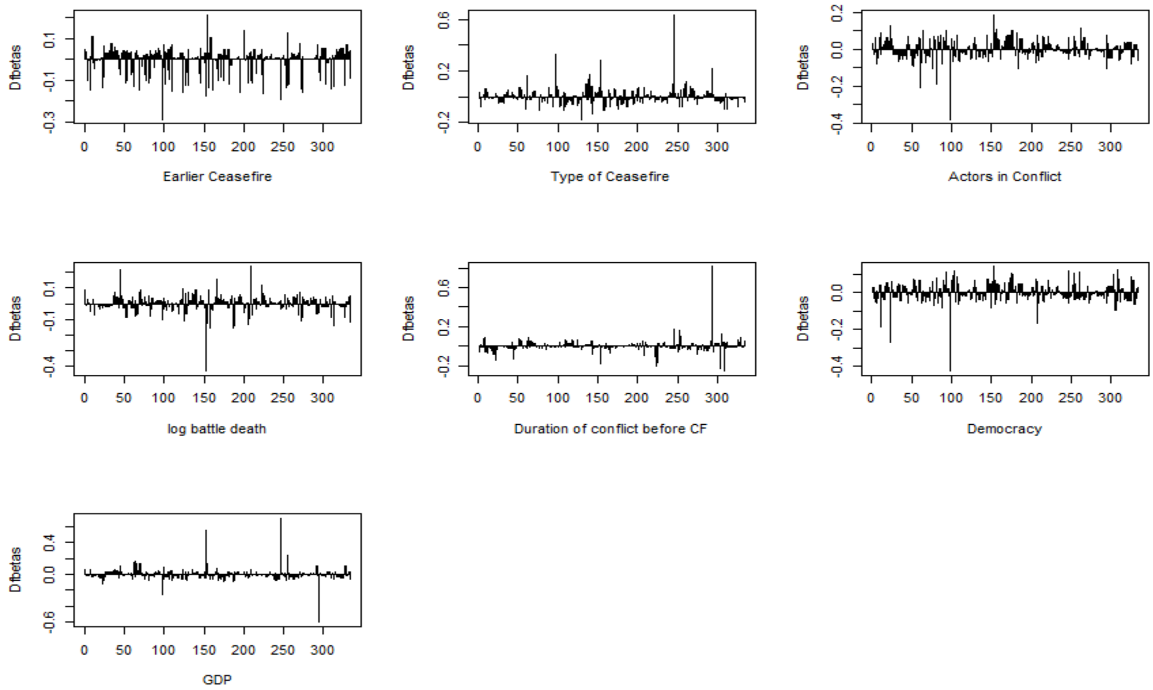


Figure A3: Outliers using DFbetas

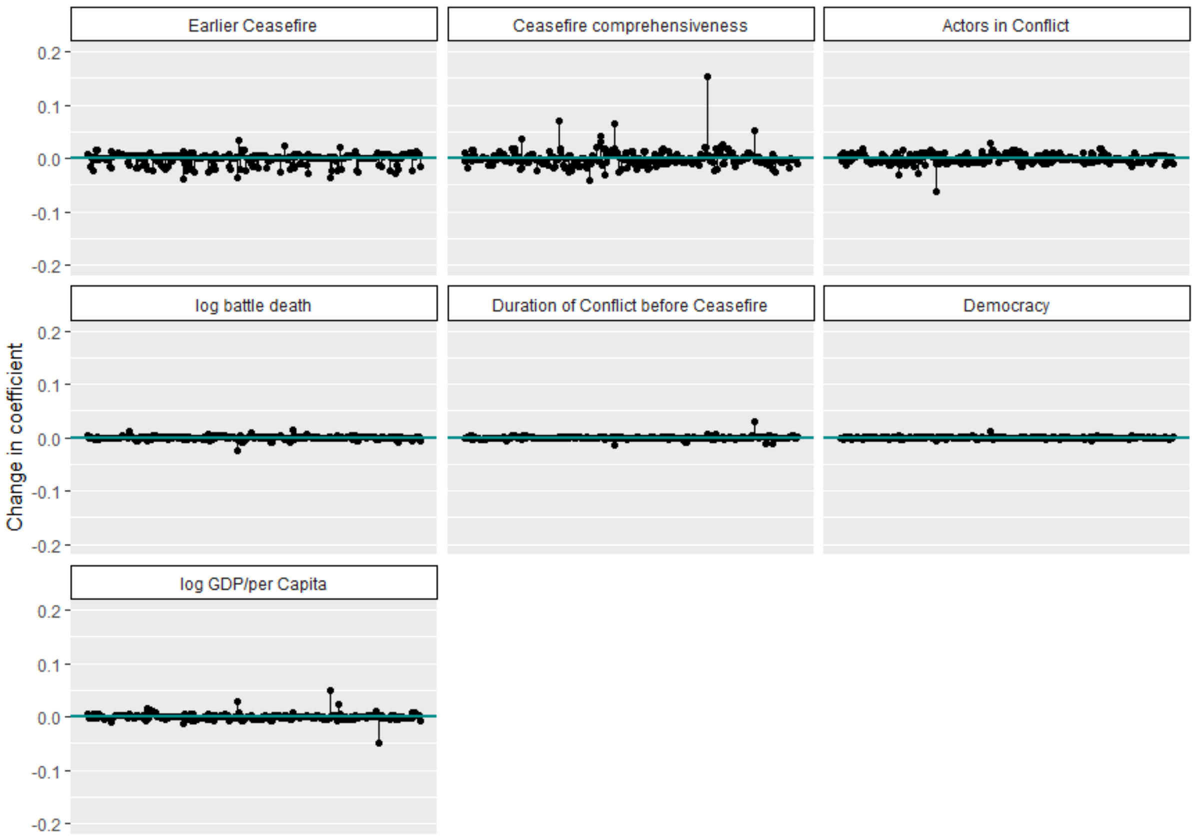


