

**The Practical Application of the Alternatives to Detention in EU
Member States**

*A Two-step Qualitative Comparative Analysis on the Application of
Alternative Measures to Detention as part of the 2008/115/EC
Return Directive*

Frank E. Hendriks

Master student Comparative Politics, Administration, and Society

*Radboud University Nijmegen: Department of Public Administration, Faculty of Management
Sciences*

Supervisory Lecturer and first reader: Dr. J.H.M.M. Tholen

Second reader: Prof. dr. S. van Thiel

June 20, 2018

Abstract

Multiple studies have already advocated the use of the Alternatives to Detention over the use of Administrative Detention. However, recent research has moved over to the evaluation of the legal implementation and practical application of the alternative to Detention measures, by mostly conducting comparative low-*N* case studies. This study took a new approach, by comparing an intermediate number of 40 cases of alternative measures for Member States of the European Union as part of the 2008/115/EC Return Directive, Article 15 (1). The outcome of non-application of ATD measures in practice was tested by using conditions from a multi-level hierarchical structure. Hypotheses from the enforcement approach and management approach were included as country-level conditions. Cross-country level conditions were added from the worlds of compliance typology. A two-step crisp-set Qualitative Comparative Analysis (csQCA) was used to analyze the non-application of the alternatives. The two-step approach meant that the four worlds of compliance were included as remote conditions, which were examined for their mediating influence as outcome-enabling factors in the first step. A second step involved the analyses of the relevant remote conditions in combination with the proximate conditions. Data was gathered for the time-period between 2008 and 2012, for when the directive came into effect until the evaluation report commissioned by the EU commission on the transposition of the directive was issued. The findings of this study indicate that the typology provides two sufficient remote conditions that function as outcome-enabling factors by their mediating influence. The analysis also shows that the different combinations of proximate conditions with each of these two remote conditions are sufficient for explaining the outcome of non-application in practice. The proximate conditions show some mixed results on their own, indicating the sufficiency for some while not finding enough evidence for others. Interestingly, this study suggests that the level of federalism of an EU Member State has a different relationship with the compliance of immigration-related policies than is generally expected by previous compliance studies. Thus, these findings contribute to the developing research on the alternatives to detention, compliance studies within the EU integration field, and the merit of using QCA methods for intermediate-*N* comparative studies.

Acknowledgements

This study is the first part of a master's thesis on the topic of the alternatives to detention within the return process of the immigration policy field. This first part provides a more theoretical study for the Radboud University. The second part consists of an evaluative study on the use of alternative measures within the return process of the Netherlands in cooperation with the Department of the Directorate-General for Migration (DGM) of the Dutch Ministry of Justice & Security. This study, therefore, is of lesser interest to the Ministry. Still, a lot of thanks and gratitude goes to Irene Ritman, Martijn ten Dam and Renée ter Hart for their supervision, guidance, and feedback for this part. The field of immigration policy is a complex and ever-changing topic to study, and their substantial knowledge on the subject was incredibly helpful for a "rookie." Further gratitude goes out to Laura Kok and Sanne Westmaas, whom both left the project at an earlier stage as well as Peter Bosch who replaced Renée ter Hart in the final stage.

In addition, special thanks and gratitude go to the following:

- Wouter Dronkers, who provided a chance to do such an exciting project with the Dutch Ministry of Justice & Security
- Dr. Berry Tholen, for providing clear and useful feedback in his role as the supervisory lecturer and first reader
- Prof. dr. Sandra van Thiel, for her willingness to take on the role of the second reader
- My family, girlfriend, and friends; who supported me through my years of study. Even when I could have made it easier for myself, they always reassured me that overcoming these challenges would be worth it in the end.

List of Contents

Abstract	2
Acknowledgements	3
List of Contents	4
List of Figures & Tables	6
1. Introduction	7
1.1 Problem Formulation.....	9
1.2 Research Questions	11
1.3 Relevance	13
1.4 Scope & Considerations	14
1.5 Key Terms & Conceptual Framework	16
1.5.1 Key terms	16
1.5.2 Conceptual framework.....	21
1.6 Structure of the Study.....	23
2. Theoretical Framework	24
2.1 Practical Application as Part of Compliance	24
2.2 Explaining the level of Compliance with EU Directives.....	27
2.2.1 Enforcement approach	27
2.2.2 Management approach	30
2.2.3 Cultural consideration: The four worlds of compliance typology.....	32
2.3 Summary & Theoretical Framework.....	35
3. Methodology	37
3.1 Qualitative Comparative Analysis.....	37
3.1.1 Terminology of QCA	38
3.1.2 Crisp-set QCA.....	39
3.1.3 Two-step approach to a QCA: Remote & proximate conditions.....	40
3.2 Population & Case-selection	41
3.3 Operationalization & Data-gathering	43
3.3.1 Outcome: Level of legal implementation and practical application.....	43
3.3.2 Proximate conditions: Enforcement approach	43
3.3.3 Proximate conditions: Management approach	45
3.3.4 Remote conditions: Four worlds of compliance typology	46
3.3.5 Raw data table.....	46
3.4 Data Transformation	47
3.4.1 Data transformation: Outcome.....	47
3.4.2 Data transformation: Proximate conditions from the enforcement approach.....	48

3.4.3	Data transformation: Proximate conditions from the management approach	49
3.4.4	Data transformation: Remote conditions from the worlds of compliance typology	50
3.4.5	Data transformation: Scope-conditions	50
3.5	Data Analysis	53
3.6	Data Procedure: Application of a two-step csQCA approach	55
3.7	Reliability & Validity	57
4.	Results	60
4.1	Descriptive results: Overview of the Transposition of ATD Measures	60
4.2	Necessary Conditions Analysis	62
4.3	Step 1: Analysis of the Remote Conditions	65
4.4	Step 2: Analysis of the Remote & Proximate Conditions	68
4.4.1	Solution formula: Remote condition FW-DL & proximate conditions	70
4.4.2	Solution formula: Remote condition FW-TN & proximate conditions	71
4.5	Combined Solution Formula	73
5.	Discussion	76
5.1	Discussion of the Remote Conditions	76
5.2	Discussion of the Proximate Conditions	78
5.2.1	Discussion of the proximate conditions: Enforcement approach	78
5.2.2	Discussion of the proximate conditions: Management approach	79
5.3	Discussion of the Combination of Remote & Proximate Conditions	81
6.	Conclusions	84
6.1	Main Findings	84
6.2	Main Limitations	86
6.3	Main Implications	88
6.4	Recommendations for Further Research	90
6.5	Final thoughts: Practical Usefulness of the Study	92
	References	93
	Appendix	100
A.	Raw data table	100
B.	Truth tables	103
C.	Oher types of solution formulas	106

List of Figures & Tables

Figures

Figure 1. Preliminary conceptual framework.....	22
Figure 2. Compliance as used in this study (marked in grey). Adapted from the reviews by Tallberg (2002) and Angelova et al. (2012).....	24
Figure 3. Definitive theoretical framework, based on the preliminary conceptual framework (Figure 1).....	35
Figure 4. Difference between remote & proximate conditions, based on Sehring et al. (2013)	40
Figure 5. The two-step approach to the csQCA of this study	56
Figure 6. Descriptive overview of the legal implementation & practical application of ATD measures. Adapted from Directorate-General for Migration and Home Affairs (2013).....	61
Figure 7. Overview of the first step of the csQCA	65
Figure 8. Overview of the second step of the csQCA.....	69
Figure 9. Overview of the combined result of the csQCA.....	73

Tables

Table 1. Overview of the hypotheses.....	36
Table 2. Example of the case-selection.....	42
Table 3. Dichotomized table of the conditions & outcome.....	51
Table 4. Necessary conditions analysis table.....	63
Table 5. Truth table, result for first step of the csQCA.....	66
Table 6. Step 1: intermediate solution for the remote conditions	67
Table 7. Step 2: Intermediate solution for the remote condition FW-DL and the proximate conditions	70
Table 8. Step 2: Complex solution for the remote condition FW-TN and the proximate conditions	72
Table 9. Intermediate solution for combined result of remote conditions and proximate conditions	74
Table 10. Raw data table of the conditions and outcome.....	100
Table 11. Truth table, result for the second step of the csQCA: remote condition FW-DL	103
Table 12. Truth table, result for the second step of the csQCA: remote condition FW-TN	104
Table 13. Truth table: combined result of remote conditions and proximate conditions.....	105
Table 14. Step 1: complex solution for the remote conditions.....	106
Table 15. Step 1: parsimonious solution for the remote conditions.....	106
Table 16. Step 2: Complex solution for the remote condition FW-DL and the proximate conditions	107
Table 17. Step 2: Parsimonious solution for the remote condition FW-DL and the proximate conditions	107
Table 18. Step 2: Intermediate solution for the remote condition FW-TN and the proximate conditions.....	108
Table 19. Step 2: Parsimonious solution for the remote condition FW-TN and the proximate conditions	108
Table 20. Complex solution for combined result of remote conditions and proximate conditions.....	109
Table 21. Parsimonious solution for combined result of remote conditions and proximate conditions.....	110

1. Introduction

Immigration policy has re-entered the scientific spotlight in recent years as a fast developing and interesting field for research. This coincided with an increase in migration globally, with an estimated increase of 78 million migrants between 1990 (154 million) and 2013 (232 million) (Sampson & Mitchell, 2013). Immigration is a dominant topic of interest within the European Union, with rising attention for immigration-related policies after the Treaty of Amsterdam came into force in 1999 (Geddes & Scholten, 2016). More recently, the European migrant crisis of 2015 resulted in even more pressure on the European Union (EU) and its countries to handle a complicated situation on a humanitarian, political and practical scale. Although uniform immigration policy is absent at the EU level, directives and EU laws are present for some specific aspects (Geddes & Scholten, 2016). An example of a such a directive is the 2008/115/EC Return Directive of the European Parliament and the Council of 16 December 2008 on common standards and procedures in the Member States for returning irregular staying migrants, which came into effect on the 13th of January 2009 (EUR-Lex, 2008). It provides a uniform approach across the EU for common standards and procedures regarding the return policies of the Member States of the EU. More specifically, Article 15(1) states that detention can only be imposed in certain circumstances if other measures cannot be applied effectively in that specific case. The purpose of Article 15(1) was to promote the necessity and prioritization of the implementation and application of alternative measures over detention in the return policy of EU Member States. These alternatives are often referred to as the “Alternatives to Detention” (ATD), encompassing several different measures such as a residence restriction or a duty to report.

Though the directive is in effect, it seems that the implementation and application of Article 15(1) as part of the Return Directive appears difficult to achieve (Basilien-Gainche, 2015). As with many EU directives, the broad and often unclear formulation provides the EU Member States the ability in the implementation to adjust these to “fit” with their national law. The wording of the 2008/115/EC Return Directive does require that EU Member States have some form of alternatives with Article 15 (1) but gives no detailed explanation. It does not explicitly obligate EU countries to implement a program containing several alternatives into their return policy. As such, a Member State that implements only one alternative can also be considered as complying with Article 15 (1). According to the study by Basilien-Gainche (2015), only 32% of the eighty-seven legal bases found for ATD in domestic legislation of the thirty-one countries of the EU were effectively applied in practice. Other studies have also shown that the implementation of ATD has been slow and a gap exists between legislative implementation and actual implementation in practice (Bloomfield, 2016; Bruycker, Bloomfield, Tsourdi, & Petin, 2015; Costello & Kaytaz, 2013; Field & Edwards, 2006)

The present study will analyze the implementation of the alternatives as part of the EU return directive in depth, by looking at the factors that cause an EU Member State to refrain from practical application. The implementation of EU directives and its parts are most often studied by scholars interested in the EU integration process, as part of the scientific field of compliance research. Two approaches are most often considered in recent studies as a theoretical basis for the different factors that influence the compliance of EU Member States with EU directives. The first being the '*enforcement approach*' or the willingness to comply, and the second being the '*management approach*' or the ability to comply (Tallberg, 2002). Another recent theory addresses the explanation for the compliance with EU directives differently. Formulated by Falkner, Hartlapp, & Treib (2007) & further specified by Falkner & Treib (2008), the '*four worlds of compliance typology*' has been developed in which there are several categories of countries that share a 'culture of law-abidingness' for compliance. The culture of each of these groups has a mediating effect on the effect of the factors from the first two approaches. Thus, a context is shaped by the cultural aspect in which some factors have a more significant impact on compliance than others. The present study aims to combine the factors from the first two approaches and compares the effects of these within the outcome-enabling context by the typology. The purpose is, therefore, to determine if it is true that the factors from these two approaches have different effects, depending on the inclusion of a country to a particular category of the typology.

The present study uses a two-step, crisp-set Qualitative Comparative Analysis (csQCA) to analyze several cases of alternatives across EU Member States. Specifically, the study investigates the instances in which countries have implemented an ATD in legislation, but not in practice. In doing so, it is the aim of the study to contribute to the increasing scientific attention for the implementation of alternatives to detention. While also providing explorative research towards the factors influencing the compliance with EU directives. The typology by Falkner & Treib (2008) is analyzed with the first step of the QCA as a remote condition, associating different countries by a shared "culture" as a mediating effect. With the factors from the other two approaches analyzed in the second step as proximate conditions that influence the compliance with Article 15(1) of the Return Directive.

1.1 Problem Formulation

As mentioned by the introduction (see Chapter 1), the present study is interested in the compliance with Article 15(1) by the relevant EU Member States that adhere to the 2008/115/EC Return Directive (EUR-Lex, 2008). The purpose of the study is the analysis of the occurrence of ATD measures that are implemented in the national legislation of a EU Member State but are not applied in practice. Four aspects are considered relevant for the present study regarding the issue of non-application of implemented ATD Measures. The first two aspects are related to the ATD measures as part of the immigration policy field. The latter two aspects are connected to the practical application of ATD as part of the compliance field on EU directives.

The first aspect involves the pattern of high variation in the practical application of alternatives in general. As mentioned in the introduction (see Chapter 1), the supervisory measures, such as administrative detention or the alternatives, are applied to two different groups for immigration-related purposes. The first group consists of the use of these measures within the entry/asylum process, while the second group contains the use of these measures within the return process. Several studies have analyzed the implementation and application of ATD measures by EU Member States and other countries. For example, Chmelickova, (2006, p. 58) concludes in her evaluation: *“The problem shared by many countries, in which systems of alternatives exist, is that they remain largely unused in practice.”* Other studies that analyzed the ATD measures include the review by Bloomfield (2016b) and the report by Field & Edwards (2006). However, most studies involve qualitative case-level analysis in an inductive manner. With little attention to investigating it systematically, or in a way that is more grounded by the incorporation of theoretical background.

The second aspect relates to the specific part of using ATD measures within the second group within the return process. Most studies have been focused on the adoption of ATD measures in general, or with the ATD measures adopted within the entry/asylum process. Data on the use of ATD measures within the return process remains limited (Bruycker et al., 2015). While other studies, such as those by Leerkes & Broeders (2010) or by Edwards (2011) consisted of a case-level analysis of one or a few countries. Together, these studies show that there is an urgent need to analyze the practical application of ATD measures, specifically regarding the use within the return process.

On the other hand, the adoption of ATD measures by EU Member States is difficult to analyze without considering it as an aspect of the level of compliance with the 2008/115/EC Return Directive. With non-compliance by having no practical application of ATD as one of the included parts for the evaluation by the EU commission of the implementation of the Return Directive (Directorate-General for Migration and Home Affairs, 2013). The cause for non-compliance with EU directives by EU Member States has been one of the key topics of interest for the European integration field (Angelova, Dannwolf, & König, 2012). Moreover, EU Member States are not obliged to adopt a specific number of alternative measures, so the absence of the practical application of some of the implemented alternative measures in national legislation poses a compelling phenomenon. Especially as the practical application is an understudied area compared to the transposition and implementation of EU directives into domestic law (Mastenbroek, 2005). Therefore, the third aspect of the issue under analysis is to analyze the non-compliance by EU Member States aimed at the stage after implementation.

As a final aspect, a search of the literature on compliance reveals that there are multiple approaches and waves of research with just as much theoretical explanations as to explain the compliance with directives. One general weakness of these theoretical findings is that there seems to be a general bias for specific policy fields (social and environmental) and countries such as Austria, Portugal, Sweden and southern states lack inclusion in comparative studies (Angelova et al., 2012). Besides, most comparative studies either use a qualitative or quantitative approach. According to Mastenbroek (2005), mixed method designs are needed to bridge the 'quantitative-qualitative divide.' Together these two aspects show that there is a need for further examination of compliance by EU Member States with EU directives, explicitly concerning mixed-method studies on less "popular" policy fields and including countries that have not been incorporated as much.

1.2 Research Questions

The following part of the study presents the aim of the present study and its research question. The purpose is to combine the aspects as described in the previous chapter (see Chapter 1.1), analyzing the issue of non-compliance by an absence of practical application of ATD measures across EU Member States. And, drawing upon studies on ATD within the immigration policy field as well as research on the compliance with EU directives to analyze possible theoretical explanations across several cases. The aim will be to uncover if any theoretical explanations could explain why, in some cases, EU Member States implement an ATD measure in national law while refraining from applying it in practice. As such, the present study includes the following research question:

Under what conditions do Member States of the European Union refrain from actual application of an ATD in practice after its legislative implementation?

Several sub-questions are formulated based on the central question of the study. The sub-questions are divided into a theoretical and empirical category. The first theoretical question that the study needs to answer involves the position that practical application takes in the research on compliance, with the aim to uncover what this means regarding the adoption of ATD measures. The second question that is necessary to address is of a theoretical nature, aimed at finding potential theories that propose factors that influence the compliance. The first empirical sub-question (sub-question 3) involves the question as to the level of implementation and application of the alternative measures by EU countries by 2013. The fourth and final question involves the comparative analysis of the influence of the theoretical explanations on the level of implementation and application. The fourth question is divided into two further questions. The first being the analysis of the mediating effect of the typology as a remote condition. The second being the influence of the other theoretical factors, within the contexts of the different categories by the typology, as proximate conditions in a QCA-analysis.

Theoretical sub-questions:

1. *What is meant by the stage of the practical application after the transposition and implementation of EU directives?*
2. *Which theories are there that can explain the non-compliance by the non-application of ATD measures in practice by EU Member States?*

Empirical sub-questions:

3. *What were the levels of legislative implementation and practical application for the cases of ATD measures by EU Member States?*
4. *Which of the theories explain the cases found of non-compliance by non-application of the ATD measures in practice?*
 - a. *Can the typology of the four worlds of compliance explain the non-compliance by the non-application of ATD measures in practice by EU Member States as an outcome-enabling condition?*
 - b. *Can the factors from the enforcement & management approach explain the non-compliance by the non-application of ATD measures in practice by EU Member States as proximate conditions?*

1.3 Relevance

In addition to the scientific challenges this study aims to contribute to as described in the problem formulation (see Chapter 1.1), the study also has a societal significance. The humanitarian aspect of the treatment of individuals in the return process has already been an essential part of studies on alternative supervisory measures. Previous studies signal that these measures provide a more humane and dignified treatment, as detention can cause physical and psychological damage to those detained (Bloomfield, 2016; Fiske, 2016). The present study hopes to contribute to the discussion on these measures, by its analysis of the actual implementation of these measures under the 2008/115/EC Return Directive. And by doing so, answering some of the questions related as to why countries in the EU differ in their implementation. In addition to providing possible information that can be used to improve the implementation and application of ATD internationally and specifically for the Member States of the EU. The immigration policy field has been developing at increasing pace at the EU-level in recent years. Providing more insight into the possible obstacles of EU-level immigration policies could be vital for European Union institutions if these want to implement integrative EU-level immigration policies. The information from the present study could be relevant to policymakers and societal organizations in EU Member States as to know where the implementation of these kinds of measures would likely struggle in the process towards application.

1.4 Scope & Considerations

Several choices have been made concerning the scope of the study and on the considerations presented during the analysis. The scope of the study limits itself to the adoption of ATD measures within the European Union. An implication of this is that the scope of this review is also limited to the conditions of the legislative framework of the European Union. Therefore, it could be entirely possible that definitions, concepts and other aspects used in the study could be different from those used in countries outside of the EU. Additionally, directives such as the Asylum Procedures Directive and the Qualification Directive already give form to some of the conditions of the immigration policy field in the EU (EU Commission, 2016a, 2016d). The scope of the study and the incorporated data is also limited to analyzing the compliance between the period of 2008-2012, since the most comprehensive overview of the compliance of EU Member States with the adoption of ATD measures is the evaluative study commissioned by the EU commission in 2013 (Directorate-General for Migration and Home Affairs, 2013). As already mentioned in the introduction and the problem formulation, the scope of the analysis is limited to the topic of ATD measures within the return process (see Chapter 1 & Chapter 1.1). The present study excludes the group of ATD measures from the entry/asylum process as these do not fall under the 2008/115/EC Return Directive.

Although the scope of the study is limited to the legislative framework of the European Union, this does not mean that the framework results in a uniform approach between EU Member States. Each Member State can differ from the other in its immigration policies and its implementation. Consequently, there exists no uniform categorization of the ATD measures and their characteristics. Instead, each Member State has its own versions of measures, processes, regulations, and classification of the target groups. A consideration is made to maintain the broad categorization of ATD measures as used by the evaluation report commissioned by the EU commission (Directorate-General for Migration and Home Affairs, 2013). While the study also refrains from distinguishing between the different categories of target groups within the return process (see also Chapter 1.5.1). The reader should bear in mind that the study uses a general approach, as it is beyond the scope of the study to examine all the variations between each Member State included in the analysis.

A theoretical consideration is made to include relevant information from studies of alternative measures within the entry/asylum procedure if necessary. These measures often share similarities with those used within the return process. It could, therefore, be possible that studies aimed at the entry/asylum procedure contain viable information. The present study incorporates that information while assessing these critically for any relevant contradictions.

A methodological consideration is found in the practical difficulty to aggregate the necessary data from the EU Member States. There are apparent issues with studies and rapports trying to get the required data for their analysis (Basilien-Gainche, 2015; European Migration Network, 2016). Although there is a clear need for more cross-country comparative analysis on aggregate case-level data (Bloomfield, 2016; Helbling & Michalowski, 2017). However, due to the scope and practical limitations of the research, a purely quantitative study is not attainable at this moment. These limitations are why the present study combines both a qualitative and quantitative approach with the use of a qualitative comparative analysis. The advantage of such an approach is that it can compare multiple cases while also enabling a more in-depth analysis than a purely quantitative study (see also Chapter 3.1). Thereby combining both the need for more mixed-method based studies by the compliance field and more comparative analysis by those involved in studies related to the ATD measures.

1.5 Key Terms & Conceptual Framework

The following chapter is divided into two sections, beginning with a description of the central concepts used in the study (Chapter 1.5.1). The second section presents the conceptual framework (Chapter 1.5.2).

1.5.1 Key terms

This section describes some of the key terms used throughout the study. The first three concepts are related to the immigration policy field. All three of these concepts lack a single general definition. Therefore, this study incorporates definitions that are most representative of the assumptions and rationale of the present study. Other concepts are related to the field of EU integration and compliance. The last concepts described in this section are connected to the QCA approach and its terminology.

Immigration policy field

The following part explains the concept of administrative detention as is adhered to in the study. Defining the concept of administrative detention before the alternatives to detention is essential, as the alternatives are considered to serve as a replacement. Some of the underlying aspects of administrative detention thus also apply to the alternatives. Multiple different definitions exist for the concept of administrative detention. This study adheres to the definition as formulated by the article of Leerkes & Broeders (2010), as it highlights the administrative nature of detention within the immigration policy field. A crucial difference from criminal detention is, that imprisonment for ‘immigration-related goals’ is not meant as a form of punishment (Leerkes & Broeders, 2010). Rather, it is an administrative measure, and not a form of detention used pre-trial or after a conviction of a crime. So, when mentioning detention for immigration-related purposes, the term ‘administrative detention’ is used to distinguish it from other forms of detention (Chmelickova, 2006). In addition, the definition as formulated by Leerkes & Broeders (2010) is appropriate for the scope of this study, as it considers the administrative detention within the return process instead of the entry/asylum process. Leerkes & Broeders (2010) see the term of administrative detention as divided into two types. The first being the detention of those not admitted to the country at the border, including asylum seekers for some countries. The second type contains the detention of those already remaining within the respective country’s border and need to leave as their stay is deemed illegitimate, which is the type under analysis by present study (see also Chapter 1.4).

For the present study, the definition for Administrative Detention is thus: “*A non-punitive, bureaucratic measure that is meant to enable the enactment of border control. It merely ensures that ‘unwanted’ migrants can be located and identified and cannot abscond while the expulsion is prepared*” (Leerkes & Broeders, 2010, p. 830-831). Throughout the study, the term ‘administrative detention’ and the abbreviation of ADM (Administrative Detention Measure) is used to refer to this definition.

The following part explains what is meant by the term of the alternatives to detention (ATD). An accepted definition of ATD mostly lacks, as there is little consensus about what these alternatives entail by different actors. For example, the International Detention Coalition (IDC) uses the broad definition of: *“Any law, policy or practice by which persons are not detained for reasons relating to their migration status”* (Bloomfield, 2016, p. 31). In the same way as with the definition of administrative detention, such a definition does not distinguish between the different purposes such an alternative can have within the immigration field. Costello & Kaytaz (2013, p. 10-11) provide a definition that distinguishes between the use of such measures at the entry/asylum process or the return process, though this definition is altered for this study to only refer to the use within the return process. As such the definition is as follows: *“Encompassing any legislation, policy or practice that allows for asylum-seekers, refugees, and migrants to reside in the community while awaiting deportation or removal from the country, albeit subject to some restrictions on movement or liberty”* (Costello & Kaytaz, 2013, p. 10-11). Throughout the current research, the term ‘*alternatives to detention (ATD)*’ (and its variations) is used to refer to this definition.

However, the definition does not include an explanation of the measures that are considered part of these alternatives. Numerous studies provide a categorization of alternative measures, each having its variations regarding the included forms of alternatives. This study adheres to the categorization as supplied by the report on the application of the Return Directive by the Directorate-General for Migration and Home Affairs (2013), as that categorization forms the basis of inclusion of ATD cases for the analysis (see also Chapter 3.2). The advantage of such a classification is that it is straightforward, improving the possibility for a comparative analysis of each variety of alternative applied by each relevant Member State.

So far, the previous sections have described the definitions the term administrative detention' and alternatives to detention used throughout the study. This section will move on to discuss for whom these measures are meant. As mentioned in the previous sections, the focus of the present study lies in the use of alternatives to detention within the return process. As the study considers the EU member states that fall under the EU directive within the European Union, the used categories for these measures can be different per country. A broad term of Third-Country National (TCN) refers to: *“Any person who is not an EU citizen and who is not a person enjoying the right of free movement under Union law”* (Directorate-General for Migration and Home Affairs, 2013). However, it then remains unclear as to whom one refers to as being the target group for ATD measures. A more specific term of ‘third-country national found to be illegally present’ is defined as : *“A third-country national who is officially found to be in the territory of a Member-state and who does not fulfil, or no longer fulfils, the conditions for stay or residence in that Member State”* (EU Commission, 2016b). This definition does not incorporate the preference for the use of ‘irregular’ as opposed to ‘illegal.’ Which is preferred as it adheres to the

administrative characteristic of the immigration field as well as the removal of any juridical or ethical objections (EU Commission, 2016b). The broader term of 'irregular migrant' is more appropriate as, in the EU context, it refers to:

A third-country national present on the territory of a Schengen State who does not fulfill, or no longer fulfills, the conditions of entry as set out in the Schengen Borders Code, or other conditions for entry, stay or residence in that Member State. (EU Commission, 2016b)

Throughout the study, the term '*irregular-staying migrant*' is used to refer to this definition, with the addition that it incorporates that the irregular staying migrant is in the return process of an EU Member State that adheres to the Return Directive.

Transposition, implementation, and application

This study is interested in why, in some instances, alternatives in EU Member States are implemented in national legislation but not applied in practice. The use of alternatives is provided within the Return Directive (EUR-Lex, 2008), with Article 15 (1) stating that:

Unless other sufficient but less coercive measures can be applied effectively in a specific case, Member States may only keep in detention a third-country national who is the subject of return procedures in order to prepare the return and carry out the removal process. In particular when: (a) there is a risk of absconding, or (B) the third-country national concerned avoids or hampers the preparation of the return or the removal process. Any detention shall be for as short a period as possible and only maintained as long as removal arrangements are in progress and executed with due diligence. (EUR-Lex, 2008, p. 105)

The part of the sentence above 'unless other sufficient but less coercive measures can be applied effectively in a specific case' provides the necessity and prioritization of applying the alternatives. The following section moves to explain the concept of implementation of such measures in the context of EU directives and immigration policy.

The summary of EU legislation by the (EU Publications Office, 2015) describes European Union directives as:

The directive is one of the legal instruments available to the European institutions for implementing European Union policies. It is a flexible instrument mainly used as a means to harmonize national laws. It requires EU countries to achieve a certain result but leaves them free to choose how to do so.

A critical aspect of such directives is that these regularly don't fit with the national legislation of Member States. Consequently, Member States are required to implement a directive such as the 2008/115/EC Return Directive into domestic law. The term of 'transposition' is used as signifying the process of

implementing a directive into domestic law. Member States can vary in their compliance with the directives, as methods and instruments can be chosen for adhering to the objective of such a directive that fit with national legislation (Angelova et al., 2012). The compliance of Member States with directives has seen much attention in the field of European integration studies, as the European Union is a compelling case of separate countries with different characteristics having to transpose central directives. According to Tallberg (2002), there are two categories of non-compliance: (1) non-compliance by failing to transpose directives into legal legislation in a correct manner or within the time frame; (2) Non-compliance with the application of EU rules. These are both based on the general concepts of regime effectiveness and compliance. In which a distinction can be made between *“measures that states take to make international accords effective in their domestic law and if countries in fact adhere to the provision of the accord and to the implementing measures they have instituted”* (Tallberg, 2002). The second form of non-compliance is the subject of the present study, as it analyses the cases in which Member States did transpose the alternatives to detention as part of the Return Directive, but which have not applied the alternative into practice. It is therefore critical to distinguish between the terms that are sometimes interchangeably used for describing the state of ATD across Member States.

The DG Home Affairs of the EU commission has analyzed the implementation and application of ATD in EU Member States as part of the ‘evaluation on the application of the Return Directive (2008/115/EC)’ report (Directorate-General for Migration and Home Affairs, 2013). The report divides the application of ATD into the legal application and practical application. Legal application means that a Member State has implemented, either before or after the 2008/115/EC Return Directive, national legislation that makes it legally possible to apply that measure. Consequently, practical application means that Member State applies the ATD measure in practice (Directorate-General for Migration and Home Affairs, 2013). The present study uses the term *‘legislative implementation’* (and its variations) to refer to the legislative application of an ATD measure as is used in the report. The term *‘practical application’* thus refers to the practical application as is used in the report, to indicate that the Member State applies the ATD measure in practice. The distinction is made to remove a possible confusion between these two forms of legal and practical application used in the report,

As already mentioned in the introduction, it is necessary to bear in mind that the Return Directive does not require a Member State to have a certain amount of different ATD. A Member State can choose their own ATD and has discretion in how its alternatives work. Consequently, the study is not interested in the ratio of alternatives used per Member States, but rather the occurrence of alternatives implemented in legislation while not being applied in practice.

Qualitative comparative analysis

The methodological approach for the present study combines both qualitative and quantitative data for a Qualitative Comparative Analysis (QCA). QCA techniques apply set-theoretic methods to analyze causal mechanisms across several cases (Rihoux & Ragin, 2009). QCA uses its own terminology, which needs to be explained further. As mentioned by Schneider & Wagemann (2010), the basis for QCA are principles such as Boolean algebra, formal logic, and set theory. As such, what would typically be referred to as an 'independent variable' is referred to by QCA as a '*condition*.' The '*outcome*' refers to the 'dependent variable,' and the '*solution formula*' refers to the 'equation.' What would normally be referred to as the 'correlations' in statistical analysis is referred to as the '*set relations*.' The present study uses a two-step approach to the QCA. Conditions are analyzed in two steps, differentiated by the terms of '*remote*' and '*proximate*' conditions. Remote conditions are relatively stable over time, are more distant to the outcome, and often consist of contextual factors. These are analyzed to look for conditions that are considered 'outcome-enabling' (Sehring, Brockhaus, & Korhonen-Kurki, 2013). Proximate conditions are more directly related to the outcome, and often show more variation over time (Schneider & Wagemann, 2006). These are some of the more general terms used in this study related to the use of QCA, with the methodology of the study explaining in more detail (see also Chapter 3.1.1).

1.5.2 Conceptual framework

The following section describes the theoretical concepts that were used in the study. A short description is included, while a detailed description of each concept can be found in the theoretical framework (see Chapter 2). Figure 1 shows the conceptual framework and how the central concepts relate to each other within the study.

Enforcement approach

Public attitude on EU membership:

A negative supportive attitude from the citizens in a EU Member State towards the EU could increase the political opposition towards the EU, thus leading to more complications during the implementation of EU policy. The public attitude towards the EU membership of the State, therefore, influences the preferences of national political actors.

Public attitude towards immigrants:

A negative public attitude toward the specific topic of EU policy, in this case towards immigration, could influence the preferences of national political actors to implement EU policy relating to this particular topic.

Ideological position of the government:

The ideological position of the government (coalition) on the right to left spectrum influences the preference for the level of restrictiveness of implementing immigration-related policies.

Management approach

Level of federalism:

When lower level federal authorities have more control over the adoption of directives, more obstacles and objections could arise. A higher level of federalism, therefore, limits the capacity of a Member State to implement EU policies from a directive.

Number of parties in government:

The ability of a government to adequately implement legislation from an EU directive becomes increasingly limited as more parties are included in a coalition.

Administrative efficiency:

The quality and capacity of the bureaucratic system limit the capability of a Member State to implement EU directives adequately.

Culture of compliance

Four worlds of compliance:

Member States differ in their fundamental propensity for compliance, with different groups of States having different typical models of a culture of compliance. These ideal-typical implementation styles

have a mediating effect, creating different contexts in which certain conditions are of influence in the non-compliance with EU directives.

As can be seen in Figure 1, the present study examines the influence of the proximate conditions from the enforcement and management approach on the outcome of non-application of ATD measures that are implemented in national law by EU Member States. The four worlds of compliance are considered as remote conditions, having a mediating influence on the other proximate conditions. The remote conditions are more spatiotemporally distant conditions in relation to the outcome. These function as outcome-enabling factors for the proximate conditions from the two approaches, which are spatiotemporally closer in relation to the outcome. In statistical analyses, the proximate conditions could be seen as similar to independent variables, and the remote conditions for the four worlds of compliance as mediating variables. Figure 1 shows the preliminary conceptual framework, while the theoretical chapter presents a detailed framework (see Chapter 2.3).

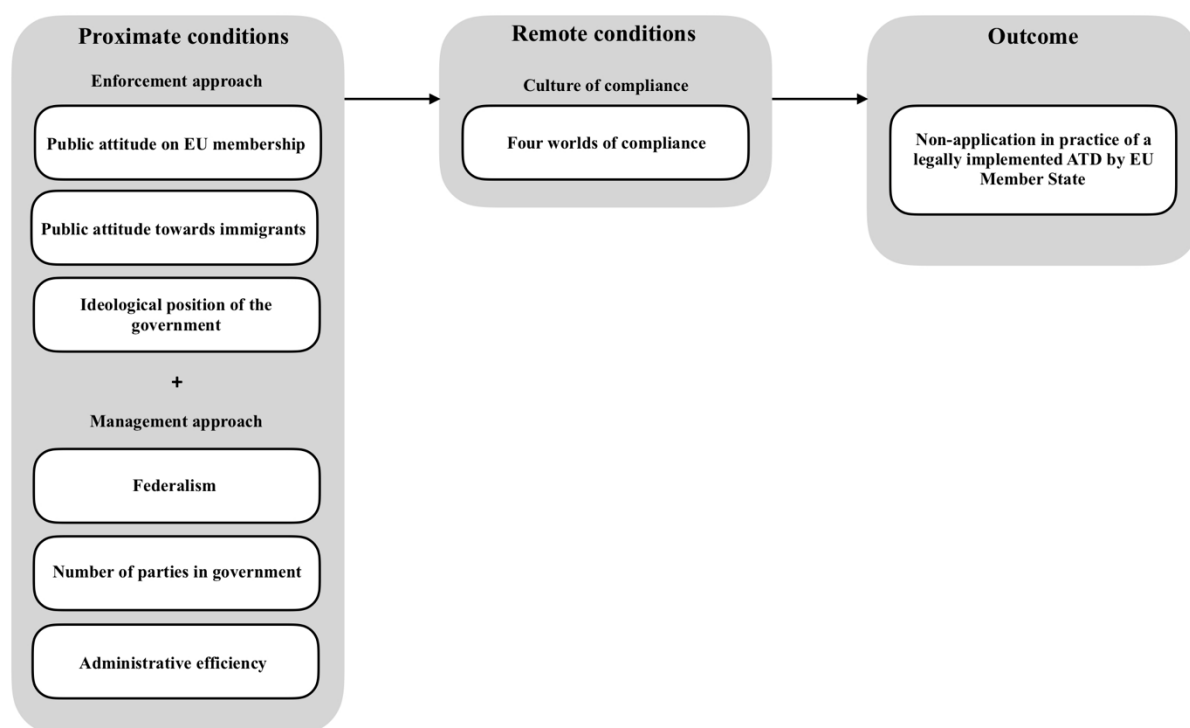


Figure 1. Preliminary conceptual framework

1.6 Structure of the Study

The structure of the study is as follows. The theoretical framework (Chapter 2) shows the relevant theories and hypotheses for the QCA and is divided into two sections. A first section describes the outcome of practical application as part of the literature on the compliance with EU directives. The second section includes the examination of existing literature for factors that could explain the non-compliance with EU directives, presenting several proximate conditions from two approaches. This section also includes an explanation of the four worlds of compliance typology as a remote condition.

The methodology (Chapter 3) gives an overview of the methods used for gathering and analyzing the data. The first section of the method clarifies the QCA method. The terminology of QCA, the use of a crisp-set form, and the inclusion of a two-step approach are also explained in the first section. Subsequent sections give an account of the data process, including the population & case-selection, operationalization & data-gathering, and data transformation. The data analysis section consists of the analytical steps, while a detailed outline of the application of the two-step approach is provided by the data procedure section. A final section of the methodology recounts the reliability & validity.

The remaining chapters proceed with a presentation of the results (Chapter 4). The results are divided into five sections. A first section describes the descriptive results, while the second section includes the analysis of necessary conditions. The last three sections involve the results from the two-step approach of the analysis.

These results are addressed in greater detail in the discussion, separated by a section on the remote conditions, proximate conditions, and the combination of both conditions.

Lastly, the conclusions (Chapter 6) provides the concluding remarks of the study, its limitations, implications recommendations for further research, and some final thoughts on the practical relevance.

2. Theoretical Framework

The following chapter of the study describes the theories that could explain the compliance with EU directives. The chapter is divided into two sections. The first section provides an overview of existing literature on compliance and the outcome of practical application as its sub-part (Chapter 2.1). Which leads to the second section, describing the relevant theories that can explain the practical application of ATD measures (Chapter 2.2). The chapter ends with an overview of the hypotheses and a detailed theoretical framework on the basis of the conceptual framework (Chapter 2.3).

2.1 Practical Application as Part of Compliance

A problem that often occurs with studies on compliance is that interpreting the findings is complicated by the unclear and varied conceptualization of compliance (Mastenbroek, 2005). Therefore, the present study carefully describes the conceptualization of compliance in the following section as well as in the conceptual framework, to limit the risk for such issues.

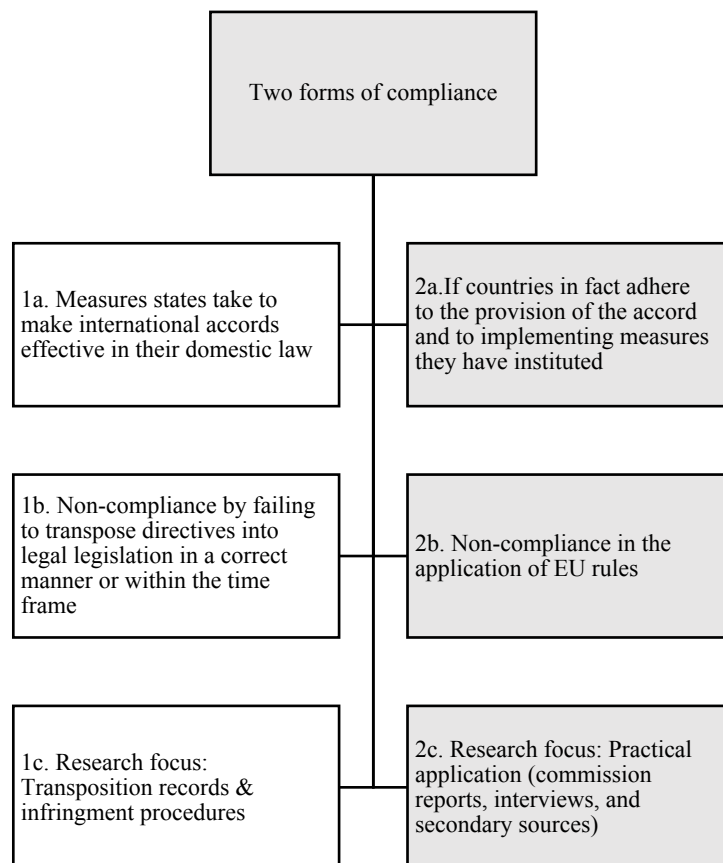


Figure 2. Compliance as used in this study (marked in grey). Adapted from the reviews by Tallberg (2002) and Angelova et al. (2012)

Over the years, compliance has become a vital aspect for those involved in European integration studies (Angelova et al., 2012). As indicated earlier (see Chapter 1.5.1), previous studies considered the practical application of EU directives as part of the overall compliance with international accords, based on the general concepts of regime effectiveness and compliance (Tallberg, 2002). Tallberg (2002) categorizes the compliance into two forms, as shown by Figure 2 (see 1a & 2a). These terms need to be adapted to the characteristics of non-compliance with EU directives (see Figure 2, 1b & 2b).

According to the study by Angelova et al. (2012), the non-compliance with EU directives has these different characteristics as the European Union consists of Member States with varied political and economic features. The challenge for each Member State of transposing a central directive into their national legislation enables a unique comparison of compliance across these countries and different policy fields. Most studies on compliance in the EU examine this by looking at cases involving either the transposition, infringements or application of directives. Whereby Angelova et al. (2012) argue that the first two aspects analyze the same form of compliance (see Figure 2, 1c), that is, the failure to transpose directives correctly or within the time frame (see Figure 2, 1b). The analysis of the present study includes the latter form, in which the practical applications of a directive fails. Source materials that provide the basis for such types of compliance studies consist of commission reports, interviews, and secondary sources (see Figure 2, 2c).