Figure A4: Number of Ties

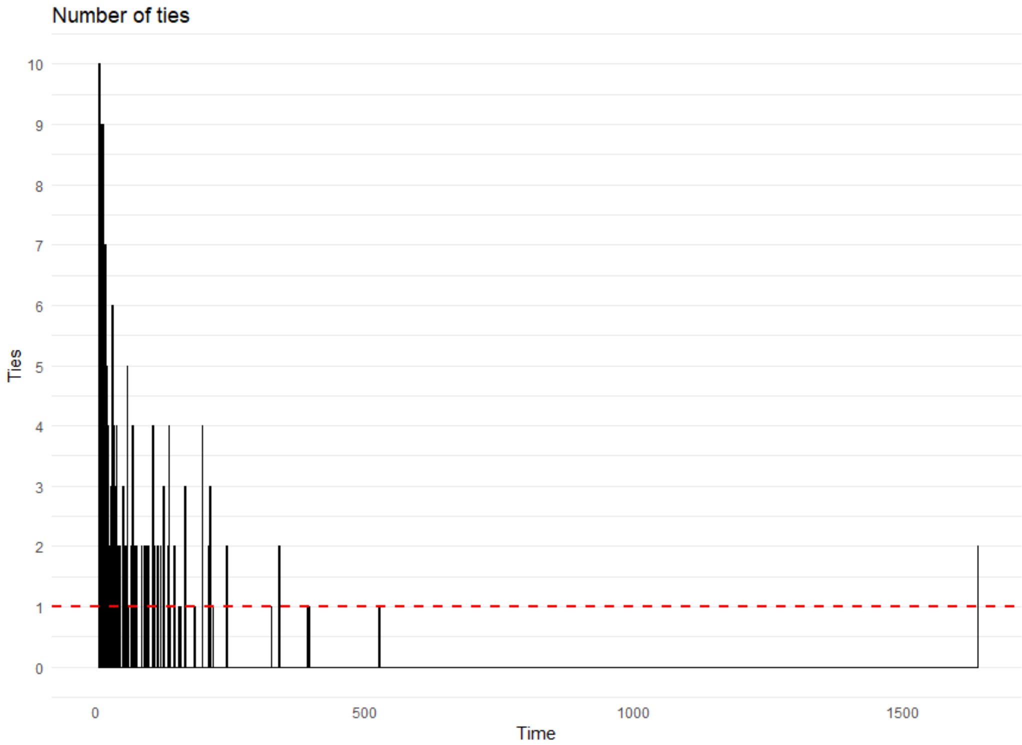
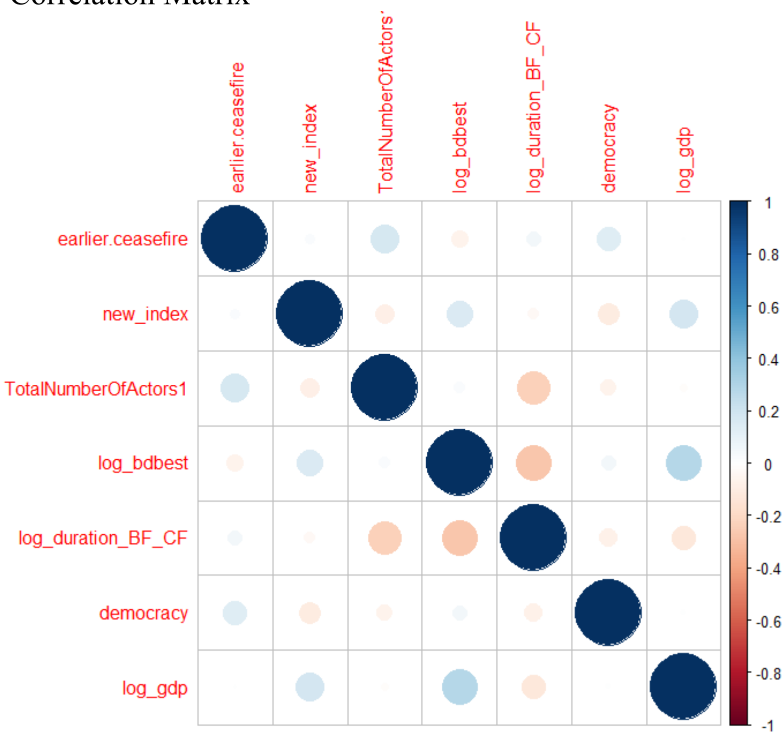


Figure A5: Correlation Matrix



Proportionality

Figure A6: Testing the Proportional Hazard assumption using the Schoenfeld residuals

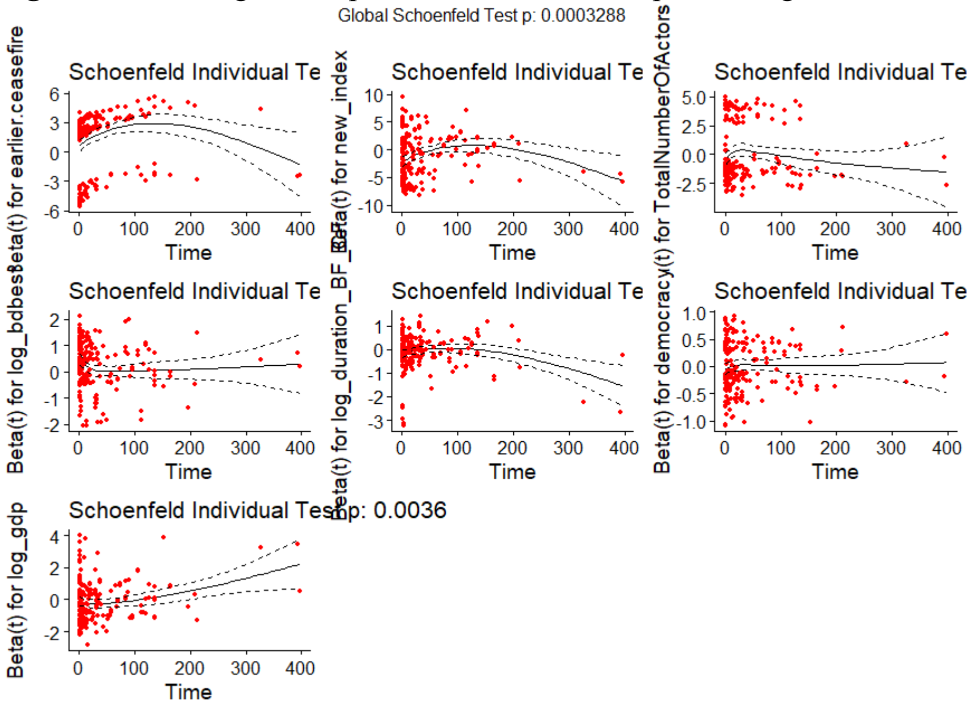


Table A2: Testing the Proportional Hazard Assumption of a Cox Regression, model 4, table 6.7