Having described the form of compliance used in the study, the following section moves on to discuss the development of theories on compliance with EU directives. Currently, it is still unclear which of the studied factors that influence the level of compliance are the most influential (Angelova et al., 2012). Mastenbroek (2005) describes three crucial "waves" of studies on compliance, that started after the compliance with EU directives came to the attention of scholars in the late 1980s. The first wave was described as an eclectic field of compliance, which includes studies advocating political, legal, and administrative variables as responsible for compliance. The field is characterized by Mastenbroek (2005) as eclectic, in that, the added variables often lacked a robust theoretical framework, leading to a random selection of aspects from several other forms of theories and research. Moreover, findings from quantitative analyses on the included variables have remained inconclusive.

The second wave of compliance studies appeared at the end of the 1990s. It involved a neo-institutionalist approach to compliance, with the compliance being dependent on the fit of EU policy with national policy. Member States try to upload their domestic policies to the EU, as to limit the costs of adopting EU legislation (Mastenbroek, 2005). Both a formulated rationalistic and sociological argument was considered part of the main aspects of adoption. Although the proposed 'goodness of fit' theory was characterized as having both a strong theoretical background as well as empirical possibilities, findings that support the argument have been limited (Mastenbroek, 2005).

The third wave followed as a reaction to the limited results of the 'goodness of fit' theory. The influence of national politics saw more attention for explaining compliance with EU directives with this

wave. The third wave includes two main approaches that compete for explaining compliance, the enforcement and management approach (Tallberg, 2002). The enforcement approach is based on rational choice and its game theory and collective action theory. Member States are considered as rational actors, assessing the costs of an option involved with a decision for its compliance (Tallberg, 2002). The two most important arguments from the enforcement approach have been the preferences and monitoring arguments (Angelova et al., 2012). The management approach is based on sociological institutionalism (Mastenbroek, 2005). It originates from qualitative case studies and considers that Member States have the general tendency to comply with EU directives. Capacity limitations and rule ambiguity limit that ability of a Member State to comply (Tallberg, 2002).

An entirely different formulated approach in the third wave considers the normative obligation of EU compliance. Which the study by Mastenbroek (2005) described as, "applications at the meta-level." This approach considers the cultural effects of individual Member States on compliance (Angelova et al., 2012). According to this approach, Member States differ in their tendency to comply with EU requirements based on their cultural characteristics. An example already mentioned by Mastenbroek (2005), is the research by Sverdrup (2004) that found a Nordic model focused on building a consensus with the commission over implementation conflicts. Another example is the more elaborate four worlds of compliance typology (Falkner et al., 2007; Falkner & Treib, 2008).

For the present study, it would be interesting to analyze the compliance of ATD measures by the three approaches of the third wave, whereby the analyses could point to the path that has the most influence regarding the practical application of ATD. The study by Angelova et al. (2012) compared the results of both qualitative and quantitative studies on variables based on the approaches as described in the section above. The comparison found robust findings for both institutional decision-making capacity of governments and goodness of fit. While studies on the culture effects found the expected outcome, though the number of reviews was too small to be conclusive. The influence of preferences also remained inconclusive, needing further research to consider its impact. A problem faced in many studies on compliance is that the data is of a hierarchical nature, making statistical analysis difficult (Mastenbroek, 2005). The approach of the present study enables the use of hierarchical data, by accounting for the interaction effects (see also Chapter 3.6). According to Mastenbroek (2005), the cultural approach by Falkner et al. (2007) could be considered as a mediating variable for others. The csQCA with a two-step approach should make it possible to incorporate such interaction effects.

2.2 Explaining the level of Compliance with EU Directives

The following section describes the explanatory factors that could explain the non-compliance for this study, based on the following three approaches: the enforcement approach, the management approach, and the cultural consideration. Hypotheses are presented which are used further in the analysis to test the explanatory factors as part of their respective approach.

2.2.1 Enforcement approach

As mentioned in the first section of the chapter, the enforcement approach was based on the rational choice theory. Two main arguments are considered as the most important on the enforcement approach. The ‘preference fit’ argument is based on the ‘goodness of fit’ argument though it considers the preferences of national politics (Mastenbroek & Kaeding, 2006). With the ‘preference fit’ fitting into the rational choice theory and enforcement approach to incorporate national politics into compliance, and the ‘dominant belief system’ fitting into the sociological institutionalism theory and management approach. The latter of which will be explained further in the section on the management approach (see also Chapter 2.2.2).

The second argument involves the ‘monitoring’ argument as an explanation for non-compliance. With national agencies and the EU commission trying to scan Member States for any instances of non-compliance (Angelova et al., 2012). Monitoring may be divided into two categories, supranational monitoring by the EU commission and national monitoring. Studies that analyze national monitoring consider the access to courts and interest group activities as an alarm mechanism for compliance. However, monitoring is an understudied argument and has so far showed only robust results in case studies (Angelova et al., 2012).

Preference fit argument

For this study, the ‘preference fit’ argument will be taken to explain the compliance level of Member States from the enforcement approach. Not in the least, because the primary challenge that still exists with this argument is to construct a valid model and variables that “capture” the preference of national actors (Mastenbroek & Kaeding, 2006). A study by Thomson, Torenvlied & Arregui (2007), found that non-compliance often takes place when Member States have little discretion in adopting a directive into national law and that the policy preferences of Member States influence the level of compliance. One studied factor as part of the preference fit argument has been the state power, which is the capacity of EU Member States to adopt EU legislation to their preferences (Toshkov, 2010). According to the review by Toshkov (2010), several studies analyzed state power, with both positive and negative effects. Furthermore, state power is not directly related to the preferences and serves more as a proxy, making it less interesting for the present study.

Another factor analyzed under the preference-fit argument are the societal EU attitudes, in which the argument is that a more EU supportive public leads to better compliance by the EU Member State. However, research on societal EU attitudes shows inconclusive results. While a more EU favorable attitude of the public does not show a positive relationship to compliance, results from some studies suggest a negative correlation (Toshkov, 2010). The reasoning behind this is that if the public attitude is negatively predisposed towards the EU, political opposition towards the EU increases (Bergman, 2000). Thus, lower support for the EU membership of a country would lead to more complications during the implementation of EU policy (Lampinen & Uusikylä, 1998). It is, therefore, a possible condition that could influence the compliance of EU Member States concerning the practical application of ATD measures. And thus, the following hypothesis is formulated:

H1: An EU Member State has an incentive not to move from legal implementation to practical application of an ATD measure if there is a negative attitude by the public regarding the country's EU membership.

A condition that has received less attention is the public attitude towards the specific subject of the EU policy. A case study by Leerkes & Broeders (2010) states that rising negative opinion of immigrants by the public could have influenced the preferences of that government for the application of ADM. Immigration policies have remained dominant subjects in the domestic political domain. With societies attitude towards immigrants being a potentially significant influence on the national government's compliance with EU immigration-related policies. Previous research could not find robust results for the predictive qualities of indirect proxies, such as economic conditions and unemployment levels, as an incentive for a state's compliance (Toshkov, 2010). However, these factors do seem connected with the incentives for a government for implementing immigration policy. As according to Haas & Natter (2015), economic growth, unemployment recent immigration levels, and other political system factors influence the implementation of more restrictive immigration policies. The influence of immigration flows on anti-immigrant opinions has also been confirmed by other studies (Kessler & Freeman, 2005; Lahav, 2004; Meyers, 2000). Altogether, these studies indicate that there could be a missing link that connects these indirect proxies with compliance in relation to EU immigration policies. The missing link could be the public attitude towards immigrants. In which, the proposed underlying causal mechanism would be that low economic growth, high unemployment levels and a rise in recent immigration flows increase the negative public predisposition towards immigrants. A high level of negative public attitude then puts pressure on a government to refrain from implementing less restrictive EU regulation.

The study by Leerkes & Broeders (2010) is based on a single-case evaluation, while the proposed link between the indirect proxies, public attitude towards immigrants, and ultimately the incentives for a government for non-compliance is untested. This makes it a compelling argument for

further analysis, testing the influence of public attitudes to immigrants across several countries from the EU. The following hypothesis is formulated as a result:

H2: An EU Member State has an incentive not to move from legal implementation to practical application of an ATD measure if there is a negative public attitude towards immigrants.

Other studies considered the impact of the ideological positions of governments. While societal attitudes of the EU examine the public perception, this involves the influence of the view of the government. One aspect considers the level of affinity of the government towards the European Union, though this is often only measurable by conducting expert surveys (Toshkov, 2010). Another factor that has been commonly proposed as an argument by ATD-related studies, is the political left/right position of governments. Koopmans, Michalowski & Waibel (2012) state that immigrant rights had been expanding and became more inclusive across European countries until 2002. Around that time, the rise of right-orientated and anti-immigrant parties negatively influenced that trend. According to that study, electoral and government configurations affect the expansion of immigrant rights. With the government composition (with conservative parties being more restrictive and liberal parties being less) as one of the factors. Another study by Welch & Schuster (2005) states that traditional social trends in the UK and the US led to a pronounced disinterest from the public for the criminalization of immigration policy. Still, the study by Haas & Natter (2015) found that the political orientation of the government does not explain the restrictiveness of a countries immigration policies. Other case studies found mixed results overall, attributed by Haas & Natter (2015) to a lack of long-term, systematic and quantitative comparative research. And factors like economic growth, unemployment, recent immigration levels and the political system were found to actually influence the level of restrictiveness.

The present study, therefore, expects that the ideological orientation of the government does not lead to the effect on the compliance with Article 15(1) of the 2008/115/EC Return Directive, in contrast to the expectation from some case studies. Which leads to the following hypothesis:

H3: An EU Member State's incentive to not move from legal implementation to practical application of an ATD measure is not influenced by the ideological orientation of the government towards the right.

2.2.2 Management approach

As explained previously, the management approach is based on sociological institutionalism, explaining compliance as originating from the administrative capacity of Member States and the often inherent vagueness and ambiguity of EU directives (Tallberg, 2002). Rule ambiguity has been part of the research by Basilien-Gainche (2015), which analyzed cases to the Court of Justice of the European Union (CJEU). And by examining these cases, found that the vague position of the CJEU on the Return Directive gave EU Member States room for interpretation. Thereby giving some importance to the interpretation discretion Member States have in adapting EU directives into national law. However, this does not explain rule ambiguity for the present study, as it is based on the vagueness of the position of the court on such a directive. For this reason, the present study directs itself towards the administrative capacity limitations as part of the management approach.

Administrative capacity limitations

As for the administrative capacity limitations, the literature distinguishes several underlying factors. An important factor considers the institutional decision-making constraints. The study by Angelova et al. (2012) mentioned three different aspects for institutional decision-making constraints: the number of veto players, federalism, and the effective number of parties. Institutional decision-making constraints showed robust findings over a substantial amount of cases in the study by Angelova et al. (2012). Which can also be attributed as the findings from studies on veto players, federalism, and the effective number of parties were all merged into one indicator for the institutional-decision making constraints. Each of these can be used as an individual indicator for institutional decision-making capacity, as each is an independent empirical phenomenon. However, these are all used to operationalize the same factor of institutional decision-making capacity (Angelova et al., 2012). Though, the study by Falkner et al. (2007) has found that the indicator of veto players only has a weak explanatory power on the level of compliance. Which is why the current research only includes federalism and the effective number of parties as indicators for institutional decision-making capacity.

The level of federalism is considered by the compliance literature to influence compliance by giving lower-level federal authorities more control over the adoption of EU directives by national governments (Tallberg, 2002). This could complicate the process of adoption by national governments, as lower level authorities could have a say on certain aspects. By way of contrast, federalism does show a positive relationship with less restrictive immigration policies in the study by Haas & Natter (2015). With the argument that negotiating laws would lead to a more liberal policy direction. Though, that effect is only found in the period before 1989, after which it turns insignificant. The comparative review by Toshkov (2010) has concluded that there is a strong argument for a significant negative relationship between federalism and compliance. As such, the present study adheres to that argument, that federalism

could be a vital cause for non-compliance with the practical application of ATD. Therefore, the following hypothesis is formulated:

H4: An EU Member State lacks the capacity to move from legal implementation to practical application of an ATD measure if there is a level of federalism.

As it remains unclear what kind of relationship federalism has for immigration-related EU policy adoption, this study also includes the number of parties in government as an indicator for institutional decision-making capacity. The effective number of parties in government considers the ability of governments to adopt and implement EU legislation. A government that consists of multiple parties has to deal with the preferences of each of these parties. As such, this makes it more difficult for these kinds of coalitions to transpose and implement EU legislation into national law. The larger the distance in preferences between parties, the more difficult it is to integrate these preferences in a unitary approach (Toshkov, 2008). The number of parties in government has been a proven concept for explaining the capacity of a government (Angelova et al., 2012). Therefore, the number of parties in government is included as part of the management approach as a condition for non-compliance with EU directives, specifically the non-compliance with practical application. The following hypothesis is thereby formulated:

H5: An EU Member State lacks the capacity to move from legal implementation to practical application of an ATD measure if the government consists of multiple parties as a coalition.

Administrative efficiency

Another factor that is part of the administrative capacity limitations is the administrative efficiency of a government (Angelova et al., 2012). With the argument that a low government capacity and quality lead to a risk of non-compliance (Toshkov, 2010). Correct transposition of an EU Directive requires an adequate bureaucratic system (Toshkov, 2008). Regarding the practical application of ATD's, the political system is dependent on the public administration for adopting, implementing and enforcing the use of ATD's. Including administrative efficiency as a condition is also of interest as previous studies only found robust findings by case study analysis (Angelova et al., 2012). A further comparative analysis could contribute to the review of its significance for EU compliance. Thus, the following hypothesis is formulated:

H6: An EU Member State lacks the capacity to move from legal implementation to practical application of an ATD measure if the administrative efficiency is low.

2.2.3 Cultural consideration: The four worlds of compliance typology

The previous approaches contain conditions that involve lower-level, proximate conditions on the multi-level structure of EU compliance data as formulated by Mastenbroek (2005). The third approach for explaining compliance under consideration is the culture of compliance approach. The study by Mastenbroek (2005) describes the culture of compliance as the difference between EU Member States in their fundamental propensity for compliance, with different country groups sharing a typical model for compliance. As such, the culture of compliance is part of a multi-level hypothesis and is argued to have a mediating effect on the effect of other lower-level conditions. The culture of compliance has been analyzed by looking at the satisfaction with democracy, prevailing norms, the rule of law, or as the cultural approach of a Member State to conflict resolution. Another theory aims at explaining the difference in implementation styles across EU countries. The studies by Falkner et al. (2007) and Falkner & Treib (2008) categorized the current EU Member States within different 'worlds of compliance' regarding their implementation of EU directives, based on their transposition, enforcement, and application levels.

The categorization was based on a qualitative study of 90 cases, in which the authors found no 'single overriding factor' that explains the compliance of a Member State. Rather, the typology of four worlds of compliance has been based on an 'ideal-typical implementation style.' Falkner et al. (2007) divide the process of transposition by an administrative and a political phase. In which the administrative system of a Member State identifies the necessary adaptations for EU law to be transposed into national law and starts that process. The political phase consists of the politicians, interest groups and other actors that influence the decision-making during the transposition process.

The first world of 'law observance' constitute of Denmark, Finland & Sweden and sees their process of transpositions and implementation as consistent.

The second world of 'domestic politics' consists of Austria, Belgium, Germany, the Netherlands, Spain and United Kingdom. These countries see some issues with transposition as domestic political concerns could cause non-compliance. When there are no political concerns, implementation is often a quick and easy process, due to their effective bureaucracy.

The world of 'dead letters' consists of Ireland, Italy, Czech Republic, Hungary, Slovakia, and Slovenia. These countries see a transposition process with some political conflict from domestic influences but show a gap between legislation and actual implementation and enforcement (Falkner et al., 2007; Falkner & Treib, 2008). For the central eastern states that are considered new Member States to the EU, this can be explained by the admission process to the EU. During the admission process of EU legislation, transposition of EU legislation was strong to meet that process of integration. After admission, national governments see fewer benefits from transposing legislation and even fewer incentives for implementation and enforcement (Falkner & Treib, 2008; Raunio, 2005). For the other countries, a possible explanation for the lack of practical application after legal implementation during

the transposition process is the concept of 'emulation.' Meaning that there was no political incentive to transpose the directive but that top-down elite-driven demands from the European Union were the reason. This leads to a difficult implementation as it lacks cultural and institutional support (Cain, Dalton, & Scarrow, 2003).

The world of 'transposition neglect', consisting of France, Greece, Luxembourg, and Portugal, considers the national level as dominant at the supranational level. Transposition is often only achieved with strong supranational pressure with inactive and ineffective administrative bureaucracy making implementation and enforcement inconsistent (Falkner & Treib, 2008).

However, studies proposing this approach fail to adequately define what the culture aspect of the four worlds consist of (Mastenbroek, 2005). This limitation is acknowledged in the article of Falkner et al. (2007) which analyzed the compliance culture by distinguishing the four worlds of compliance by a 'culture of law-abidingness.' The culture of law-abidingness is proposed by Toshkov (2007) as consisting of the value towards the rule of law that is expressed by the majority of citizens, preference to abide explicit rules, favor rule obedience, the trust in authority, and general trust of the citizens of a Member State. Though, only the general trust of citizens provides robust findings in that study. The limited explanation behind the typology thus remains. Falkner et al. (2007) advocate that more systematic empirical research is necessary to see if the typology is also usable in all or other EU-related policies. While also explicitly promoting the use of qualitative methods, uncovering the causal mechanisms influencing the level of compliance with EU directives.

It is therefore of significant value to see if this theory holds up in comparison with alternatives to detention being part of the migration policy field. In particular, the typology is meant to function as a form of a mediating filter as to which conditions are more dominantly present in some EU Member States within one of the categories. For example, non-compliance for countries that pertain to the world of 'transposition neglect' can be attributed to the dominance of national standards over EU level standards as well as administrative inefficiency. These countries could see problems in the enforcement and application of EU directives (Falkner & Treib, 2008). The analysis of this study can thus consider the value of the typology as a mediating condition in the first step. Especially, as Toshkov (2007) states that the culture of compliance is a stable factor over time, making it an appropriate remote condition with a two-step qualitative comparative analysis (see also Chapter 3.6). Which in turn makes it possible to see if the practical application of ATD measures shows certain patterns across the four worlds. It can then zoom in on the conditions of the first two approaches, determining which conditions are dominant across and within the different worlds of the typology. Accordingly, the following hypotheses are formulated, based on the theory of four 'worlds of compliance' and their influence on the level of compliance with the practical non-application of ATD measures:

H7a: An EU Member State's culture of compliance does not lead to non-compliance with the practical application of ATD measures if it is part of the first world of 'law observance' or the second world of 'domestic politics.'

H7b: An EU Member State's culture of compliance does lead to non-compliance with the practical application of ATD measures if it is part of the third world of 'dead letters' or the fourth world of 'transposition neglect.'

H7c: The appearance and influence of the factors from the enforcement & management approach on the outcome are different for each of the relevant worlds of compliance.

2.3 Summary & Theoretical Framework

The chapter began by describing the different aspects of compliance, resulting in the explanation of the position of practical application within the literature on EU compliance. The following section described the three approaches and their underlying conditions that the present study incorporates for further analysis on the practical application. Figure 3 gives an overview of the previously described conditions in Chapter 2.2. A symbol indicates the relation with the outcome of each of the proximate and remote conditions. Based on the theoretical framework, hypotheses were formulated for each of these conditions. Table 1 on the following page gives an overview of these.

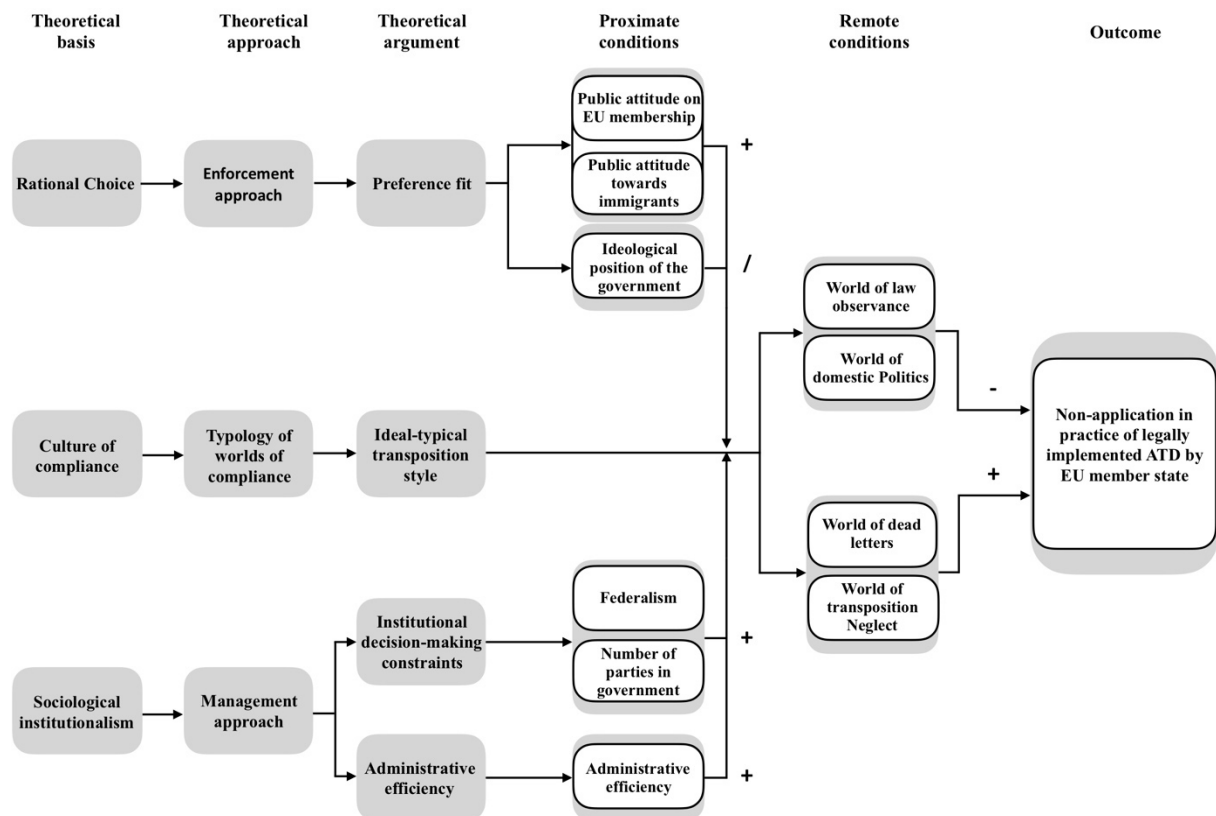


Figure 3. Definitive theoretical framework, based on the preliminary conceptual framework (Figure 1)

Table 1. Overview of the hypotheses

Approach	Hypothesis
Enforcement approach	<i>H1: An EU Member State has an incentive not to move from legal implementation to practical application of an ATD measure if there is a negative attitude by the public regarding the country's EU membership.</i>
	<i>H2: An EU Member State has an incentive not to move from legal implementation to practical application of an ATD measure if there is a negative public attitude towards immigrants.</i>
	<i>H3: An EU Member State's incentive to not move from legal implementation to practical application of an ATD measure is not influenced by the ideological orientation of the government towards the right.</i>
Management approach	<i>H4: An EU Member State lacks the capacity to move from legal implementation to practical application of an ATD measure if there is a level of federalism.</i>
	<i>H5: An EU Member State lacks the capacity to move from legal implementation to practical application of an ATD measure if the government consists of multiple parties as a coalition.</i>
	<i>H6: An EU Member State lacks the capacity to move from legal implementation to practical application of an ATD measure if the administrative efficiency is low.</i>
Four worlds of compliance typology	<i>H7a: An EU Member State's culture of compliance does not lead to non-compliance with the practical application of ATD measures if it is part of the first world of 'law observance' or the second world of 'domestic politics.'</i>
	<i>H7b: An EU Member State's culture of compliance does lead to non-compliance with the practical application of ATD measures if it is part of the third world of 'dead letters' or the fourth world of 'transposition neglect.'</i>
	<i>H7c: The appearance and influence of the factors from the enforcement & management approach on the outcome are different for each of the relevant worlds of compliance.</i>

3. Methodology

This section of the study describes the methodological process. The first section of the chapter explains the QCA approach, its terminology, and assumptions (Chapter 3.1). The case-selection is described thereafter (Chapter 3.2). The next section gives an overview of the operationalization and data-gathering of each of the conditions and the outcome (Chapter 3.3). The data transformation section shows how the data was converted to dichotomized values (Chapter 3.4). A section is then included that clarifies the analytical process (Chapter 3.5), while another section describes the two-step procedure used by the study (Chapter 3.6). Finally, some remarks are included in the final section regarding the reliability and validity of the study (Chapter 3.7).

3.1 Qualitative Comparative Analysis

This study aims to explain the appearance of cases in which EU Member States implemented an ATD in national law yet have not applied these in practice. The analysis of this form of non-compliance with an EU Directive was done with the use of a qualitative comparative analysis. One of the main advantages of a QCA is that it can combine other qualitative and quantitative methods and data (Rihoux, 2008). A QCA provides a possibility for an in-depth investigation of the relevant patterns and variations that the study tries to explain (Schneider & Wagemann, 2010). This is another advantage of QCA, as it can verify hypotheses based on existing theoretical expectations (Schneider & Wagemann, 2010; Vis, 2009). However, the actual interpreting of the causal inference should be made explicit and with due consideration as the generalization can only apply to specific combinations of cases within a particular context (Rihoux, 2008). A common criticism of the use of QCA is that it is vulnerable to selection bias and other data problems (Krook, 2010). These issues necessitate the need for substantial (theoretical) knowledge to conclude if the cases and their conditions and outcomes represent a causal inference to a certain extent (Schneider & Wagemann, 2010). For the present study, the QCA method was meant to give further indications for important (combinations) of hypotheses leading to the outcome, given that there were not a lot of specific theoretical explanations on the practical application of ATD measures. This approach to analysis was also quite useful for this study as it assumes 'equifinality,' which proposes that different combinations of the theoretical expectations could lead to the same outcome. Furthermore, it does not share some of the assumptions that quantitative methods do have, such as linearity, additivity or homogeneous units (Rihoux, 2008).

Lastly, there are two strategies described that are central to QCA. The first being the identification of conditions (formal language in QCA for independent variables) that are shared by the cases with the same outcome, which involves the determination of necessary conditions. The second strategy is to identify which cases with the same causal conditions share comparable outcomes and consists of the determination of sufficient cases (Rihoux, 2008). Regarding the practical application of

the QCA approach, it is mainly suitable for small-*N* to intermediate-*N* studies. Most intermediate-*N* research include between ten and fifty cases, which is similar with the number of available cases for this study. The analysis of the different cases is based on both qualitative and quantitative sources and is both theory and case informed, going back and forth between the data and the QCA (Rihoux, 2008; Vis, 2011).

3.1.1 Terminology of QCA

As has been mentioned in the description of the key terms (see Chapter 1.5.1), QCA is based on formal logic, Boolean algebra, and set-theoretic language. Moreover, it does not use terms such as independent and dependent variables, as often is one of the leading causes of confusion with QCA (Rihoux, 2008). The different terminology of QCA was maintained in the present study as to not confuse the differences between the QCA approach and its logical and methodological approach with conventional statistical terminology (Schneider & Wagemann, 2010). Therefore, this study used the term of ‘*condition*’ instead of an independent variable. The term of ‘*outcome*’ was used instead of the dependent variable, with the result termed as the ‘*solution formula*.’ The analysis of the conditions and outcome were done with the aid of a ‘*truth table*’ (Table 3) that organized and aided in determining the conditions that (together) lead to the outcome of the implementation of ATD measures (Rihoux, 2008). The resulting solution formula was presented with the use of Boolean algebra. If a condition was found to be present, uppercase letters were used. Alternatively, small letters indicated the presence of a negated version of a condition (meaning the absence of a condition). A ‘+’ indicated a logical OR, used to indicate that the outcome can be caused by either a (subset of) condition(s) or another (subset of) condition(s). A ‘*’ indicated a logical AND, used to indicate that the outcome can be caused by the combination of either a (subset of) condition(s) and another (subset of) condition(s) (Schneider & Wagemann, 2010). ‘→’ was used to indicate sufficient conditions, while ‘←’ indicated a necessary condition. As an illustration, $A \rightarrow Y$ indicates that A is a sufficient condition for the outcome Y. $A \leftarrow Y$ indicates that A is a necessary condition for Y (Corcaci, 2017). Another possibility was the use of ‘=’ to indicate a sufficient and necessary condition for the outcome. For example, $A=Y$ indicates that A is a necessary and sufficient condition for outcome Y. According to Schneider & Wagemann (2010), this only rarely happens as it should be based on an entirely specified truth table with absent contradictory rows and logical remainders are a non-issue.

3.1.2 Crisp-set QCA

The study used a crisp-set QCA to explain how the conditions lead to the non-compliance by non-application of ATD measures as the outcome. A csQCA uses dichotomous scores to denote set membership. Meaning that crisp-set QCA scores set-membership as being 0 (no membership) or 1 (full membership). For example, the outcome of non-compliance of the practical application of an ATD by an EU Member State scored either a 1 or 0. As shown by ‘Table 3: legal and practical application of alternatives to detention’ of the evaluation report by the Directorate-General for Migration and Home Affairs (2013, p. 30), Greece had implemented residence restrictions legally but not in practice. In a csQCA, that case was scored a 1 for full membership. Slovenia had a form of residence restrictions implemented in legislation as well as having it applied in practice, resulting in a membership score of 0 (no membership).

Another method would be to use a fuzzy-set QCA (fsQCA), where membership to a particular condition could vary between zero and one. The zero and one are the qualitative distinctions with such a calibration, while any value in between is of a quantitative distinction. For this study, however, csQCA was more appropriate. As with a fsQCA, the outcome also needed to include a fuzzy set-membership calibration (Krook, 2010). Such a calibration was not possible with the outcome as operationalized for this research, and thus csQCA was used for the further analysis. Thus, the data for the study was operationalized and calibrated to indicate the dichotomous set scores of a crisp-set QCA (see also Chapter 3.3).

3.1.3 Two-step approach to a QCA: Remote & proximate conditions

Just like previous studies that used a QCA approach, the study incorporated theories on a multi-level hierarchy (Rihoux, 2008). In that way, the approach enabled the inclusion of both the lower-level conditions from the enforcement & management approach and the higher level conditions from the four worlds of compliance typology (Vis, 2011). The inclusion of both levels was done by incorporating a two-step structure of QCA, which was developed to limit the complexity of the research by reducing the number of conditions in one QCA (Schneider & Wagemann, 2006). The conditions in such a structure are analyzed in two steps, differentiated by the terms of 'remote' and 'proximate' conditions as is shown in Figure 4.

Remote conditions are those that are spatiotemporally distant to the outcome, stable over time, and out of the manipulative reach of the involved actors

- Remote conditions

Proximate conditions are those that are spatiotemporally close to the outcome, can vary easily over time, and can be manipulated by the involved actors

- Proximate conditions

Figure 4. Difference between remote & proximate conditions, based on *Sehring et al. (2013)*

The use of these terms signifies a difference in the hierarchical level of the conditions, similar to the use of the terms of micro- and macro-level variables in statistical terminology. The difference is that it depends on the approach of study if a condition is regarded as remote or proximate, as it can be both.

The first step of the two-step structure is meant to result in one or multiple configurations, which are analyzed further with the relevant proximate factors. As such, several "pathways" of remote factors and their proximate factors are examined. The purpose is to make it possible to determine which proximate factors play a role within a specific outcome-enabling condition (see also Chapter 3.6).

3.2 Population & Case-selection

The population for the included cases of the csQCA consisted of the 25 European countries in addition to Denmark, Ireland and the United Kingdom and their implementation of ATD measures in 2012. The latter three countries are allowed to opt in or out of EU legislation when these are part of Asylum policies (Walter-Franke, 2017). And all three countries had opted out of the Return Directive. These countries were still included in the present study, as their implementation and application of ATD had been analyzed (Bruycker et al., 2015; Directorate-General for Migration and Home Affairs, 2013). The advantage of including these countries was that these had not been bound to apply EU law from the Return Directive, making it possible to examine if these countries showed a different combination of causes than the other included Member States.

Of the initial group of 28 countries, 19 countries were selected based on their inclusion in the worlds of compliance by Falkner & Treib (2008). For each of the 19 states, cases were selected across their implementation and application of ATD measures from four categories (see also Chapter 1.5.1). These consisted of: residence restrictions, regular reporting to authorities, obligation to surrender passport and documents, and the deposit of adequate financial guarantee. The report by the Directorate-General for Migration and Home Affairs, (2013) was the source of the included categories of ATD measures. The fifth category of electronic monitoring was excluded as these kinds of alternatives were and are still almost never used as a viable option for nearly all EU Member States (Bruycker et al., 2015).

With the removal of the category of electronic monitoring, the resulting population consisted of 76 potential cases. 31 of the possible cases did not have enough information for an evaluative score on the practical application of the ATD measure, (Directorate-General for Migration and Home Affairs, 2013). These 31 cases were thus also excluded. Moreover, the analysis revealed that there were several instances of different outcomes on the practical application within a country. The present study could not account for these, which is why Member States that contained cases of ATD measures with a score that conflicted with an ATD measure that showed the outcome of being not applied in practice or implemented in legislation were also excluded. The reasoning was that the Member States are not required by Article 15(1) of the 2008/115/EC Return Directive to have implemented measures from each category (Basilien-Gainche, 2015). It is thus the assumption that for these 5 cases the implementation of the ATD measure was not considered, making its exclusion necessary as the conditions did not influence these cases. The cases that were excluded on the basis of that argument consisted of: (1) Czech Republic (residence restrictions, obligation to surrender passport and documents); (2) France (deposit of adequate financial guarantee); Luxembourg (obligation to surrender passport and documents, deposit of adequate financial guarantee) (Directorate-General for Migration and Home Affairs, 2013).

Altogether, this resulted in a total of 40 included cases for further analysis. Table 2 provides an example of the case-selection, while Table 3 of the data transformation offers an overview of all the included cases (see Chapter 3.4).

Table 2. Example of the case-selection

<i>Countries</i>	<i>Category of ATD</i>	<i>Example of cases</i>
Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Portugal, Slovakia, Slovenia, Spain, Sweden	1. Residence restrictions 2. Regular reporting to authorities 3. Obligation to surrender passport and documents 4. Deposit of adequate financial guarantee	Implementation of: 1. Austrian ATD of Residence Restrictions 2. Austrian ATD of Regular reporting to authorities 3. Austrian ATD of Obligation to surrender passport and documents 4. Austrian ATD of Deposit of adequate financial guarantee

Note. Adapted from (Directorate-General for Migration and Home Affairs, 2013).

3.3 Operationalization & Data-gathering

After the case-selection, the next stage of the two-step csQCA was to operationalize and gather the necessary data on the indicators that are used to measure the outcome and conditions. The following sections of the chapter describe that process more extensively for the outcome and the conditions from the three approaches. A final section describes the raw data table, as is required for a QCA, which includes all the data from the outcome and conditions for the included cases of the study.

3.3.1 Outcome: Level of legal implementation and practical application

The outcome under analysis was the *practical non-application of ATD measures by EU Member States*. As mentioned in the theoretical chapter (Chapter 2.1), the specific type of non-compliance under the scope of the present study is the situation in which an ATD measure was implemented legally but with non-application in practice. The data for the outcome was based on the evaluation of the application of the ATD measures by the Directorate-General for Migration and Home Affairs (2013, p. 30). The assessment of the application of ATD's combined stakeholder interviews and previous legal studies from 2008 to 2012. For each of the five categories in that rapport, the legal implementation (noted as legal application in the report, see also Chapter 1.5.1 Key terms) and practical application for each Member State was determined. This study maintained the same classification of the outcome as either showing no legal implementation and practical application, legal implementation but no practical application, or legal implementation and practical application. As already mentioned in the previous section of the chapter (see Chapter 3.2), the outcomes that were included in the analysis were part of four categories. Furthermore, there were cases for which no information was available on the outcome. Cases for which there was no information available on the outcome were excluded in the case-selection. The evaluation rapport also omitted a determination of the practical application of two ATD categories for Spain, as there was a disagreement between the government and other stakeholders on the application (Directorate-General for Migration and Home Affairs, 2013). The outcomes of these cases were also left out of the analysis.

3.3.2 Proximate conditions: Enforcement approach

Three conditions were incorporated in the study which aimed to measure the 'preference-fit argument' within the enforcement approach: the *public attitude on EU membership*, the *public attitude towards immigrants*, and the *ideological position of the government*.

The *public attitude on EU membership* was measured by using survey data from the Eurobarometer (European Commission, 2009, 2010, 2013a, 2013b, 2015). The total percentage of the participants that answered, 'a bad thing' to the question 'Generally speaking, do you think that (your country's) membership of the EU is ...?' was used to measure the indicator. The data were collected from 2008, 2009, 2010, 2011, and 2012. After which these percentages were divided, to provide an

average rate. Only the total percentage of the participants that gave that specific answer was used, excluding the other categories. According to Bergman (2000), such a category has proven to be a suitable indicator for EU opposition.

The *ideological position of the government* was measured by using the Comparative Political Data Set (CPDS). The CPDS collected political and institutional data from 36 OECD and/or EU Member State countries between 1960 and 2015 (Armingeon et al., 2017). The cabinet composition, as calculated by the Schmidt-Index, was used as the indicator for the *ideological position of the government* (Schmidt, 1992). The Schmidt-Index calculates the cabinet composition by the government composition of social democratic and other left parties as a percentage of the total cabinet posts. These are then weighted by the number of days in office in a given year to account for any party composition changes during that year. Normally, the Schmidt-Index assigns a code between 1 and 5 for each year (Schmidt, 1992). The present study combined the government composition of EU Member States from 2008 to 2012, to account for any fluctuations during that period. The collected government composition percentages were averaged and then classified by the Schmidt-Index, which resulted in an average cabinet composition between 2008 and 2012. However, the database missed the data for Italy in 2012, as this was an entirely technocratic government (Armingeon et al., 2017). The average of Italy was adjusted accordingly in this study, as the data from 2012 was excluded from the calculation of the averages.

The *public attitude towards immigrants* was measured by using survey data from the fourth round of the European Value Study (EVS) in 2008 (EVS, 2016). Another option was to use an indicator from the European Social Survey (ESS), which would have made it possible to retrieve data from 2008 and 2012 (Norwegian Centre for Research Data, 2016). Additionally, the same question was used in the case study of Leerkes & Broeders (2010) to analyze the negative public attitude on immigration from non-Western countries in the Netherlands. However, the ESS did not include any data from Luxembourg after 2004. The present study, therefore, obtained data from the EVS instead. The percentages of participants from each Member State that provided an answer to the statement 'In the future the proportion of immigrants will become a threat to society' were collected for 2008 (EVS, 2016). The answer possibilities covered a scale from 1 to 10, with 1= strongly agree, and 10= strongly disagree. For the current analysis, an average value was calculated for each Member State and assigned to the included ATD cases of that State. The total percentage of the responses from 1 to 4 for the respondents of each country were added together and divided to provide an average. The same step was also done for the percentages from 5 to 10.

3.3.3 Proximate conditions: Management approach

Two conditions were incorporated to measure the ‘institutional capacity limitations argument’ within the management approach: the *level of federalism* and the *number of parties in government*.

The *level of federalism* was measured by using the Comparative Political Data Set, which was also used for the *ideological position of the government* condition (Armingeon et al., 2017). The data from the CPDS on the *level of federalism* is based on the data from the Comparative Welfare States dataset ((Huber, Ragin, Stephens, Brady, & Beckfield, 2004). The *level of Federalism* has been a consistent indicator, rarely changing between years for a state. Still, data was gathered from the period of 2008 to 2012. An examination of that data found no relevant changes for the period under consideration. The scale indicating the level of federalism used by the CPDS goes from 0 (indicating no form of federalism, to 1 (indicating a weak level of federalism), and to 2 (indicating a strong level of federalism) (Armingeon et al., 2017). There were no missing data from the CPDS, although it should be noted that Spain has been classified as a strong federal state while officially being classified as a unitary state. The present study adheres to the classification of Spain as determined by the CPDS (Armingeon et al., 2017).

The *number of parties in government* was measured by using the European Representative Democracy data Archive (ERDA). The ERDA database consists of cabinet-level data of 29 European democracies, starting from 1945 (Andersson, Bergman, & Ersson, 2014). The present study includes the number of government parties from each country from 2000 to 2012, while accounting for the differences in the length of time that each government was active. Subsequently, the aggregated data for each Member State were averaged to provide an average of the number of government parties.

Another condition from the management approach was the *administrative efficiency* as part of the similarly named ‘administrative efficiency’ argument. The *administrative efficiency* was measured by using the indicator of government effectiveness from the Worldwide Governance Indicators (WGI) research project. The WGI project has measured the government effectiveness of 212 countries on an annual basis since 2002. The included dimensions are based on a high number of variables across 35 sources and aggregated by using a components model (Kaufmann, Kraay, & Mastruzzi, 2011). For each of the cases for the present study, the data on the government effectiveness was retrieved from the years of 2008, 2009, 2010, 2011, and 2012. These values were then averaged to give an average indicator of the administrative effectiveness of each EU Member State between 2008 and 2012. These scores were based on a scale that ranges from -2.5 to +2.5, with higher scores indicating better government effectiveness.

3.3.4 Remote conditions: Four worlds of compliance typology

The indicators for the *ideal-typical transposition style* were based on the classifications made by (Falkner, Hartlapp, & Treib, 2007) and Falkner & Treib (Falkner & Treib, 2008). EU Member States were classified accordingly, as either belonging to the first world of '*law observance*,' the second world of '*domestic politics*,' the third world of '*dead letters*,' or the fourth world of '*transposition neglect*.'

3.3.5 Raw data table

The operationalization and data-gathering process resulted in a raw data, which provides an overview of the data and cases of the present study, as can be found in the Appendix (see Appendix, Table 10). As described in the case-selection (Chapter 3.2), cases were included from four categories of ATD measures per Member State. The first column of Table 10 lists the resulting cases. These are organized per category and sorted by their land-code (Eurostat, 2018). Presenting the gathered data with a raw data table is a required part of the csQCA process, as it provides a transparent overview of the data that was used for the analysis (Schneider & Wagemann, 2010). The table was created with the use of the software program fsQCA (version 3.0) (Ragin & Davey, 2016). The data was then transformed to dichotomous values for the crisp-set QCA, for which the process is described in the subsequent section.

3.4 Data Transformation

The following part of the methodology gives an account of the transformation of the data after the operationalization and data-gathering process of the previous section (see Chapter 3.3). As mentioned earlier (see Chapter 0), the data needed to be transformed into dichotomous values to account for the crisp-set nature of the QCA. Some of the indicators were more easily converted to a dichotomous scale, while others that consist of ordinal or interval scales needed calibration into separate dichotomous values for each of the categories (Ragin, 2010). Thresholds for denoting the membership scores for the condition to be present were then applied with the aid of the threshold setter function of another QCA software program TOSMANA software (version 1.54) (Cronqvist, 2017). Using the function from Tosmana was done as the fsQCA program used for the further analysis does not include such a feature. Such a threshold setter function was used to aid in the setting of thresholds, but as is required with conducting a QCA, the final decision for the threshold was based on either theoretical, empirical or case-level knowledge whenever possible. These choices have also been described further to be transparent towards the reader as to how the scores were determined (Schneider & Wagemann, 2010). The dichotomized table (see Table 3) provides the resulting overview of the transformation of all the conditions and the outcome. The following section describes the choices that were made for the transformation.