	Rho	Chisq	p
History Ceasefire	0.1230	4.405	0.035823
Ceasefire Comprehensiveness	0.0949	2.483	0.115048
Total Actors	-0.0520	0.692	0.405588
Log battle death	-0.0862	1.818	0.177538
Log conflict duration	-0.0852	1.904	0.167671
Democracy	0.0373	0.362	0.547212
Log GDP/per capita	0.1822	8.497	0.003558
GLOBAL	NA	26.539	0.000329

Table A2A: Testing the Proportional Hazard Assumption of a Cox Regression, after stratification, model 4, table 6.7

	Rho	Chisq	p
History Ceasefire	0.02061	0.11308	0.73666
Ceasefire Comprehensiveness	0.09962	2.71408	0.09947
Total Actors	-0.06146	0.95074	0.32953
Log battle death	-0.07901	1.54666	0.21363
Log conflict duration	-0.02680	0.18164	0.66996
Democracy	0.04400	0.50858	0.47575
History Ceasefire: strata tgroup 2	0.00429	0.00454	0.94630
History Ceasefire: strata tgroup 3	-0.06472	1.00883	0.31518
Log GDP/per capita	-0.00557	0.00891	0.92478
Log GDP/per capita: strata tgroup 2	0.00718	0.00564	0.94011
Log GDP/per capita: strata tgroup 3	0.20390	7.09539	0.00773
GLOBAL	NA	16.90781	0.11063

Table A3: Testing the Proportional Hazard Assumption of a Cox Regression, model 1, table 6.7

	Rho	Chisq	p
History Ceasefire	0.1380	5.458	0.01948
Log battle death	-0.0874	1.803	0.17935
Log conflict duration	-0.1243	4.281	0.03854
Democracy	0.0380	0.399	0.52776
Log GDP/per capita	0.1740	8.103	0.00442
GLOBAL	NA	26.542	0.00007

Table A3A: Testing the Proportional Hazard Assumption of a Cox Regression, after stratification, model 1, table 6.7

	Rho	Chisq	p
History Ceasefire	0.01918	0.09728	0.75511
Log battle death	-0.07949	1.51444	0.21846
Log conflict duration	-0.06383	1.08730	0.29707
Democracy	0.04663	0.60865	0.43529
History Ceasefire: strata tgroup 2	0.00410	0.00419	0.94840
History Ceasefire: strata tgroup 3	-0.07345	1.30481	0.25334
Log GDP/per capita	-0.00689	0.1418	0.90522
Log GDP/per capita: strata tgroup 2	0.00532	0.00336	0.95378
Log GDP/per capita: strata tgroup 3	0.19104	7.13811	0.00766
GLOBAL	NA	14.62630	0.10173

Table A4: Testing the Proportional Hazard Assumption of a Cox Regression, model 2, table 6.7

	Rho	Chisq	p
Ceasefire Comprehensiveness	0.0707	1.32	0.251253
Log battle death	-0.1048	2.70	0.100072
Log conflict duration	-0.1056	2.75	0.097526
Democracy	0.0420	0.42	0.517044
Log GDP/per capita	0.1683	6.83	0.008958
GLOBAL	NA	20.56	0.000982

Table A4A: Testing the Proportional Hazard Assumption of a Cox Regression, after stratification, model 2, table 6.7

	Rho	Chisq	p
Ceasefire Comprehensiveness	0.0882	2.0372	0.1535
Log battle death	-0.0850	1.7946	0.1804
Log conflict duration	-0.0824	1.6136	0.2040
Democracy	0.0739	1.3026	0.2537
Log GDP/per capita	-0.0150	0.0545	0.8155
Log GDP/per capita: strata tgroup 2	-0.0351	0.1946	0.6591
Log GDP/per capita: strata tgroup 3	0.1603	3.5443	0.0597
GLOBAL	NA	11.8400	0.1059

Table A5: Testing the Proportional Hazard Assumption of a Cox Regression, model 3, table 6.7

	Rho	Chisq	p
Total Actors	-0.0584	0.834	0.361009
Log battle death	-0.1138	2.994	0.083550
Log conflict duration	-0.1160	3.340	0.067597
Democracy	0.0216	0.117	0.732738
Log GDP/per capita	0.1761	8.027	0.004610
GLOBAL	NA	23.561	0.000264

Table A5A: Testing the Proportional Hazard Assumption of a Cox Regression, after stratification, model 3, table 6.7

	Rho	Chisq	p
Total Actors	-0.0482	0.5699	0.4503
Log battle death	-0.0842	1.6597	0.1976
Log conflict duration	-0.0930	2.0658	0.1506
Democracy	0.0590	0.8676	0.3516
Log GDP/per capita	-0.0087	0.0193	0.8894
Log GDP/per capita: strata tgroup 2	-0.0429	0.3290	0.5662
Log GDP/per capita: strata tgroup 3	0.1448	2.9355	0.0867
GLOBAL	NA	11.4427	0.1204

Table A6: Testing the Proportional Hazard Assumption of a Cox Regression, model 2, table 6.8

	Rho	Chisq	p
History Ceasefire	0.137	3.30	0.0693
Ceasefire Comprehensiveness	0.145	3.31	0.0691
Total Actors	0.132	2.98	0.0845
Log battle death	-0.116	2.23	0.1360
Log conflict duration	-0.155	4.56	0.0327
Democracy	-0.104	1.76	0.1850
Log GDP/per capita	0.282	14.23	0.000162
GLOBAL	NA	33.05	0.0000259

Table A6A: Testing the Proportional Hazard Assumption of a Cox Regression, after stratification, model 2, table 6.8

	Rho	Chisq	p
History Ceasefire	0.03004	0.15076	0.6978
Ceasefire Comprehensiveness	0.14244	3.14669	0.0761
Total Actors	0.13181	2.80822	0.0938
Log battle death	-0.10998	2.08859	0.1484
Log conflict duration	-0.10890	2.26429	0.1324
Democracy	-0.08785	1.24270	0.2650
History Ceasefire: strata tgroup 2	0.00568	0.00517	0.9427
History Ceasefire: strata tgroup 3	-0.07496	0.88101	0.3479
Log GDP/per capita	0.07730	1.18490	0.2764
Log GDP/per capita: strata tgroup 2	0.01110	0.00863	0.9260
Log GDP/per capita: strata tgroup 3	0.07886	0.83463	0.3609
GLOBAL	NA	16.95333	0.1093

Table A7: Testing the Proportional Hazard Assumption of a Cox Regression, model 3, table 6.8

	Rho	Chisq	p
History Ceasefire	0.1313	2.108	0.14653
Ceasefire Comprehensiveness	0.1541	2.013	0.15597
Total Actors	0.1175	1.444	0.22944
Log battle death	-0.1042	1.292	0.25560
Log conflict duration	-0.0877	0.927	0.33567
Democracy	-0.1215	1.421	0.23323
Log GDP/per capita	0.2738	9.311	0.00228
GLOBAL	NA	18.474	0.01001

Table A7A: Testing the Proportional Hazard Assumption of a Cox Regression, after stratification, model 3, table 6.8

	Rho	Chisq	p
History Ceasefire	0.1284	2.052	0.152
Ceasefire Comprehensiveness	0.1550	2.025	0.155
Total Actors	0.1047	1.111	0.292
Log battle death	-0.0896	0.973	0.324
Log conflict duration	-0.0535	0.340	0.560
Democracy	-0.1256	1.515	0.218
Log GDP/per capita	0.0662	0.572	0.449
Log GDP/per capita: strata tgroup 2	0.0127	0.011	0.916
Log GDP/per capita: strata tgroup 3	0.0832	0.666	0.414
GLOBAL	NA	9.324	0.408

Figure A7: Kaplan-Meier survival estimate of Earlier Ceasefire Failure, whole period