3.4.1 Data transformation: Outcome

The raw data on the outcome for each of the cases was transformed to a dichotomous value. The data was transformed to a dichotomous value of '1' when the raw data indicated the legal implementation of an ATD, but with no practical application. Cases with no legal implementation and no practical application, as well as, cases with both legal implementation and practical application, were both assigned a '0' for no membership to the outcome of non-application of an ATD measure in practice (Directorate-General for Migration and Home Affairs, 2013).

3.4.2 Data transformation: Proximate conditions from the enforcement approach

Three columns of Table 3 represent the conditions from the enforcement approach. Each of these conditions was assigned a code, starting with the first part of 'EA' as the designating for the inclusion in the enforcement approach. The *public attitude on EU membership* was assigned 'EA-PAEU.' The *ideological position of the government* was assigned 'EA-IO.' The *public attitude towards immigrants* was assigned 'EA-PAIM.'

The raw data on the influence of the *public attitude on EU membership* had a value expressed in percentages. Transforming these values to a dichotomous value based on a theoretical basis proved difficult, as the study found no theoretically defined categorization as to what kind of percentage indicates a cut-off point from existing literature. The threshold-setter function of the Tosmana software (version 1.54) (Cronqvist, 2017) found that the average percentage of the public attitude that indicates that EU membership is a bad thing for their respective country was 13.74%. The threshold-setter function provided a cut-off point of 16.31%. Following that cut-off point, the values for each of the cases were assigned a '0' for no membership that falls below 16.31%. Values that were higher than 16.31% were assigned a '1' for full membership.

The raw data on the *ideological position of the government* consisted of the average ideological position of the government between 2008 and 2012, classified according to the Schmidt-index (Schmidt, 1992). The transformation to dichotomous values included scoring a '1' for cases in which the average score on the composition of the EU Member State fell in the first or second category of the Schmidt-index. The composition of the first category consists of a percentage of 0%, indicating a hegemony of right-wing and center parties. The second category consists of the composition between 0% and 33,33%, indicating the dominance of right-wing and center parties. The average composition of EU Member States that showed higher averages for left-wing governments, and are therefore part of the other categories, were scored a '0' for no membership (Armingeon et al., 2017).

The raw data on the *public attitude on immigrants* also consisted of average values expressed in percentages. The present study incorporated a similar operationalization as the study of Valentova & Alieva (2010). That study transformed the individual responses as 'full membership' of a negative attitude if their responses were a score between 1 and 4 on the 10-point scale of the question in the EVS. Responses scoring between 5 and 10 were registered as 'no membership.' According to that study, such a form of dichotomization has been a common approach (Valentova & Alieva, 2010). The present study incorporated that operationalization but included the average percentage of responses between 1 and 4 as a first category and the average percentage of responses between 5 and 10 as a second category. Cases for which their respective county had a higher percentage for the first category compared with the second category were assigned a '1' for full membership. Vice versa, cases with a higher percentage for the second average value were assigned a '0' for no membership. For example, 61,50% of the responses

from Greek respondents were within the range of 1 to 4 on the scale in 2008. 38.30% of the responses were in the range of 5 to 10. Cases of ATD measures in Greece were, therefore, assigned a '1' for membership to indicate a negative public attitude towards immigrants.

3.4.3 Data transformation: Proximate conditions from the management approach

The *level of federalism*, the *number of parties in government*, and *administrative efficiency* were included in Table 3 as the three conditions from the management approach. These conditions were also assigned a code, similarly to the conditions from the enforcement approach. The first part 'MA' was assigned to all three as to designate these as being part of the management approach. The full code for the *level of federalism* was 'MA-FE.' For the *number of parties in government*, the full code was 'MA-PG.' Lastly, the full code for the *level of administrative efficiency* was 'MA-AD.'

The raw data on the *level of federalism* had a value ranging from 0 (indicating no federalism), 1 (indicating weak federalism), and 2 (indicating strong federalism) (Huber et al., 2004). The codes for 1 (weak federalism) and 2 (strong federalism) were transformed to the dichotomous value of '1' for the presence of a level of federalism. A '0' for no membership was assigned when there was a raw data value of 0 for no level of federalism.

As for the *number of parties in government*, the raw data on the average number of government parties between 2008 and 2012 were transformed to a '0' for no membership if the average number consisted of more than two parties (Andersson et al., 2014). Raw data that had an average higher than 1 were assigned a '1' for full membership, indicating that some form of a coalition consisting of more than two parties had occurred within the time frame.

The raw data on *administrative efficiency* consisted of values on a range from -2.5 to +2.5. As there was no apparent theoretical cut-off score found in the literature, a threshold was calculated and assigned with the use of the threshold-setter function of the TOSMANA software (version 1.54) (Cronqvist, 2017). The suggested cut-off when the function was set to assigning one threshold was 1,35. A closer examination of the values indicates that such a threshold was too moderate. Therefore, the suggested cut-off was therefore used for when the function was set to assigning two thresholds, resulting in a more accurate cut-off point. This indicated a cut-off point of 1.07, though this excluded some of the values at that specific edge. Instead, values of the raw data for this condition lower than 1.09 were assigned a '1' for full membership, indicating a relatively low administrative efficiency. Values higher than 1.09 were assigned a '0' for no membership, indicating a relatively higher administrative efficiency.

3.4.4 Data transformation: Remote conditions from the worlds of compliance typology

The final four included conditions were part of the four worlds of compliance typology by Falkner & Treib (2008). The first part of the assigned code for each of the conditions was 'FW.' There were four included conditions for the typology. Each was assigned a separate code: 'FW-LO' for the world of '*law observance*'; 'FW-DP' for the world of '*domestic politics*'; 'FW-DL' for the world of '*dead letters*'; 'FW-TN' for the world of '*transposition neglect*.' As the data for the worlds of compliance was based on the typology made by Falkner & Treib (2008), a '0' was assigned indicating no membership if the case was part of a Member State that is not considered part of that specific world. A '1' was assigned for the instances in which a case was part of a Member State that is included in that specific world. For example, the Netherlands is not part of the world of '*dead letters*' and had thus been assigned a score of '0' for no membership to that category. Subsequently, it was only assigned a '1' for full membership for the category of the world of '*domestic politics*.'

3.4.5 Data transformation: Scope-conditions

In addition to the outcome and the regular conditions, two scope-conditions were added to account for the contradictory rows and logical remainders (see also Chapter 3.5 Reliability & Validity).

'SC-EUM' was added, assigning a '1' for states that were part of the EU enlargement round of 2004 and assigned a '0' for states that were not part of that enlargement round.

'AD-CC' was added to indicate cases from a Member State for which the outcome differs from cases of that state of non-application as an outcome. Adding the negated version of that condition 'ad-cc' as a scope-condition enabled the exclusion of the within-country influences from the dataset, which was necessary as the scope of this study was limited to cross-group (the remote) and within-group (the proximate) conditions. A '1' was assigned for each case that had a conflicting outcome within a Member State to a case of non-application. A '0' was assigned to cases in which that did not appear.

Table 3 on the following page provides the resulting dichotomized table of all the included conditions and the outcome. The next part of the methodology describes the data analysis process of the study.

Table 3. Dichotomized table of the conditions & outcome

Case	<i>Enforcement approach</i>			<i>Management approach</i>			<i>Ideal-typical transposition style</i>				<i>Scope conditions</i>		Outcome
	EA-PAEU	EA-IO	EA-PAIM	MA-FE	MA-PG	MA-AD	FW-LO	FW-DP	FW-DL	FW-TN	SC-EUM	AD-CC	
AT-RS	1	0	1	1	0	0	0	1	0	0	0	0	0
BE-RS	0	1	1	1	1	0	0	1	0	0	0	0	0
DK-RS	0	1	0	1	0	0	1	0	0	0	0	1	0
EL-RS	1	0	1	0	0	1	0	0	0	1	0	0	1
FI-RS	1	1	0	0	1	0	1	0	0	0	0	0	0
LU-RS	0	0	0	0	1	0	0	0	0	1	0	0	1
SI-RS	0	0	0	0	1	1	0	0	1	0	1	1	0
SK-RS	0	0	1	0	1	1	0	0	1	0	1	0	1
AT-RRtA	1	0	1	1	0	0	0	1	0	0	0	0	0
BE-RRtA	0	1	1	1	1	0	0	1	0	0	0	0	0
CZ-RRtA	1	1	1	0	1	1	0	0	1	0	1	1	0
DE-RRtA	0	1	1	1	0	0	0	1	0	0	0	0	0
DK-RRtA	0	1	0	1	0	0	1	0	0	0	0	1	0
EL-RRtA	1	0	1	0	0	1	0	0	0	1	0	0	1
FR-RRtA	1	1	0	0	0	0	0	0	0	1	0	0	1
IE-RRtA	0	1	1	0	1	0	0	0	1	0	0	1	0
LU-RRtA	0	0	0	0	1	0	0	0	0	1	0	0	1
SI-RRtA	0	0	0	0	1	1	0	0	1	0	1	1	0
SK-RRtA	0	0	1	0	1	1	0	0	1	0	1	0	1
AT-SPD	1	0	1	1	0	0	0	1	0	0	0	0	0
BE-SPD	0	1	1	1	1	0	0	1	0	0	0	0	0
DE-SPD	0	1	1	1	0	0	0	1	0	0	0	0	0
DK-SPD	0	1	0	1	0	0	1	0	0	0	0	1	0
EL-SPD	1	0	1	0	0	1	0	0	0	1	0	0	1

FR-SPD	1	1	0	0	0	0	0	0	0	1	0	1	0
SI-SPD	0	0	0	0	1	1	0	0	1	0	1	0	1
SK-SPD	0	0	1	0	1	1	0	0	1	0	1	1	0
AT-FG	1	0	1	1	0	0	0	1	0	0	0	0	0
BE-FG	0	1	1	1	1	0	0	1	0	0	0	0	0
CZ-FG	1	1	1	0	1	1	0	0	1	0	1	0	1
DE-FG	0	1	1	1	0	0	0	1	0	0	0	0	0
DK-FG	0	1	0	1	0	1	1	0	0	0	0	0	1
EL-FG	1	0	1	0	0	1	0	0	0	1	0	0	1
ES-FG	0	0	0	1	0	1	0	1	0	0	0	0	0
HU-FG	1	0	1	0	0	1	0	0	1	0	1	0	0
IE-FG	0	1	1	0	1	0	0	0	1	0	0	0	1
NL-FG	0	1	0	0	1	0	0	1	0	0	0	0	0
SE-FG	1	1	0	0	1	0	1	0	0	0	0	0	0
SI-FG	0	0	0	0	1	1	0	0	1	0	1	0	1
SK-FG	0	0	1	0	1	1	0	0	1	0	1	0	1

Note. Adapted from (Directorate-General for Migration and Home Affairs, 2013) & (Eurostat, 2018) for Case, from European Commission (2009, 2010, 2013a, 2013b, 2015 for EA-PAEU, from Armingeon et al. (2017) for EA-IO, from (EVS, 2016) for EA-PAIM, from (Armingeon et al., 2017) for MA-FE, from Andersson et al. (2014) for MA-AD, from Kaufmann et al. (2011) for MA-AD, from Falkner & Treib (2008) for FW-LO; FW-DP; FW-DL; FW-TN, from the Directorate-General for Migration and Home Affairs (2013) for Outcome.

3.5 Data Analysis

The data analysis of the two-step csQCA of this study started with the provision of a descriptive overview of the outcome results for all the cases (see Figure 6). The Member State's inclusion sorted these into a world of compliance for each of the cases. The aim was to uncover an initial assessment of relevant patterns of the outcome-enabling influence of each of these worlds and the overall results on the application of ATD measures in practice in between 2008 and 2012.

The following aspect of the analysis was to look for conditions that were deemed necessary or singularly sufficient for the outcome. The fsQCA program has a function that calculates the set relational coverage and consistency of each of the conditions for the outcome. The definition of the consistency for a set relation is *"the degree to which instances of an outcome agree in displaying the causal condition thought to be necessary."* Set relational coverage is considered as *"assessing the "relevance" of the causal condition, or the degree to which instances of the causal condition are paired with instances of the outcome'* (Ragin, 2006, p. 2). A condition needs to meet a high threshold before it can be considered necessary. This study maintained a threshold of 0.90 for consistency to indicate a necessary causal condition for the outcome (Lilliefeldt, 2012).

The next aspect of the two-step csQCA was to use the truth table algorithm method of the fsQCA software, which incorporates the Quine-McCluskey algorithm (Ragin, 2010). That method was used multiple times for the two-step approach of this study, although the process was the same each time.

The truth table provided all the possible combinations of causal conditions, with 2^k rows based on the number of causal conditions (k). Membership was numbered with a '1' for full membership of the causal condition and a '0' for no membership. The truth table displayed the number of cases that showed the number of cases included for each of the combinations (Rihoux, 2008). For a csQCA, the raw consistency value represents the number of cases in that row that showed the outcome with the causal conditions that belong in that specific row (Ragin, 2010, p. 39).

Before analyzing the data, two assessments were done to prepare the truth table for analysis. The first assessment involved the classification of irrelevant and relevant combinations of conditions for the outcome, based on the frequency of cases in which the combination occurs. The second assessment was to determine which combinations were a subset of the outcome, by analyzing the raw consistency value. There is no pre-determined threshold for such a value, although a threshold lower than 0.75 is considered to indicate substantial inconsistency (Ragin, 2010; Schneider & Wagemann, 2010). What is vital in such an assessment is to be transparent about the threshold maintained for the analysis. A lower threshold is entirely possible but must be mentioned and determined by a reasoned choice (Schneider & Wagemann, 2010). Rows that were deemed a subset were assigned a '1', while those that did not meet the threshold were denoted with a '0'. Two possible ways of analyses were possible with the fsQCA program, a single analysis that had to be specified or a standard analysis consisting of three versions. The first "complex"

analysis gives the most complex solution, giving no simplifying assumptions for logical remainders. Logical remainders occur because of the limited diversity of working with small N analysis. The assessing of multiple causal conditions often results in the inclusion of more possible combinations of causal conditions in the cases. The three forms of analyses provide different ways to deal with these logical remainders, with increasing precision when the degree of complexity is higher. The "complex" analysis does not include any simplifying assumptions, thereby resulting in a complex but accurate solution (For a more in-depth explanation, see Ragin (2010, p. 42-45)). The second "parsimonious" analysis uses a computational simplification of the assumptions for logical remainders leading to the solutions that are considered the least complex. Another third option gives an "intermediate" analysis (Schneider & Wagemann, 2010). In which, conditions need to be manually selected that would contribute to the outcome when present, contributing to the outcome when absent, or contributing to the outcome when either present or absent. Based on a theoretical interpretation of the different versions in the current analysis, the forms that were most relevant for the study were included in the results, while the other versions can be found in the Appendix (

Appendix, Table 14, Table 15, Table 16, Table 17, Table 18, Table 19, Table 20, Table 21). The result of the truth table method was a solution formula that presented the set-theoretic relationship, which was denoted with the use of the Boolean language as explained previously (see Chapter 3.1.1). The truth table algorithm method was used multiple consecutive times for each of the steps of the two-step approach. That process is described further in the next section.

Another vital part of any QCA, a discussion of the results (Chapter 5), comes after the mechanical analysis (by using software). Such a discussion is meant to provide a closer examination of the mechanical results in relation to the cases (Schneider & Wagemann, 2010).

3.6 Data Procedure: Application of a two-step csQCA approach

The two-step approach of this study's csQCA demanded multiple separate applications of the truth table algorithm method as described in the previous section (Chapter 3.5). As can be seen by Figure 5, the first step of the two-step approach involved the use of the truth table algorithm of the fsQCA software (version 3.0) (Ragin, 2010; Ragin & Davey, 2016), and applying it on the remote conditions of the four different categories from the typology by Falkner & Treib (2008). The remote conditions were considered as having a mediating influence on the underlying proximate conditions. The remote conditions, in that case, would function as providing the context necessary for enabling the outcome. The four worlds would serve as having a different mediating influence on the included proximate conditions, in that different proximate conditions are of significance in each of the worlds as remote conditions. From the first step, remote conditions that scored high enough on the consistency and coverage scores, in addition to being part of the resulting solution-formula, were included for further analyses in the second step. That second step again used the truth table algorithm of the fsQCA software (version 3.0) (Ragin, 2010; Ragin & Davey, 2016). However, the second step was to analyze the different combinations of remote conditions and their proximate conditions separately, as can be seen in Figure 5. Thereby scrutinizing the consistency and coverage of the combinations, determining which combinations are relevant for the cases considered for this study. These two steps were to lead to a combined result, whereby a solution formula would show the combination(s) that explain the outcome under consideration of the study.

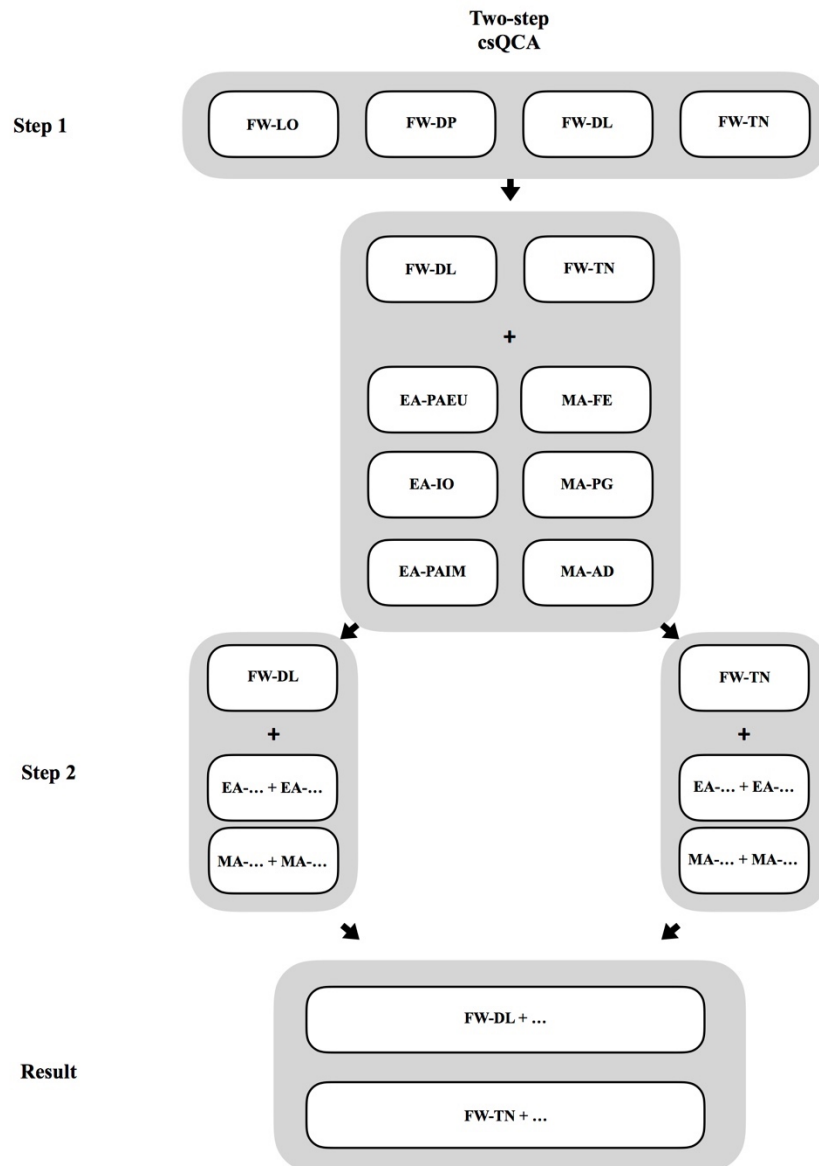


Figure 5. The two-step approach to the csQCA of this study

3.7 Reliability & Validity

The reliability and validity of a QCA have been somewhat of a contested area, receiving critique on its nature and features (Krook, 2010). As for its reliability, objections have been made to the dangers of limited generalization with the use of a QCA. The aim of quantitative statistical analysis is generally to enable predictive and generalizing reasoning, while QCA is more aimed at uncovering more in-depth knowledge about conditions and their dynamics within a specific group of cases (Krook, 2010). The present study accounted for these critiques by ensuring that the conditions were never analyzed purely in isolation, but rather as part of the causal pathway (Schneider & Wagemann, 2010). According to Rihoux (2008), interpretation of a solution formula expresses a potential causal connection. An assessment needs to be made on the basis of the results of the analysis in combination with theoretical knowledge to determine the level of causality that can be made in the interpretation. The discussion includes such a determination of the two-step csQCA and its results (see Chapter 5).

Another issue that was both relevant for the reliability and validity of the study is the regularly expressed concern for the dangers of data manipulation while conducting a QCA. The selection of the cases and the condition could interfere or influence the analysis, while the iterative nature of QCA means several choices need to be made to deal with any difficulties. Being transparent with the selection, decisions, and considerations on the trade-offs of the method and its strategies are potentially even more critical for a QCA than for statistical methods (Schneider & Wagemann, 2010). As the steps made in a QCA could appear a little abstract compared to the more typical forms of qualitative or quantitative research, the results were presented in such a way as to make the process towards obtaining these results more transparent and clear. Each step of the csQCA of this study included a brief explanation of the step as part of the two-step approach as described in the data procedure section (see Chapter 3.6). Additionally, some relevant choices have also been described in the following part of this section.

A first significant choice concerned itself with the selection of the cases used in the QCA. As already described in the scope & considerations chapter of the introduction (see Chapter 1.3) and the case-selection of the methodology (see Chapter 3.2), the analysis limits itself to ATD measures used within the EU. The results of the analysis only hold for the cases of ATD measures used by EU Member States, excluding to a large extent the generalization to other countries and their use of ATD measures. Being clear about the case-selection criteria is an integral part of being transparent in describing the QCA (Schneider & Wagemann, 2010). The two mentioned chapters were meant to contribute towards the transparency of the case-selection.

The selection of the conditions, outcome, and indicators was an important aspect as well. A trade-off was made between conducting a QCA with many conditions, producing complex results that could capture the “whole” issue under investigation (Schneider & Wagemann, 2010). Or conversely, a QCA that included fewer conditions that limit the generalizability, but enabled a better comparison (Sehring, Brockhaus, & Korhonen-Kurki, 2013). Introducing multiple conditions into a QCA can also increase the risk of including more logical remainders. Which leads to a QCA with limited diversity (Schneider & Wagemann, 2010).

There are several strategies to address these issues. The present study incorporated a two-step QCA to investigate a limited range of different proximate conditions. Determining the influence of these proximate conditions on the outcome under outcome-enabling, remote conditions. These remote conditions were hypothesized to function as mediating factors on the proximate conditions (Schneider & Wagemann, 2006; Sehring et al., 2013). Furthermore, the conditions and outcome were added and operationalized based on theoretical concepts and established databases. QCA is an iterative process, in which it is possible that conditions need to be changed, removed or included. For the present study, the condition AD-CC was added during the QCA to address some contradictory rows (where a configuration of causal conditions included cases with different outcome values) (Schneider & Wagemann, 2010).

Unfortunately, the inclusion of ad-cc as a negated scope condition did not completely solve the issue with the contradictory rows and logical remainders. And so, the case-selection was changed, and SC-EUM was added as another scope addition. Each of these strategies has its consequences. The exclusion of cases demands a substantial theoretical, case-related or empirical justification. While adding conditions leads to more complex results and increases the issue of possible logical remainders. However, leaving the inclusion of contradictory rows unaffected lowers the consistency value of the resulting solution formula (Schneider & Wagemann, 2010). As described in the results, the choice for the use of any of these strategies was based on considering the potential consequences for each of the relevant solutions formulas (see also Chapter 4.3 & Chapter 4.4).

Regarding these solution formula's, another way to address the problem of logical remainders is the choice for the different solution formulas that are presented in a fsQCA. As already mentioned in the methodology, three possible solutions are presented with the use of the fsQCA software (see Chapter 3.5). It depends on the analysis as to which solution is most applicable. However, the intermediate solution is most often preferred, as it ensures a balance between the trade-offs of the parsimonious solution (that includes all logical remainders, enabling broader generalization) and the complex solution (containing only the empirically observed cases, limiting the analysis only to these cases) (Lilliefeldt, 2012). These choices were also made for this study, which is mentioned in the results whenever relevant to increase the transparency.

There were also some choices made that were more related to the validity of this study. Which was related to the way the conditions and outcome were operationalized and how the necessary scores for the QCA were determined.

An important choice was made concerning the transformation of raw data from the conditions and outcome to a dichotomous value for a crisp-set QCA. As described earlier, a crisp-set QCA involves the conversion to dichotomous values of '1' for full membership and '0' for no membership. A fuzzy-set QCA includes more thresholds between '1' and '0'. As such, a fuzzy-set QCA enables a more accurate understanding of the condition and their values as these are often more complicated than can be captured by a dichotomous value (Sehring et al., 2013). Dichotomization is more subjective, demanding more of a theory based and empirically validated argument (Schneider & Wagemann, 2006). The choices for the thresholds from the transformation of the raw data for the conditions were, therefore, based as much as possible on theoretical and empirical knowledge as described in the methodology (see Chapter 3.3).

Lastly, the results list all the scores for the consistency and coverage measures from the csQCA. Thresholds indicated when a score was considered valid. As mentioned earlier in the methodology (see Chapter 3.1), consistency and coverage levels are research specific and to be determined on the basis of the number of cases, case & theoretical knowledge, and the quality of the data (Schneider & Wagemann, 2010). Relevant scores were described in the results (Chapter 4), including the thresholds maintained for the present study.

4. Results

The following chapter presents the results of the csQCA of the study and is divided into five sections. The first section includes a descriptive overview of the transposition of the ATD measures in the selected EU Member States (see Chapter 4.1). The second section addresses the necessity analysis of the conditions for the outcome (see Chapter 4.2). The subsequent sections discuss the two-step approach of the csQCA and its findings. The results from the first step of the approach are presented in Chapter 4.3. The results from the second step can be found in Chapter 4.4. Chapter 4.5 provides the results from the combination of the remote and proximate conditions. An overview of the methodological process and procedure can be found in the methodology of the study (see Chapter 3.5 & Chapter 3.6).

4.1 Descriptive results: Overview of the Transposition of ATD Measures

Figure 6 provides an overview of the level of legal implementation and practical application for the selected cases. Each case is designated by its country label and ATD measure and categorized by their inclusion in the typology by Falkner & Treib (2008). The categorization by the typology was made as if the assumptions from the theory were correct that the remote conditions from the typology function as outcome-enabling conditions, patterns of the spread of transposition would be likely to appear between these worlds of compliance.

Please note that this figure still includes the 5 cases with contradictory outcomes, while the further analysis was based on a further selection from 45 towards 40 cases, as is described in the methodology (see Chapter 3.2 Population & Case-selection).

As can be observed from Figure 6, a noticeable pattern appears. Both the cases from the world of ‘dead letters’ and ‘transposition neglect’ show almost all of the appearances of the outcome under consideration. Only the case of DK-FG shows the outcome for the first world. The observed pattern gave the first indication that the hypotheses for the typology appeared to have some merit. Although interestingly, the first two groups showed both the most ATD measures fully applied in practice as well as the least amount of any implementation or application of ATD measures. The first world of ‘law observance’ showed 3 cases of ATD measures fully applied in practice, while the second world of ‘domestic politics’ showed five cases. Also, the second world provided a relatively low level of the level of implementation and application of ATD measures, with 7 cases. However, a possible explanation was the fact that 4 of these cases consist of ATD measures within the same country (BE-RS, BE-RRtA, BE-SPD, BE-FG). Comparing the figure with the findings from the evaluation report as commissioned by the EU commission, Belgium had not applied any of the alternatives before the end of the rapport (Directorate-General for Migration and Home Affairs, 2013). Instead, Belgium provided special housing

and counseling for families as an ATD measure since 2008. That measure was deemed a best practice by NGO's, which made it unfortunate that the evaluation could not incorporate these forms (showing some of the limitations with these kinds of evaluation reports). It should also be noted that the figure does not include the cases for which information on the application is missing, which are found relatively more so in the second world of 'domestic politics.' The latter two groups showed a more balanced spread, with 14 cases of the outcome, 7 cases of no implementation or application, and 5 cases of full application.

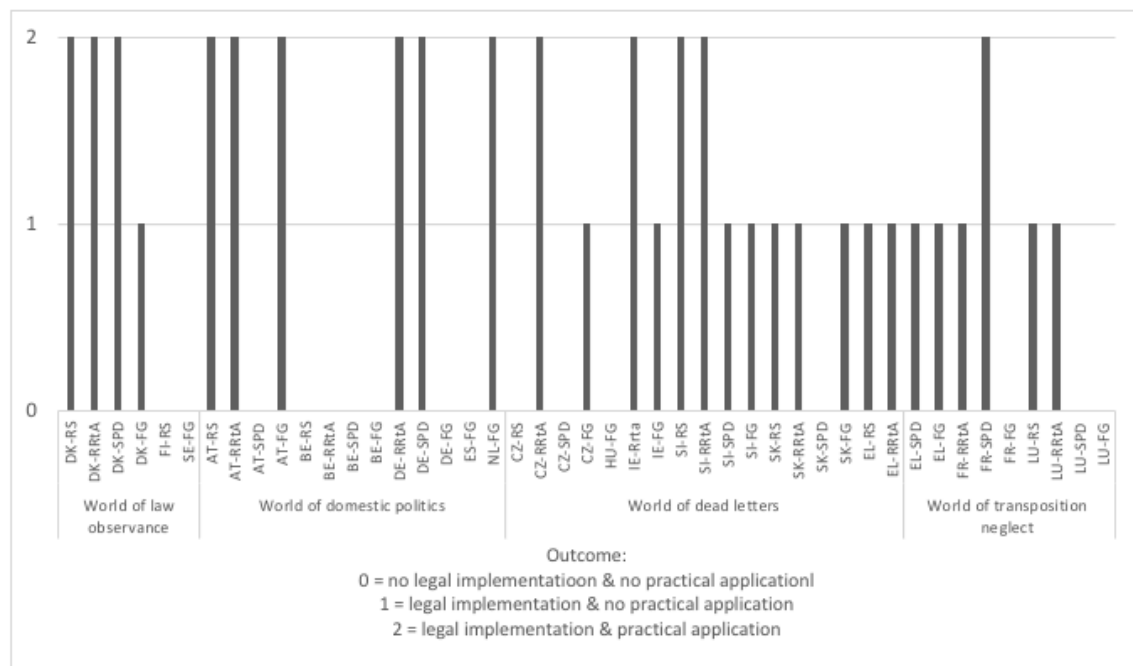


Figure 6. Descriptive overview of the legal implementation & practical application of ATD measures. Adapted from Directorate-General for Migration and Home Affairs (2013)

Altogether, these descriptive results showed some indication towards the expectation that the outcome of no practical application is more likely to appear within the worlds of 'dead letters' and 'transposition neglect.' Providing a first indication that these two worlds have a mediating influence on the proximate conditions, providing the contextual factors for the proximate factors to influence the outcome of non-application. In contrast, the figure does not provide a clear indication for the argument that there would be more cases of applied ATD measures in the first two worlds of 'law observance' and 'domestic politics.' Mostly because there were too many cases within these two groups that lacked information on the practical application.

4.2 Necessary Conditions Analysis

As described in the methodology, the following part of the two-step csQCA process concerned the examination of the conditions and their necessity for the outcome, based on their consistency and coverage scores (see also Chapter 3.5). Based on the literature of the theoretical chapter, the expectation would be that none of these is a necessary or condition for the outcome on its own. Due to the fact none of these conditions have been hypothesized as needed to be entirely present for the occurrence of non-application (see also Chapter 2.2). Table 4 provides the consistency and coverage scores for each of the conditions. The scope conditions were not included since these were not based on any theoretical expectations. Please note that if a condition is stated in uppercase letters, this means that the condition is present. Alternatively, a condition in lowercase letters indicates the negated version of the condition (the absence of a condition).

Table 4 shows that there were no conditions present which had a higher consistency than the threshold of 0.90. Present conditions with relatively high consistency and coverage were the conditions of EA-PAIM and EA-PAEU as part of the enforcement approach. The condition EA-IO showed a lower consistency, which provided a first indication that the expectation of the weak influence of the ideological position of a government on the non-compliance by non-application holds true. As for the conditions within the management approach, both the condition MA-PG and the condition MA-AD showed relatively high levels of consistency and coverage. Whereas the condition MA-FE showed a more surprising result for its consistency and coverage, as the results from Table 4 show. Closer inspection of the scores of that condition from the dichotomized table (see Table 3) revealed that for all the cases where the outcome of non-application is present, federalism was scored as absent. Based on the theory on the influence of federalism as part of the compliance field, such a relation was not to be expected. This did not automatically mean that the opposite was true, in that the absence of federalism was of influence. However, recent statistical research on the variables that influence the restrictiveness of immigration policy found that a higher level of federalism could lead to more liberal immigration policy, as was already mentioned in the theoretical framework (see Chapter 2.2.2 The management approach). It could, therefore, be possible that the absence of federalism causes a government to maintain or increase the restrictiveness of its immigration policy, which could explain the results obtained for this study. This explanation was interpreted with some caution as it did not match the expectation from the compliance field. Accordingly, MA-FE was not included in the further process of the csQCA. The negated version of ma-fe was instead selected to investigate the proposed relationship. Ma-fe was not considered a necessary condition, despite surpassing the threshold of 0.90. Such a cautious consideration was made given the unexpected reversal of the relationship. Especially, because of the argument that the findings of the study on the restrictiveness of immigration policy only indicated

the possibility of a reversed relationship of federalism regarding the EU compliance with ATD measures.

Two observations can be made from Table 4 regarding the results from the remote conditions of the typology. First of all, both the presence of the conditions for the FW-LO and FW-DP showed a low consistency for being a subset of the outcome. This was to be expected based on the theoretical framework and provides further evidence for the argument that the typology has a mediating influence on the proximate conditions as outcome-enabling conditions. The second observation that was made involved consistency levels for the remote conditions of FW-DL and FW-TN. As shown in Table 4, both conditions scored a relatively high consistency score. Especially when compared to the other two conditions for the typology, the two conditions together include a significant subset of the outcome. In addition, these also scored a high coverage, indicating that these conditions were present in most of the cases.

Table 4. Necessary conditions analysis table

<i>Condition</i>	<i>Consistency</i>	<i>Coverage</i>
EA-PAEU	0.400000	0.400000
ea-paeu	0.600000	0.360000
EA-IO	0.266667	0.200000
ea-io	0.733333	0.550000
EA-PAIM	0.600000	0.375000
ea-paim	0.400000	0.375000
MA-FE	0.066667	0.062500
ma-fe	0.933333	0.583333
MA-PG	0.600000	0.428571
ma-pg	0.400000	0.315789
MA-AD	0.733333	0.647059
ma-ad	0.266667	0.173913
FW-LO	0.066667	0.166667
fw-lo	0.933333	0.411765
FW-DP	0.000000	0.000000
fw-dp	1.000000	0.555556
FW-DL	0.466667	0.538462
fw-dl	0.533333	0.296296
FW-TN	0.466667	0.875000
fw-tn	0.533333	0.250000

Note. the conditions EA-PAEU (Public attitude EU), EA-IO (Ideological position government), EA-PAIM (Public attitude Immigrants, MA-FE (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), FW-LO (World of law observance), FW-DP (World of domestic politics), FW-DL (World of dead letters), FW-TN (World of transposition neglect). A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

In sum, Table 4 shows a combination of unexpected and expected results. A first expectation that no single condition is a necessary condition for the outcome on its own was confirmed. However, the QCA incorporated the unexpected and reversed relationship of federalism, instead of the present form of the condition. Consequently, that surprising finding was also why the negated conditions for federalism is not regarded as a necessary condition, as a precaution for any over-interpretation on its influence. Lastly, the table confirms the expectation that the first two remote conditions of the typology were of less importance than the third and fourth remote conditions. The results on these remote conditions provided an early indication that the latter two remote conditions have a mediating effect on the proximate conditions as outcome-enabling factors. The two-step csQCA was used to confirm the initial findings of the descriptive results and the necessary conditions analysis in the following parts of the chapter.

4.3 Step 1: Analysis of the Remote Conditions

The analysis of the remote conditions was the first step in the two-step csQCA approach as described in the methodology (see Chapter 0). The dichotomized data from the outcome and the in addition to the remote conditions of FW-LO, FW-DP, FW-DL, and FW-TN were included as input for this step. Initial analyses included several contradictory rows. The most significant issue was the appearance of several cases in which different cases of ATD measures in the Member States displayed different outcome results. As explained in the methodology (see Chapter 3.7), several strategies were applied to solve these issues. Some cases were removed, while the scope condition AD-CC was added for cases with contradictory results on the ATD measures within a Member State. The condition was added as it proved too complex to include conditions that explain the differences within a country within the scope of the present study. Though, this does lead to a model with higher scores for its consistency and coverage scores when compared to models without the scope condition. Possible distortions of these scores were taken into account in the further analyses and will also be addressed in the limitations of the study (see Chapter 6.2). Figure 7 shows an overview of the first step as part of the two-step approach (see also Figure 5).

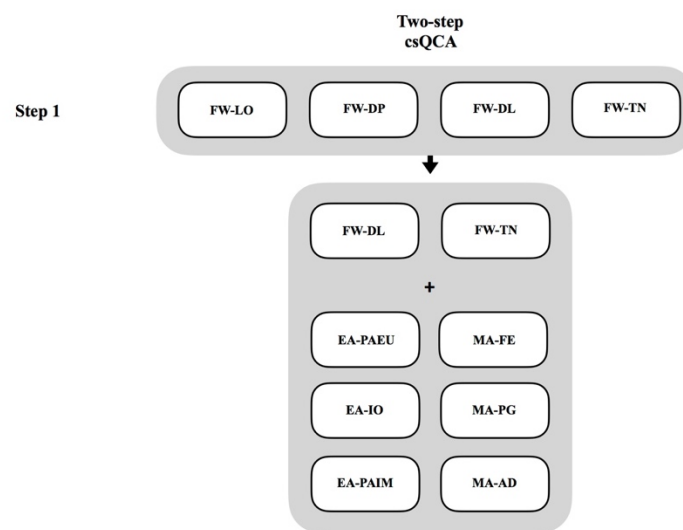


Figure 7. Overview of the first step of the csQCA

The input of the remote conditions and outcome in the fsQCA program resulted in a truth table (Table 5) that shows all the possible combinations of causal conditions for the outcome. The configuration of the outcome (indicated with a 0 for its absence and 1 for its presence) was done manually. The manual configuration involved the selection of relevant cases, based on their distribution over the rows of a causal pathway. The number of cases in each causal pathway is listed in the table by the column of *number*, providing the number and the combined total percentage of the cases that a specific cell and the cells above it represents of the total cases. The threshold for the frequency of cases that needed to

be present was set to 1 (see also Chapter 3.5 Data Analysis). Consequently, any combinations that fell below that threshold were removed. A further selection for the manual configuration was made based on the raw consistencies of the causal pathways. A threshold of 0.75 was maintained for the raw consistency (see also Chapter 3.5 Data Analysis). As can be observed from Table 5, this leads to a consistency cutoff of 0.875 with the inclusion of the scope condition ad-cc. The column of *cases* gives all of the included cases for each causal pathway.

Table 5. Truth table, result for first step of the csQCA

<i>FW-LO</i>	<i>FW-DP</i>	<i>FW-DL</i>	<i>FW-TN</i>	<i>AD-CC</i>	<i>Number</i>	<i>Outcome</i>	<i>Raw consistency</i>	<i>Cases</i>
0	1	0	0	0	13 (33%)	0	0	AT-RS, BE-RS, AT-RRtA, BE-RRtA, DE-RRtA, AT-SPD, BE-SPD, DE-SPD, AT-FG, BE-FG, DE-FG, ES-FG, NL-FG
0	0	1	0	0	8 (54%)	1	0.875	SK-RS, SK-RRtA, SI-SPD, CZ-FG, HU-FG, IE-FG, SI-FG, SK-FG
0	0	0	1	0	7 (71%)	1	1	EL-RS, LU-RS, EL-RRtA, FR-RRtA, LU-RRtA, EL-SPD, EL-FG
1	0	0	0	0	6 (87%)	0	0.166667	DK-RS, FI-RS, DK-RRtA, DK-SPD, DK-FG, SE-FG
0	0	1	0	1	4 (97%)	0	0	SI-RS, CZ-RRtA, IE-RRtA, SI-RRtA
0	0	0	1	1	1 (100%)	0	0	FR-SPD

Note. Including the remote conditions FW-LO (World of law observance), FW-DP (World of domestic politics), FW-DL (World of dead letters), FW-TN (World of transposition neglect), and the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4.

Different configurations for the solution formulas were compared, with and without the addition of the scope conditions for ad-cc. Eventually, this led to the most appropriate and relevant configuration, which is based on the table above. The three types of solutions were compared, and no significant differences in consistency or coverage scores were found between the three. The intermediate solution was incorporated in the results, as it provided a more precise overview of the solution formula. For example, the complex solution provided a convoluted solution formula, by including the negated (absent) versions of each of the conditions. As can be seen from the data in Table 6, the intermediate solution presents two causal pathways of remote conditions. Both the condition FW-DL and the condition FW-TN function as outcome-enabling conditions for the considered cases and the outcome. Given these results,

step two of the csQCA tested several models with different combinations of proximate conditions for each of the two remote conditions.

Table 6. Step 1: intermediate solution for the remote conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
ad-cc*FW-DL	0.466667	0.466667	0.875	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), HU-FG (1,0), IE-FG (1,1), SI-FG (1,1), SK-FG (1,1)
ad-cc*FW-TN	0.466667	0.466667	1	EL-RS (1,1), LU-RS (1,1), EL-RRtA (1,1), FR-RRtA (1,1), LU-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)
Solution coverage:	0.933333			
Solution consistency:			0.933333	

Note. Includes the remote conditions FW-LO (World of law observance), FW-DP (World of domestic politics), FW-DL (World of dead letters), FW-TN (World of transposition neglect), and the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent

4.4 Step 2: Analysis of the Remote & Proximate Conditions

The second step of the two-step csQCA used the same analytical process as with the first step (see also Chapter 3.5 Data Analysis). The difference was the inclusion of the conditions of FW-DL and FW-TN as the remote conditions as a result of the first step, with the addition of the proximate conditions of EA-PAEU, EA-IO, EA-PAIM, MA-FE, MA-GP, MA-AD. Several configurations were tested, split between a combination of the proximate conditions with either of the remote conditions (see Figure 8). The first step found that the two remote conditions of FW-DL and FW-TN display the outcome-enabling characteristics of a remote condition. The second step was meant to uncover for which combinations of the proximate conditions these two remote conditions served as a mediating factor on their influence on the outcome.