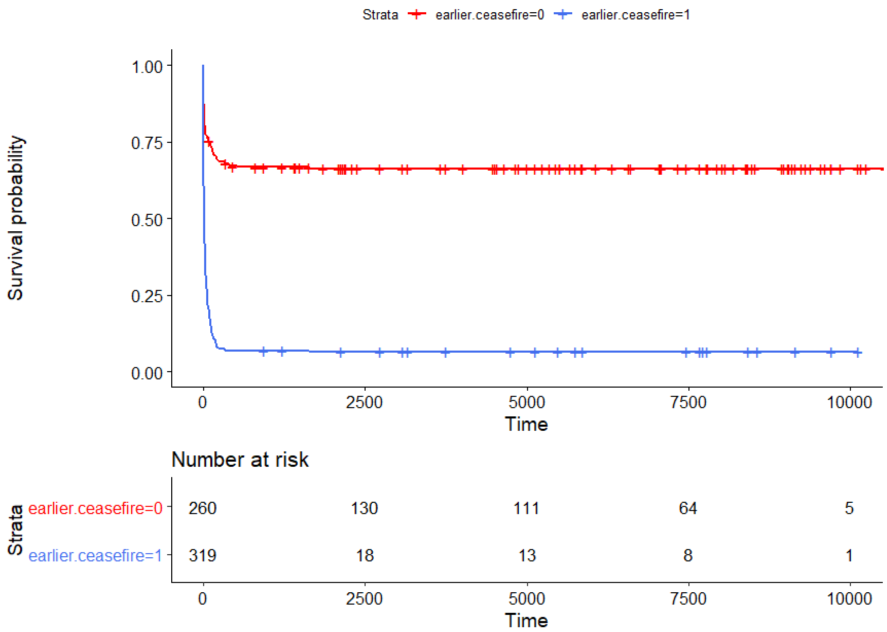


Figure A8: Kaplan-Meier survival estimate of Ceasefire Comprehensiveness, whole period

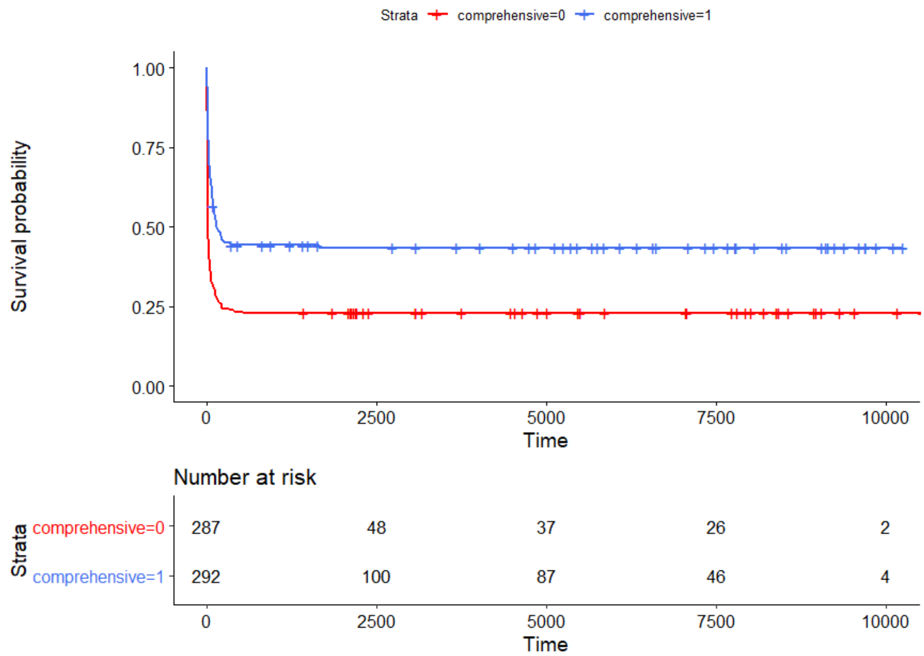
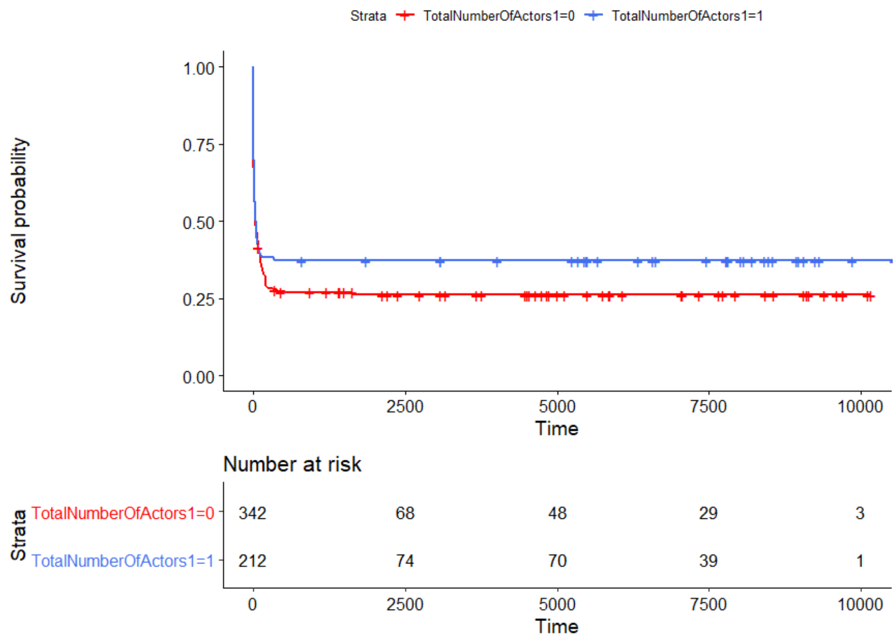


Figure A9: Kaplan-Meier survival estimate of Number of Actors, whole period



Empirical Analysis

Table A8: Log-rank test of Earlier Ceasefire Failure

Earlier Ceasefire	Events observed	Events Expected
0	87	230
1	298	155
Total	385	385

Chisq = 248 on 1 degrees of freedom, p= 2e-16

Table A9: Log-rank test of Comprehensiveness of Ceasefire

Type of Ceasefire	Events observed	Events Expected
0	95	44.2
1	58	42.7
2	68	68.8
3	62	72.9
4	46	82.8
5	37	44.6
6	19	29
Total	385	385

Chisq = 92 on 6 degrees of freedom, p= 2e-16

Table A10: Log-rank test of Number of Actors

Number of Actors	Events observed	Events Expected
0	251	233
1	133	151
Total	384	384

Chisq = 3.7 on 1 degrees of freedom, p= 0.06

Table A11: Model including independent variables Earlier Ceasefire Failure and Number of Actors

	Dependent variable:
Number of Actors	0.873
	(0.149)
Battle death (ln)	1.317***
	(0.054)
Duration of conflict (ln)	0.903*
	(0.051)
Democracy	0.997
	(0.028)
Strata: Earlier Ceasefire: tgroup 1	2.977***
	(0.202)
Strata: Earlier Ceasefire: tgroup 2	17.331***
	(0.622)
Strata: Earlier Ceasefire: tgroup 3	6.896***
	(0.406)
Strata: GDP/per capita (ln): tgroup 1	0.895
	(0.107)
Strata: GDP/per capita (ln): tgroup 2	0.885
	(0.191)
Strata: GDP/per capita (ln): tgroup 3	1.373
	(0.160)
Observations	599
Log Likelihood	-1,207.989
Wald Test	128.810*** (df = 10)
LR Test	160.010*** (df = 10)
Note:	*p<0.05; **p<0.01; ***p<0.001

Table A12: Model including independent variables Ceasefire Comprehensiveness and Number of Actors

	Dependent variable:
Number of Actors	0.698*
	(0.154)
Ceasefire Comprehensiveness	0.302***
	(0.250)
Battle death (ln)	1.313***
	(0.055)
Duration of conflict (ln)	0.875**
	(0.044)
Democracy	1.016
	(0.027)
Strata: GDP/per capita (ln): tgroup 1	0.798*
	(0.108)
Strata: GDP/per capita (ln): tgroup 2	0.816
	(0.167)
Strata: GDP/per capita (ln): tgroup 3	1.230
	(0.140)
Observations	599
Log Likelihood	-1,244.497
Wald Test	90.850*** (df = 8)
LR Test	86.994*** (df = 8)
Note:	*p<0.05; **p<0.01; ***p<0.001