Similar issues with contradictory rows and logical remainders appeared as with the first step (see Chapter 4.3). And this is why the second step incorporated the same selection of cases as well as the inclusion of the scope conditions for ad-cc. Another scope conditions of SC-EU was added for the Member States for the combinations of proximate conditions with remote condition FW-DL, as an indicator for the Member States that became an EU member in the enlargement round of 2004 (see Chapter 3.4.5). The truth tables for both combinations of FW-DL and FW-TN with the proximate conditions can be found in the Appendix (see Appendix, Table 11 & Table 12). These tables present all the possible causal pathways for the separate analysis of the remote conditions of FW-DL and FW-TN and the proximate conditions. Only the proximate conditions that showed sufficient scores for each of the two combinations were included in these tables, as including the other conditions would have produced an overly convoluted and unnecessary analysis.

Similar to the first step, the manual configuration of the outcome maintained the threshold of a minimum of one case per causal pathway. Any combinations that fell below that threshold was removed. A further selection was made by keeping the threshold of 0.75 for the raw consistencies of the causal pathways. This led to a frequency cutoff of 1 and a consistency cutoff of 0.8 for the analysis with the remote condition FW-DL. As for the remote condition FW-TN, the 0.75 threshold resulted in both a frequency- and consistency cutoff of 1. The included cases of each of the causal pathway are listed in the *cases* column of the truth tables (see Appendix, Table 11 & Table 12). Figure 8 gives an overview of the second step. The following two parts of this section describe the results for each of the two remote conditions and their combinations of proximate conditions.

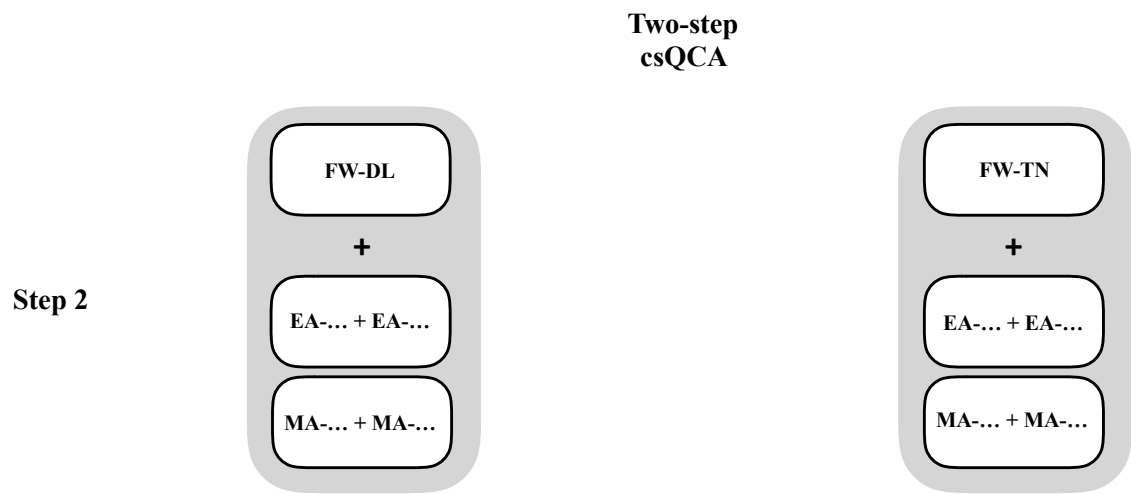


Figure 8. Overview of the second step of the csQCA

4.4.1 Solution formula: Remote condition FW-DL & proximate conditions

The first analysis of the second step determined which proximate conditions are part of the combination with the remote conditions for FW-DL as the outcome-enabling factor. The intermediate solution of the csQCA found two pathways containing the remote condition FW-DL and different combinations of the proximate conditions (see also Appendix, Table 16 & Table 17 for the other solution types). The intermediate solution had a coverage of 0.4666667 and a consistency of 1. As can be observed in Table 7, the first causal pathway contained most of the cases for this solution. The second combination was only represented by one case (IE-FG). That pathway also showed a noticeably different combination of proximate conditions. This might indicate that IE-FG is a “black swan” case, a possibility that is further analyzed in the discussion (see Chapter 5).

Two further observations can be made from in Table 7. Both causal pathways included the negated conditions for ma-fe. This observation provided further evidence that the conflicting result of the necessary condition analysis and the subsequent change of MA-FE to ma-fe was consistent for the current analysis. A second observation is that the first causal pathway mainly consisted of conditions from the management approach, with the inclusion of both the condition MA-PG and MA-AD. While the second pathway also includes MA-PG, conditions from the enforcement approach were also present in its causal pathway. These observations are further analyzed in the discussion (Chapter 5).

Table 7. Step 2: Intermediate solution for the remote condition FW-DL and the proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-DL*ma-fe*MA-PG*MA-AD* SC-EUM*ad-cc	0.4	0.4	1	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), SI-FG (1,1), SK-FG (1,1)
FW-DL*EA-PAIM*ma-fe* MA-PG*ma-ad*sc-eum*ad-cc	0.0666667	0.0666667	1	IE-FG (1,1)
Solution coverage:	0.466667			
Solution consistency:			1	

Note. Includes the remote condition FW-DL (World of dead letters) and proximate conditions: EA-PAIM (Public attitude Immigrants, ma-fe (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), the scope condition ad-cc (Alternative to Detention, conflicting case) and SC-EUM (Scope condition European membership in 2004). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

4.4.2 Solution formula: Remote condition FW-TN & proximate conditions

Table 8 displays the results from the second remote condition FW-TN and the proximate conditions. All three solutions were again compared based on their consistency, coverage and their ability to explain the pathways that lead to the outcome. The complex solution was chosen for this part of the analysis, as it gave a complete representation of the different causal pathways (see also Appendix, Table 18 & Table 19 for the other solution types). The complex solution has a coverage score of 0.4666667 and a consistency score of 1. As can be seen from the table, all three pathways show mixed results on the conditions of EA-PAEU, ea-paeu, EA-PAIM, ea-paim from the enforcement approach. EA-PAIM is only present in the third pathway of the table, although it is the most consistent pathway that covers most of the solution formula. The mixed findings can be explained by looking at the cases, which included cases of non-application of ATD measures from the same country. Though, EA-PAIM was also present for a single case in the first pathway. The second pathway was unusual, as it only included the present conditions for MA-PG (besides the scope conditions and the remote condition). The inclusion of the absent form of MA-FE for all these pathways is similar to the combination of the first remote condition FW-DL. However, the second pathway did not show any more differentiating conditions besides MA-PG. These findings suggested that the mediating influence of the remote conditions for FW-TN results in a more balanced influence of proximate conditions from both the enforcement and management approach than with the first remote conditions for FW-DL. Though, the conditions from the management approach appeared to be of more influence than those of the enforcement approach. As Table 8 shows, the negated version of ma-fe is present in all the pathways, while the conditions for MA-PG, and the conditions for MA-AD are both present in another but significant pathway of the complex solution. With the conditions for EA-PAEU being the only consistent influence from the enforcement approach.

Table 8. Step 2: Complex solution for the remote condition FW-TN and the proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-TN*EA-PAEU* ea-paim*ma-fe* ma-pg*ma-ad*ad-cc	0.0666667	0.0666667	1	FR-RRtA (1,1)
FW-TN*ea-paeu*ea-paim*ma-fe*MA- PG*ma-ad*ad-cc	0.133333	0.133333	1	LU-RS (1,1), LU-RRtA (1,1)
FW-TN* EA-PAEU *EA-PAIM*ma- fe*ma-pg*MA-AD*ad-cc	0.266667	0.266667	1	EL-RS (1,1) EL-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)
Solution coverage:	0.466667			
Solution consistency:			1	

Note. Including the remote condition FW-TN (World of transposition neglect), EA-PAEU (Public attitude EU), EA-PAIM (Public attitude Immigrants, ma-fe (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Together, it appeared that the two separate solution formulas from step two provide further evidence that each of the remote conditions lead to different combinations of proximate conditions being more important. The next section provides the csQCA with the solution formulas of both remote conditions and the proximate conditions integrated in one combined solution formula (see Chapter 4.5).

4.5 Combined Solution Formula

The following section provides the combined result of the two-step csQCA. These findings indicated that two solution formulas could explain the outcome of non-application of ATD in practice by EU Member States between 2008 and 2012. These solution formulas included one of the remote conditions and a different combination of proximate conditions, as a result of the mediating influences of FW-DL and FW-TN as outcome-enabling factors for these conditions. The purpose of the following part is to present the combined solution formula that merged these two into one complete solution.

The csQCA for the combined solution formula followed the same steps of analysis as the previous analyses. A truth table can be found in the Appendix (see Appendix, Table 13), which includes a frequency threshold of 1 for the number of cases in each causal pathway and a raw consistency threshold of 0.75. These thresholds led to a frequency cutoff of 1 and a consistency cutoff of 0.75. The included conditions were: FW-DL, FW-TN, EA-PAEU, EA-PAIM, MA-FE, MA-GP, and MA-AD. Additionally, the scope conditions of ad-cc and SC-EUM which addressed the issues of contradictory rows and logical remainders for the previous analyses were also included (see also Chapter 3.7). Figure 9 provides an overview of this part of the analysis as part of the two-step QCA as described in the methodology (see Chapter 3.6).

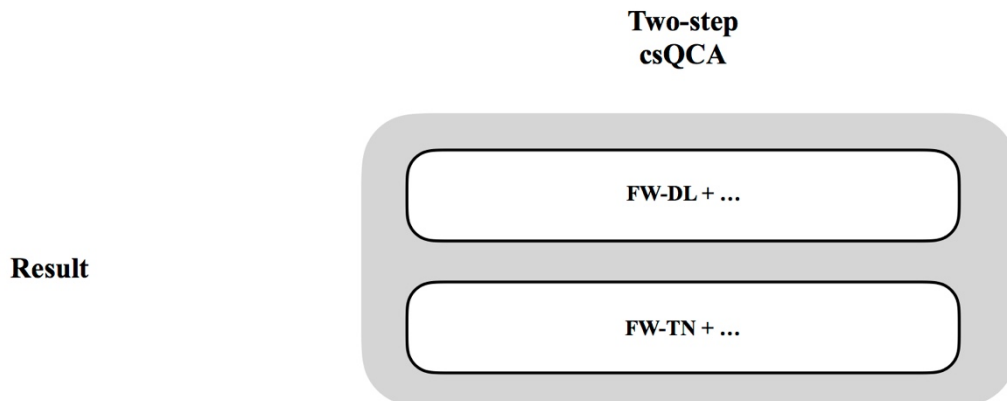


Figure 9. Overview of the combined result of the csQCA

Table 9 shows the result of the combined analysis of the remote and proximate conditions. The combined results were relatively similar to the solution formulas of the second step. One difference was the exclusion of ea-paeu and ea-paim for the second causal pathway. The exclusion was a consequence of using the intermediate solution type for the combined solution formula instead of the complex solution type as was used in for FW-TN in the second step. The intermediate solution was incorporated as the complex solution became unnecessarily complex, while the parsimonious solution was too simple of a representation. The intermediate solution was determined to be the best choice as a compromise between the other two, while also providing results more suitable for generalization (see also Chapter 3.5). A second difference was the exclusion of sc-eum in the fifth causal pathway. Sc-eum is the negated (absent) form of the scope condition that was used for the combination of the remote condition FW-DL and the proximate conditions. The absence of the scope condition did not affect the solution formula, as the ‘present’ form of the condition SC-EUM was still functional in the combined result.

Table 9. Intermediate solution for combined result of remote conditions and proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-TN*EA-PAEU* ea-paim*ma-fe* ma-pg*ma-ad*ad-cc	0.0666667	0.0666667	1	FR-RRtA (1,1)
FW-TN *ma-fe*MA-PG*ma-ad*ad-cc	0.133333	0.133333	1	LU-RS (1,1), LU-RRtA (1,1)
FW-TN* EA-PAEU *EA-PAIM* ma-fe*ma-pg*MA-AD*ad-cc	0.266667	0.266667	1	EL-RS (1,1) EL-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)
FW-DL*ma-fe*MA-PG*MA-AD* SC-EUM*ad-cc	0.4	0.4	1	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), SI-FG (1,1), SK-FG (1,1)
FW-DL*EA-PAIM*ma-fe* MA-PG*ma-ad*ad-cc	0.0666667	0.0666667	1	IE-FG (1,1)
Solution coverage:	0.933333			
Solution consistency:			1	

Note. Including the remote conditions FW-DL (World of dead letters) & FW-TN (World of transposition neglect). Including the proximate conditions: EA-PAEU (Public attitude EU), EA-PAIM (Public attitude Immigrants, ma-fe (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency). Including the scope condition ad-cc (Alternative to Detention, conflicting case) and SC-EUM (Scope condition European membership in 2004). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

The intermediate solution has a coverage score of 0.933333 and a consistency score of 1. As can be observed from Table 9, three causal pathways have the highest unique consistency and coverage scores. These were reformulated with the use of an expression based on Boolean algebra as described in the methodology chapter (see Chapter 3.1.1). In short, a ‘+’ indicates a logical OR and a ‘*’ indicate a logical

AND. A condition described in uppercase letters indicates its presence and a description in lowercase letters indicates its absence. A '→' from the condition to the outcome expresses a sufficient relation, while a '←' from the condition to the outcome expresses a necessary relation. The three most dominant causal pathways have been:

$$FW-TN*ma-fe*MA-PG*ma-ad*ad-cc \rightarrow Outcome$$

$$FW-TN*EA-PAEU*EA-PAIM*ma-fe*MA-AD*ad-cc \rightarrow Outcome$$

$$FW-DL*ma-fe*MA-PG*MA-AD*SC-EUM*ad-cc \rightarrow Outcome$$

The two casual pathways that were less dominant, showing only a single case for each, have been:

$$FW-TN*EA-PAEU*ea-paim*ma-fe*MA-PG*ma-ad*ad-cc \rightarrow Outcome$$

$$FW-DL*EA-PAIM*ma-fe*MA-PG*ma-ad*ad-cc \rightarrow Outcome$$

As one might have noticed, the scope conditions of ad-cc and SC-EUM were removed from these representations, as to present the combined solution formula. However, the removal of these scope-conditions from the solution formula does not exclude the effect these conditions had. The following solution formula, therefore, only represented the solution for causal pathways for the selection of cases of ATD measures in EU Member States and the differences for the outcome between the remote and proximate conditions. Additionally, the solution did not include the assessment of the differences in the outcome of the non-application of ATD measures within a specific country. The removal of SC-EUM had less of an effect on the interpretation of the solution formula, as it was used to differentiate between the cases from Member States within FW-DL that have recently become a member of the EU and the ones that do not (only Ireland in the current selection of cases). The condition was included as the analysis would otherwise remove the conditions for MA-AD and ma-ad from the solution formula. Which could be attributed to the case from Ireland potentially being a "black swan" case (outlier). The last section of the chapter (Chapter 5) discusses this further. The result of the combined solution formula was:

$$(FW-TN* EA-PAEU*ea-paim*ma-fe*MA-PG*ma-ad + FW-TN*ma-fe*MA-PG*ma-ad + FW-TN*EA-PAEU*EA-PAIM*ma-fe*MA-AD) + (FW-DL*ma-fe*MA-PG*MA-AD*SC + FW-DL*EA-PAIM*ma-fe*MA-PG*ma-ad) \rightarrow Outcome$$

The following chapter discusses the combined result of the csQCA more in-depth (see Chapter 5). It provides an interpretation of the complex combined solution, as well as a discussion in relation to the cases. The chapter also includes the assessment of the hypotheses that were formulated in the theoretical chapter and if these should be accepted or rejected (see also Table 1).

5. Discussion

The following chapter discusses the results from the two-step csQCA analysis. As described in the methodology, the discussion is meant to compare the mechanical analysis of the QCA in relation to the cases (see Chapter 3.5). The chapter begins with a review of the results from the remote conditions as outcome-enabling conditions with the first step of the csQCA (Chapter 5.1). After which the influence of the proximate conditions on the outcome in general are examined (Chapter 5.2). The last section discusses the two separate solution formulas of the remote conditions and their different combinations of proximate conditions (Chapter 5.3). Thereby examining the different mediating influences of the remote conditions, which lead to different outcome-enabling factors in which certain proximate conditions are of influence on the outcome.

5.1 Discussion of the Remote Conditions

The first step of the analysis aimed to determine which of the four categories from the Four Worlds of Compliance Typology by Falkner & Treib (2008) could be seen as remote conditions. Such a remote condition would provide, as a mediating influence, the contextual factors in which certain proximate conditions lead to the outcome under study. It appears that the condition for the world of *'dead letters'* as well as the condition for the world of *'transposition neglect'* can indeed be seen as these remote conditions. Both of these conditions contained 7 cases, meaning that these conditions combined were present in 14 of the 15 cases for which the outcome of non-application was present. With the other case being the ATD measures of a financial guarantee in Denmark, with Denmark belonging to the condition for the world of *'law observance.'* A first indication for the findings was already provided by the descriptive overview of the legal implementation & practical application of ATD measures in the selected EU Member States (see Chapter 4.1). Furthermore, the necessary & sufficient condition analysis gave additional evidence for the potential of the third and fourth category of worlds as remote conditions compared with the other two categories. The resulting causal pathways from the world of *'dead letters'* and *'transposition neglect'* showed high unique coverage scores as well as a high level of consistency. The coverage and consistency scores for the combined solution formula of these two remote conditions were also high. However, this was with the inclusion of the scope condition for conflicting outcomes within EU Member States, and these results should thus be interpreted with some caution, as explained in the methodology (see Chapter 3.7). Still, these findings suggest that the world of *'dead letters'* and the world of *'transposition neglect'* function as outcome-enabling conditions. Specifically, this indicates that these two remote conditions have a mediating influence on other proximate conditions, thus providing the contextual factors under which these proximate conditions could lead to the outcome of non-application.

An unexpected finding was the high number of cases from the world of '*law observance*' and '*domestic politics*' for which the information on the outcome of the practical application of ATD measures was missing. A possible explanation for the lack of information is given by the evaluation report by the Directorate-General for Migration and Home Affairs (2013). The report indicates that the lack of information could be attributed to the low level of awareness concerning ATD measures in these countries and a lack of attention towards the monitoring of legislative implementation and practical application thereof. For this reason, some uncertainty remains present as to how the results for these two conditions would be if more information were available.

Altogether, these findings show a strong indication of the mediating influence of the worlds of '*dead letters*' and '*transposition neglect*' on other proximate conditions that lead to the outcome. Therefore, the hypothesis (H7b) for the third and fourth category can be accepted, in which the world of '*dead letters*' or the world of '*transposition neglect*' could function as outcome-enabling, remote conditions for the outcome of non-application of ATD measures in EU Member States. The analysis does find some evidence that neither the world of '*law observance*' or the world of '*domestic politics*' operate as outcome-enabling conditions, though some uncertainty remains. As such, it can be concluded that the findings on the absence of such influences from the first and second category are less strong, which is why hypothesis (H7a) can only be partly confirmed.

5.2 Discussion of the Proximate Conditions

This section of the discussion examines the proximate conditions and their impact on the outcome in general, addressing their influence as part of the causal combinations with the remote conditions in the following section (see Chapter 5.3). The combined solution formula as presented in the last section of the results included the proximate conditions for *public attitude of EU membership*, *negative public attitude towards immigrants*, *absence of federalism*, *number of government parties*, and *administrative efficiency*. The discussion of these conditions is divided between those from the enforcement and management approach.

5.2.1 Discussion of the proximate conditions: Enforcement approach

None of the applied csQCA analyses found a causal pathway which includes the condition for the *ideological orientation of the government*. It thus appears that the outcome is not influenced by an ideological orientation of a government towards the right of the political spectrum. As such, the hypothesis (H3) that the orientation of a government towards an ideological ‘right’ position does not influence the non-application of an ATD measure is accepted.

Regarding the other two conditions that were included from the enforcement approach, both did appear in the causal pathways. However, a further interpretation of the results and the related cases provides a mixed assessment of the influence of these conditions. The condition for *negative public attitude of EU membership* gave the strongest indication for its impact, but only as a part of the causal pathways with the remote conditions for the world of ‘*transposition neglect*.’ The only included cases found for that pathway were the residence reporting requirements in France and all four cases of ATD measures in Greece. The results for the conditions for *negative public attitude towards immigrants* showed a similar trend. That condition included the same cases of ATD measures in Greece, in addition to a case of financial guarantee in Ireland from one of the causal pathways from the combination of the remote conditions for the world of ‘*dead letters*’ and the proximate conditions. These findings are consistent with the results from previous studies, for which the results related to these conditions also remained inconclusive. Nevertheless, it appears that these conditions play a role in the level of compliance with the legislative implementation and practical application of ATD measures. That role is especially apparent for the cases of ATD measures in Greece, as can be observed from the raw data table (Table 10), which shows that Greece has the highest score on the conditions for *negative public attitude towards immigrants* as well as having a high score on the condition for *negative public attitude of EU membership*. However, Greece also scores the lowest on the conditions for *administrative efficiency* out of all the selected EU Member States. Determining the exact level of influence of these conditions over the other was limited by the csQCA approach, thus making it impossible to distinguish for this study (see also Chapter 6.2 for further explanation on the limitations of this study). The membership to a low

administrative efficiency was also present across other causal pathways, which will be addressed in the following part of this section.

Given these assessments, the hypotheses that the willingness of a EU Member State is influenced by a negative public attitude on the membership of the EU Member State of the case (H1) or likewise by the negative public attitude towards immigrants (H2) are partly confirmed within the scope of the present study. Although, due to the limited generalizability of a QCA in general and specifically for these conditions, these findings may only be found in the presence of the other conditions and selected cases.

5.2.2 Discussion of the proximate conditions: Management approach

Having described the conditions for the enforcement approach, the conditions for the management approach are discussed next. As previously mentioned, *administrative efficiency* appears to be a relatively dominant condition within the combined solution formula. An early indication of this was already observed in the results from the necessary conditions analysis (Chapter 4.2), as indicated by the scores of consistency and coverage of the condition. Furthermore, the condition for *administrative efficiency* was also present in the two combinations (the third and fourth causal pathway) with the highest individual coverage and consistency from the second step of the csQCA. What is surprising is that each of these causal pathways was part of the different combinations of proximate conditions with the remote conditions. Providing even more evidence for the hypothesis that a low level of *administrative efficiency* leads to a lack of capacity for an EU Member State to apply an ATD measure in practice. The related hypothesis (H6) is therefore accepted within the scope of the current research.

One finding that was more unexpected, based on the theoretical framework, involved the results from the condition for *the level of federalism*. A higher level of federalism has been reported as having a strong relationship with non-compliance by previous studies, while the results from this study strongly indicate that this was not the case for the non-application in practice of ATD measures for the selected cases. Initial evidence for this was found during the analysis for necessary conditions, where the presence of *federalism* showed remarkably low scores for its consistency and coverage in explaining the outcome. Further inspection of the cases revealed that the present form of the condition was missing for every case with the outcome of non-application that belonged to the remote conditions of step two. Only in the case of the ATD measure of financial guarantee in Denmark was the condition present. Closer inspection of the raw data showed that the level of federalism was scored a 1, classifying it as a weak form of federalism (Huber, Ragin, Stephens, Brady, & Beckfield, 2004). Conversely, a previous study on the influences on the restrictiveness of immigration policy found that higher levels of *federalism* would lead to the implementation of less restrictive immigration policies. That finding was already mentioned in the theoretical framework and incorporated in the csQCA of this study, as the article by Haas & Natter (2015) gives a possible explanation for the reversed relationship of *federalism*. Hence,

the replacement of the ‘present’ form of the condition for the ‘negated’ (absent) version in the subsequent analyses. With the expectation that a high level of *federalism* would make the full application of a measure more likely, as the necessary negotiating between government institutions at higher and lower government level leads to more liberal immigration policy (Haas & Natter, 2015). The capacity to apply an ATD measure in practice would thus be limited, as the absence of a form of federalism makes it more likely that less negotiation takes place between the different levels of government. The findings of the present study show that the negated conditions for *federalism* is consistently present in each of the causal pathways, and thus it can safely be assumed that the original hypothesis (H4) for *federalism* needs to be rejected. *Federalism*, in its present form, has not been a present condition in any of the causal pathways in either of the combinations of remote conditions. Furthermore, the results from the included absent version of the condition indicate that in these cases it could be quite the opposite. That is to say; the findings of this study indicate a reversed relationship of *federalism* compared to previous studies, specifically in relation to immigration-related policies.

A final condition analyzed as part of the management approach has been the condition for the *number of government parties*. The csQCA results for the *number of government parties* corroborate with the findings of earlier studies. An early indication of which was already present in the necessary conditions analysis. More importantly, the condition was included in three of the causal pathways of the combined solution formula, across both combinations of remote conditions with the proximate conditions. Although Angelova, Dannwolf, & König (2012) considers it as a robust indicator for compliance in her comparative study, another comparative review by Toshkov (2010) states that previous findings are mixed. The present study finds further evidence for the influence of the number of government parties on non-compliance. Indicating that a government coalition of multiple parties could influence the capacity of a government to apply an ATD measure in practice. Thus, the related hypothesis (H5) for the condition is accepted within the scope of the present study.

5.3 Discussion of the Combination of Remote & Proximate Conditions

The previous sections of this chapter included separate discussions of the remote conditions and the proximate conditions. Two conditions from the compliance typology appear to function as remote conditions, having a mediating effect and thus providing the context in which certain proximate conditions could play a part in the outcome. The discussion from the second condition determined which of the conditions included in the study could be considered as proximate conditions. The following section discusses the possible differences in the combinations of proximate conditions under each of the outcome-enabling context of the two remote conditions.

The theoretical framework included the expectation that different proximate conditions are important under each of the different remote conditions for explaining the outcome. The results from the two-step csQCA provided a strong indication for that expectation. In that, the results show a distinctive combination of proximate conditions for both the world of '*dead letters*' and the world of '*transposition neglect*.'

The analysis indicates that the conditions from the management approach are more dominant in the world of '*dead letters*.' Especially, for the causal pathway that includes ATD measures of residence restrictions for Slovakia, regular reporting for Slovakia, surrendering of passports and documents for Slovenia, and financial guarantee for the Czech Republic, Slovenia, and Slovakia. These findings broadly reflect those by the study of Falkner & Treib (2008), who suggested that these recently joined EU Member States generally show high levels of legal implementation but lack in their practical application. The influence of domestic politics is the dominant factor at the start of the transposition phase. Falkner & Treib (2008) therefore expected that the level of legal implementation would be low, when in fact this appeared to be the opposite. The given explanation is that these countries were actively monitored by the EU commission surrounding their accession process. However, after legal implantation has occurred, several obstacles exist that can interfere with the application in practice. These obstacles consist of a lack of institutions and processes to apply these laws in practice (Falkner & Treib, 2008). The present study finds that the three conditions from the management approach are all present in the causal pathways for the Member States that were included in the world of '*dead letters*.' Though, the causal pathway that consisted of the case of an ATD measure of financial guarantee in Ireland included a less dominant presence of the conditions from the management approach, with the absence of the condition for *administrative efficiency*. The enforcement condition for *public attitude towards immigrants* was also included in the pathway. These findings differ from the other causal paths from the analysis of the remote condition for the world of '*dead letters*'. The somewhat contradictory result for the Irish case can be explained by the fact that Ireland is not bound to the 2008/115/EC Return

Directive (EUR-Lex, 2008). Similar to the UK and Denmark, Ireland has a specific arrangement within the EU regarding immigration policies. These arrangements enable these states to opt out of EU legislation (Walter-Franke, 2017). As a result of this, one would expect to see more cases of non-application of ATD measures for these Member States. In that, EU Member States that are bound to the Return Directive are more inclined to implement ATD measures in national legislation as to meet the goal of the Directive compared to these three countries. The investigation of these findings provides no substantial evidence for such a trend. The data on the legal implementation and the practical application shows that all the included categories of ATD measures are at least implemented in national law for the three countries (Directorate-General for Migration and Home Affairs, 2013). In Denmark, only the aforementioned case of an ATD measure of financial guarantee shows the outcome of non-application in practice. The other categories were all implemented and applied in practice. In regard to the cases from Ireland and the UK, not enough information was found for the practical application of most of the ATD measures. Consequently, the explanation that specific arrangements of these countries explain the appearance of the conflicting cases from Ireland and Denmark remains a suggestion. Due to the fact that there was not enough information for most of the cases from Ireland and all of the cases for the UK.

Continuing with the world of '*transposition neglect*,' the findings can only partly confirm the expectation from the theoretical framework (see Chapter 2.2.3). The causal pathways that included that remote condition showed a similar influence of the conditions from the management approach. The absence of the condition for *federalism* remained consistent for these paths as with the paths from the other remote condition. Although the condition for *number of government parties* was of less importance in these combinations, only being relevant for the ATD measures from the category of residence restrictions and regular reporting within Luxembourg. Likewise, the conditions for *low administrative efficiency* was only present for the cases within Greece. However, all the measures of ATD within Greece showed the presence of the outcome, giving more weight to the influence of *low administrative efficiency* in comparison to the conditions for *the number of government parties*. The findings indicate that the proximate conditions from the enforcement approach play a larger role in the world of '*transposition neglect*' when compared to the other world. However, the evidence for such an indication may be limited. Comparing these findings with the expectations based on the characteristics of the world of '*transposition neglect*' also shows some mixed results. According to the characteristics of the world, one would expect to see more issues at the start of the transposition phase. Which would lead to more instances of absent cases for ATD measures within the category, caused by 'administrative inertia by conflicting administrative interests and inefficient processes' (Falkner & Treib, 2008, p. 298). As a consequence of that inertia, countries from this category consider the practical application as being of less importance. The present study cannot find any evidence for the expected results for the first part of the process, as it lacks a condition that indicates the presence of an arduous process of legal implementation. Be that as it may, the results showed a relatively high distribution of cases of non-

application for states from this world. These results might indicate that this distribution is caused by the second stage of disinterest in the practical application of these states. Altogether, though the present study cannot demonstrate that these states saw a problematic process of legislative implementation, the findings of this study do show the presence of both conditions that demonstrate the willingness and capacity of the government for non-compliance for this world. It seems possible that their presence interfered with their willingness and capacity for the process of legislative implementation instead of the process of practical application. After which these states show less importance to the practical application, leading to a high level of non-compliance by non-application in practice.

Taken together, these findings confirm that the expected difference in the combination of proximate conditions for each of the incorporated remote conditions appears to be present under the scope of the present study. It appears that the conditions from the management approach are more important in the combination of proximate conditions within the world of '*dead letters*.' The present study finds a weaker link for the expectation that a balanced influence between the conditions from the enforcement- and management approach is important within the world of '*transposition neglect*.' The related hypothesis (H7c) can therefore only be partly accepted, as the expectation for the first world of '*dead letters*' is confirmed while the results provided inconclusive evidence for the second world of '*transposition neglect*.'

6. Conclusions

The present study set out to analyze the practical application of the ‘Alternatives to Detention’, supervisory measures used within the return process of the immigration policies of EU Member States. More specifically, it examined the transposition of the use of these measures under Article 15(1) of the 2008/115/EC Return Directive between 2008 and 2012. Moreover, the present study incorporated a two-step approach to a crisp-set Qualitative Comparative Analysis. The first step was aimed at the examination of the Four Worlds of Compliance Typology by Falkner & Treib (2008) as remote conditions, and if these categorized worlds of EU Member States function as outcome enabling factors for other conditions. The second step incorporated proximate conditions based on two approaches for explaining compliance with EU directives and legislation. The enforcement approach involves the willingness of a government to comply, while the management approach consists of the capacity of a government to comply (Tallberg, 2002). The results were based on a methodological two-step process of testing the remote conditions first, and subsequent testing of different combinations of proximate conditions with the remote conditions. This culminated in a combined solution formula of two of the remote conditions with each having a different configuration of proximate conditions. The following sections present some conclusive arguments for this study, including the main findings & limitations (Chapter 6.1 & Chapter 6.2), the implications of the study (Chapter 6.3), and recommendations for further research (Chapter 6.4). The final section presents some last thoughts on the study and its practical relevance for those involved in the immigration policy field (Chapter 6.5).

6.1 Main Findings

The findings of this study show that no remote or proximate condition is a necessary condition for non-application of ATD measures in practice on its own. Though, the level of federalism showed an unexpected result, resulting in a further inclusion of the reversed or absent version of the condition in the other analyses.

The worlds of ‘*dead letters*’ and ‘*transposition neglect*’ emerge as reliable remote conditions that function as outcome-enabling factors for the outcome of non-application. However, the results indicate that the worlds of ‘*Law observance*’ and ‘*domestic politics*’ are less likely to function as outcome-enabling factors.

As for the proximate conditions, the present study finds strong indications that the conditions of the presence of *multiple parties in a government*, and the presence of *low administrative efficiency* played a significant role in the non-application of ATD measures in general. Some evidence is found for the absence of *federalism* and the presence of a *negative public attitude towards immigrants*. Weak evidence is found for the influence of a *negative public attitude on EU membership*. The present study determined that the *ideological position of a government* did not play a sufficient role in the outcome.

Though these findings need to be seen in light of their inclusion in the causal pathways that were found in the second step.

The two-step csQCA of this study found a total of five causal pathways that were deemed sufficient for the non-application of ATD measures. Two of these combinations included the remote conditions for the world of '*dead letters*.' The first combination included the absence of a *level of federalism*, the presence of *multiple parties in a government*, and the presence of low *administrative efficiency* for cases in which the EU Member State was part of the enlargement round of the EU in 2004. The second combination included the presence of a negative *public attitude towards immigrants*, the absence of a *level of federalism*, the presence of *multiple parties in a government*, and the absence of low *administrative efficiency*. The second combination included cases for which the EU Member State was not a part of the enlargement round of the EU in 2004.

Three of these combinations included the remote conditions for the world of '*transposition neglect*.' The first combination consisted of the presence of a negative *public attitude on EU membership*, the absence of a negative *public attitude towards immigrants*, the absence of a *level of federalism*, the absence of *multiple parties in a government*, and the absence of low *administrative efficiency*. The second combination included the absence of a *level of federalism*, the presence of *multiple parties in a government*, and the absence of low *administrative efficiency*. The third combination included the presence of a negative *public attitude on EU membership*, the presence of a negative *public attitude towards immigrants*, the absence of a *level of federalism*, the absence of *multiple parties in a government*, and the presence of low *administrative efficiency*.

More importantly, the combined results provide an initial confirmation of the expectation that different combinations of proximate conditions from the enforcement and management approach are of importance for each of the remote conditions. For the world of '*dead letters*,' conditions from the management approach were dominant. A more balanced combination of conditions from both approaches was found to be important for the world of '*transposition neglect*.'

While these results are promising, the generalizability of these findings is first and foremost limited to the selected categories of ATD measures in the selected EU Member States within the period of 2008-2012. Additionally, the generalizability of the influences of the conditions should primarily be assessed based on their presence with the other presented conditions and with the operationalization used for the current research. The following section discusses these main limitations further.

6.2 Main Limitations

The availability of relevant data for the outcome and conditions showed several issues that should be mentioned. In general, finding accurate information on the application of EU directives proved more difficult compared to information on the legislative implementation. The data sources for legislative implementation are often more accessible and usually consisted of data from transposition records and infringement procedures (Mastenbroek, 2005). Reports, interviews, and secondary sources are often the sources used to provide information on the practical application, which are more vulnerable to interpretation. Such a source was also used for the present study on the practical application of ATD measures, as it used the evaluation report on the Return Directive (Directorate-General for Migration and Home Affairs, 2013). The assessment of the legal implementation in that report was based on country profiles and secondary sources, while the estimate on the practical application was based on interviews with national stakeholder interviews. Details of these assessment procedures remained unmentioned, providing no further detail on the criteria used for the evaluation. Some of the cases were determined by the report as lacking necessary information, while national stakeholders in Spain disagreed on the practical application of some of the alternatives (Directorate-General for Migration and Home Affairs, 2013). The cases that included these uncertain assessments were excluded from the analysis. Though, it is recognized that these kinds of assessments by an evaluation report leave something to be desired compared to thoroughly scientific sources of data. As there were no better options available, the data of the report was incorporated just as it was used by other literature (Bruycker et al., 2015; Walter-Franke, 2017). This also reaffirms that data based on rigorous scientific procedures is still a problem found for these kinds of research.

Similar, though less pronounced, issues also appeared with the data collection for the conditions. Several different data sources were considered, with some including missing data on some of the included countries, some for the years considered, and some included both types of missing data. In the end, the inclusion of data on all states was prioritized over the inclusion of data across all of the years. Such a choice was based on the consequences of missing data on some of the Member States for running a QCA. Most of the data on the conditions included as much yearly data points for the years between 2008-2012, as long as this was deemed necessary or relevant. The gathering of data from one source that could have affected the measurement of that condition was the use of the data from the European Value study for the condition for *public attitude towards immigrants* (EVS, 2016). The only possible year of measurement within the scope of the study was the data from 2008, as a newer round was yet to be completed (which started in 2017). The decision to use the data from the EVS was based on its inclusion of Luxembourg, which was not included in other possible sources for the period under review of this study.

Other limitations of the study were inherent to the characteristics of the use of a crisp-set QCA with a two-step approach. A first limitation was related to the two-step approach, while the second limitation involved the use of a crisp-set QCA. A final limitation concerned the general limitations of QCA when interpreting the results.

A two-step approach can only incorporate conditions from two hierarchical levels, which made it unable for this study to account for the difference between the outcomes within states. Several Member States included cases of ATD measures with different outcomes on the non-application in practice. These differences within the Member States could not be assessed within the scope of this study, as it included conditions for explaining the differences between groups of states and different countries. Consequently, these differences also resulted in several cases of contradictory rows. The present study addressed these contradictory rows by including a scope condition that acted as a filter for instances of conflicting outcomes within a Member State. Consequently, this led to higher consistency and coverage scores than would appear if these contradictory cases were included. The potential bias for putting too much importance on these scores had been taken into account in the further interpretation. Additionally, the distortions of these scores were frequently mentioned throughout the results when necessary.

A second limitation consisted of the used crisp-set approach of the QCA. A crisp-set QCA can only distinguish between the presence or absence of a condition in contrast to a fuzzy set QCA that includes the possibility of distinguishing between multiple degrees of absence or presence. As mentioned in the methodology (Chapter 3.1), using such an approach is difficult for the dichotomous operationalization of the outcome in this study. The outcome is operationalized based on its presence or absence, leaving the influence of varying levels of presence or absence of the outcome out of the present scope. Furthermore, the threshold for the determination of the presence or absence of csQCA is more vulnerable to interpretation. The present study aimed to address this by being sufficiently transparent for the choice of a threshold whenever necessary. While also incorporating theoretical, empirical and case-related knowledge for the determination of the threshold.

The last limitation involves a general limitation of interpreting the results from a QCA (see also Chapter 3.7 for further explanation of the reliability and validity of the study). In that, a QCA comprises of an assessment of conditions within a causal pathway. Interpretation of the individual influences of conditions should thus be made with due caution.

Notwithstanding these limitations, some indications for further implications regarding research on the compliance with EU directives and the application of ATD measures as part of the field of immigration policy can still be made. These are addressed in the following section.

6.3 Main Implications

Despite some of the previously mentioned limitations, this study offers further insights for the research on the compliance with EU directives, the application of ATD measures within the immigration policy field, and the use of the Qualitative Comparative Analysis methodology.

The present study has taken a different approach to the analysis of the application of ATD measures. Most of the literature on the alternatives to detention have recognized and discussed the inclusion of ATD measures as part of Article 15(1) of the 2008/115/EC Return Directive (EUR-Lex, 2008). However, this study considered the Return Directive as a vital aspect of the application of ATD measures within EU Member States.

The present study indicates that two of the four worlds of compliance function as outcome-enabling conditions. Even more so, these findings provide further evidence for the theory by the study of Falkner et al. (2007) that most compliance theories are 'only sometimes-true theories.' This study determined that the different proximate conditions vary in their importance for explaining the outcome, mediated by the influence of the outcome-enabling characteristics of the worlds of compliance. Prior to this study it was difficult to find any research that could confirm the typology. A study by Toshkov (2008) was unable to find significant evidence regarding its influence on transposition delays or non-transposition. Still, it could not reject the typology either. Even the original authors of the typology considered it as imperfect, stating that these categories were dependent on their interpretation of the findings (Falkner & Treib, 2008). The implication of the present study is that it provides further evidence of the explanatory power of the typology, by considering its merit in the compliance with the 2008/115/EC Return Directive as having a mediating influence on other factors.

Regarding the other approaches from the compliance field, this study should provide a further understanding of the predictive qualities of conditions from both the enforcement as well as the management approach for non-application of the aspects of an EU Directive. It indicates that the conditions from the management approach are more dominant in the practical application than in the process of legislative implementation. The observed reversal of the causal relationship of federalism with the non-application of ATD measures also signals that some condition might behave differently between policy fields.

In addition, the inability of the analysis to account for the substantial impact of within-country differences shows that the influence of EU directive-related compliance theories only partially explains the non-application in practice.

A final implication of the study concerns the merit of using a QCA. The method has several advantages and disadvantages. The previous section includes a discussion of some of the disadvantages of the

method (see Chapter 6.2). The insights gained from incorporating a two-step approach to the csQCA may be of assistance for further studies that are interested in applying such a method.

One implication of this study is that it is indeed so that a QCA can be used to examine hierarchical data on conditions from a multi-level structure multi-level. However, a two-step QCA already proves difficult to adequately examine, as the complexity of the models increases very quickly with the addition of remote conditions with their combinations of proximate conditions. It is thus likely that a QCA remains limited to the inclusion of conditions from two levels of hierarchical data.

The advantage of using a QCA is that it enables a comparative analysis of an intermediate number of cases in contrast to a purely qualitative case study. It also makes it possible to gain more in-depth knowledge of the included cases as compared to studies of large N analyses with statistical methods. Maybe even more so than with these methods, QCA stands and falls by the transparency on the choices made during its process. As it is more vulnerable to interpretation and data manipulation, it should be able for others to scrutinize the decision-making process. The present study attempted to provide a clear and extensive explanation of all of its choices. Hopefully, this also makes it possible for other relevant studies to adopt and further develop the QCA method.

6.4 Recommendations for Further Research

The present study is of an explorative nature, providing some first insights into the influence of several conditions from the compliance field for the non-application of ATD measures in practice. Considerably more work is needed to give further evidence for the indications provided by this study.

Several questions regarding the four worlds of compliance typology still exist. Due to the limitations of the csQCA, the findings cannot be generalized beyond the scope of this study. Further research is needed to determine if the typology does, in fact, explain non-application with EU directives. Considering the non-application of ATD measures, this study could not determine to the full extent the influence of the first two worlds of the compliance typology. Further research on the typology could be aimed at examining these two worlds. This would only be possible as long as additional information becomes available on the application level of ATD measures for these countries. In addition, the typology does not include the more recent additions to the EU of Bulgaria, Romania (both in 2007) and Croatia (in 2013). Further research for these countries should determine if these countries are part of one of the existing worlds or that these have enough distinctive characteristics as to consider the inclusion of another world of compliance.

Certain factors from both the enforcement and management approach were included as proximate conditions. The precise relationship of the *level of federalism* in relation to its possible reversed role with the non-application of ATD requires further research. Possible approaches could include further the analysis of the evaluation report by the Directorate-General for Migration and Home Affairs (2013), including the data from other evaluated measures or parts of the 2008/115/EC Return Directive. Another option is to consider sources based on the transposition of other immigration-related EU directives.

The included conditions from the enforcement approach were all based on the ‘preference fit’ argument, leaving out the role of the other ‘monitoring’ argument (Angelova, Dannwolf, & König, 2012). The ‘monitoring’ argument can either be incorporated by including a condition for the level of monitoring by the EU commission or the national agencies. The findings from the present study found less support for the influence of the included preference conditions compared to those from the management approach. Besides the ‘monitoring’ argument, other “less eligible” conditions may also be incorporated in further research. These other potential conditions could include the bargaining power, government EU position, and conflict-handling conditions (Toshkov, 2010).

The present study limited its scope to the presence of cases of non-application in practice. Further research considering the included conditions could determine if there are any differences related to the legislative implementation or the presence of practical application regarding cases of ATD measures. For example, one could hypothesize that if the indication of the reversed relationship on

federalism is correct, a present level of federalism would appear as a sufficient condition for the legislative implementation as well as practical application of ATD measures.

In addition, more extensive research is needed on the conditions accounting for within-country differences. The present study was unable to assess these differences within its scope. More information on the within-country differences would aid the further interpretation of the results from this study. Several studies and reports already argue that some measures are more widely used by states, though a more extensive examination of such conditions provides a more precise overview of the determinants for the implementation and application of ATD measures within EU Member States.

6.5 Final thoughts: Practical Usefulness of the Study

The previous sections of this chapter presented the concluding remarks, predominately regarding the scientific impact of the study. However, these findings could also have practical implications for those involved with immigration policies and the use of ATD measures. The relevance of the study described the societal importance of the present study. Further research on the implementation of ATD measures remains essential, as it is expected that the use of these measures provides better treatment of irregularly-staying migrants in the return process of (EU) countries. The immigration policy field has been rapidly developing in recent years, with the field being of important focus to the European Union. Directives, such as the 2008/115/EC Return Directive, were some of the attempts to provide a more unified approach for the EU Member States to the immigration process (Geddes & Scholten, 2016). As the study indicates, there are several influences and other obstacles which inhibit the implementation and application of ATD measures as part of EU directives. Together with previous and future research, that information could be used by relevant policymakers and other stake-holders for conducting retroactive analyses of previous immigration-related policies. For example, these could include (parts of) the 2013/32/EU Asylum Procedures Directive and the 2013/33/EU Reception Conditions Directive (EUR-Lex, 2013a, 2013b). In addition to a retroactive approach, such information can also be used proactively. Research on the obstacles and reasons for non-compliance could be used in a predictive manner regarding the development of new EU-level directives. Knowing which barriers are likely to inhibit the transposition for the different EU Member States is the first step in removing, circumventing or adapting to these obstacles by relevant stakeholders.

References

- Andersson, S., Bergman, T., & Ersson, S. (2014). The European Representative Democracy Data Archive (Version Release 3). Retrieved from www.erdda.se
- Angelova, M., Dannwolf, T., & König, T. (2012). How Robust Are Compliance Findings? A Research Synthesis. *Journal of European Public Policy*, 19(8), 1269–1291. <https://doi.org/10.1080/13501763.2012.705051>
- Armingeon, K., Wenger, V., Wiedemeier, F., Isler, C., Knöpfel, L., Weisstanner, D., & Engler, S. (2017). Comparative Political Data Set 1960-2014. Bern: Institute of Political Science, University of Berne. Retrieved from <http://www.cps-data.org/index.php/data>
- Basilien-Gainche, M.-L. (2015). Immigration Detention under the Return Directive: The Cjeu Shadowed Lights. *European Journal of Migration and Law*, 17(1), 104–126. <https://doi.org/10.1163/15718166-12342073>
- Bergman, T. (2000). The European Union as the next step of delegation and accountability. *European Journal of Political Research*, 37(3), 415–429. <https://doi.org/10.1023/A:1007085408779>
- Bloomfield, A. (2016). Alternatives to Detention at a Crossroads: Humanisation or Criminalisation? *Refugee Survey Quarterly*, 35(1), 29–46. <https://doi.org/10.1093/rsq/hdv018>
- Broeders, D., & Leerkes, A. (2010). A Case of Mixed Motives?: Formal and Informal Functions of Administrative Immigration Detention. *British Journal of Criminology*, 50(5), 830–850. <https://doi.org/10.1093/bjc/azq035>
- Bruycker, P. de, Bloomfield, A., Tsourdi, E., & Petin, J. (2015). *Alternatives to Immigration and Asylum Detention in the EU - Time for Implementation*. European Union: European Commission. Retrieved from <http://odysseus-network.eu/wp-content/uploads/2015/02/FINAL-REPORT-Alternatives-to-detention-in-the-EU.pdf>
- Cain, B. E., Dalton, R. J., & Scarrow, S. E. (2003). *Democracy Transformed?* Oxford, U.K: Oxford University Press. <https://doi.org/10.1093/0199264996.001.0001>
- Chmelickova, N. (2006). *Survey on Alternatives to Detention of Asylum Seekers in EU Member States*. (The regional coalition 2006). Retrieved from <https://idcoalition.org/wp-content/uploads/2008/12/survey-on-alternatives-to-detention-of-asylum-seekers-in-eu-member-states.pdf>
- Corcaci, A. (2017). Conceptual considerations on compliance in the European Union. In R. Lhotta, O. Lembcke, & V. Frick, *Politik und Recht: Umrisse eines politikwissenschaftlichen Forschungsfeldes* (pp. 117–133). Baden-Baden: Nomos Verlagsgesellschaft.
- Costello, & Kaytaz. (2013). *Building Empirical Research into Alternatives to Detention: Perceptions of Asylum-Seekers and Refugees in Toronto and Geneva*. Geneva: UN High Commissioner for Refugees (UNHCR). Retrieved from <http://www.refworld.org/docid/51a6fec84.html>

- Cronqvist, L. (2017). Tosmana (Version 1.54) [QCA software]. University of Trier. Retrieved from <http://www.tosmana.net>
- Directorate-General for Migration and Home Affairs. (2013). *Evaluation on the application of the Return Directive (2008/115/EC)* (Final report No. DR-02-13-466-EN-N). Luxembourg: Publications Office of the European Union.
- Edwards, A. (2011). Measures of First Resort: Alternatives to Immigration Detention in Comparative Perspective. *The Equal Rights Review*, 7, 117–142.
- EU Commission. (2016). *Asylum procedures*. Retrieved March 26, 2018, from https://ec.europa.eu/home-affairs/what-we-do/policies/asylum/common-procedures_en
- EU Commission. (2016b). *Irregular migrant*. Retrieved January 24, 2018, from https://ec.europa.eu/home-affairs/content/irregular-migrant-0_en
- EU Commission. (2016c). *Third-country national found to be illegally present*. Retrieved January 24, 2018, from https://ec.europa.eu/home-affairs/content/third-country-national-found-be-illegally-present_en
- EU Commission. (2016d). *Who qualifies for international protection*. Retrieved March 26, 2018, from https://ec.europa.eu/home-affairs/what-we-do/policies/asylum/refugee-status_en
- EU Publications Office. (2015). *Summaries of EU legislation: European Union directives*. Retrieved January 25, 2018, from <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:114527>
- EUR-Lex. (2008). *Directive 2008/115/EC of the European Parliament and of the Council of 16 December 2008 on common standards and procedures in Member States for returning illegally staying third-country nationals*. Retrieved February 26, 2017, from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:348:0098:0107:en:PDF>
- EUR-Lex. (2013). *Directive 2013/32/EU of the European Parliament and of the Council of 26 June 2013 on common procedures for granting and withdrawing international protection*. Retrieved April 4, 2018, from <http://data.europa.eu/eli/dir/2013/32/oj/eng>
- EUR-Lex. (2013). *Directive 2013/33/EU of the European Parliament and of the Council of 26 June 2013 laying down standards for the reception of applicants for international protection*. Retrieved April 4, 2018, from <http://data.europa.eu/eli/dir/2013/33/oj/eng>
- European Commission. (2009). Eurobarometer 69.2: National and European Identity, European Elections, European Values, and Climate Change, March-May 2008. (Version 3). Inter-university Consortium for Political and Social Research. <https://doi.org/10.3886/ICPSR25021.v3>

- European Commission. (2010). Eurobarometer 71.3: Globalization, Personal Values and Priorities, European Identity, Future of the European Union, Social Problems and Welfare, and European Elections, June-July 2009. (Version 1). Inter-university Consortium for Political and Social Research. <https://doi.org/10.3886/ICPSR28184.v1>
- European Commission. (2013a). Eurobarometer 73.4: Financial and Economic Crisis, the Future of the European Union, Globalization, and European Citizenship, May 2010. (Version 1). Inter-university Consortium for Political and Social Research. <https://doi.org/10.3886/ICPSR34384.v1>
- European Commission. (2013b). Eurobarometer 75.3: Financial and Economic Crisis, European Union Budget, and the Common Agricultural Policy. (Version 1). Inter-university Consortium for Political and Social Research. <https://doi.org/10.4232/1.11768>
- European Commission. (2015). Eurobarometer 77.4: European Parliament, Development Aid, Social Climate and Discrimination. (Version 1). Inter-university Consortium for Political and Social Research. <https://doi.org/10.4232/1.12049>
- European Migration Network. (2016). *Guidelines for monitoring and evaluation of AVR(R) programmes*. Directorate General Migration and Home Affairs, European Commission. Retrieved November 8, 2017, from http://emn.ie/files/p_20160129124148guidelines_for_monitoring_and_evaluation_final_jan2016.pdf
- Eurostat. (2018). *Glossary: Country codes - Statistics Explained*. Retrieved March 26, 2018, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Country_codes
- EVS. (2016). European Values Study 2008: Integrated Dataset [EVS 2008]. (Version 4.0.0). Cologne: GESIS Data Archive. <https://doi.org/10.4232/1.12458>
- Falkner, G., Hartlapp, M., & Treib, O. (2007). Worlds of compliance: Why leading approaches to European Union implementation are only "sometimes-true theories" *European Journal of Political Research*, 46(3), 395–416. <https://doi.org/10.1111/j.1475-6765.2007.00703.x>
- Falkner, G., & Treib, O. (2008). Three Worlds of Compliance or Four? The EU-15 Compared to New Member States. *JCMS: Journal of Common Market Studies*, 46(2), 293–313. <https://doi.org/10.1111/j.1468-5965.2007.00777.x>
- Field, O., & Edwards, A. (2006). *Alternatives to Detention of Asylum Seekers and Refugees*. Geneva: UN High Commissioner for Refugees (UNHCR). Retrieved from <http://www.refworld.org/docid/4472e8b84.html>
- Fiske, L. (2016). *Human rights, refugee protest and immigration detention*. London: Palgrave Macmillan.
- Geddes, A., & Scholten, P. (2016). *The Politics of Migration and Immigration in Europe*. London: SAGE Publications Ltd. <https://doi.org/10.4135/9781473982703>

- Haas, H. de, & Natter, K. (2015). The determinants of migration policies. Does the political orientation of governments matter?, *IMI Working Paper Series*, 117. Retrieved from <https://www.imi.ox.ac.uk/publications/the-determinants-of-migration-policies-does-the-political-orientation-of-governments-matter>
- Helbling, M., & Michalowski, I. (2017). A New Agenda for Immigration and Citizenship Policy Research. *Comparative Political Studies*, 50(1), 3–13. <https://doi.org/10.1177/0010414016666864>
- Huber, E., Ragin, C., Stephens, J. D., Brady, D., & Beckfield, J. (2004). Comparative Welfare States Data Set. Retrieved from <http://www.lisproject.org/publications/welfaredata/welfareaccess.htm>
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law*, 3(02), 220–246. <https://doi.org/10.1017/S1876404511200046>
- Kessler, A. E., & Freeman, G. P. (2005). Public Opinion in the EU on Immigration from Outside the Community*. *JCMS: Journal of Common Market Studies*, 43(4), 825–850. <https://doi.org/10.1111/j.1468-5965.2005.00598.x>
- Koopmans, R., Michalowski, I., & Waibel, S. (2012). Citizenship Rights for Immigrants: National Political Processes and Cross-National Convergence in Western Europe, 1980–2008. *American Journal of Sociology*, 117(4), 1202–1245. <https://doi.org/10.1086/662707>
- Krook, M. L. (2010). Women’s Representation in Parliament: A Qualitative Comparative Analysis. *Political Studies*, 58(5), 886–908. <https://doi.org/10.1111/j.1467-9248.2010.00833.x>
- Lahav, G. (2004). Public Opinion Toward Immigration in the European Union: Does It Matter? *Comparative Political Studies*, 37(10), 1151–1183. <https://doi.org/10.1177/0010414004269826>
- Lampinen, R., & Uusikylä, P. (1998). Implementation Deficit? Why Member States do not Comply with EU directives? *Scandinavian Political Studies*, 21(3), 231–251. <https://doi.org/10.1111/j.1467-9477.1998.tb00014.x>
- Lilliefeldt, E. (2012). Party and gender in Western Europe revisited: A fuzzy-set qualitative comparative analysis of gender-balanced parliamentary parties. *Party Politics*, 18(2), 193–214. <https://doi.org/10.1177/1354068810380094>
- Mastenbroek, E. (2005). EU compliance: Still a ‘black hole’? *Journal of European Public Policy*, 12(6), 1103–1120. <https://doi.org/10.1080/13501760500270869>
- Mastenbroek, E., & Kaeding, M. (2006). Europeanization Beyond the Goodness of Fit: Domestic Politics in the Forefront. *Comparative European Politics*, 4(4), 331–354. <https://doi.org/10.1057/palgrave.cep.6110078>

- Meyers, E. (2000). Theories of International Immigration Policy-A Comparative Analysis. *International Migration Review*, 34(4), 1245. <https://doi.org/10.2307/2675981>
- Norwegian Centre for Research Data. (2016). European Social Survey Cumulative File, ESS 4 & 6 (Version Data file edition 1.0). Norway: Data Archive and distributor of ESS data for ESS ERIC. Retrieved from <http://www.europeansocialsurvey.org/downloadwizard/>
- Ragin, C. C. (2006). Set Relations in Social Research: Evaluating Their Consistency and Coverage. *Political Analysis*, 14(03), 291–310. <https://doi.org/10.1093/pan/mpj019>
- Ragin, C. C. (2010). User's Guide to Fuzzy-Set/Qualitative Comparative Analysis 2.0. Irvine, California: Department of Sociology, University of California. Retrieved from <http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml>
- Ragin, C. C., & Davey, S. (2016). Fuzzy-Set/Qualitative Comparative Analysis (Version 3.0). Irvine, California: Department of Sociology, University of California. Retrieved from <http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml>
- Raunio, T. (2005). Holding governments accountable in European affairs: Explaining cross-national variation*. *The Journal of Legislative Studies*, 11(3–4), 319–342. <https://doi.org/10.1080/13572330500348307>
- Rihoux, B. (2008). Case-Oriented Configurational Research: Qualitative Comparative Analysis (Qca), Fuzzy Sets, and Related Techniques. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Retrieved from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-31>
- Rihoux, B., & Ragin, C. C. (Eds.). (2009). *Configurational comparative methods: qualitative comparative analysis (QCA) and related techniques*. Thousand Oaks: Sage.
- Sampson, R., & Mitchell, G. (2013). Global Trends in Immigration Detention and Alternatives to Detention: Practical, Political and Symbolic Rationales. *Journal on Migration and Human Security*, 1(3), 97–121. <https://doi.org/10.14240/jmhs.v1i3.14>
- Schmidt, M. G. (1992). Regierungen: Parteipolitische Zusammensetzung. In M. G. Schmidt (Ed.), *Lexikon der Politik, Band 3: Die westlichen Länder* (pp. 393–400). Munchen: C.H. Beck.
- Schneider, C. Q., & Wagemann, C. (2006). Reducing complexity in Qualitative Comparative Analysis (QCA): Remote and proximate factors and the consolidation of democracy. *European Journal of Political Research*, 45(5), 751–786. <https://doi.org/10.1111/j.1475-6765.2006.00635.x>
- Schneider, C. Q., & Wagemann, C. (2010). Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets. *Comparative Sociology*, 9(3), 397–418. <https://doi.org/10.1163/156913210X12493538729793>

- Sehring, J., Brockhaus, M., & Korhonen-Kurki, K. (2013). *Qualitative Comparative Analysis (QCA): An application to compare national REDD+ policy processes*. Center for International Forestry Research (CIFOR). <https://doi.org/10.17528/cifor/004278>
- Silverman, S. J., & Massa, E. (2012). Why Immigration Detention is Unique. *Why Immigration Detention Is Unique*, 18(6), 677–686. <https://doi.org/10.1002/psp.1720>
- Sverdrup, U. (2004). Compliance and Conflict Management in the European Union: Nordic Exceptionalism. *Scandinavian Political Studies*, 27(1), 23–43. <https://doi.org/10.1111/j.1467-9477.2004.00098.x>
- Tallberg, J. (2002). Paths to Compliance: Enforcement, Management, and the European Union. *International Organization*, 56(3), 609–643. <https://doi.org/10.1162/002081802760199908>
- Thomson, R., Torenvlied, R., & Arregui, J. (2007). The Paradox of Compliance: Infringements and Delays in Transposing European Union Directives. *British Journal of Political Science*, 37(04). <https://doi.org/10.1017/S0007123407000373>
- Toshkov, D. (2007). In search of the worlds of compliance: culture and transposition performance in the European Union: In search of the worlds of compliance: a brief reply, by Gerda Falkner, Miriam Hartlapp, Simone Leiber and Oliver Treib In search of the worlds of compliance? A response, by Dimitar Toshkov. *Journal of European Public Policy*, 14(6), 933–959. <https://doi.org/10.1080/13501760701497956>
- Toshkov, D. (2008). Embracing European Law: Compliance with EU Directives in Central and Eastern Europe. *European Union Politics*, 9(3), 379–402. <https://doi.org/10.1177/1465116508093490>
- Toshkov, D. (2010). *Taking stock: a review of quantitative studies of transposition and implementation of EU law*. Institute for European integration research (EIF). Retrieved from <https://EconPapers.repec.org/RePEc:erp:eifxxx:p0009>
- Valentova, M., & Alieva, A. (2010). Immigration as a Threat: The Effect of Gender Differences Among Luxembourg Residents with and without a Migration History. *CEPS/INSTEAD, Working Paper Series*.
- Vis, B. (2009). Governments and unpopular social policy reform: Biting the bullet or steering clear? *European Journal of Political Research*, 48(1), 31–57. <https://doi.org/10.1111/j.1475-6765.2008.00783.x>
- Vis, B. (2011). Under which conditions does spending on active labor market policies increase? An fsQCA of 53 governments between 1985 and 2003. *European Political Science Review*, 3(2), 229–252. <https://doi.org/10.1017/S1755773910000378>
- Walter-Franke, M. (2017). *Asylum Detention in Europe: State of play and ways forward*. Berlin: Jacques Delors Institut. <https://doi.org/10.13140/RG.2.2.27985.63844>

Welch, M., & Schuster, L. (2005). Detention of asylum seekers in the US, UK, France, Germany, and Italy: A critical view of the globalizing culture of control. *Criminology and Criminal Justice*, 5(4), 331–355. <https://doi.org/10.1177/1466802505057715>

Appendix

A. Raw data table

Table 10. Raw data table of the conditions and outcome

<i>Country & category of ATD</i>	<i>Enforcement approach</i>			<i>Management approach</i>			<i>Ideal-typical transposition style</i>	<i>Outcome</i>	
	<i>Public EU attitude</i>	<i>Ideological orientation</i>	<i>Public immigrants attitude</i>	<i>Federalism</i>	<i>Parties in government</i>	<i>Administrative efficiency</i>	<i>Four worlds of compliance</i>	<i>Legal implementation ATD</i>	<i>Practical application ATD</i>
<i>Residence restrictions</i>									
AT (Austria)	24.80%	3	55.10%	1	2	1.7	Domestic politics	1	1
BE (Belgium)	11.22%	2	54.30%	2	5	1.3	Domestic politics	0	0
CZ (Czech Republic)	16.84%	2	57.10%	0	3	0.92	Dead letters	0	0
DK (Denmark)	13.42%	2	38.50%	1	2	2.14	Law observance	1	1
EL (Greece)	20.34%	3	61.50%	0	1	0.52	Transposition neglect	1	0
FI (Finland)	22.08%	2	42.50%	0	4	2.18	Law observance	0	0
LU (Luxembourg)	8.96%	3	35.60%	0	3	1.68	Transposition neglect	1	0
SI (Slovenia)	15.74%	4	48.20%	0	4	1.08	Dead letters	1	1
SK (Slovakia)	7.82%	3	58.60%	0	3.5	0.84	Dead letters	1	0
<i>Regular reporting to authorities</i>									
AT (Austria)	24.80%	3	55.10%	1	2	1.7	Domestic politics	1	1
BE (Belgium)	11.22%	2	54.30%	2	5	1.3	Domestic politics	0	0
CZ (Czech Republic)	16.84%	2	57.10%	0	3	0.92	Dead letters	1	1
DE (Germany)	13.32%	2	51.90%	2	2	1.58	Domestic politics	1	1
DK (Denmark)	13.42%	2	38.50%	1	2	2.14	Law observance	1	1
EL (Greece)	20.34%	3	61.50%	0	1	0.52	Transposition neglect	1	0
FR (France)	19.00%	2	39.20%	0	2	1.44	Transposition neglect	1	0
IE (Ireland)	11.12%	2	56.00%	0	3	1.42	Dead letters	1	1

LU (Luxembourg)	8.96%	3	35.60%	0	3	1.68	Transposition neglect	1	0
SI (Slovenia)	15.74%	3	48.20%	0	3.5	0.84	Dead letters	1	1
SK (Slovakia)	7.82%	4	58.60%	0	4	1.08	Dead letters	1	0
<i>Obligation to surrender passport and documents</i>									
AT (Austria)	24.80%	3	55.10%	1	2	1.7	Domestic politics	0	0
BE (Belgium)	11.22%	2	54.30%	2	5	1.3	Domestic politics	0	0
CZ (Czech Republic)	16.84%	2	57.10%	0	3	0.92	Dead letters	0	0
DE (Germany)	13.32%	2	51.90%	2	2	1.58	Domestic politics	1	1
DK (Denmark)	13.42%	2	38.50%	1	2	2.14	Law observance	1	1
EL (Greece)	20.34%	3	61.50%	0	1	0.52	Transposition neglect	1	0
FR (France)	19.00%	2	39.20%	0	2	1.44	Transposition neglect	1	1
LU (Luxembourg)	8.96%	3	35.60%	0	3	1.68	Transposition neglect	0	0
SI (Slovenia)	15.54%	3	48.20%	0	3.5	0.84	Dead letters	1	0
SK (Slovakia)	7.82%	4	58.60%	0	4	1.08	Dead letters	0	0
<i>Deposit of adequate financial guarantee</i>									
AT (Austria)	24.80%	3	55.10%	1	2	1.7	Domestic politics	1	1
BE (Belgium)	11.22%	2	54.30%	2	5	1.3	Domestic politics	0	0
CZ (Czech Republic)	16.84%	2	57.10%	0	3	0.92	Dead letters	1	0
DE (Germany)	13.32%	2	51.90%	2	2	1.58	Domestic politics	0	0
DK (Denmark)	13.42%	2	38.50%	1	2	2.14	Law observance	1	0
EL (Greece)	20.34%	3	61.50%	0	1	0.52	Transposition neglect	1	0
ES (Spain)	13.74%	4	35.20%	2	1	0.98	Domestic politics	0	0
FR (France)	19.00%	2	39.20%	0	2	1.44	Transposition neglect	0	0
HU (Hungary)	21.82%	3	54.80%	0	1.33	0.68	Dead letters	0	0
IE (Ireland)	11.12%	2	56.00%	0	3	1.42	Dead letters	1	0
LU (Luxembourg)	8.96%	3	35.60%	0	3	1.68	Transposition neglect	0	0

NL (Netherlands)	9.88%	2	42.10%	0	2.5	1.74	Domestic politics	1	1
SE (Sweden)	18.74%	1	34.00%	0	4	2	Law observance	0	0
SI (Slovenia)	15.74%	3	48.20%	0	3.5	0.84	Dead letters	1	0
SK (Slovakia)	7.82 %	4	58.60%	0	4	1.08	Dead letters	1	0

Note. Data for country & category of AD from the Directorate-General for Migration and Home Affairs, (2013), for the public EU attitude from (Kaufmann et al., 2011), for the ideological orientation from Armingeon et al. (2017), for the public immigrants attitude from EVS (2016), for the federalism from Armingeon et al. (2017), for the parties in government (Andersson et al., 2014), for the administrative efficiency from Kaufmann et al. (2011), for the four worlds of compliance from Falkner & Treib (2008), for the legal implementation and practical application from the Directorate-General for Migration and Home Affairs (2013)

B. Truth tables

Table 11. Truth table, result for the second step of the csQCA: remote condition FW-DL

<i>FW-DL</i>	<i>EA-PAIM</i>	<i>MA-FE</i>	<i>MA-PG</i>	<i>MA-AD</i>	<i>SC-EUM</i>	<i>AD-CC</i>	<i>Number</i>	<i>Outcome</i>	<i>Raw consistency</i>	<i>Cases</i>
0	1	1	0	0	0	0	7 (18%)	0	0	AT-RS, AT-RRtA, DE-RRtA, AT-SPD, DE-SPD, AT-FG, DE-FG
0	0	0	1	0	0	0	5 (31%)	0	0.4	FI-RR, LU-RS, LU-RRtA, NL-FG, SE-FG
1	1	0	1	1	1	0	4 (41%)	1	0.8	SK-RS, SK-RRtA, CZ-FG, SK-FG
0	1	1	1	0	0	0	4 (51%)	0	0	BE-RS, BE-RRtA, BE-SPD, BE-FG
0	1	0	0	1	0	0	4 (61%)	0	1	EL-RS, EL-RRtA, EL-SPD, EL-FG
0	0	1	0	0	0	0	3 (69%)	0	0	DK-RS, DK-RRtA, DK-SPD
0	0	1	0	1	0	0	2 (74%)	0	0.5	DK-FG, ES-FG
1	0	0	1	1	1	0	2 (78%)	1	1	SI-SPD, SI-FG
1	0	0	1	1	1	1	2 (82%)	0	0	SI-RS, SI-RRtA
0	0	0	0	0	0	0	1 (85%)	0	1	FR-RRtA
1	1	0	1	0	0	0	1 (88%)	1	1	IE-FG
0	0	0	0	0	0	1	1 (91%)	0	0	FR-SPD
1	1	0	1	0	0	1	1 (94%)	0	0	IE-RRtA
1	1	0	0	1	1	0	1 (97%)	0	0	HU-FG
1	1	0	1	1	1	1	1 (100%)	0	0	CZ-RRtA

Note. Including the remote condition FW-DL (World of dead letters) and proximate conditions: EA-PAIM (Public attitude Immigrants, MA-FE (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), the scope condition ad-cc (Alternative to Detention, conflicting case) and SC-EUM (Scope condition European membership in 2004). An explanation of the codes for the cases can be found in Chapter 3.4.

Table 12. Truth table, result for the second step of the csQCA: remote condition FW-TN

<i>FW- TN</i>	<i>EA- PAEU</i>	<i>EA- PAIM</i>	<i>MA- FE</i>	<i>MA- PG</i>	<i>MA- AD</i>	<i>AD- CC</i>	<i>Number</i>	<i>Outcome</i>	<i>Raw consistency</i>	<i>Cases</i>
0	1	1	1	0	0	0	4 (10%)	0	0	AT-RS, AT-RRtA, AT-SPD, AT-FG
0	0	1	1	1	0	0	4 (20%)	0	0	BE-RS, BE-RRtA BE-SPD, BE-FG
0	0	1	0	1	1	0	4 (30%)	0	1	EL-RS, EL-RRtA, EL-SPD, EL-FG
1	1	1	0	0	1	0	3 (38%)	1	1	SK-RS, SK-RRtA, SK-FG
0	0	0	1	0	0	0	3 (46%)	0	0	DK-RS, DK-RRtA, DK-SPD
0	0	1	1	0	0	0	3 (54%)	0	0	DE-RRtA, DE-SPD, DE-FG
0	1	0	0	1	0	0	2 (59%)	0	0	FI-RS, SE-FG
0	0	0	1	0	1	0	2 (64%)	0	0.5	DK-FG, ES-FG
0	0	0	0	1	1	0	2 (69%)	0	1	SI-SPD, SI-FG
0	0	0	0	1	1	1	2 (74%)	0	0	SI-RS, SI-RRtA
1	0	0	0	1	0	0	2 (79%)	1	1	LU-RS, LU-RRtA
0	0	0	0	1	0	0	1 (81%)	0	0	NL-FG
0	0	1	0	1	0	0	1 (84%)	0	1	IE-FG
0	1	1	0	0	1	0	1 (87%)	0	0	HU-FG
0	1	1	0	1	1	0	1 (90%)	0	1	CZ-FG
0	0	1	0	1	0	1	1 (93%)	0	0	IE-RRtA
0	1	1	0	1	1	1	1 (96%)	0	0	CZ-RRtA
1	1	0	0	0	0	0	1 (98%)	1	1	FR-RRtA
1	1	0	0	0	0	1	1 (100%)	0	0	FR-SPD

Note. Including the remote condition FW-TN (World of transposition neglect), EA-PAEU (Public attitude EU), EA-PAIM (Public attitude Immigrants, MA-FE (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4.

Table 13. Truth table: combined result of remote conditions and proximate conditions

<i>FW-DL</i>	<i>FW-TN</i>	<i>EA-PAEU</i>	<i>EA-PAIM</i>	<i>MA-FE</i>	<i>MA-PG</i>	<i>MA-AD</i>	<i>SC-EUM</i>	<i>AD-CC</i>	<i>Number</i>	<i>Outcome</i>	<i>Raw consistency</i>	<i>Cases</i>
1	0	0	1	0	1	1	1	0	4 (10%)	1	1	EL-RS, EL-RRtA, EL-SPD, EL-FG
0	0	1	1	1	0	0	0	0	4 (20%)	0	0	AT-RS, AT-RRtA, AT-SPD, AT-FG
0	0	0	1	1	1	0	0	0	4 (30%)	0	0	BE-RS, BE-RRtA, BE-SPD, BE-FG
0	1	1	1	0	0	1	0	0	3 (38%)	1	1	SK-RS, SK-RRtA, SK-FG
0	0	0	0	1	0	0	0	0	3 (46%)	0	0	DK-RS, DK-RRtA, DK-SPD
0	0	0	1	1	0	0	0	0	3 (54%)	0	0	DE-RRtA, DE-SPD, DE-FG
1	0	0	0	0	1	1	1	0	2 (59%)	1	1	SI-SPD, SI-FG
1	0	0	0	0	1	1	1	1	2 (64%)	0	0	SI-RS, SI-RRtA
0	0	1	0	0	1	0	0	0	2 (69%)	0	0	FI-RS, SE-FG
0	0	0	0	1	0	1	0	0	2 (74%)	0	0.5	DK-FG, ES-FG
0	1	0	0	0	1	0	0	0	2 (79%)	1	1	LU-RS, LU-RRtA
1	0	1	1	0	0	1	1	0	1 (82%)	0	0	HU-FG
1	0	1	1	0	1	1	1	0	1 (85%)	1	1	CZ-FG
1	0	1	1	0	1	1	1	1	1 (87%)	0	0	CZ-RRtA
0	0	0	0	0	1	0	0	0	1 (90%)	0	0	NL-FG
1	0	0	1	0	1	0	0	0	1 (92%)	1	1	IE-FG
0	1	1	0	0	0	0	0	0	1 (95%)	1	1	FR-RRtA
1	0	0	1	0	1	0	0	1	1 (97%)	0	0	IE-RRtA
0	1	1	0	0	0	0	0	1	1 (100%)	0	0	FR-SPD

Note. Including the remote conditions FW-DL (World of dead letters) & FW-TN (World of transposition neglect). Including the proximate conditions: EA-PAEU (Public attitude EU), EA-PAIM (Public attitude Immigrants), MA-FE (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency). Including the scope condition ad-cc (Alternative to Detention, conflicting case) and SC-EUM (Scope condition European membership in 2004). An explanation of the codes for the cases can be found in Chapter 3.4.

C. Other types of solution formulas

Table 14. Step 1: complex solution for the remote conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
ad-cc*FW-DL*fw-tn	0.466667	0.466667	0.875	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), HU-FG (1,0), IE-FG (1,1), SI-FG (1,1), SK-FG (1,1)
ad-cc*FW-TN*fw-dl	0.466667	0.466667	1	EL-RS (1,1), LU-RS (1,1), EL-RRtA (1,1), FR-RRtA (1,1), LU-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)
Solution coverage:	0.933333			
Solution consistency:			0.933333	

Note. Including the remote conditions FW-LO (World of law observance), FW-DP (World of domestic politics), FW-DL (World of dead letters), FW-TN (World of transposition neglect), and the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent

Table 15. Step 1: parsimonious solution for the remote conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
ad-cc*fw-lo*fw-dp	0.933333	0.933333	0.933333	EL-RS (1,1), LU-RS (1,1), EL-RRtA (1,1), FR-RRtA (1,1), LU-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1), SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), HU-FG (1,0), IE-FG (1,1), SI-FG (1,1), SK-FG (1,1)
Solution coverage:	0.933333			
Solution consistency:			0.933333	

Note. Including the remote conditions FW-LO (World of law observance), FW-DP (World of domestic politics), FW-DL (World of dead letters), FW-TN (World of transposition neglect), and the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent

Table 16. Step 2: Complex solution for the remote condition FW-DL and the proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-DL*ma-fe*MA-PG*MA-AD* SC-EUM*ad-cc	0.4	0.4	1	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), SI-FG (1,1), SK-FG (1,1)
FW-DL*EA-PAIM*ma-fe* MA-PG*ma-ad*sc-eum*ad-cc	0.0666667	0.0666667	1	IE-FG (1,1)
Solution coverage:	0.466667			
Solution consistency:			1	

Note. Including the remote condition FW-DL (World of dead letters) and proximate conditions: EA-PAIM (Public attitude Immigrants, ma-fe (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), the scope condition ad-cc (Alternative to Detention, conflicting case) and SC-EUM (Scope condition European membership in 2004). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Table 17. Step 2: Parsimonious solution for the remote condition FW-DL and the proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-DL*MA-PG* ad-cc	0.466667	0.466667	1	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), IE-FG (1,1), SI-FG (1,1), SK-FG (1,1)
Solution coverage:	0.466667			
Solution consistency:			1	

Note. Including the remote condition FW-DL (World of dead letters) and proximate conditions: MA-PG (Parties in government), the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Table 18. Step 2: Intermediate solution for the remote condition FW-TN and the proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-TN*EA-PAEU* ea-paim*ma-fe* ma-pg*ma-ad*ad-cc	0.0666667	0.0666667	1	FR-RRtA (1,1)
FW-TN*ea-paeu*ea-paim*ma-fe*MA- PG*ma-ad*ad-cc	0.133333	0.133333	1	LU-RS (1,1), LU-RRtA (1,1)
FW-TN* EA-PAEU *EA-PAIM*ma- fe*ma-pg*MA-AD*ad-cc	0.266667	0.266667	1	EL-RS (1,1) EL-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)
Solution coverage:	0.466667			
Solution consistency:			1	

Note. Including the remote condition FW-TN (World of transposition neglect), EA-PAEU (Public attitude EU), EA-PAIM (Public attitude Immigrants, ma-fe (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency), the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Table 19. Step 2: Parsimonious solution for the remote condition FW-TN and the proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-TN*ad-cc	0.466667	0.466667	1	EL-RS (1,1) EL-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1), FR-RRtA (1,1), LU-RS (1,1), LU-RRtA (1,1)
Solution coverage:	0.466667			
Solution consistency:			1	

Note. Including the remote condition FW-TN (World of transposition neglect), the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Table 20. Complex solution for combined result of remote conditions and proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-TN*fw-dl*EA-PAEU* ea-paim*ma-fe* ma-pg*ma-ad*sc-eum*ad-cc	0.0666667	0.0666667	1	FR-RRtA (1,1)
FW-TN *fw-dl*ea-paeu*ea-paim*ma-fe*MA-PG*ma-ad**sc-eum ad-cc	0.133333	0.133333	1	LU-RS (1,1), LU-RRtA (1,1)
FW-TN*fw-dl*EA-PAEU *EA-PAIM*ma-fe*ma-pg*MA-AD*sc-eum*ad-cc	0.266667	0.266667	1	EL-RS (1,1) EL-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)
FW-DL*fw-tn*ea-paeu*ea-paim*ma-fe*MA-PG*MA-AD*SC-EUM*ad-cc	0.4	0.4	1	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), CZ-FG (1,1), SI-FG (1,1), SK-FG (1,1)
FW-DL*fw-tn*ea-paeu*EA-PAIM*ma-fe*MA-PG*ma-ad*sc-eum*ad-cc	0.0666667	0.0666667	1	IE-FG (1,1)
Solution coverage:	0.933333			
Solution consistency:			1	

Note. Including the remote conditions FW-DL (World of dead letters) & FW-TN (World of transposition neglect). Including the proximate conditions: EA-PAEU (Public attitude EU), EA-PAIM (Public attitude Immigrants, ma-fe (Federalism), MA-PG (Parties in government), MA-AD (Administrative efficiency). Including the scope condition ad-cc (Alternative to Detention, conflicting case) and SC-EUM (Scope condition European membership in 2004). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Table 21. Parsimonious solution for combined result of remote conditions and proximate conditions

<i>Causal pathway</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases</i>
FW-TN *ad-cc	0.466667	0.466667	1	EL-RS (1,1) EL-RRtA (1,1), EL-SPD (1,1), EL-FG (1,1)FR-RRtA (1,1), LU-RS (1,1), LU-RRtA (1,1)
FW-DL*MA-PG* ad-cc	0.466667	0.466667	1	SK-RS (1,1), SK-RRtA (1,1), SI-SPD (1,1), IE-FG (1,1), CZ-FG (1,1), SI-FG (1,1), SK-FG (1,1)
Solution coverage:	0.933333			
Solution consistency:			1	

Note. Including the remote conditions FW-DL (World of dead letters) & FW-TN (World of transposition neglect). Including the proximate condition MA-PG (Parties in government). Including the scope condition ad-cc (Alternative to Detention, conflicting case). An explanation of the codes for the cases can be found in Chapter 3.4. A condition in uppercase letters means it is present, while a condition in lowercase letters means it is absent.

Ready or Not?

An Evaluability Assessment on the Readiness of the Directorate-General for Migration to Evaluate the Less-coercive Supervisory Measures in the Netherlands

Frank Hendriks

Master student Comparative Politics, Administration, and Society
*Radboud University Nijmegen: Department of Public Administration, Faculty of
Management Sciences*

Supervisory lecturer: DR. J.H.M.M. Tholen

Second reader: Prof. dr. S. van Thiel

In cooperation with the Department of the Directorate-General for Migration (DGM) of the Dutch Ministry
of Justice & Security

Supervisors: Irene Ritman, Martijn ten Dam, Renée ter Hart

June 20, 2018

Abstract

The current study used an evaluability assessment (EA) to examine the implementation of new supervisory measures and the improvements made to already applied supervisory measures as part of the return process of the migration policy in the Netherlands. The present study analyzed the “readiness” of the implementation and improvements to the supervisory measures policy program. The six-step EA-model by Wholey (2015a) was adapted to determine if indeed the policy program was ready for useful evaluation.

The analysis consisted of five aspects. The first two aspects formed the results of the study. The first aspect of ‘program theory’ involved the identification of the components of the policy program and the construction of a logic model. These findings were compared with the perceptions of relevant stakeholders on the implementation of the policy program in practice as the second aspect of ‘program reality.’ The combined findings from these aspects were further examined in the discussion of the summative and formative aspects of this study. The third aspect involved the summative analysis, as to provide information on the measurements and analysis of the outcomes and impact of the policy program in the ‘program plausibility & data availability’ aspect. The fourth aspect consisted of the formative analysis, as to determine in which way the policy program needs to be improved to enable further evaluation in the ‘program utility’ aspect. A final aspect of ‘program evaluation’ presented several evaluation options. The necessary information was gathered by conducting a document review and eight semi-structured interviews with relevant stakeholders.

The summative results from this study indicate that that the general intention of the policy, to enable more possibilities to apply less-coercive measures in the return process, has clearly improved. However, gathering accurate and substantive information on the effect of the policy on the expected outcomes of the proportionality, effectiveness, and necessity of the policy will be difficult. With most of the information being based on the interpretations and assumptions of the stakeholders. The formative results found some discrepancies between the ‘program theory’ and ‘program reality.’ Several recommendations are included in the present study to remove these discrepancies, so that the policy program is ready to be evaluated. Several evaluation options are presented on the basis of the summative and formative discussion. These consist of options directed towards the effectiveness and efficiency of the adapted policy program, as well as the treatment of the irregular-staying migrants.

Acknowledgments

This study is the second part of a master's thesis on the topic of the alternatives to detention within the return process of the immigration policy field in the Netherlands. And it was done in cooperation with the Department of the Directorate-General for Migration (DGM) of the Dutch Ministry of Justice & Security. The present study is the culmination of the cooperation between the research student, those involved in the workgroup, and the other stakeholders. The immigration policy field can be daunting to study, as it is continuously changing due to external influences. A lot of thanks and gratitude thus goes to Irene Ritman, Martijn ten Dam and Renée ter Hart for their supervision, guidance, and feedback specifically for this part and the thesis as a whole. Their understanding of the subject matter and their balanced view provided helpful access to the policy field for a "rookie" on the subject. Further gratitude goes out to Laura Kok, and Sanne Westmaas, whom both left the project at an earlier stage as well as Peter Bosch who replaced Renée ter Hart in the final stage. In addition, special thanks and gratitude go out to Wouter Dronkers, who provided a chance to do such an exciting project with the Dutch Ministry of Justice & Security, and Dr. Berry Tholen, who provided clear and useful feedback in his role as the supervisory lecturer and first reader. Additional thanks and gratitude for their input goes towards the following:

- Prof. dr. Sandra van Thiel, for her willingness to take on the role of the second reader
- The legal advisor from the INS
- The policy data analyst from the Directorate-General for Migration
- The operational specialist from the National Police
- The Data analyst and senior departure supervisors from the Return & Departure Service (also for showing the freedom-restricting center (FRC) center in Ter Apel)
- The data analyst from the Royal Marechaussee
- The senior employee from Amnesty International
- Those involved that made the site-visit to the closed family centre (CFF) at Detention Center Zeist possible
- And last but not least, towards my family, girlfriend, and friends, who supported me through my years of study. Even when I could have made it easier for myself, they always reassured me that overcoming these challenges would all be worth it in the end.

List of Contents

Abstract.....	2
Acknowledgments.....	3
List of Contents	4
List of Figures & Tables	6
1. Introduction.....	7
1.1 Research Questions	9
1.2 Relevance.....	11
1.3 Scope & Delimitation.....	13
1.4 Key Terms.....	14
1.4.1 Immigration policy field.....	14
1.4.2 Evaluation Field	16
1.5 Structure of the Study.....	18
2. Theoretical Framework.....	19
Program Evaluation Research.....	19
2.1 Type of Program Evaluation Research.....	20
2.2 Evaluability Assessment Model.....	23
2.3 Six-step EA-Model.....	25
2.3.1 Program theory & logic model.....	27
2.4 Summary of the Theoretical Chapter	30
3. Methodology	31
3.1 Research Design	31
3.2 Population & Selection	32
3.3 Data Collection	34
3.4 Data Analysis	36
3.5 Procedures.....	37
3.5.1 Step 1: Involve intended users and other key stakeholders.....	37
3.5.2 Step 2: Clarify the program design.....	38
3.5.3 Step 3: Explore program reality.....	40
3.5.4 Step 4: Assess the plausibility of the program.....	40
3.5.5 Step 5: Reach agreement on any needed changes.....	41
3.5.6 Step 6: Reach agreement on evaluation focus and intended use	41
3.6 Ensuring Reliability & Validity	42
3.7 Summary of the methodology.....	44
4. Results.....	45
4.1 Program Theory & Logic Model.....	45
4.1.1 Contextual factors.....	45
4.1.2 Rationale.....	47
4.1.3 Input.....	50
4.1.4 Activities.....	52
4.1.5 Output.....	59
4.1.6 Outcomes.....	63
4.1.7 Objectives.....	65
4.1.8 Summary: Logic model of adapted policy program.....	66
4.2 Program Reality.....	68
4.2.1 The contextual factors in practice.....	68
4.2.2 The rationale in practice	70
4.2.3 Input in practice.....	73
4.2.4 Activities as implemented in practice.....	74
4.2.5 Output of the adapted policy program in practice	79
4.2.6 Outcomes & objectives.....	81
5. Discussion	83
5.1 Program Plausibility & Data Availability: Summative Analysis.....	83
5.1.1 Program Plausibility.....	83
5.1.2 Data Availability.....	89
5.2 Program Utility: Formative analysis.....	93

5.2.1	Recommendations regarding program components.....	93
5.2.2	Recommendations regarding the collection of data.....	94
5.2.3	Recommendations regarding the terminology of ATD	94
5.2.4	Recommendations regarding maximum application duration.....	95
5.3	Program Evaluation.....	97
5.3.1	The status quo: Perception based supervision.....	97
5.3.2	Overview of the evaluation options.....	99
6.	Conclusions.....	105
6.1	Main Findings	105
6.2	Main Limitations.....	106
6.3	Main Implications.....	108
6.4	Main Recommendations	109
	References.....	110
	Appendix.....	117
A.	Example of an interview guide	117
B.	Codebook for Data analysis	120
C.	Quotes for the Results.....	124
D.	Evaluation Options	129
E.	Executive Summary for the Department of the Directorate-General for Migration (in Dutch).....	137

List of Figures & Tables

Figures

Figure 1. Structure of the theoretical chapter	19
Figure 2. Six continua for evaluation approaches. Adapted from Newcomer et al. (2015)	20
Figure 3. Four standards required for useful evaluation. Adapted from Wholey (2015a)	22
Figure 4. Summative and formative aspects of Evaluability Assessment, based on Trevisan & Walser (2015, p. 5 & 7).....	22
Figure 5. Generic Logic Model. Adapted from McLaughlin & Jordan (2015).	28
Figure 6. Six-step EA-model. Adapted from Wholey (2015a).	37
Figure 7. Five steps for constructing a logic model. Adopted from McLaughlin & Jordan (2015).....	38
Figure 8. Methodological requirements. Adopted from Newcomer et al. (2015)	42
Figure 9. Overview of the rationale.....	49
Figure 10. Logic model for the intended policy program of the implementation and improvement of the supervisory measures in the Netherlands.....	67
Figure 11. Statements made regarding the use of administrative detention as an 'ultimum remedium' in practice	70
Figure 12. Overview of the structure of the supervisory measures in practice.....	78
Figure 13. Hypothetical process of an irregular-staying migrant in the return process.....	79
Figure 14. Overview of the different structures of the supervisory measures.....	96

Tables

Table 1. The six-step model in comparison to other EA-models.....	24
Table 2. Interview respondents list.....	33
Table 3. Components used to describe the program and elements of the program theory & logic model...	40
Table 4. Included elements of the evaluation options	42
Table 5. Overview of the methodology	44
Table 6. Overview of the results of the pilots in 2012 in comparison to administrative detention in 2010	61
Table 7. Overview Program Plausibility	84
Table 8. Overview of the Evaluation Options.....	102
Table 9. Mixed codebook for data analysis	120
Table 10. Overview of the quotes used in the Results	124
Table 11. Necessity evaluation option 1	129
Table 12. Necessity evaluation option 2	130
Table 13. Effectiveness evaluation option 1.....	131
Table 14. Effectiveness evaluation option 2.....	132
Table 15. Proportional evaluation option 1	133
Table 16. Proportional evaluation option 2	134
Table 17. Proportional evaluation option 3	135
Table 18. Proportional evaluation option 4	136

1. Introduction

In previous years, most of the attention on the implementation of the alternatives to detention (ATD) over administrative detention measures (ADM) has been paid to the expected advantages of these measures over detention. The scientific field, relevant non-governmental organizations (NGO's) and governments have started to shift away from detention as the most used and most viable measure. According to the 2008/115/EC Return Directive of the European Union, ADM should only be applied if necessary and less-coercive measures are to be the preferred choice for providing supervision in the return process (EUR-Lex, 2008). ADM, in that case, is considered an '*ultimum remedium*,' or a measure of last resort (Leerkes & Broeders 2010; Leerkes 2011). Recent research has moved on from advocating the shift towards the use of less-coercive measures, towards the analysis of the implementation of these measures in practice. Several studies point out that the development of alternative measures to ADM has been slow and the effects of implementation of these are difficult to identify (Bloomfield, 2016; Costello & Kaytaz, 2013; Field & Edwards, 2006). While there is still a lack of empirical data to assess the evaluations of these measures (Field & Edwards, 2006; The European Migration Network, 2014).

The present study aims to contribute to the analysis of ATD by examining the implementation of new supervisory measures and the improvement of already existing measures in the Netherlands (also referred to in the present study as the 'adapted policy program (of supervisory measures)'). The development of several pilots for less-coercive measures started in 2010 (Leers, 2011). In 2013, the State Secretary of Justice & Security informed the Dutch House of Representatives of the commitments on implementing or improving the use of less-coercive measures in the return process (Teeven, 2013). Some of the measures were already described in some form in the Aliens Act 2000, and thus, were adjusted to fit with the other measures in the range of supervisory measures that could be applied (Ministry of the Interior and Kingdom Relations, 2000a). The development of less-coercive supervisory measures received renewed attention after the suicide of Alexander Dolmatov in the detention center of Rotterdam. The investigations and reports into the circumstances that led to his suicide further increased the desire of the Ministry of Justice & Security to embed the alternatives further into the return policy (Ministry of Justice & Security, 2013). The Advisory Committee on Alien Affairs provided advice on the renewed desire to embed the alternatives, while Amnesty International provided a critical report 'Detention of immigrants in the Netherlands: It should and could be done differently' (Vreemdelingendetentie in Nederland: Het moet en kan anders) (Advisory Committee on Alien Affairs, 2013; Amnesty International, 2011). These reports concluded that the application of the alternatives have previously been lacking and the results from the pilots indicate that these could improve that application. The Government of the Netherlands has incorporated the advice from

the Advisory Committee on Alien Affairs and other actors and used the encouraging results from the pilots to change the policy program concerning the use of supervisory measures and administrative detention. The 2000 Aliens act and the 2000 Aliens decree provide the relevant rules and regulations concerning the immigration policies of the Netherlands (Ministry of the Interior and Kingdom Relations, 2000a, 2000c). Article 59c of the Aliens Act mentions that less-coercive measures have priority over detention unless these cannot be applied (Global Detention Project, 2016). The regularly used alternatives are a duty to report with return supervision, the freedom-restricting measure, and the subsidizing of projects from NGO's (Global Detention Project, 2016; Teeven, 2013).

The Ministry of Justice & Security has thus made increased efforts in improving the circumstances of irregular-staying migrants in the return policy and now wishes to evaluate the implementation of the related policy program. The present study, therefore, goes more in-depth on the practical part of gathering the necessary data to make a worthwhile evaluation of the implementation of ATD as less-coercive measures in the return process. By analyzing the case of the Netherlands with an Evaluability Assessment (EA), the purpose is to see if a policy is ready for evaluation, what needs improvement and how further evaluation should proceed. For this study, this consists of an analysis with both a summative- and formative aspect, based on the findings from the policy program as intended and the policy program as applied in practice. These findings include the formulation of an overview of the policy, designed to clarify the intentions, objectives, and components of the policy program. These are then analyzed and compared with the implementation of the actual policy program, with the use of in-depth semi-structured interviews with relevant actors. The summative aspect of the study examines these findings, to determine which outcomes are important and how further evaluation should measure these. The formative aspect of the study incorporates the findings from the result and the summative discussion, to provide recommendations for changes in the policy program to enable further evaluation and present suggestions for such evaluations. The study is conducted in cooperation with the Directorate-General for Migration (DGVZ) of the Ministry of Justice & Security from the Netherlands and relevant stakeholders from other government agencies.

1.1 Research Questions

The present study investigates the possibilities for a government to analyze the implementation of less-coercive measures into the return policy. As such, it involves exploratory research on the evaluation of these kinds of policies before the actual evaluation. This study adheres to two crucial assumptions of evaluation research, that everything can be evaluated but that not everything should be evaluated (Finckenauer, Margaryan, & Sullivan, 2005). It is thus not the question *if* it is possible to evaluate the adapted policy program of the supervisory measures, but rather *what* should be evaluated, *how* should this be evaluated, and *what* needs to change to enable such an evaluation. Together, these aspects should determine *if* the Directorate-General for Migration is ready for such an assessment. The present study, therefore, includes the following central research question:

To what extent is the Directorate-General for Migration of the Netherlands ready to evaluate the adapted policy program of the use of supervisory measures in the return process?

The following section presents several research questions and sub-questions formulated for the study as to answer the central research question. These represent the necessary aspects the study must include to provide a conclusive answer to the research question. The first part relates to the use of the EA as the approach that fits with the aim of the study, based on the theoretical research (see also the Theoretical research questions). The following part is related to providing the necessary information for the aforementioned aspects of the central research question. It is aimed at uncovering the components of the policy program as it was intended and the subsequent comparison between the policy as intended and its implementation in practice (see also the Empirical research questions: Program comparison). The three aspects that form the basis for the research question are divided into three further research questions. The fourth of which involves the summative aspect of the study, which considers the assessment of any preliminary results and the availability of additional information. The fifth question is part of the formative aspect, which considers if any changes are necessary to the policy program to enable further evaluation and provides evaluation options. The following research questions and sub-questions are therefore formulated for the present study:

- **Theoretical research questions**

1. How can one determine when the adapted policy program of supervisory measures is ready to be evaluated?
 - A) *What kind of evaluative design is most suitable to determine when a policy program is ready to be evaluated?*
 - B) *What kind of an evaluability assessment model can be adapted to fit the characteristics of the adapted policy program of supervisory measures?*
 - C) *Of what steps does a suitable evaluability assessment model consist of for this study?*

- **Empirical research questions**

2. What are the contextual factors, program structure, and outcome structure of the adapted policy program of supervisory measures as intended?
 - A) *What are the relevant antecedent- and mediating contextual factors?*
 - B) *What rationale, objectives, input, activities, and output are relevant for the program structure?*
 - C) *What short- and intermediate-term outcomes and long-term outcome objectives are included in the outcome structure?*
3. How does the intended policy program compare with the perceptions and assumptions of the policy program as implemented in practice?
4. *Summative:* Is it plausible that the policy program and its outputs lead to the objectives of the program?
 - A) *What feasible measures are there and to what extent is data available to assess these further?*
 - B) *What are the preliminary results of the policy program in practice?*
5. *Formative:* What recommendations can be made that increase the evaluability and performance of the adapted policy program of the supervisory measures?
 - A) *Is there an agreement on the approach and substance of further evaluation between the stakeholders?*
 - B) *What recommendations can be made concerning the adapted policy program to enable further evaluation?*
6. What kind of evaluations are suitable options for the actual evaluation?

1.2 Relevance

The relevance of the study is spread across the scientific, institutional and social domain. The scientific significance consists of the use of the evaluability assessment model. Such an approach to the analysis has its origin in the early 1980s, as to improve the evaluation of complex federal policies in the United States. To date, the model is still used, even more so than expected by one of its advocates. Trevisan (2007) finds several studies that used an EA and made their own adaptations to these. The present study is particularly relevant for the development of the EA-model. The use of an EA-model towards immigration-related policies could contribute to a better understanding of the possibilities and obstacles for those that wish to investigate or evaluate the recent implementation of less-coercive measures in the return process or other immigration-related policies.

In addition, the findings from this study can also be relevant for the development of quantifiable data on ADM & ATD measures. The demand for comprehensive data on immigration is shared by both governments, civil society organizations, and the scientific field alike (Welch & Schuster, 2005). Several studies from all domains have tried aggregating and subsequently analyzing quantifiable data on ADM & ATD measures. The study by Basilien-Gainche (2015) noted that there is a lack of detailed data from most countries. While a review by the Global Detention Project (GDP) states that data on migrants and asylum seekers in European countries is difficult to retrieve from the respective governments or the data is not complete or accurate (Global Detention Project, 2015). Furthermore, the data that can be aggregated is often found to be difficult to compare across countries. This is because it is difficult to determine where the data received is based on, what definition of detention is utilized, or on which level the data is aggregated. As stated in the article by Bloomfield (2016), it is crucial for civil society to be able to examine the use of ADM and ATD critically. Thus, there is a need for aggregated and accurate information.

The need for that kind of information is especially relevant for the practical domain, in specific for the government of the Netherlands, as to see if their implementation of ATD related policies has the desired outcome. A new government was formed in the Netherlands not too long ago. The new government formation intends to increase the rate of effective return. This should be done by increasing the capacity of the Repatriation and Departure Service to ensure actual departure. Furthermore, possibilities for forced return by for example ADM should be broadened within the legal framework of the EU. The example of criminal migrants is used in that case, but it is yet unclear which other categories would also be considered part of the expanded application of forced return (Rutte, Van Haersma Buma, Pechtold, & Segers, 2017). It is still unclear what effect ATD measures have on the preparedness to cooperate with the return process. The current study contributes to the evaluation of the ATD measures and as such contributes to the development of

understanding the effects and outcomes of ATD measures. These contributions might also be relevant for a government from other countries if they wish to evaluate their implementation of the less-coercive measures in their return policies.

As for the relevance of the study for the social domain, it is especially relevant on a humanitarian level. The use of ATD as part of the return procedure is preferred over the ADM measure, as ADM can cause physical and psychological damage to those detained (Bloomfield, 2016). It is expected by the scientific field, governments, and NGO's alike that ATD measures provide a more humane and efficient treatment of irregular-staying migrants. By aiding in the evaluation of policies to improve the implementation of ATD, the present study hopes to contribute to the implementation of these measures as part of the return process.

1.3 Scope & Delimitation

The present study investigates the implementation of ATD measures and the improvement of the use of (less-coercive) supervisory measures in the Netherlands. Examining the implementation in the Netherlands increases the attention for the specific situation in the Netherlands, as well as providing possible interesting results for other (EU) countries to adapt for their evaluation of ATD.

As already mentioned in the introduction, the Netherlands is one of the countries which has seen substantive developments in the implementation of ATD's. The present study involves the case of the adapted policy program of supervisory measures in the Netherlands. Therefore, the analysis limits itself to the characteristics of these measures as implemented and applied by the Ministry of Justice & Security of the Netherlands. The less-coercive measures under consideration all pertain to the category of irregular-staying migrants in the return process and not to those measures applied to those who are refused entry to the Netherlands. The included measures are the implementation of return supervision, bail, and subsidizing of NGO's as mentioned by the letter to the House of Representatives from the State Secretary for Justice & Security (Teeven, 2013). Other measures that were already in place or were not a part of the pilots that are included in the study are the expulsion order used by the National Police and Royal Netherlands Marechaussee, the freedom-restricting measure with a possible stay in a freedom-restricting center (FRC). The family locations (GL) and closed family facility (CFF) are special facilities meant for vulnerable target groups, such as families with underage children and unaccompanied minors.

1.4 Key Terms

The following section gives a description of the central concepts that were used throughout the study. The study incorporates the definitions that are most representative of the assumptions and rationale of the characteristics of the present study and the immigration policy field in the Netherlands. The first group of central concepts is part of the immigration policy field. The second group includes the concepts from the evaluation field.

1.4.1 Immigration policy field

To begin with, the following part explains the concept of administrative detention as is adhered to in the study. Defining the concept of administrative detention before the alternatives to detention is important, as the alternatives serve as a replacement. Several of the underlying aspects of administrative detention are, therefore, similar for the alternatives. Multiple different definitions exist for the concept of administrative detention. This study adheres to the description as formulated by the article of Leerkes & Broeders (Broeders & Leerkes, 2010), as this highlights the administrative nature of detention within the immigration policy field. A crucial difference from criminal detention is, that detention for ‘immigration-related goals’ is not meant as a form of punishment (Broeders & Leerkes, 2010). Rather, it is an administrative measure, and not a form of detention used pre-trial or after a conviction of a crime. So, when mentioning detention for immigration-related purposes, the term ‘administrative detention’ is used to distinguish it from other forms of detention (Chmelickova, 2006). The definition by Leerkes & Broeders (2010) is also appropriate for the scope of this study as it relates to the return process instead of its use in the entry/asylum process. Leerkes & Broeders (Broeders & Leerkes, 2010) see the term of administrative detention as divided into two types. The first being the detention of those not admitted to the country at the border, including asylum seekers for some countries. The second type contains the detention of those already remaining within the respective country’s border and need to leave as their stay is deemed illegitimate, which is the type under analysis by present study (see also Chapter 1.3).

This is why, for the present study, the definition for Administrative Detention is: *“A non-punitive, bureaucratic measure that is meant to enable the enactment of border control. It merely ensures that ‘unwanted’ migrants can be located and identified and cannot abscond while the expulsion is prepared”* (Leerkes & Broeders, 2010, p. 830-831). Throughout the study, the term ‘administrative detention’ and the abbreviation of ADM (Administrative Detention Measure) is used to refer to this definition.

The following part explains what is meant in the study when it refers to the alternatives to detention (ATD). There is no universally accepted definition of ATD, as there is little consensus about what these alternatives entail by different actors. For example, the civil society organization the International Detention Coalition (IDC) uses the broad definition of: *"Any law, policy or practice by which persons are not detained for reasons relating to their migration status"* (Bloomfield, 2016, p. 31). In the same way as with the definition of administrative detention, such a definition does not distinguish between the different purposes such an alternative can have within the immigration field. Costello & Kaytaz (2013, p. 10-11) provide a definition that distinguishes between the use of such measures at the entry/asylum process or the return process, though this definition is altered for this study to only refer to the use within the return process. As such the definition is as follows: *"Encompassing any legislation, policy or practice that allows for asylum-seekers, refugees, and migrants to reside in the community while awaiting deportation or removal from the country, albeit subject to some restrictions on movement or liberty"* (Costello & Kaytaz, 2013, p. 10-11). Throughout the study, the term 'alternatives to detention (ATD)' (and its variations) is used to refer to this definition.

So far, the previous sections have described the definitions the term administrative detention' and alternatives to detention used throughout the study. This section will move on to discuss for whom these measures apply. As mentioned in the previous sections, the focus of the present study lies in the use of alternatives to detention within the return process. As the scope of the study is the EU member states that fall under the EU directive within the European Union, the used categories for these measures can be different per country. A broad term of Third-Country National (TCN) is used by the EU commission as to refer to: *"Any person who is not an EU citizen and who is not a person enjoying the right of free movement under Union law"* (Directorate-General for Migration and Home Affairs, 2013). However, it remains unclear with this definition as to whom one refers to as being the target group for ATD measures. A more specific term of 'third-country national found to be illegally present' is provided by the Migration and Home Affairs of the EU Commission (2016b) and is defined as: *"A third-country national who is officially found to be in the territory of a Member-state and who does not fulfil, or no longer fulfils, the conditions for stay or residence in that Member State."* This definition does not incorporate the preference for the use of 'irregular' as opposed to 'illegal.' Which is preferred as it adheres to the administrative characteristic of the immigration field as well as removing any juridical and ethically objections (EU Commission, 2016a).

The broader term of 'irregular migrant' is more appropriate as, in the EU context, it is defined as:

A third-country national present on the territory of a Schengen State who does not fulfill, or no longer fulfills, the conditions of entry as set out in the Schengen Borders Code, or other conditions for entry, stay or residence in that Member State. (EU Commission, 2016a)

Throughout the study, the term 'irregular-staying migrant' is used to refer to this definition, with the addition that this means that the irregular-staying migrant is in the return process of an EU Member State that adheres to the Return Directive.

1.4.2 Evaluation Field

The first thing to consider with an evaluation is to specify what it means in relation to the characteristics of the study. According to Rossi, Lipsey, & Freeman (2004), evaluation as a concept includes both a description of the performance of the entity being evaluated and the standards/criteria used to form a decision on its performance. While evaluation also consists of a determination of the aspects that are the basis for measuring the performance. Assessing the value of government-funded programs are vital, as governments use it to increase the transparency, strengthen accountability, improve the performance, and determine the value of public programs (Newcomer, Hatry, & Wholey, 2015; Wholey, 2015b). The evaluation of public policy programs can also function to determine the continuation of a policy or program, determine the usefulness of new programs and could contribute to expanding the knowledge on theoretical and methodological aspects in practice (Rossi et al., 2004). Programs in that respect are defined by Newcomer et al. (2015, p. 5) as:

A set of resources and activities directed toward one or more common goals, typically under the direction of a single manager or management team. A program may consist of a limited set of activities in one agency or a complex set of activities implemented at many sites by two or more levels of government and by a set of public, non-profit, and even private providers.

When combining the two terms into the evaluation of these programs, program evaluation is considered as:

The application of systematic methods to address questions about program operations and results. It may include ongoing monitoring of a program as well as one-shot studies of program processes or program impact. The approaches used are based on social science research methodologies and professional standards. (Newcomer et al., 2015, p. 5-6)

The present study uses an Evaluability Assessment as an approach to a preliminary form of program evaluation. The article by Trevisan & Walser (2015, p.15) defines the EA as:

The systematic investigation of program characteristics, context, activities, processes, implementation, outcomes and logic to determine the extent to which the theory of how the program is intended to work aligns with the program as it is implemented and perceived in the field, the plausibility that the program will yield positive results as currently conceived and implemented, and the feasibility of and best approaches for further evaluation of the program.

Several models exist as to how an EA can be implemented in research. The theoretical chapter explains these in more detail (see Chapter 2.2).

1.5 Structure of the Study

The previous section defined the central concepts used in the study. The following section describes the structure of the study. A theoretical chapter outlines the evaluative approach to answer the research questions (see Chapter 2).

The subsequent methodology provides a detailed explanation of the resulting evaluative approach of the six-step EA-model by Wholey (2015a), as adapted for this study (see Chapter 3.5). The methodology also includes the analytical process for the data collection and analysis, as well as a review of the reliability and validity of the study.

The results are described in the fourth chapter (see Chapter 4). That chapter consists of two sections, related to the aspects of the program theory and the program reality. The program theory provides an overview of the relevant components of the adapted policy program of the supervisory measures in the return process, and the causal links between these components. The second section on program reality summarizes the perceptions, needs, and concerns for evaluation of the program by relevant stakeholders. That data from the stakeholders is then compared with the data from the program theory.

The discussion comprises of the analysis of the findings from the results, regarding the summative and formative aspects of the analysis (see Chapter 5). The first section of program plausibility & data availability provides the determination of the objectives, outcomes, and indicators that are deemed relevant for the actual evaluation of the policy program. In addition, this section also includes a part on the data-availability for such an evaluation and a preliminary indication of the results. The second section of program utility combines the data from the previous chapters, offering (possible) recommendations that enable the actual evaluation. The third part of program evaluability presents several options for such a comprehensive assessment.

The last chapter provides a discussion of the main findings, main limitations, main implications, and recommendations for further research (see Chapter 6). An executive summary for the DGM can be found in the appendix (see Appendix. E).

2. Theoretical Framework

The theoretical framework comprises of the examination of the broad field of program evaluation research (Chapter 0), the types of evaluation research (Chapter 2.1), the Evaluability Assessment model (Chapter 2.2), the specific model of EA for the present study (Chapter 2.3), the program theory & logic model of an EA (2.3.1), and a summary of the chapter (see Chapter 2.4). Figure 1 illustrates the structure of the chapter.

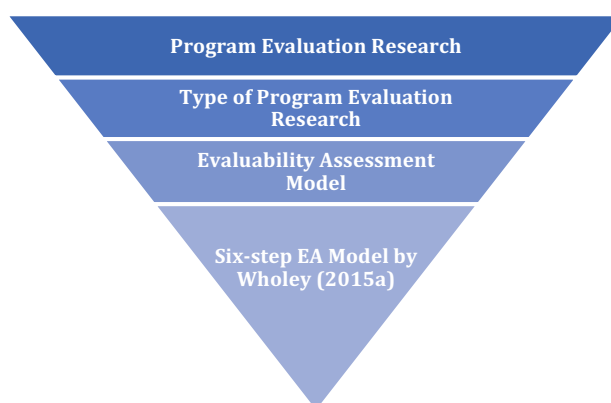


Figure 1. Structure of the theoretical chapter

Program Evaluation Research

Systematic program evaluation first began in the educative and public health domains before the first world war. This led to increased attention to the “knowledge of results” and developments in social science on statistical and other methods (Rossi et al., 2004). Parallel developments of evaluation in governments led to the applied research field of policy science/analysis (Rossi et al., 2004). Two major developments made the evaluation of large-scale programs possible. The first being the enhancement of measurement and survey procedures, improving the possibility of systematic data collection. The latter being the introduction of computers that can process vast amounts of statistical data. According to Rossi et al. (2004), the introduction of computers greatly advanced the field of evaluation research to this day. For governments, another important development for the use of evaluation was the introduction of performance management systems in the 1970s. Performance management systems include: “the formation of outcome-oriented goals, performance targets, monitor progress, stimulate performance improvements and communicate results to higher policy levels and the public (Wholey, 2015b, p. 653-654)”. Recent years saw a shift to transparency for evaluation, as data on performance measurement and evaluations are increasingly made public online (Wholey, 2015b).

2.1 Type of Program Evaluation Research

An important consideration is the position of the evaluation between a scientific versus pragmatic position (Rossi et al., 2004). The scientific posture has first been propositioned by Campbell (1969), which states that decisions on programs and policy should originate from social experimentation and by testing manners in which to improve social conditions. Campbell (1969) considered evaluation research part of the scientific research paradigm. Evaluations in that regard should adhere to the rigorous standards of scientific research (Rossi et al., 2004). This was countered by the views of Cronbach & Shapiro (1982), who argued that evaluations differ from scientific research. Although the logic of inquiry and research procedures might be similar, the aim of evaluation differs from that of scientific research. Evaluation research is tailor-made to the specific circumstances of the involved actors. Evaluations should meet expectations and provide useful information for the involved actors, instead of achieving the scientific quality standards solely. Even more so, the current study adheres to a new combined approach. This approach entails the identification of similar elements across different perceptions to develop an '*Evaluation Theory*.' According to Rossi et al. (2004), an evaluation theory should: "describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront (Shadish, Cook, & Leviton, 1991, p. 30-31). Such an ideal theory, which the authors mention is never entirely achievable, consists of three parts:

- (1) It would clarify activities, processes, and goals coming of the evaluation.
- (2) Explicate relationships among evaluative activities and the processes and goals they facilitate.
- (3) Empirically test propositions to identify and address those that conflict with research or other critically appraised knowledge about evaluation (Shadish et al., 1991, p. 30-31).

Evaluation research exists of several types or forms of evaluation that can be used depending on the evaluation. Newcomer et al. (2015) has distinguished six different '*continua*' on which evaluation approaches may differ (see Figure 2).

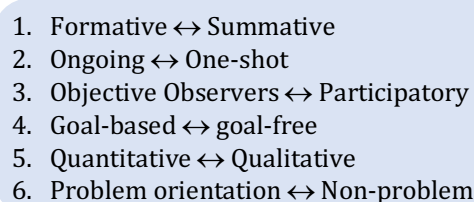
- 
1. Formative ↔ Summative
 2. Ongoing ↔ One-shot
 3. Objective Observers ↔ Participatory
 4. Goal-based ↔ goal-free
 5. Quantitative ↔ Qualitative
 6. Problem orientation ↔ Non-problem

Figure 2. Six continua for evaluation approaches. Adapted from Newcomer et al. (2015)

A *summative evaluation* measures and analyses the outcomes and impacts of a program, while a *formative evaluation* intends to analyze the way in which a program can be improved (Trevisan & Walser, 2015). Regarding the purpose of the present study, the Ministry of Justice & Security is mostly interested in the summative aspect, as the Ministry is primarily interested in the determination of the effect of the implemented program. Though, incorporating a formative aspect is often necessary for such an evaluation (Newcomer et al., 2015).

A problem that has often been found which evaluations of a summative kind are that the surrounding environments around these are often complicated (Wholey, 2015a). Large federal programs in the 70s and 80s were evaluated using complex summative evaluations. Often, those evaluating the programs faced difficulties as they discovered that most programs had vague strategies, goals or objectives, and stakeholders had different perceptions of the intended purposes. One of the reasons for this is that most programs or policies are part of an incremental decision-making process. The objectives stated in policy are often vague, ambiguous and have inadequate relevance or connection to the overall program. This made these programs unsuitable for evaluation in their current form (Smith, 1990). Stakeholders were often found to be supportive of evaluation on an abstract level, but not to the actual assessment of specific programs due to vested interests (Finckenauer et al., 2005; Trevisan & Walser, 2015). In the end, these summative evaluations have been considered to be expensive and time-consuming and a misuse of resources (Smith, 1990). While not delivering useful information to stakeholders on the review of the program (Trevisan & Yi Min, 2003). Exploratory evaluations are intended to alleviate these problems, by conducting small, low-cost evaluations intended to uncover the intended objectives of the program and the feasibility and usefulness of these comprehensive summative evaluations. Examples of explorative research are the *evaluability assessment* (EA), *rapid feedback evaluation* (RFE), *Evaluation synthesis*, and *Small-sample studies* (Wholey, 2015a).

Small-sample studies are used to test the intended performance measures for pre-determined objectives for the program. It does so by using smaller samples than normally used and base their evaluation findings on these for the intended measures. *Evaluation Synthesis* collects and reports on the known effectiveness of the program, based on all the research and studies previously executed. *Rapid feedback evaluation* is based on the evaluability assessment but can only start when there are pre-determined goals. RFE is a form of rapid evaluation that produces additional designs for further evaluation. Most importantly, it is a lesser researched form of explorative evaluation compared to EA (Wholey, 2015a).

The purpose of an *evaluability assessment* is to analyze to what extent a program is prepared for evaluation. It is used to decide on the realistic objectives for the program, examine the logic of the program (resources, activities, causal links), as well as evaluation criteria and ways to identify and incorporate evaluation information into a coherent evaluation (Wholey, 2015a;

Zimmerman & Holden, 2009). A worthwhile EA should answer if the program is organized in such a way that useful evaluation is possible. The usefulness of an evaluation is based on four standards as presented in Figure 3.

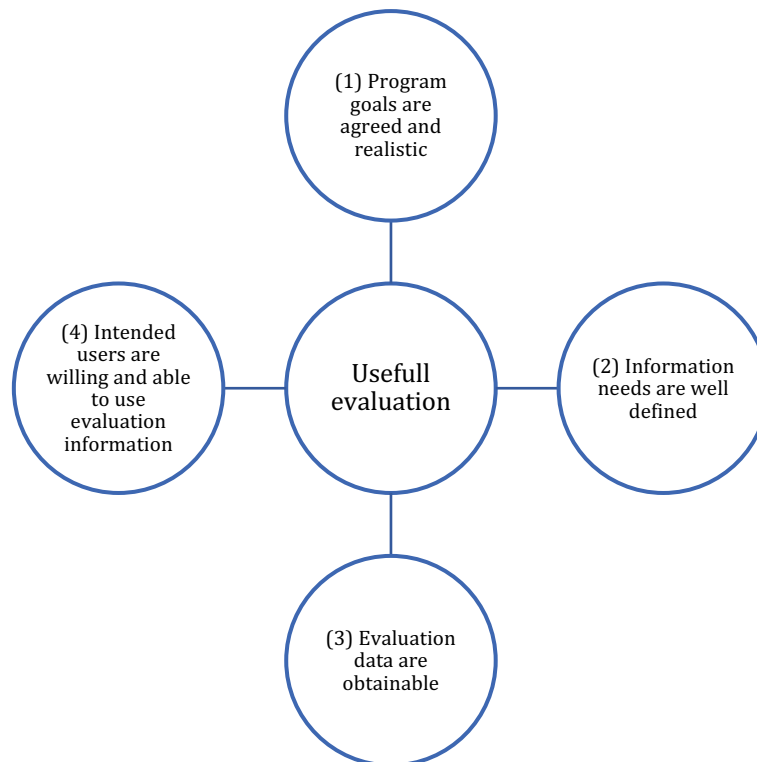


Figure 3. Four standards required for useful evaluation. Adapted from Wholey (2015a)

Evaluability assessment is designed to clarify the program design and provide advice on the necessary changes to meet these four standards. EA is especially useful for the evaluation of the program of alternative supervisory measures to administrative detention. As the program is of a broad decentralized nature, with different organizations responsible for policy and management as well as no readily apparent results or evaluation criteria (Wholey, 2015a). Furthermore, an EA can include both a summative and formative aspect. The summative and formative aspects of an EA can be found in Figure 4.

"Although EA will not provide information about the overall worth of a program (its effectiveness or allocative efficiency), it can provide a foundation in support of quality summative evaluation, increasing the likelihood that results will be useful and acted on"

•*Summative aspect of an EA*

"EA can take on a formative evaluation purpose when results are used to make recommendations for filling in gaps in alignment between program theory and program reality and for increasing program plausibility"

•*Formative aspect of an EA*

Figure 4. Summative and formative aspects of Evaluability Assessment, based on Trevisan & Walser (2015, p. 5 & 7)

2.2 Evaluability Assessment Model

The preceding parts of the chapter have explained the development of evaluation as a research field as well as the choice for the evaluability assessment approach. The Evaluability assessment approach is placed a step prior to a fully-fledged evaluation. The first framework for an evaluability assessment had been conceived by Wholey (1979) and consisted of four conditions (D'Ostie-Racine, Dagenais, & Ridde, 2013). The first condition of *program intent and logic model* describes the structure of the policy. It shows the intent of the policy and that the objectives, indicators, and priorities are clear. The second condition of the *plausibility of the program* lists the objectives which should be measured and how these are supposed to be evaluated. The third condition of the framework, the *equivalency model*, analyses the policy in its practical implementation and to uncover the feasible measures and data for evaluation. The final condition of the framework is the *evaluable model* which presents the policy or its parts of which the objectives can be measured and evaluated. This is meant to assess the sensibility of implementing an evaluation (D'Ostie-Racine et al., 2013; Smith, 1990). The original framework included an eight-step model to conduct such an EA (Trevisan & Walser, 2015). Several authors have since developed their process or systematic approach for conducting such an EA, such as Smith (1989), Trevisan & Walser (2015) and Wholey by introducing an adapted model (Wholey, 2015a). These models have all incorporated the original four conditions in some way, though these may differ in the sequence of these conditions and include different perspectives on the EA.

The model by Smith (1989) has been an extension and adaption of an earlier eight-step model by Wholey (1979). The model of Smith differs from that of Wholey (1979) as it incorporates the stakeholders of a policy as vital parts of the assessment. The EA-model of Wholey (1979) is also mostly used in preparation for summative evaluation, determining the merit a particular policy has had on its objectives (Trevisan & Walser, 2015). The model of Smith (1989) can also be used in a formative way, as it can also point out the parts of a policy that needs to be changed to close the possible difference between the program in theory and its practical implementation (Trevisan & Walser, 2015). The model by Trevisan & Walser (2015) is primarily aimed at the formative aspect of EA's and further enlarges the role of the stakeholders.

However, a newer six-step model by Wholey (2015a) would be a more appropriate fit for the present study, as it combines both a primary focus on the EA as preparation for summative evaluation as well as incorporating a formative use (See Table 1). The advantage over the other models is that it provides the right balance between the summative and formative aspects, making it more adept for the present study and its approach. The six-step model assigns a more prominent role to the involvement of the stakeholders in comparison to the original eight-step model (Wholey, 2015a). In addition, the formative aspect of the six-step model is provided further by the fifth step. Recommendations are given in that step for removing any relevant disparities between

the program design and the implemented program, aimed at enabling possible fully-fledged evaluations. The summative aspect of the six-step model is provided by the third step, which comprises of a rough estimation of the odds to which the outputs lead to the outcomes. It can thus be considered an initial form of an effectiveness evaluation (Wholey, 2015a). Another advantage of this model is that it presents a set of ‘evaluation options’ as part of the sixth step. A ‘*status quo*’ option is also offered, which explains the benefits of the evaluation by describing the consequences of maintaining the current situation without any review (Wholey, 2015a). Providing multiple options for evaluation by using this model provides a clear overview for the Ministry on the possibilities for evaluation.

Table 1 provides an overview of the six-step model in comparison to the other three models. The following section of the chapter describes the six steps of the EA-model by Wholey (2015a) more extensively.

Table 1. The six-step model in comparison to other EA-models.

Six-step model by Wholey (2015a)	<i>Eight-step model by Wholey (1979)</i>	<i>Ten-step model by Smith (1989)</i>	<i>Four-step model by Trevisan & Walser (2015)</i>
1. Involve intended users and other key stakeholders	1. Define the program to be evaluated	1. Determine purpose, secure commitment, identify workgroup members	1. Focusing the EA
2. Clarify the program design	2. Collect information on the intended program through document review and stakeholder interviews	2. Define boundaries of program under study	2. Developing an initial program theory
3. Explore program reality	3. Develop a program model	3. Identify and analyze program documents	3. Gathering feedback on program theory
4. Assess the plausibility of the program	4. Analyze the extent to which stakeholders have identified goals, objectives, activities and so forth in measurable terms	4. Develop/clarify program theory	4. Using the EA
5. Reach agreement on any needed changes in program design or program implementation	5. Collect information on program reality through site visits and document reviews	5. Identify and interview stakeholders	
6. Reach agreement on evaluation focus and intended use	6. Synthesize findings to determine the plausibility of program goals	6. Describe stakeholder perceptions of program model	
	7. Identify options for evaluation and management	7. Identify stakeholder needs, concerns and differences in perceptions of program model	
	8. Present conclusions and recommendations to management	8. Determine plausibility of program model	
		9. Draw conclusions and make recommendations	
		10. Plan specific steps for utilization of EA data	

Note. Six-step model by Wholey (2015a). Eight-step model by Wholey (1979), adapted from Trevisan & Walser (2015). Ten-step model by Smith (1989), adapted from Trevisan & Yi Min (2003). Four-step model by Trevisan & Walser (2015).

2.3 Six-step EA-Model

The six-step model by Wholey (2015a) begins with the first step to *'Involve intended users and other key stakeholders.'* The first step aims to increase the commitment of those involved. Relevant documentation is to be reviewed, and the evaluator meets with the working group to identify possible stakeholders. Such documents include authorization legislation, regulations & guidelines, subsidy applications, internal memoranda, progress reports, evaluative studies, monitoring reports, and audit reports (Trevisan & Yi Min, 2003; Wholey, 2015a; Zimmerman & Holden, 2009).

The second step to *'Clarify the program design'* is meant to examine the components of the program and its design more extensively. The information from initial talks with the workgroup, stakeholders and the information from relevant documents are combined to construct a program theory & logic model (Wholey, 2015a). The development of a program theory & logic model is a vital aspect of an EA, which is further explained in the following part of this section (see Chapter 2.3.1).

After the second step of the model, the third step of *'explore program reality'* aims to compare the constructed program theory & logic model with the actual implementation. The comparison is most often made by using the findings from the previous document analysis in combination with interviews and site visits. Different perceptions of the inner workings of the program, objectives, problems, and needs for evaluation are described and analyzed. Subsequently, such a comparison offers an improved insight into the program itself and any discrepancies between the intended and actual policy program that are relevant for further evaluation (Trevisan & Yi Min, 2003).

The fourth step of *'assess the plausibility of the program'* combines all of the information from the previous steps, producing a preliminary indication on the results from the program and if these lead to the intended outcomes of the program. The resulting analysis leads to step five and step six of the model.

Step five of *'reach agreement on any needed changes in program design or program implementation'* consists of a final determination of the program design and implementation. Some intended components could be removed or adjusted if the previous step determined these to be of less importance. Additionally, necessary changes can be identified that could improve the implementation if necessary (Wholey, 2015a).

The sixth step of *'reach agreement on evaluation focus and intended use'* presents several evaluation options containing a description based on four aspects. As previously stated, a status quo option is also included, which describes the consequences when no further evaluation is implemented. This option is added to activate the involved stakeholders by showing the benefits of further evaluations (Wholey, 2015a).

Considerations regarding the use of an evaluability assessment

Every methodological approach has its limitations and potential weaknesses, and this also applies to the evaluability assessment approach. Trevisan & Yi Min (2003) have mentioned two issues that could interfere with the use of an EA. The first issue states the importance of having a well-functioning work group. The work group should consist of representative members from each of the relevant stakeholders. Team building and the monitoring of the cooperation between members of the group is essential for a successful EA. The second issue has been the risk of a time-consuming process by using an EA. Detailed time management, supervision, and task management have been considered necessary to limit that risk (Trevisan & Yi Min, 2003).

Besides controlling the time it takes for conducting an EA, Hare & Guetterman (2014) also warned that organizational and material support are essential for conducting a successful EA. The EA needs that support if it wants to achieve actual change. Finding and involving relevant members of the group performing the EA and stakeholders for interviews is important to gather enough organizational and material support. Another consideration that was mentioned by Hare & Guetterman (2014) involved the availability of the necessary data. An EA needs accurate, complete, and consistent data. The data also needs to be accessible, relevant, and it should be possible to integrate information from different systems to meet the contextual data considerations (Hare & Guetterman, 2014).

A final aspect to consider is the “reality” of the program. The EA attempts to compare the formulated program theory & logic model with “reality” by interviewing stakeholders. The assumptions of how the program was intended to work should not diverge too far from the policy program as implemented in practice. Furthermore, there should not be too much difference between perceptions on the intended policy program and the actual program between the interviewed stakeholders. Monitoring that “implementation integrity” is important, and those conducting the EA should request the assistance of the relevant decision-makers as to how much room between the differences in perceptions is tolerated for the EA (Smith, 1990).

2.3.1 Program theory & logic model

The development of a program theory & logic model is an essential aspect of an EA, as was mentioned in the previous part of this section. Program theories and logic models have seen increased usage in recent years, due to its usefulness for researchers, program managers and others for conceptualizing, planning, communicating and framing of evaluation reports (McLaughlin & Jordan, 2015). The causal links, mechanisms, and assumptions between the components of the program and their relation to the intended outcomes form the program theory (McLaughlin & Jordan, 2015). A program theory can be considered as akin to a hypothesis, meaning that when the policy program is implemented, it should follow the causal links of these components towards the intended outcomes. A logic model is a graphical representation of that program theory, presenting a clear and simplified representation of the program theory. Several approaches have been developed for formulating a program theory & logic model. The article by Leeuw (2003) distinguishes between a '*policy-scientific*' approach, '*strategic assessment*' approach and an '*elicitation*' approach. The policy-scientific approach is regarded as the most empirical approach. As it tests the assumptions of how the program is intended to function through the analysis of relevant scientific studies, stakeholder-interviews, and documents (McLaughlin & Jordan, 2015). The other two approaches revolve around the perceptions of the staff and are meant to promote debate, which is why the present study adheres to the policy-scientific approach.

Program theory is not the same as a traditional theory. A program theory is not based on some larger theoretical hypothesis. Rather, the program theory follows a more literal interpretation: *"Among the dictionary definitions of theory however we find 'a particular conception or view of something to be done or the method of doing it'"* (Rossi et al., 2004). Program theory involves the assumptions of how the program and its components work and interrelate to lead to the expected outcomes.

There are several advantages to having a program theory developed. First, the data from the different sources used are collected and presented by the components, so the different sources per component can be analyzed for their consistency (Cooksy, Gill, & Kelly, 2001). A program theory provides a framework for the triangulated collection of data collected from theory, documents, and interviews. Each component of a policy within a program theory is also analyzed in relation to the other components. Another advantage is that the program theory assists in the data collection and performance measurement by finding any obstacles for evaluation (McLaughlin & Jordan, 2015). Additionally, it is also used to improve the program by analyzing which parts of the objectives are critical, irrelevant or have incorrect assumptions on how the program components interact. Finally, it also aides in the communication of the program by

providing a generally accepted basis of the components and the program as a whole (McLaughlin & Jordan, 2015).

The resources, activities, outputs, and outcomes of the policy program are regarded as the basic components of any program theory & logic model (Cooksy et al., 2001; McLaughlin & Jordan, 2015). The outcomes range from short-term, to intermediate, and to long-term outcomes. Moreover, additional elements can be added, such as the contextual influences. Figure 5 gives an example of the components of the program theory, based on a generic logic model by McLaughlin & Jordan (2015). The following part of this section examines these components further.

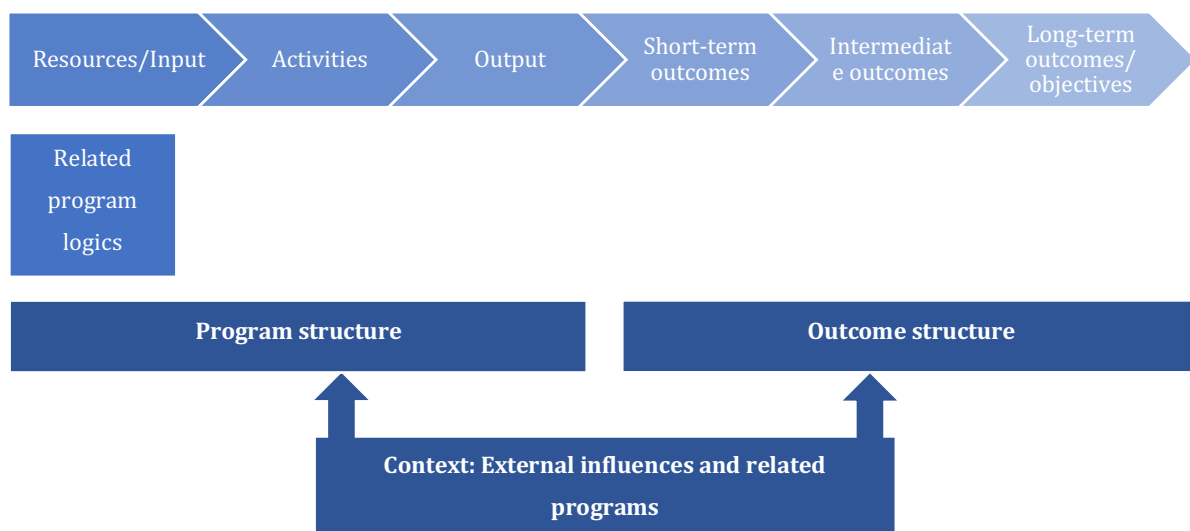


Figure 5. Generic Logic Model. Adapted from McLaughlin & Jordan (2015).

Components of a program theory & logic model

The *context* of a logic model contains any external (positive or negative) influences on the program outside of the control of the program. There are two forms of influences. Antecedent influences are the basic characteristics that are present during the start of the program. Economic and geographical variables are examples of these kinds of influences. Mediating influences are those that become apparent during the implementation of the program, such as political changes and the start of other relevant programs. Identifying the context clarifies what possible external influences could affect the outcomes (McLaughlin & Jordan, 2015).

The *rationale* consists of a description of the general intention and goals behind the need to implement the program (Centers for Disease Control and Prevention, 1999; Rossi et al., 2004). These would be considered part of the related program logics by McLaughlin & Jordan (2015), though these are not mentioned explicitly in that model.

The *input* describes the necessary resources to support the program. A crucial aspect of an EA for this component is to analyze if and where a unbalance may appear between the expected

input and the resources used in practice (Trevisan & Walser, 2015). Trevisan & Walser (2015) mention the component as the input of the program, stating that it consists of all the resources used to carry out the activities. Both McLaughlin & Jordan (2015), as well as Rossi et al. (2004), consider these elements as part of the resource of the program. This study uses the term of input from Trevisan & Walser (2015) to distinguish the component (input) from its parts (the resources).

The *activities* include those services that the program implements to produce the necessary program outputs (McLaughlin & Jordan, 2015). The activities should instinctively lead to the outputs and outcomes (Trevisan & Walser, 2015). Additionally, describing the activities and their interrelationship shows how these interact with each other and the effect the interaction has on the outputs and outcome (Centers for Disease Control and Prevention, 1999).

The *outputs* contain the resulting products provided by the activities. The difference with the activities is that the outputs are what the program produces, while the activities are what the program does to achieve the production (McLaughlin & Jordan, 2015). Most often, these are regarded as numerical outputs, though other outputs are also possible (Trevisan & Walser, 2015).

The *outcomes* are the effects of the activities and outputs that should be present if the program is implemented successfully (Centers for Disease Control and Prevention, 1999; McLaughlin & Jordan, 2015). The outcomes are divided, based on the effect these have over time, with short-term, intermediate, and long-term outcomes. The short-term outcomes are the immediate effects as a result of the activities. Intermediate outcomes are those that follow from these short-term effects.

The *objectives* represent the outcomes that are the long-term result of the activities, output, and outcomes (Zimmerman & Holden, 2009). Or, to put it differently, the desired outcomes that follow from the benefits of the intermediate outcomes to the extent that the objectives are achieved (McLaughlin & Jordan, 2015; Zimmerman & Holden, 2009). These long-term outcomes (or impact) consist of the Objectives of the program (Centers for Disease Control and Prevention, 1999; Rossi et al., 2004). These are related to the goals from the rationale, though these illustrate the precise long-term outcomes of the policy program (Rossi et al., 2004).

Consideration regarding the formulation of a program theory & logic model

As already mentioned in the introduction to the program theory, the formulation of the program theory has several advantages for an EA to use. There are, however, also some potential dangers to the use of a program theory & logic model that need to be addressed. Forming a working group is recommended, as it is important to receive input on its development. The theory and model should use clear and understandable language and avoid too much use of jargon. The program theory & logic model are part of an iterative process, as it could be adapted when new information

makes it necessary. However, this could lead to a lot of time being spent into its development. A program theory & logic model should, therefore, be considered as a general concept of the program. In which not all things *can* or *should* be described (McLaughlin & Jordan, 2015).

2.4 Summary of the Theoretical Chapter

In summary, the theoretical chapter started by examining the evaluation research field. The evaluability assessment approach is shown to be a viable choice for the present study. The six-step EA-model by Wholey provides the necessary balance between the summative and formative aspects, which is why that model was examined in more detail. Additionally, the critical part of formulating a program theory and logic model is explained, as it is meant to provide the necessary information for the subsequent steps. Lastly, a review of the weaknesses of the EA ensures a critical consideration of the potential of the approach. The following chapter describes the application of the EA for the present study, as well as the data process.

3. Methodology

The methodological chapter of the study describes the way the aggregation of the data and analysis were conducted. This includes: the research design (Chapter 3.1), the population & selection (Chapter 3.2), the data collection (Chapter 3.3), the data analysis (Chapter 3.4), the procedure of the study (Chapter 3.5), the reliability & validity (Chapter 3.6), and a summary (Chapter 3.7).

3.1 Research Design

The research design of this study was directed towards a qualitative case-study on the adapted policy program of supervisory measures in the returns process of the Netherlands. The theoretical framework gave an evaluative approach to analyze if the Directorate-General for Migration is ready to evaluate the policy program. As mentioned in that chapter, the first EA-model included four aspects (see Chapter 2.2). This study incorporates these four aspects in an adapted form with five aspects, as it uses a newer EA-model by Wholey (2015a).

As such, the first aspect of this research comprises of the *program theory*. The program theory of the study describes the structure of the policy program. It identifies the intent of the policy and its components and provides a graphical representation of these.

The second aspect of *program reality* incorporates the original equivalency model, for analyzing the policy program in its practical implementation in comparison to the program theory. Such a comparison should result in the necessary information on any issues that would inhibit effective program performance, the identification of feasible measures, and what data is needed to evaluate the policy program (D'Ostie-Racine et al., 2013). These two aspects, therefore, provide the basis for the subsequent discussion of the other aspects.

The third 'summative' aspect of *program plausibility and data availability* of the present study gives a preliminary indication on the odds that the activities lead to the outputs and eventually, that the outputs lead to the intended outcomes (Wholey, 2015a). Consequently, this aspect comprises of the provision the relevant output, outcomes, and objectives as made possible by the comparison of the previous two aspects. In addition, this aspect includes a determination of the data availability as an addition to the preliminary indication.

A fourth 'formative' aspect of *program utility* is part of the original evaluable model and is meant to assess the sensibility of implementing an evaluation (D'Ostie-Racine et al., 2013; Smith, 1990). The aspect determines if the policy program is ready to be evaluated and to provide recommendations for any changes to the policy program which enable further evaluation if necessary.

The final aspect of *program evaluation* is meant to produce several evaluation options. By providing these options, involved stakeholders are motivated and encouraged to the use of the information of this study for further evaluation (Wholey, 2015a).

The following section of the methodology explains the necessary data process for getting the required information for these aspects. While the step-for-step procedure of the study to analyze these aspects was based on the six-step EA-model by Wholey (2015a), as described in the section after the data analysis section (see Chapter 3.3).

3.2 Population & Selection

Data from two different types of sources were gathered to provide the necessary information for the present study, relevant documents, and stakeholders. The research student has asked those in the working group and stakeholders for documents that describe the intended program. The documents were scanned using the *snowball sampling method* for further documents (D'Ostie-Racine et al., 2013). Documents from related programs, scientific research and NGO's were also collected and reviewed to get the information from multiple sources (McLaughlin & Jordan, 2015).

The identification and interviewing of stakeholders of step five began by consulting the workgroup members for relevant stakeholders using the *intensity sampling method*. Relevant stakeholders were selected based on their thorough understanding of the subject (Patton, 2002). These stakeholders were then asked for other stakeholders who could have the potential to provide additional information as is characteristic for using the *snowball sampling method* (D'Ostie-Racine et al., 2013). The resulting selection is presented in Table 2 (anonymized to protect the privacy of the respondents).

Table 2. Interview respondents list

<i>Organization</i>	<i>Respondent</i>	<i>Function</i>	<i>Purpose for study</i>
Ministry of Justice and Security	JV1	Policy advisor	Responsible for ATD policy
	JV2	Policy data analyst	Responsible for data collection, monitoring, and analysis
	INS1	Legal advisor INS (Immigration and Naturalization Service)	Legal advisor and knowledge of background contextual information
National Police	NP1	Operational specialist	Application of ATD measures, specifically the duty to report measure
R&DS (Repatriation and Departure Service)	RDS1	Data analyst	Data collection, monitoring, and analysis, specifically the freedom-restricting and family location measures
	RDS2, RDS3	Two senior departure supervisors	Application of ATD measures, specifically the freedom-restricting and family location measures
Royal Netherlands Marechaussee	KM1	Data analyst	Data collection, monitoring, and analysis, specifically the taking of documents measure
Amnesty International	AI1	Senior employee Human Rights Netherlands/Migration	Outside perspective on implementation and application of ATD measures

Most of the respondents were divided into those responsible in their organization for the application of the ATD measures that fall under that organization and those responsible for data collection, monitoring, and analysis. That choice was made to present different perspectives on the program within each organization, both from the program structure and outcome structure (McLaughlin & Jordan, 2015). Respondents from the Ministry of Justice & Security also contain these two perspectives, as well as a respondent that provided contextual information. A respondent of Amnesty International was selected to give an essential perspective from outside of the government organizations, thereby also limiting the potential for a biased interpretation (Goodrick & Rogers, 2015).

3.3 Data Collection

The necessary data from the documents were gathered with the use of the qualitative documentation and analysis software ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, 2016). Data from the respondents was collected with the use of semi-structured interview (SSI). An SSI uses both closed-ended and open-ended questions in interviews and is an intermediate form between surveys and open interviews (Adams, 2015). These questions can then be followed up with why or how questions. A disadvantage of SSI's is that it can be time-consuming in preparation, administering, and analysis. Another issue is the difficulty of drawing a representative sample of respondents out of the population due to the time-consuming nature of conducting interviews. One of the reasons for the use of SSI's as a means to gather information from the respondents was that it has been mentioned explicitly as a usable way to obtain information from key stakeholders (Adams, 2015). SSI's make it possible to formulate broad questions to guide the interview but leave enough flexibility to ask follow-up questions and to have further discussions (Trevisan & Walser, 2015).

Interview guides were formulated for each of the respondents. These guides were made, so each respondent has a custom interview guide while sharing a general basis (Trevisan & Walser, 2015). The interview guides used the basic questions for EA as described by Trevisan & Walser (2015) as a starting point. An example of an interview guide can be found in the Appendix (see Appendix A). The EA questions that have been considered as the most important for an EA are the program perspectives, context, implementation, research logic and methodological scoping. These are combined with relevant documents and the program theory to build interview guides. These are directed at the perceptions and knowledge of the stakeholders of the program (Finckenauer, Margaryan, & Sullivan, 2005). Several questions were also based on the study by Bloomfield (2016). Such as questions about the policy resulting in a reduction of the detained number of migrants, if the ATD measure assists in the obligation of the third-country to leave the country, the cost-effectiveness of the ATD measure, and the intent of the state by implementing the ATD measure (Bloomfield, 2016). The interview guides received feedback from the working group, as well as a reassessment after the first interview and further continuous assessment (Adams, 2015). Any modification to the guides while conducting the interviews are mentioned in the following '*Experience of administering SSI's*' part of the current section.

The interviews were recorded to collecting the information for further transcriptions. The advantage to using a recorder is that every detail of the conversation is registered and the interviewer can remain centered on the questions at hand (Adams, 2015). Notes were also made, to highlight important parts of the interview or any behavior that cannot be collected using a recorder. Additional relevant information from e-mails or feedback on the transcriptions by the respondents were included in the transcription. Every respondent has been anonymized to

protect their privacy. Each respondent was asked before interviewing if they would allow the recording of the interview.

The interviews were conducted by an interviewer with experience in conducting SSI's. An important technique used to increase the validity of the interview was the use of '*member-checking*' or otherwise known as '*active listening*' techniques. These techniques involve listening, summarizing and repeating the viewpoint or answer of the respondent to check if the interviewer understood the answer correctly (Adams, 2015; D'Ostie-Racine et al., 2013).

Experience of conducting SSI's

The following section describes any relevant situations or consequences from the process of conducting the interviews. As mentioned in the section above, a reassessment of the interview guides was done after the first interview. A change was made to include sub-questions for the program components on the situation before the program and what aspects were changed by the implementation of the program. Furthermore, some questions were rephrased, and concepts were explained or better defined.

There were several instances in which the interview had to be briefly interrupted. The interviewer and respondent agreed to pause the interview whenever the respondent felt it necessary to answer in each of these cases. The recording was then paused and resumed when the respondent indicated that it could be resumed. It was left to the respondent to choose a suitable location for the interview if it took place in their workplace, as long as the room was quiet enough to reduce any interfering noise for the recording and did not result in any distractions for the interviewer and respondent.

One notable problem occurred after a brief interruption of the interview when a respondent received a phone call. The recording of the second part of the interview, after the interruption, was lost due to unknown problems. However, the interviewer had written down some field notes as with all the interviews. As the problem was found immediately after the interview, the interviewer was able to make a more extensive version of these field notes, with the '*inscription and transcription*' strategy for writing field notes (Gibbs, 2007). The conversation and key points of the interview were described and presented to the interviewer after that for confirmation. That information was included while recognizing that these field notes are to be regarded as secondary data. Any further interpretation was thus cautiously.

3.4 Data Analysis

The necessary data from the documents were analyzed with the use of the qualitative documentation and analysis software ATLAS.ti version 8.0 (ATLAS.ti Scientific Software Development GmbH, 2016), which provided a database of relevant information for the program policy from the gathered documents.

The collected recordings from the interviews were transcribed using the principles for transcription of Edwards & Lampert (1993) and a thematic content analysis approach (D'Ostie-Racine et al., 2013). Recordings were transcribed verbatim, writing down everything necessary for understanding the conversation. A balance was found between the principle of *authenticity* and *practicality*. That means that some adaptations for improving readability were made while remaining as close as possible to the recording (Bloor & Wood, 2006). The transcriptions were then further analyzed by categorizing the transcriptions with the use of *coding*. Both *descriptive (first-level)* coding, as well as *Pattern (second level)* coding, was used to categorize the data (Goodrick & Rogers, 2015).

Descriptive coding was used to explicitly label segments of text based on categories developed both before and during the coding. As such a mixed codebook has been formed consisting of both *concept-driven* coding and *data-driven* coding (Gibbs, 2007). The concept-driven coding process includes codes based on scientific research and relevant government documents that were gathered beforehand (D'Ostie-Racine et al., 2013). Categories developed before the coding were mainly based on the elements used in the program theory & logic model and the domains for the research questions by Trevisan & Walser (2015). The data-driven coding process includes any codes that are relevant for the analysis based on the data (Gibbs, 2007). Transcripts were coded serially (one at a time), and constant checks were made for overlapping, redundant or similar codes. All codes were either listed *in vivo*, as content paraphrases or based on definitions from scientific research (Goodrick & Rogers, 2015). Typically, the coding should be done by multiple coders to ensure *intercoder-reliability*. The present study only used one coder due to time and labor constraints. The working group checked the resulting codes and preliminary analyses to ensure the validity of the coding (D'Ostie-Racine et al., 2013).

Pattern coding was then used to compare and distinguish relationships, patterns, and differences between the perspectives of the stakeholders on the program theory & logic model (Adams, 2015). The transcripts were collected and coded with the use of the qualitative documentation and analysis software ATLAS.ti version 8.0 (ATLAS.ti Scientific Software Development GmbH, 2016). The resulting analysis can be found in the program reality section of the results (see Chapter 4.2). An overview of the mixed codebook can be found in the Appendix (see Appendix B). The codebook also includes the descriptions for each of the codes and the coding hierarchy (Gibbs, 2007).

3.5 Procedures

The following section describes the methodological process of the present study. Figure 6 provides an overview, based on the of the application the six-step model by Wholey (2015a). The subsequent parts of this section give a detailed description of each of these steps.

Step 1	•Involve intended users and other key stakeholders
Step 2	•Clarify the program design
Step 3	•Explore program reality
Step 4	•Assess the plausibility of the program
Step 5	•Reach agreement on any needed changes in program design or program implementation
Step 6	•Reach agreement on evaluation focus and intended use

Figure 6. Six-step EA-model. Adapted from Wholey (2015a).

3.5.1 Step 1: Involve intended users and other key stakeholders

The purpose of the first step was to find relevant stakeholders. The first step also incorporated some aspect of the EA-model by Smith (Smith, 1989), with the formation of a group for conducting the EA, as well as forming a basic understanding of the objectives and boundaries for the EA and the gathering of relevant documents. The formation of the work-group and other relevant stakeholders aims to increase the commitment of those involved (Trevisan & Yi Min, 2003). The aim of the EA had been formulated, and a workgroup had been formed consisting of three members from the Directorate-General for Migration and one research-student. Based on the considerations on EA from the theory, the three members from the DGM represented the different areas of expertise within the organization that were relevant for this study (see Chapter 2.3). The working group provided a guiding role and gave feedback on the progress made by the research student of the Radboud University. In addition, the working group also aided in setting up the boundaries and objectives of the study.

3.5.2 Step 2: Clarify the program design

The second step of the model by Wholey (2015a) consisted of gathering the necessary information to clarify the program design. Key documents and consultations with the working group provided an outline of the intended program, which provided the basis for uncovering the intent and objectives of the policy program of the study (see also Chapter 3.3). That information was used to formulate the program theory & logic model (see also Chapter 2.3.1).

Program theory and logic model of the study

As mentioned in the theoretical framework, the program theory describes the assumptions on how the program and its components work and interrelate to lead to the expected outcomes (Rossi, Lipsey, & Freeman, 2004). A logic model is a graphical representation of that program theory (see also Chapter 2.3.1, for further explanation of the program theory & logic model). The program theory & logic model for this study have been constructed following the five steps as described by McLaughlin & Jordan (2015) (see Figure 7).

The program theory of the study was described as to provide a reasonably accurate overview of the policy, as the considerations for a program theory include that a complete and detailed description of a program is often difficult to achieve and unnecessary for an EA (Trevisan & Yi Min, 2003).

As there is no rigid standard for which components it should contain, the present study included aspects from several descriptions of program components best suitable for the policy program under consideration. Some basic components are found in every program theory or logic model, while other features differ per policy program. Basic elements that were included were the *resources*, *activities*, *outputs* and, *outcomes* as described by McLaughlin & Jordan (2015) and Trevisan & Walser (2015). In addition, these basic components were adapted for this study and other components were added. What follows is an account of the adaptations and choices that were made for the program theory & logic model of the present study.

1. Collecting relevant information
2. Clearly defining the problem that the program intends to solve and its context
3. Defining the elements of the model in a table
4. Drawing the model
5. Verifying the model with stakeholders

Figure 7. Five steps for constructing a logic model. Adopted from McLaughlin & Jordan (2015)

Adaptations and choices made for the program theory and logic model

Certain choices were made during the formulation of the program theory and construction of a logic model. A framework for the evaluation of Assisted Voluntary Return Programs by the European Migration Network (2016) functioned as the basic outline for the logic model, adapting it further to the specific details of the current program under analyses. Contextual factors were analyzed as well as the objectives and outcomes of the program as part of the second step. The

formulation was done using *forward and backward mapping* as well as feedback from the working group on the accuracy of the formation of the program theory & logic model (McLaughlin & Jordan, 2015).

The component of *related program logics* has been mentioned by McLaughlin & Jordan (2015) while not expanding that element further. The present study added the component of *rationale* to replace it.

The Centers for Disease Control and Prevention (1999) framework for program evaluation highlights the importance of the inclusion of the mission and goals of the program, for which *the rationale* was used. The significance of the formation of the goals have also been mentioned in step two of McLaughlin & Jordan (2015) but is not explicitly included in the related program logics part of its underlying model. As the framework for the evaluation of AVR(R) by the European Migration Network (2016) also includes these parts, this study incorporates this component of *the rationale* in the program description and logic model.

The theoretical chapter also described the differences in the interpretation of the *resources* or *input* element (see Chapter 2.3.1). The present study incorporated the more intuitive feeling interpretation by Trevisan & Walser (2015). Therefore, the element of *input* is included, which consists of the different resources.

Lastly, the present study divided the *outcomes* in those on a short to intermediate category and a long-term outcome component of *objectives*. This division was made to differentiate more between the outcomes that follow from the activities and outputs and the expected long-term outcomes or objectives' (Zimmerman & Holden, 2009).

Table 3 on the next page lists the components that were used to describe the program theory and construct a logic model for the present study. The resulting program theory & logic model of the second step is described in the first section of the results (see Chapter 4.1).

Table 3. Components used to describe the program and elements of the program theory & logic model

<i>Part of:</i>	<i>Component:</i>	<i>Description of component:</i>	<i>Source:</i>
Contextual factors	Antecedent influences	The basic characteristics that are present during the start of the policy program	McLaughlin & Jordan (2015)
	Mediating influences	Influences that become apparent during the implementation of the policy program	McLaughlin & Jordan (2015)
Related program logics	Rationale	The intention and goals behind the need to implement the policy program	Centers for Disease Control and Prevention (1999)
Program structure	Input	The necessary resources to support the program	Trevisan & Walser (2015)
	Activities	The services that the program implements to produce the necessary program outputs	McLaughlin & Jordan (2015)
	Outputs	The products of the services provided by the activities	McLaughlin & Jordan (2015)
Outcome structure	Outcomes	The expected short-term and intermediate effects as a result of the Activities and outputs	McLaughlin & Jordan (2015)
	Objectives	The extent to which the activities and outputs have achieved the long-term change as intended for the policy program	Centers for Disease Control and Prevention (1999); Zimmerman & Holden (2009)

3.5.3 Step 3: Explore program reality

The resulting program theory & logic model from the second step functioned as a kind of hypothesis, showing how the program is intended to work. Step Three of the EA-model by Wholey (2015a) consisted of the other part of the comparative results, by comparing the expected program with its implementation in practice. Interviews were conducted with relevant stakeholders (see also Chapter 3.3). Information obtained from the interview data was analyzed and compared with the formulated program theory & logic model. Relevant differences in the intended use and implementation were described in the second part of the results (see also Chapter 4.2).

3.5.4 Step 4: Assess the plausibility of the program

The fourth ‘summative’ step of the EA-model was used to determine the plausibility of the program. That included a determination of any preliminary findings on the outcomes of the policy, the key components of the policy and what data is available for further research. As stated by Wholey (2015a, p. 87):

With what has been learned in steps 1 to 3, the evaluator now may be able to make rough estimates of the likelihood that intended outputs (products or services) will be delivered to intended recipients and the likelihood that intended outcomes will occur—thus producing a preliminary evaluation of program effectiveness.

The plausibility of the program thus clarifies to what extent the outputs and outcomes as intended have indeed occurred in practice and therefore tries to provide a preliminary assessment of the odds that the components of the policy program lead to the objectives. As such, this step combines the data from the previous steps, in addition to further interview data for this aspect of the study (see also Chapter 3.1). The first section of the discussion presents the findings from the fourth ‘summative’ step (see also Chapter 5).

3.5.5 Step 5: Reach agreement on any needed changes

The fifth ‘formative’ step involved the formulation of recommendations for any changes to the implemented policy program. These were based on the information retrieved from the first four steps and includes the relevant documents, scientific research, and interview data in addition to further interview data for this aspect of the study (see also Chapter 3.1). It was used to determine the readiness of the adapted policy program, and if not, what changes can be recommended within the current program that possible (Wholey, 2015a). The second section of the discussion describes the findings from this ‘formative’ step (see also Chapter 5).

3.5.6 Step 6: Reach agreement on evaluation focus and intended use

The sixth step included the development of a plan for further evaluation, based on the information from the aspects of the result (step 2 & 3) and the formative and summative aspects of the discussion (step 4 & 5). Evaluation options for the assessment of the program and the specific outcomes were made. The elements of which were based on those formulated by Wholey (2015a), as described in the theoretical chapter (see Chapter 2.3). These were divided into the data that needs to be collected, the data analysis method most appropriate, an estimate of the costs (time, labor, financial), and how the evaluation should be used (Wholey, 2015a). In addition, the present study included two more aspects. The data availability aspect was covered as to provide an overview of the data that is already available, and the additional data that needs to be gathered. The usability aspect was covered as to include a brief estimation of the evaluation option. Estimates of the *priority* of the option, the *complexity* of implementing the evaluation, and the number of *resources* needed to implement the option were included. Furthermore, the purpose of the evaluation has been moved to the first aspect of each option to be discussed. Table 4 on the next page provides an overview of the included content in such a proposal. The third section of the discussion gives an overview of the resulting evaluation options (see also Chapter 5.3).

Table 4. Included elements of the evaluation options

Aspect	Content
(1) Purpose	How the information from the option can be used
(2) Data collection	What data needs to be collected on the relevant components of the programs
(3) Data availability	What data is already available
(4) Data analysis	What kind of analysis should be conducted
(5) Costs	The level of resources required for the option: financial, time, political, bureaucratic
(6) Usability	Estimation of the usability of the option

Note. Adapted from Wholey (2015a)

3.6 Ensuring Reliability & Validity

The previous sections of the methodology chapter described how each of the sub-questions was answered in the study. The following section clarifies how the methodology ensured the reliability and validity of the study. More importantly, several issues were recognized that could have influenced the quality of the study. These are also presented together with the choices made to counteract these issues.

Newcomer, Hatry, & Wholey (2015) have described six requirements essential for upholding the methodological quality of evaluative research designs. Figure 8 presents these six elements. With the credibility of the design increasing with every requirement towards the top.

Measurement validity entails if the approaches for measuring the EA are accurate enough. The continuous feedback by the working group ensured the ‘*face validity*’ of the study (Newcomer et al., 2015). The ‘*reliability*’ and ‘*validity*’

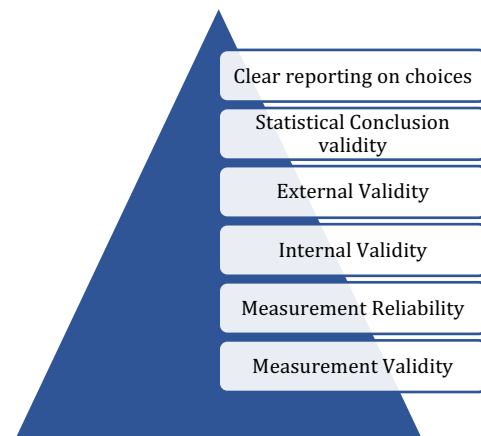


Figure 8. Methodological requirements. Adopted from Newcomer et al. (2015)

of the measurements means the extent to which another study using these measures on the same conditions gets similar results. This study ensured the reliability of the interviews by using interview guides. These were continually reassessed, especially after the first interview, for any inconsistencies or necessary modifications. Furthermore, the program theory, logic model, and codebook were all checked with the working group (Goodrick & Rogers, 2015). Still, an essential part of the requirement was that there should be consistency between the coders when coding

the same sample. The study only used one coder, resulting in some questions on the reliability of the coding. Validating the first results with the working group was used to counteract some of the limitations of having only one coder (D'Ostie-Racine et al., 2013). In addition, the transcriptions were sent to each of the respondents for a final control, and any additions or adaptations were marked and incorporated in the results (Gibbs, 2007).

The '*internal validity*' of evaluation research directs itself to whether the study has been able to determine if the program has led to an outcome. The present study incorporated the program theory & logic model and compared these with stakeholder interviews. Documents and scientific research related to the evaluation of previous/similar programs were collected to ensure a '*triangulation*' of sources to further improve the internal validity (D'Ostie-Racine et al., 2013). A respondent from an outside organization was also included to include the analysis of a "negative" case as to compare the findings from the other interviews within the present study. The purpose is to enhance the quality of the analysis, by providing another perspective and thereby also limiting the risk of a bias from solely collecting the positions from government agencies (Goodrick & Rogers, 2015).

The '*external validity*' or '*generalizability*' of an evaluation concerns the extent to which (similar) research can use the results of this study. Newcomer et al., (2015) advocate the analyses be conducted with a selection of the sites and respondent's representative of the population. The study used a case study on the implementation of ATD measures within the Ministry of Justice & Security of the Netherlands. The representativeness of the findings of the study for other countries could thus be limited. The present study recognizes that limitation by expressing the limitations for the generalizability of the findings in the main limitations of the conclusions (see also Chapter 6.2).

A final consideration for the study was the (unintended) bias towards quantitative data from the Ministry of Justice & Security. Just as in other governments, the Ministry mostly uses statistical data for summative evaluation. This study accounted for that tendency, while also refrain from limiting itself to more qualitative approaches for further evaluation. Especially if previous or similar studies point out that these could be more viable to evaluate certain outcomes. The desire for the examination and incorporation of other approaches in the study also explicitly expressed by the DGM during the first step of the EA (see also Chapter 3.5.1).

3.7 Summary of the methodology

In summary, the methodology started by presenting its research design consisting of five aspects. The subsequent sections described the data process, from the selection to the analysis. The procedures of the study were explained after, which followed the six-step model from the theoretical framework. The last section of the chapter discussed the reliability & validity of the study.

The five aspects of the research design were related back to the empirical research questions of this study and were answered by step two to step six. Table 5 gives an overview of the methodology. It shows the links between the research questions, the aspects, and the steps of the study. Additionally, the table also lists which data was used and where the results of these can be found. These are distinguished by their inclusion in the following chapters and section of the results and discussion.

Table 5. Overview of the methodology

<i>research question</i>	<i>Aspect of the study</i>	<i>Step of the EA</i>	<i>Data used</i>	<i>Findings</i>	<i>Chapter</i>	<i>Section</i>
RQ 2	Program theory & logic model	Step 2	Document study: describing the intended program and its components & feedback by working group by using the qualitative software program ATLAS.ti	Description of program theory and constructed logic model	4. Results	4.1 Program Theory & Logic Model
RQ 3	Program reality	Step 3	Semi-structured interviews: the perceptions of relevant stakeholders on the inner workings of the program in practice. Analyzed by using descriptive and pattern coding	Description of the comparison between the program theory and the program as implemented in practice	4. Results	4.2 Program Reality
RQ 4	Program plausibility & data availability	Step 4	Combining the data from the literature study, document study and interview analysis	<i>Summative</i> description of the preliminary program effectiveness assessment	5. Discussion	5.1 Program Plausibility & Data Availability
RQ 5	Program utility	Step 5	Combining the data from the literature study, document study and interview analysis	<i>Formative</i> description of recommendations for changes to the policy program as to enable further evaluation	5. Discussion	5.2 Program Utility
RQ 6	Program evaluation	Step 6	Combining the data from the literature study, document study and interview analysis	Presenting several evaluation options for further evaluation of the policy program	5. Discussion	5.3 Program Evaluation

4. Results

The following chapter presents the findings of the study. The results cover the first two aspects of the program theory & logic model (Chapter 4.1) and program reality (Chapter 4.2) of the study and their related research questions (see also Chapter 3.7, Table 5).

4.1 Program Theory & Logic Model

The program theory & logic model was based on the examination of key documents and explorative conversations with stakeholders and the workgroup. This section of the chapter was the product of the second step of the EA-model by Wholey (2015a), and more information on the procedure and explanation of the program theory & logic model can be found in the theoretical and methodological parts of the study (see Chapter 2.3 & Chapter 3.5.2). The program theory gives the examination of the intended policy program and its components, describing the policy as intended to be implemented (see Chapter 4.1). The section ends with a summary of the program theory by presenting the logic model (see Chapter 4.1.8).

4.1.1 Contextual factors

The contextual factors consisted of both characteristics apparent at the start (antecedent), as well as, those appearing during the program (mediating) (McLaughlin & Jordan, 2015). As described in the methodology, only the most relevant contextual factors were included as to provide a useful overview of the policy program (see also Chapter 3.5.2).

1a. Antecedent contextual factors

Several contextual factors were obtained from the data as antecedent influences on the policy program.

Focus on cost-efficiency: The financial crisis of 2008 and the following budget cuts in 2010 put pressure on the use of administrative detention of irregular-staying migrants. External political influence showed an intention to decrease the costs associated with the use of detention as a consequence (Amnesty International, 2011).

Influence of EU legislation: A significant contextual influence has been the 2008/115/EC Return Directive (EUR-Lex, 2008).. It introduced common regulations and standards for detaining migrants and the circumstances for detention on a EU-level (Teeven, 2013). The expectation from the DGM was that this would lead to an increase in attention and oversight by the European Union on the implementation of alternative measures. As stated in the internal report on the alternative measures: *“It is expected that, more frequently, it needs to be made transparent what alternatives*

for detention are considered and why these are not applied.” (Directorate for Migration Policy, 2011). Other relevant directives that influenced the increase in attention have been the recast 2013/32/EU Asylum Procedures Directive and the 2013/33/EU Reception Conditions Directive (EUR-Lex, 2013a, 2013b).

The suicide of Alexander Dolmatov: On a national level, the suicide of Alexander Dolmatov in a detention center renewed the attention on the situation in detention centers and the emphasis on the return policy. Inquiries and evaluations of the return policy were executed as well as additional reports by organizations such as the Advisory Committee on Alien Affairs and Amnesty International provided other feedback on the use of ADM and ATD measures (Advisory Committee on Alien Affairs, 2013; Amnesty International, 2011; Teeven, 2013).

Government desire for effective return policy: The previous administration intended to improve the effectiveness of the return policy and to prevent and decrease illegal stay (Directorate for Migration Policy, 2011). The authority of the Minister of Justice & Security over the involved police services for the supervision of aliens was also strengthened by a subsequent formation of the centralized National Police.

1b. Mediating contextual factors

The following contextual factors were identified as mediating influences on the policy program since its implementation in 2013.

New government coalition reinforces aim for effective return policy & use of ADM: After the previous government, the recently formed cabinet Rutte III has further reinforced the aim for improving effective return, which should be done by increasing the capacity of the Repatriation and Departure Service to ensure actual departure. Furthermore, possibilities for forced return by for example ADM should be broadened within the legal framework of the EU (Rutte et al., 2017).

Demands for a flexible & integral approach to immigration after migration crisis of 2015: The Syrian civil war and resulting migration crisis of 2015 led to an increase in the return procedures, as a parallel rise of asylum-seekers from other countries occurred. The total cases of return had risen with 66% from 10.300 cases in 2015 to 17.090 cases in 2016 (Ministry of Justice & Security, 2016a). The new government coalition has been of the opinion that these sudden and unforeseen situations necessitate a flexible asylum system, which improves the effectiveness of taking short-term measures. And thus, an integral approach to the reception, asylum procedure, integration, and return is needed (Rutte et al., 2017).

4.1.2 Rationale

The rationale of the adapted policy program examines the intention and goals behind its implementation as part of the program theory & logic model (see Chapter 3.5.2). The document analysis found three underlying rationales that were the basis for the main rationale.

Main rationale

The main rationale of the adapted policy program, as intended, was part of the overall rationale behind the immigration and subsequent return policy of the government at the time. As stated in the letter to the House of Representatives (Teeven, 2013, p. 3): *“The coalition agreement states that a restrictive and just immigration policy also includes an active and consistent return police.”* The aim has been to promote assisted voluntary and sustainable return. Sustainable return is seen as that which: *“deters new irregular migration of the returnee, by consolidating the position of returnees in their home countries”* (European Migration Network, 2016, p. 9). With the government resorting to a forced return when it is not possible to effectuate voluntary return. The rationale behind the supervisory measures was stated as:

A process of forced return is started when an irregular-staying migrant is not willing to work on their return, despite the aim for voluntary return and the conversations with the R&DS. Several measures exist to effectuate this, among which ADM is the most far-reaching measure of removal. (Teeven, 2013, p.3)

The contextual factors indicated that several influences were relevant for the implementation of the adapted policy program (see also Chapter 4.1.1). These culminated towards the following main rationale for the adapted policy program:

More assisted voluntary return is the aim of this cabinet. Thus, I will embed the possibility to impose an alternative supervisory measure into policy, in addition to the conditions in which these can be applied for individual cases, to contribute to this. The irregular-staying migrant that is willing to work on their return can, in principle, be imposed a less-coercive measure. These measures are directed towards effectuating return of those that have yet to return voluntarily, in abidance to the proportionality and subsidiarity principles. Important to note is that it does not involve implementing the largest possible number of alternatives but to impose the most appropriate measure on the proportionality and effectiveness for different categories of irregular-staying migrants as to effectuate return. (Teeven, 2013, p. 8)

The document analysis and preliminary conversations pointed towards several underlying rationales, which were related to the contextual factors.

2a. Practical rationale

According to the article by Sampson & Mitchell (2013), a practical rationale for a supervisory measure includes its use as to meet certain practical goals for a government. One external influence has been the focus on improving the cost-efficiency of the return process in the wake of the financial crisis, as mentioned as an antecedent contextual factor (see also Chapter 4.1.1). As such, the practical rationale behind the adapted policy program was that these alternative measures are more cost-efficient to impose (Costello & Kaytaz, 2013). The internal report on alternatives for ADM stated this aim for ATD as: *“Cheaper, ADM costs between €160 million and €200 million per year in total across the cooperating organizations”* (Directorate for Migration Policy, 2011).

2b. Practical condition

Another intended goal that was identified as pertaining to the practical rationale was a condition for the adapted policy program. The primary function of any supervisory measure in the return process is to provide supervision on the irregular-staying migrant and their progress for return (Broeders & Leerkes, 2010). One of the reasons administrative detention had been the measure of choice of the return policy before the program was the doubt from the Ministry of Justice & Security that the intended ATD could achieve such a goal. Therefore, the rationale of the adapted policy program included that these alternative measures are to be imposed on the condition that these achieve the same practical policy outcome as with the use of detention (Directorate for Migration Policy, 2011). Namely, to provide the same level of effectiveness in effectuating return by maintaining oversight on the irregular-staying migrant and that the public safety and national security are not at risk. The aim of detention is then as it is originally intended, as an ‘ultimum remedium’ to provide supervision as to effectuate return.

2c. Political rationale

According to the article by Sampson & Mitchell (2013), the political rationale for ATD measures is that these serve to relieve political pressure for governments. The political rationale for the adapted policy program was to reduce the overall restrictiveness of the application of supervisory measures, as the alternative measures are considered less-coercive measures in limiting the freedom of movement of irregular-staying migrants in comparison to ADM (Broeders & Leerkes, 2010; Bruycker, Bloomfield, Tsourdi, & Petin, 2015; Mangiaracina, 2016). The antecedent factors of the suicide of Alexander Dolmatov and the influence of EU legislation, specifically the 2008/115/EC Return Directive, were identified as the driving forces behind the political rationale (see also Chapter 4.1.1). As stated in the letter to the House of Representatives: *“The*

implementation and application of the ADM have, even more so in recent months, become part of the political and public debate.” (Teeven, 2013, p. 2)

The main rationale behind the adapted policy program thus included a practical rationale, practical condition, and a political rationale (see Figure 9). It enables the choice for a supervisory measure based on an assessment of each individual case, under the practical condition that the practical policy objective of effective oversight can be maintained. The practical rationale also includes that the intended policy program will lead to a lower expenditure as alternatives are cheaper to impose. Subsequently, the political rationale is that the expected increase in the use of less-coercive measures leads to a more humanitarian treatment of those in the return process, relieving the political pressure felt by the government.

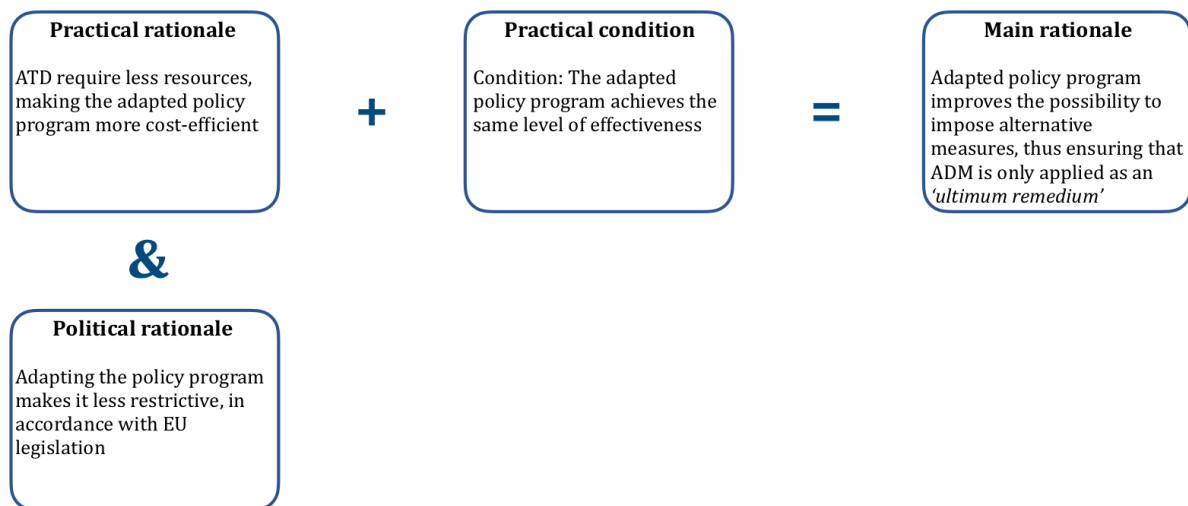


Figure 9. Overview of the rationale

4.1.3 Input

The input consisted of the resources that were needed to support the adapted policy program, distinguished by financial and non-financial inputs (see also Chapter 3.5.2). Providing the necessary resources has been one of the major obstacles for the implementation of ATD measures in existing policy of countries according to Sampson & Mitchell (2013). The following section describes the different resources incorporated into the adapted policy program.

3a. Financial resources

A lack of financial resources was one of the main obstacles identified by the internal report for further implementation of alternative measures after the pilots (Directorate for Migration Policy, 2011). However, the only specific financial resource identified from the documents has been the input of a government subsidy of €1 million per year in total for return-related projects of NGO's. As stated by Article 1 in the legislation for the subsidy scheme: *"The Minister can grant a subsidy for the benefit of activities related to supporting assisted voluntary departure for irregular-staying migrants with a duty to depart the Netherlands"* (Ministry of the Interior and Kingdom Relations, 2014).

For the other alternatives, no specific mentioning was found for the input of additional financial resources, though the budgets for each of the measures would fall under its respective organization (Teeven, 2013). Therefore, the budgets would be adjusted according to the number of applications of the measures from the adapted policy program.

3b. Non-financial resources

According to Sampson & Mitchell (2013), governments often do not need to modify their existing legislation or policies for supervisory measures to a large extent to enable the implementation of ATD measures. Still, the adapted policy program in the Netherlands did require some adaptations to the Alien Act (Mulder, 2012). For the duty to report, separate categories now exist for the admission procedure and the return procedure. With section A2 10.3.2 of the Aliens Act giving an outline of the duty to report as part of the return procedure. The measure of imposing a bail was added to the alien circular under section 10.4, based on its implementation in the pilots. The subsidizing of NGO's was incorporated in the regulation 'subsidy regulation for assisted-voluntary departure' (Ministry of the Interior and Kingdom Relations, 2014). The measures of the taking of documents and the expulsion order did not require any formal policy or legislative changes. While section A5 5, which specifies the use of a freedom-restricting measure, was modified to remove the restrictions for the target group (Mulder, 2012).

Regarding the target group, another non-financial input had been to broaden the target group for the ATD measures. Both the target group for the duty to report and the freedom-restricting measure were expanded to apply to a broader group of irregular-staying migrants. While these measures were previously only used for those asylum seekers issued with a return decision (Teeven, 2013).

Another non-financial input for the current program was the adjustments to the capacity of specific supervisory measures. The capacity of the freedom-restricting center was increased to 400 as a result of the other input of broadening the target group. The capacity of the ADM was decreased from 2000 to 933 as a consequence (Teeven, 2013).

A vital part of the input has been to improve the possibilities for an individual assessment for imposing a supervisory measure. With the intention that the involved organizations regularly assess the process of return, while the R&DS provides the case management for each irregular-staying migrant. The Local Repatriation Consultations (LTO) that were already in operation in several municipalities were to be used by the R&DS, AVIM, Immigration and Naturalization Service (INS) and the municipality to review the cases. The LTO is also meant to discuss the choice of most effective and efficient supervisory measure and the progress of an imposed measure (Directorate for Migration Policy, 2011).

Lastly, the pilots for the ATD measures that ran in 2012 and consisted of pre-determined measures gave crucial information for the further development of the adapted policy program (Teeven, 2013). Not in the least as these showed that the practical condition of the rationale could be achieved with these measures. In addition, the input of these findings provided essential insights into the other resources and what was needed to further implement these pilots as supervisory measures into the return policy.

4.1.4 Activities

The activities included those services which were implemented or adapted to lead to the further components of the program theory & logic model (see also Chapter 3.5.2). Several activities were identified on the basis of the document analysis and preliminary conversations with stakeholders and the work group.

4a. Individual assessment

The first activity entails the assessment for a supervisory measure, which is made for each case of an irregular-staying migrant. This step had been made possible by the input, as can be found in the non-financial input (see Chapter 0). As mentioned in the main rationale behind the adapted policy program (see Chapter 4.1.2), determining the application of supervisory measures should be based on the necessity, proportionality, and effectiveness of a measure regarding each case (Directorate for Migration Policy, 2011; Teeven, 2013). This is in conformance with the 2008/115/EC Return Directive as it is not the intention of the Return Directive for a government to decide based on a simple list of options. Instead, a government should only impose ATD measures that are deemed effective and proportional to the objective that it wants to achieve (Bruycker et al., 2015).

The supervisory measures could either be imposed by the assistant public officer of the National Police, the Royal Marechaussee, or the supervisor departure of the R&DS. Imposing a supervisory measure is based on the previously mentioned principles, as well as the following three conditions for the irregular-staying migrant:

- *The degree of cooperation by the irregular-staying migrant working towards their return*
 - *The degree of danger for the public safety by his/her behavior or background*
 - *If the irregular-staying migrant has already had a history of evading supervision with a previous supervisory measure*
- (Teeven, 2013, p. 8)

The Ministry of Justice & Security could proceed with a forced return if the irregular migrant does not meet these three conditions in a sufficient manner. ADM is then the most restrictive measure for forcing a return. However, ADM can also be imposed for certain situations in which someone is not actively working on their return or shows no real risk of absconding, but who can be returned within a short time frame. In these cases, supervisory measures such as the duty to report or the freedom-restricting measure are also a possibility for a forced return (Teeven, 2013).

The described assessment leads to the possible application of a supervisory measure, which are listed in the following parts of this section.

4b. Alternative supervisory measures

The adapted policy program implemented the measures from the pilots of the duty to report, bail, and the subsidizing of NGO's as explicit alternatives for detention as described by the non-financial input component (see Chapter 0). In addition, (legislative) adaptations were made to other measures such as the freedom-restricting measure. Other included measures of the adapted policy program have been the expulsion order and taking of documents.

Expulsion order

The letter to the House of Representatives stated the following on the expulsion order: *“An expulsion order is not a measure for working on the return (by an irregular-staying migrant) on its own”* (Teeven, 2013, p. 13). This could also be why no comparable measure was found in relevant literature. It is used when ADM or other measures do not result in a departure. Additionally, it is used for those cases in which a forced return is not possible at the time to the involved country, and the migrant does not intend to actively work on their return (Teeven, 2013). The expulsion order may result in a return decision, which starts the return process, in the following three ways (Mulder, 2012):

- The irregular-staying migrant did not adhere to the order
- The expectation is that the irregular-staying migrant will not adhere to the order based on their statements or behavior
- If it is necessary to depart immediately for reasons of safeguarding the public safety and national security

Subsidizing NGO's

The subsidizing of NGO's is another measure that was actually not implemented as a supervisory measure but rather as a part of the possibilities for promoting voluntary return (Teeven, 2013). Finding comparable measures in other reports or literature proved difficult. However, a similar form of measures was found, in which NGO's provide case workers that assist in the return procedures. That form is most often directed at vulnerable target groups such as families and those with medical problems (Field & Edwards, 2006). Such a function has also been considered by the Ministry of Justice & Security of the Netherlands, though further incorporated as a possibility for the subsidy scheme (Teeven, 2013). Comparing the measure internationally to that of the Netherlands, it could be seen best as part of the uncategorized category of measures from the article by Mangiaracina (2016). Based on the findings of the pilots (see Chapter 0), it has been included in the adapted policy program as a subsidy scheme for assisted voluntary departure. With the intention that this contributes to the aim of increasing the rate of voluntary return (Teeven, 2013). The subsidy scheme subsidizes projects that:

- *Aim to impel irregular-staying migrant towards an assisted voluntary departure of the Netherlands*
- *Aim to prepare irregular-staying migrants for assisted voluntary return*
- *Aim to provide a perspective on the reintegration in the country of repatriation*

(Ministry of the Interior and Kingdom Relations, 2014, p. 1)

The latter of these aims could increase the chance of sustainable return, as was an identified aim in the rationale (see Chapter 4.1.2). The way in which this measure works as a part of the adapted policy program is that the projects provide a service for the benefit of a maximum sum of €1500 per individual. The subsidy scheme includes the agreements made between the NGO and R&DS to ensure the rationales of cost-efficiency and effectiveness behind the adapted policy program (Ministry of the Interior and Kingdom Relations, 2014).

Taking of documents

The taking of documents was already in place before the adapted policy program and saw no adaptations from the input (see Chapter 0). Internationally, the taking of documents as a measure takes many different forms. The article of Field & Edwards (2006) described a similar measure of the deposit of travel and identity documents, which has been a widely-imposed measure in western countries. The taking of documents is often combined with other supervisory measures such as a duty to report or the obligation to remain in a designated area (Mangiaracina, 2016).

The letter to the House of Representatives described the measure as:

The police can already take the documents of an irregular-staying migrant so that these can be kept and not lost, whereby the migrant can work towards (assisted) voluntary return with the aid of the R&DS, IOM or other NGO's. (Teeven, 2013, p. 14)

It was identified on the basis of the preliminary conversations as a relatively “soft” measure. In which the documents are returned to the pre-specified address when the person reports with the Royal Marechaussee at the airport. Consequently, the measure was identified as being the responsibility of the Royal Marechaussee, as they report on the processes of the measure. The measure is often imposed in combination with a duty to report. The measure is meant for those that have valid documents and are willing to cooperate with their return. If the irregular-staying migrant shows no progress in their return, the documents can be transferred to the R&DS to proceed with a forced return procedure. If the situation requires it, the combined duty to report could subsequently be revised and changed to a more restrictive supervision such as administrative detention (Teeven, 2013).

Bail (with departure period of 28 days)

Generally, the term ‘bail’ in relation to the ATD measures has been interpreted as the actual financial deposit given to a government agency organization of a country that is responsible for its application in the asylum or return process (Field & Edwards, 2006). Requesting a financial surety of a bail is not often used in practice in the EU, both for the asylum process nor the return process (Directorate-General for Migration and Home Affairs, 2013). Essentially, because the presumption is that irregular-staying migrants lack the necessary financial means to qualify for a bail (Bruycker et al., 2015).

The letter to the House of Representatives gives the following interpretation of bail as a supervisory measure: *“A bail has been chosen in combination with a departure period of 28 days, as it appears from practical experience that assisted voluntary departure can be effectuated within this term”* (Teeven, 2013, p. 13). A default sum of €1750 was implemented, although both that sum and the departure period of 28 days, could be altered for the individual circumstances of a case. The sum is returned when the migrant reports to the Royal Marechaussee at the airport before departure (Ministry of the Interior and Kingdom Relations, 2000b). The measure can be imposed by the R&DS for irregular-staying migrants with an obligation to depart, on the condition that the individual works actively on their return, are in possession of travel documents, and sign a contract that details the involved rights and duties of the bail. Additionally, the measure can also be imposed from within detention, it then again depends on the circumstances and willingness of the individual if the ADM is lifted in favor for a return-contract with a bail. While other countries

provided the possibility of designating a guarantor, this was not implemented in the adapted policy program. The feeling was that a guarantor would shift the responsibility away from the irregular-staying migrant (Teeven, 2013).

The implementation of the bail measure was the result of the input of the pilots (see also Chapter 0), though these showed that most migrants opting for the pilot were motivated by the prospect of leaving detention instead of their willingness to cooperate with the return process. However, the measure was still implemented in the adapted policy program because:

I definitely think that this instrument has added value in supporting irregular-staying migrants with a duty to depart the Netherlands, while also appealing to their own responsibility. This is why I will also give a bail a place within the supervisory measures for return. (Teeven, 2013, p. 6)

Duty to report

The reporting duty has been one of the most widely adopted measures used by countries in their return process (Mangiaracina, 2016). It is generally considered as an obligation to report to either the police or another relevant organization at regular intervals. The article of by Bruycker et al. (2015) the reporting duty as a cost-efficient, effective, and one of the least coercive measures. Although, this depends on the location and frequency of reporting along with the included sanctions for absconding, as these aspects of a duty to report can vary across countries.

The input on the individual assessment and broadening the eligible target group for the measure in the adapted policy program were aimed at strengthening the duty to report as a relatively mild supervisory measure (see also Chapter 0). The Chief of police can impose a duty to report onto the irregular-staying under article 54, paragraph 1.F of the Aliens Act 2000 and article 4.51 of the Aliens decree 2000 (Ministry of the Interior and Kingdom Relations, 2000b). The R&DS supervises those imposed with the measure. The location, time and frequency of reporting to be decided by the chief of police, based on the key principles of necessity and proportionality as mentioned by the individual assessment aspect of the activities. Just like with reporting measures in other countries, the duty to report can either be imposed as a standalone measure or in combination with other measures (The European Migration Network, 2014). In the Netherlands, the taking of documents and the freedom-restricting measure are often part of such a combination (Teeven, 2013). When an irregular-staying migrant has not reported at the specified time and location for two times in a row, that person is considered to have left the Netherlands or evaded supervision. He/she will then be signed off in the Aliens records/register. The police carry out a check at the last known address or the reception center where the irregular-staying migrant remained, to determine if the actual return has taken place (Ministry of the Interior and Kingdom Relations, 2000b). The police then inform the INS and the R&DS of the (assumed) departure.

The former secretary expressed the intention of the measure as part of the adapted policy program as follows:

I expect that this supervisory measure can mainly be used with people that have not been known previously by the R&DS, who are prepared to work on their return, have some form of accommodation, and for which it suffices to apply such a relatively mild supervisory measure as to effectuate return. (Teeven, 2013, p. 13)

Freedom-restricting measures

The freedom-restricting measure has been one of the more commonly used ATD by different countries besides the reporting duty. A term often used in the literature for such a measure is the obligation to remain at a designated residence (Bruycker et al., 2015). In general, freedom-restricting measures are designated as one of the most coercive measures, as it limits the freedom of movement of the individual in more ways than other measures (Chmelickova, 2006). It is more often found as a supervisory measure in the asylum/entry process of countries, and thus its implementation in the return process has been less developed in comparison. Still, it is considered as a more cost-efficient and possibly more effective measure than detention (Bruycker et al., 2015). Additionally, relocating someone to a return center sends a strong signal to the individual or family that it is the final step in the return process.

The freedom-restricting measure was already in use in the return process of the Netherlands before the adapted policy program. During which it received some criticism from both a rapport by the Advisory Committee on Alien Affairs (ACVZ) as well as a rapport by Amnesty International. They lamented the use of freedom-restricting measures as they see no significant difference between the freedom-restricting centers and ADM (Amnesty International, 2011; Teeven, 2013). The former secretary disagreed with this criticism in the letter to the House of Representatives. Arguing that the freedom-restricting measure and its centers are supposed to accommodate the category of irregular-staying migrants which actively work on their return or those that can depart within twelve weeks by voluntarily return under the supervision of the R&DS (Teeven, 2013). Subsequently, no legislative adaptations were deemed necessary to article 56 of the Alien Act 2000 (Ministry of the Interior and Kingdom Relations, 2000a). As mentioned in the non-financial input, resources that have been implemented where the broadening of the categories of irregular-staying migrants eligible for being imposed the measure and increase the available places in freedom-restricting centers (FRC's). As such, the freedom restricting measure can be imposed by the R&DS, which can opt for a designation to a freedom-restricting center if the irregular-staying migrant does not have an accommodation to stay when working on their return (which is usually the case). These individuals are expected to show some progress in returning

(voluntarily or forced). The default term of stay was still a maximum of 12 weeks. The default term would be terminated when the irregular-staying migrant has departed to the country of repatriation or if not enough progress had been made in their return process. In the latter case, the Deputy officer of Justice examines if a more restrictive measure is needed (R&DS, 2015).

Measures for vulnerable target groups

Many studies have given particular attention to the use of alternatives to detention with vulnerable target groups in many studies on the alternatives to detention. These include either unaccompanied minors as well as underage children as part of a family. Other categories often regarded as vulnerable target groups are those with physical or mental health issues or the elderly (Field & Edwards, 2006).

In the Netherlands, the input of the individual assessment was meant to ensure that the choice for a measure fits with the characteristics of vulnerable irregular-staying immigrants (see also Chapter 0). As explained in the letter to the House of Representatives: *“Special attention is paid to vulnerable people regarding the application of supervisory measures. Administrative detention will, even more so than among non-vulnerable people, be imposed only as a measure of last resort”* (Teeven, 2013, p. 16). Furthermore, a court decision in 2011 obligated the Dutch government to provide special accommodations for families with underage children in the FRC’s (R&DS, 2015). Another addition has been the inclusion of a closed family center as of 2014 for families with underage children and unaccompanied minors. Administrative detention is only imposed on families or unaccompanied minors that show a high risk of absconding supervision before a forced departure. This is because the government of the Netherlands considers it undesirable for children to reside illegally, with no provision of education, health services, and other support. The closed family center in Zeist is used as a location for the application of an administrative detention measure, and families are typically sent there two weeks before their departure (R&DS, 2015).

4c. After imposing a supervisory measure

A regular re-assessment of the application of a supervisory measure was also identified as part of the activities for after being imposed with a supervisory measure, as part of the activities of the adapted policy program. Depending on the individual case and if the current measure is still the most effective and efficient measure. A different or adjusted measure could be either less or more restrictive, or it could be that new developments require a different approach (Directorate for Migration Policy, 2011; Teeven, 2013).

4.1.5 Output

The indented outputs of the program are the resulting products that originate from the activities (see also chapter 2.3.1). Several outputs were identified as being expected from the activities of the adapted policy program.

5a. Efficiency

The efficiency output contains the output expected from the adapted policy program and its practical rationale of reducing the number of resources necessary to effectuate return (see Chapter 4.1.2). The Administrative detention is considered a resource-intensive measure, requiring detention centers, personal and the provision of utilities for those detained (Bruycker et al., 2015). Whereas, the expectation is that ATD measures require fewer resources to meet the same purpose of effectuating return. The current analysis identified three outputs regarding the efficiency of the adapted policy program.

The first expected output was that of a reduction in costs of the adapted policy program, based on the estimations of the internal report (Directorate for Migration Policy, 2011). A rough estimate has been given of the intended costs per day of ATD measures and comparing these with the costs per day of the ADM. The costs per day for the detention measure had been estimated to be €197,00, while two included alternatives had been estimated to cost €56,00 for the freedom-restricting measure and €10,00 per day for the duty to report. These estimations were based on a maximum of 6500 irregular migrants and historical data. Furthermore, these did not incorporate the costs for decreasing the detention capacity or other investments (Directorate for Migration Policy, 2011).

A second identified output was that the adapted policy program requires fewer resources, as the letter to the House of Representatives explicitly described the expectation that these alternatives are more straightforward to impose (Teeven, 2013). The included ATD measures are supposed to require less bureaucracy and resources and instead rely more on promoting the responsibility of the irregular-staying migrant and their self-sufficiency.

A third expected output involves the duration to which an irregular-staying migrant is imposed with the measure. The average process time for the administrative detention measure was 74 days, while the process time for the freedom-restricting measure and duty to report was 84 days at the time of the internal rapport (Directorate for Migration Policy, 2011). As such, the expected overall efficiency output was that the activities result in a decrease in costs and the use of resources. However, the process time of such less-coercive measures was expected to be a little longer.

5b. Practical condition

The first practical output that was expected to follow from the previous components of the adapted policy program is of a practical nature. The rationale identified that the aim of is to implement “*the possibilities to make more use of alternatives to administrative detention*” (Teeven, 2013, p. 4). The input made it possible to implement ATD and adapt existing supervisory measures (see also Chapter 0). As such, the first expected practical output was a shift in the application numbers between the ADM and ATD measures. If the intended policy program works as it is supposed to, the result would be that the use of detention would decrease as it is no longer the ‘de facto’ choice.’ This shift would equate to a parallel rise in the application of ATD measures as these are now several viable options to consider. The output should, therefore, be that the use of ATD compared to ADM increases in the ratio relative to the ratio before the changes were made, providing that all other factors remain the same (Teeven, 2013). This expectation also functioned as the basis for decreasing the capacity of ADM, as identified in the input (see also Chapter 0).

Another practical output that was intended to follow from the activities was the difference in case resolution rates between detention and alternative measures. The document analysis and preliminary conversations from the second step of the EA-model found three different types of case resolutions: (1) Confirmed departure, for those registered by a relevant agency as departed either by assisted voluntary or forced return; (2) Non-return, for those that are imposed with another supervisory measure as they were not able to return on time; (3) Non-demonstrable departure, for those that are released from supervision or have absconded between measures (Teeven, 2013). The non-demonstrable departure registration thus comprises of those that could remain illegally in the Netherlands, traveled to another EU country, or to another destination (R&DS, 2015). As such, monitoring and interpreting the case resolution rates in the Netherlands is mostly done by examining the confirmed departure rates. The non-demonstrable departure rate can also be taken into consideration, though it is more difficult to analyze and therefore generally less preferred. As identified in the rationale, a practical condition was that the implementation of ATD would result in a similar level of effectiveness of return (Directorate for Migration Policy, 2011). The input of the pilots was aimed at the examination of this condition (see also Chapter 0).

Table 6. Overview of the results of the pilots in 2012 in comparison to administrative detention in 2010

	<i>Administrative detention</i>	<i>Duty to report</i>	<i>Bail</i>	<i>NGO's</i>
<i>population</i>	7900	75	15	180
<i>Still in process</i>	Not mentioned	10 (13%)	5 (20%)	85 (47%)
<i>Alternative terminated</i>	Not mentioned		<5	5 (3%)
<i>Total departure</i>	4350 (55%)	60 (80%)	<10	90 (50%)
<i>Of which: Actual voluntary departure</i>	450 (10,3%)	45 (60%)	<5	75 (83%)
<i>Of which: Actual forced departure</i>	3900 (89,7%)	<5 (4%)	0	0
<i>Of which: Non- demonstrable departure</i>	Not mentioned	10 (13%)	<5	15 (17%)
Note. Information retrieved from the Directorate for Migration Policy (2011) and the Directorate-General for Immigration (2013).				

Comparing these numbers from Table 6 would give the impression that the pilots were successful. With the subsidy of NGO's and a bail having a similar result as the detention measure on total departure, and the duty to report leading to an 83% departure rate. Even more so, the (preferred) confirmed departure of 60% and 83% for the duty to report and subsidy for NGO's were promising in comparison to the 10,3% confirmed departure rate for ADM.

However, several caveats were identified with the current analysis that limited the impact of these promising findings. These included the fact that the compared target groups differ on a critical aspect, their willingness to cooperate. One of the conditions for imposing an ADM is that the irregular-staying migrant is not willing to cooperate. Even more so, strict inclusion requirements for the pilots meant that those included showed a higher willingness to cooperate than would probably be required with the adapted policy program (Teeven, 2013). In addition, the preliminary conversations pointed out that individuals can be imposed different supervisory measures during their return process in the regular return process. This makes the analysis of the measures influence on the case resolution in practice far more difficult. Lastly, the data quality also left something to be desired, as the internal report and biannual reports on the immigration process provided only limited information on the resolution rates of ADM. Even more so, the appendix of the letter to the House of Representatives solely comprised of a limited and

inconsistent summary of the results of the pilots. Consequently, the expected output of the adapted policy program was that the ATD measures should lead, at a minimum, to similar confirmed departure rates.

The compliance rate was identified as the third practical output of the adapted policy program. Field & Edwards (2006) classified the compliance rate into two broad types: The rate in which the irregular-staying migrant complies to the obligations that are part of the procedures and the rate to which irregular-staying migrants abscond the supervision of the government. The second kind of compliance rates was used to monitor the results for the pilots (Directorate-General for Immigration 2013). Precisely determining the absconding rates is difficult to achieve with the use of the non-demonstrable departure category as it combines the individuals that have absconded supervision as well as those that have departed to an unknown destination, as already mentioned with the case resolutions. The second condition for meeting the practical policy objective of effectuating return for the Netherlands is that the risk of absconding remains low enough as to not lead to any danger to the public safety and national security (Directorate for Migration Policy, 2011). The non-demonstrable rates of the duty to report in the pilots are 13% and from the subsidizing of NGO programs 17% as can be seen in Table 6. On the basis of these results, the expectation was that the absconding rate would be low enough to meet that second conditional outcome for achieving the practical condition (Teeven, 2013).

5c. Humanitarian

The third expected output was that the use of lighter supervisory measures should result in improvements being made to the treatment of the irregular-staying migrants, and is thus related to the political rationale (see Chapter 4.1.2). Several studies and reports, such as the one by the ACVZ (2013), Amnesty International (2011) and Fiske (2016) had expressed concerns on the ADM and its effect on the freedom, human rights and physical and psychological health of irregular-staying migrants. The adapted policy program should lead to more use of the less-coercive measures and thus ensure better treatment of those in the return process. Precisely comparing these populations could be difficult, as already stated by the former Secretary of Justice & Security in the Letter to the House of Representatives (Teeven, 2013). As there is a difference in the willingness to cooperate with the return process between those imposed with an ATD measure and those imposed with an ADM, the second group will already have a more negative predisposition towards their treatment. Prospective evaluations by qualitative (e.g., interviews) or quantitative (e.g., questionnaires) had already been used by for example Costello & Kaytaz (2013) and could be the most viable option to investigate the outcome.

4.1.6 Outcomes

The outcomes are the expected benefits and downsides of the improved integration of the ATD measures. These can be either on a short-term or an intermediate period and are the result of the outputs (see also Chapter 3.5.2). Three expected outcomes were identified as part of the adapted policy program.

6a. Necessity

The first outcome expected from the adapted policy program is based on the principle of necessity from the individual assessment of the activities (see also Chapter 0). The primary output for this outcome were the outputs of cost-efficiency, simplicity, and time of the efficiency output, together with the other outputs in the background. The necessity outcome thus forms the efficiency of the adapted policy program, which can be seen as: *“The extent to which the resources/input (funds, expertise, time) were justified, given the changes/effects which have been achieved”* (European Migration Network, 2016, p. 8). These should be compared to the changes achieved, which are the expected differences in the use of ATD over ADM and the influence this has on the return rates (Directorate for Migration Policy, 2011; Directorate-General for Immigration, 2013). As well as the intended effect of using less-coercive measures, decreasing the impact on the freedom of movement and its associated negative consequences.

Therefore, the intended outcome of the program was that there are fewer resources necessary for the adapted policy program, ensuring that only the necessary resources are used to effectuate return (Teeven, 2013).

6b. Effectiveness

The second outcome of the adapted policy program was identified as the effectiveness as a result of the effectiveness outputs. The effectiveness can be seen as: *“The extent to which planned results are expected to be achieved”* (European Migration Network, 2016, p. 8). As mentioned by the rationale, the practical condition was that the adapted policy program results in a similar level of effectiveness in effectuating return and an acceptable risk for the public safety & national security (see Chapter 4.1.2). The relevant output of the effectiveness involves the case resolution and compliance rates, while the other practical output of more application of ATD measures is related to the main objective (see Chapter 4.1.7). There is still debate in the Netherlands and internationally on the influence of less-coercive (ATD) measures on these rates. The following arguments were identified from the analyzed literature:

- *Compliance* increases when measures persuade individuals instead of coercion (Costello & Kaytaz, 2013)

- *Case resolution & compliance* rates can increase with the use of ATD as (Bloomfield, 2016);
 - These measures improve the level of trust between the individual and the government agency and gives them more control over their decisions
 - Improve the access to resources and (legal)assistance
- *Willingness to return*: Less-coercive measures improve the assessment of the individual on the *procedural justice* of the decision (Costello & Kaytaz, 2013). This makes it likelier that the individual is more willing to return and accept the return decision (Alphen, Molleman, Leerkes, & Hoek, 2013). However;
 - The study by Esaiasson, Persson, Gilljam, & Lindholm (2016) have found that procedural justice only plays a limited role on the willingness to accept a decision. With similar findings being found in regard to the Dutch immigration process by the study of Klaver, Telli & Witvliet (2015)
 - The study by Leerkes, Galloway, & Kromhout (2011) in the Netherlands showed that other influences are more important regarding the willingness to return, such as;
 - Their own perceived health (Leerkes, Galloway, & Kromhout, 2011)
 - The safety situation in the Netherlands and country of origin, background characteristics, and practical limitations (Alphen et al., 2013; Klaver, Telli, & Witvliet, 2015)

In spite of these inconclusive findings, the input of the pilots indicated that the case resolution rates for ATD measures can be comparable to those of the ADM as is also shown by Table 6 from the outputs (see also Chapter 0 & 4.1.5). These findings also revealed that the compliance rate, as measured by the rate of non-demonstrable return, was low enough as to meet the other aspect of the effectiveness (Teeven, 2013). As all ATD measures showed rates lower than 20%, which is the suggested standard by Field & Edwards (2006) as to consider these as effective in their compliance.

As such, the intended outcome of the effectiveness of the intended program consisted of the combined results on the case resolution and compliance rates of the adapted policy program. With the expectation that the case resolution rates would be similar to that of the detention measure. The compliance rate would be lower than it is under detention but will remain at an acceptable level.

6c. Proportionality

The third and final outcome of the adapted policy program consist of the humanitarian treatment of irregular-staying migrants in the return process and originates from the political rationale (see Chapter 4.1.2). As mentioned in the output section, the use of lighter supervisory measures should have less impact on the mental and physical health of the irregular-staying migrants than the use of detention (see Chapter 4.1.5). The improved incorporation of ATD measures in return policy should result in an actual perceived improvement of the treatment and effects of supervisory measures in the return process of the Netherlands. The intended policy program has provided several alternative measures, with different levels on the humanitarian output and the other outputs. The inclusion of improved alternative measures as a viable option to detention gives more room to impose a supervisory measure based on the principle of proportionality. With proportionality meaning that the negative consequences of a decision by an administrative agency should not be disproportionate in relation to the goals associated with that decision (Ministry of the Interior and Kingdom Relations, 1992). In the case of imposing a supervisory measure, the negative consequences on the health and restriction in the freedom of movement should not be disproportionate in relation to the other goals from the rationale.

4.1.7 Objectives

The objectives of the adapted policy program consist of the outcomes and their impact on achieving long-term objectives (see also Chapter 3.5.2). These objectives are related to the goals from the rationale, although these signify the more precise objectives that are intended from the adapted policy program. The main objective of the adapted policy program was that the adapted policy program provides more possibilities by the individual assessment of the activities to impose the right measure for each case. This ensures that the less-coercive measures are applied more often as is an expected practical output of the adapted policy program (see chapter 4.1.5), and thus that ADM is to be imposed as a true '*ultimum remedium*.'

By doing so, the main objective also ensured that the following secondary objectives could be achieved:

- *7a. Necessity objective:* Enabling the use of alternative measures should result in a more efficient return policy
- *7b. Effectiveness objective:* The adapted policy program maintains a similar or slightly better effectiveness in effectuating return and acceptable risks for the public safety & national security
- *7c. Proportionality objective:* The use of less-coercive measures when possible, ensures a more humane and dignified return procedure for the irregular-staying migrant

These objectives are, as of now, not backed up by any data on the actual effectiveness, proportionality, and efficiency of the adapted policy program in practice. Evaluating these outcomes could provide a more informed consideration of these outcomes, based on the exact results of the measures. The summary below presents the resulting logic model of the program theory (see also Chapter 3.5.3).

4.1.8 Summary: Logic model of adapted policy program

The following part presents the logic model, which gives a visual representation of the adapted policy program on the basis of the program theory. The logic model shows how the different aspects relate to each other and provided the framework for the further analysis of the study (see also Chapter 3.5.2). As the logic model of Figure 10 shows, several contextual factors have been identified that influenced the program as antecedent and moderating influences. The rationale comprises of the main intention and three underlying goals that were intended to be the result of the adapted policy program. The described input and activities were needed to realize the rationale into practice. The expected outputs and outcomes of the input and activities should lead to the objectives, which explicitly stated the rationale of the adapted policy program into the intended long-term outcomes. The second section of the results describes the comparison of the program theory with the perception of the adapted policy program as implemented in practice by relevant stakeholders. The logic model served as the framework for that comparison.

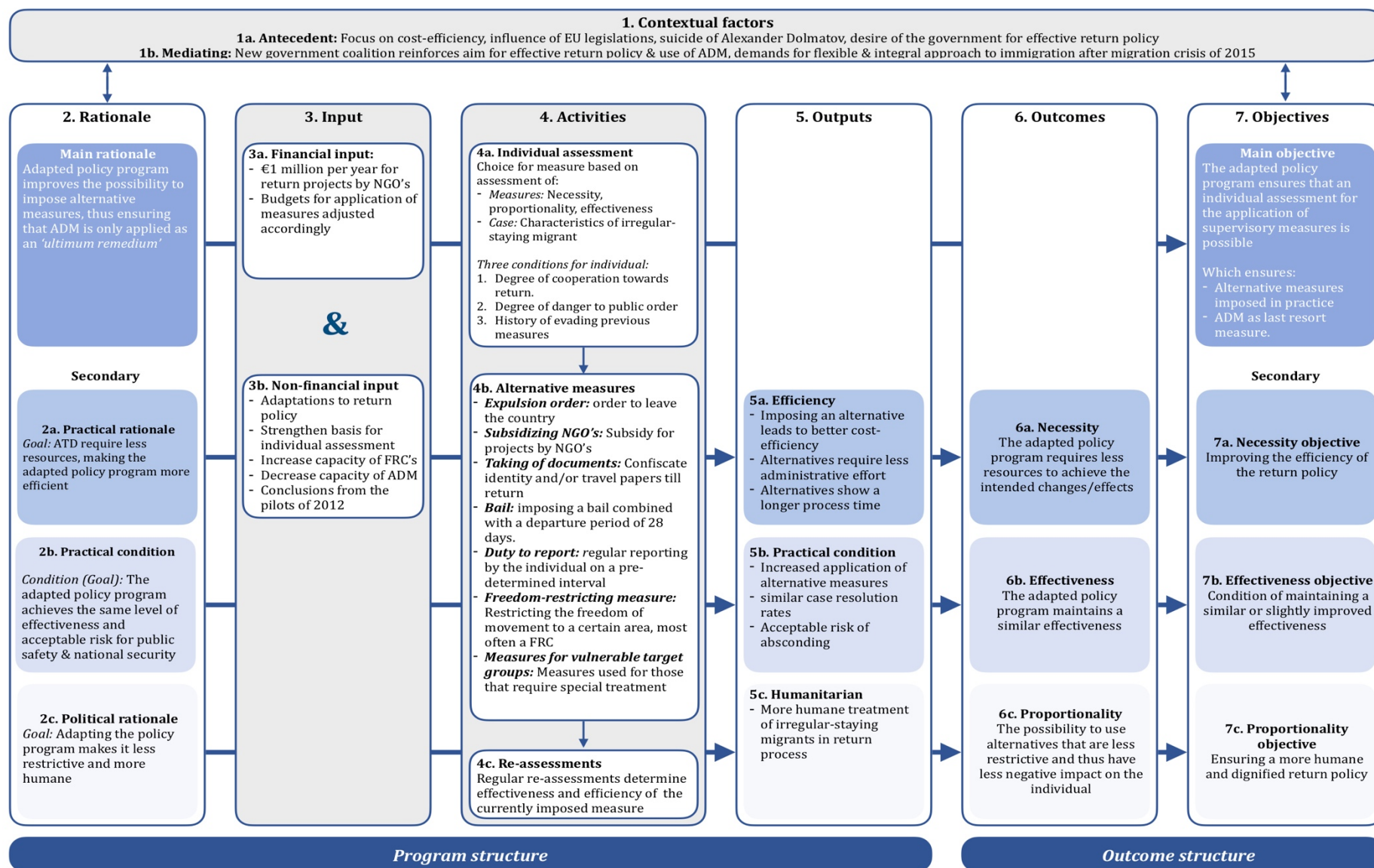


Figure 10. Logic model for the intended policy program of the implementation and improvement of the supervisory measures in the Netherlands

4.2 Program Reality

The following section of the results describes the comparison that is made between the program theory & logic model and its implementation in practice as part of step three of the EA-model by Wholey (2015). The described parts of this section are structured around the components of the logic model (see Figure 10).

The incorporated quotes from the interviews were translated from Dutch into English. The original version of the quote can be found in the appendix (Appendix C). Some quotes include some clarifications to improve the understandability of the quote in question, and the use of parentheses designates these. A '[. . .]' between one part of the quote and the following indicates that an intermediate part of the quote is left out.

4.2.1 The contextual factors in practice

The contextual factors included any influences that are outside of the control of the program. These were divided into two categories for the program under study. Any influences that were present at the start of the program were regarded as antecedent contextual factors. Any influences that became apparent during the program were included as mediating contextual factors. The statements from the respondents made it clear that such a clear-cut distinction does not appear in practice, with many influences showing an overlap between these two categories.

One such influence that was mentioned as a broad theme of continuing significance by many stakeholders has been the influence of legislation at the European level. The program theory already identified the influence, with the 2008/115/EC Return Directive as the most important aspect of the adapted policy program under consideration. As one respondent put it: *"I would argue that European legislation more and more has an influence on migration policies at a national level as well"* (NP, Q1). In addition, the comparison showed that the mediating factor of the 2015 migration crisis also increased the attention to EU-wide regulations. Consequently, the involved government agencies in the Netherlands felt that effect to a large extent.

As already had been identified for this mediating factor, the sudden increase in the number of asylum seekers and other immigrants demanded increased flexibility and integration of the organizations responsible (see Chapter 4.1.1). More importantly, the respondents described the effects and consequences this had in the Netherlands. For example, the Aliens Police Department, Identification and Human Trafficking (AVIM) had to transfer additional capacity to their responsibilities within the asylum procedure. This led to a decrease in their capability to provide supervision within the return process. However, the crisis also led to increased flexibility and faster reactions within the government agencies cooperating within the immigration field. On this issue, one of the respondents stated: *"We also see a fast response to irregular-staying migrants from*

certain countries” (DTV2, Q2). Meaning that the respondent indicated that the involved agencies react faster to increased immigration from specific countries that are not an immediate priority based on their country circumstances. With agencies resorting to forced return more often when necessary in such cases. The result is that these asylum procedures are handled in a faster process, so more asylum applications from prioritized countries, such as Syria, could be handled.

Another influence on the capacity of the Aliens Police Department has been the reorganization of the police organization, as mentioned in the program theory (Chapter 4.1). A consequence of the restructuring was a further decrease in the capacity of the police to conduct routine checks for irregular-staying migrants. Instead, the prioritization of the National Police in the Netherlands was directed towards other aspects.

Another contextual factor from the program theory & logic model that was confirmed by several respondents has been the suicide of Alexander Dolmatov. Which was a significant incident that led to a re-evaluation of the immigration policy as a whole, even going so far as being a wake-up call for NGO's as to increase their focus on the Dutch immigration system. Thus, putting more societal pressure on the government to change their immigration policy. That focus has mostly maintained over the years, due to the other factors like the migration crisis.

Furthermore, the political developments were mentioned by several of the respondents as well. One noticeable concern shared by some was the diminished attention for sustainable return. With the term being explained by one respondent as: “Sustainable return, of course, means that people should be willing and able to re-integrate into the society of which they came from” (IND1, Q3). While another respondent expressed the following concern:

Yes, that was a very important one then (regarding sustainable return). Nowadays you don't hear as much about it, but sustainable return means that someone returns with some means and possibilities [...] not only that, but it is also so an individual can return with some dignity. And that is something you don't hear about anymore. (JV1, Q4)

A last concern many shared was the troubled and lengthy process of the new law ‘Return and Alien Detention,’ that has yet to be debated in the House of Representatives (*Dutch House of Representatives, 2015*). With many expressing their frustrations, as it improves the framework surrounding the application of administrative detention and the supervisory measures. Thus, differentiating these kinds of measures from the criminal context as it is then placed in a separate regime. It should be mentioned that the respondent from Amnesty International did agree that the new law is a definite improvement but that it could and should have included more.

4.2.2 The rationale in practice

As described by the program theory & logic model (see also Chapter 4.1), the idea was that the improved embeddedness of the supervisory measures would ensure that executive officials had more and better-implemented options for less restrictive measures that could be imposed based on an individual assessment of each case. Thereby, making a determination on several conditions in combination with the personal circumstances. The majority of the respondents recognized that that intention still holds in practice, while also stating that the administrative detention measure can now really be seen as a '*ultimum remedium*.' As one respondent summarized:

What you see now with the alternatives and their use within the police... where previously an administrative detention measure was imposed fairly regularly, this has changed significantly. Where previously one would say; 'administrative detention unless it is not applicable,' nowadays we look much more towards the possible alternatives that we have. Or which measures we can impose, which lighter measures we can impose. And administrative detention can thus really be applied as an ultimum remedium. (NP1, Q5)

Figure 11 shows more examples of the statements made regarding the rationale behind the program towards making the administrative detention measure a '*ultimum remedium*.' The most significant change in practice has been that it is now necessary to provide more substantiated arguments for the choice of a particular measure.

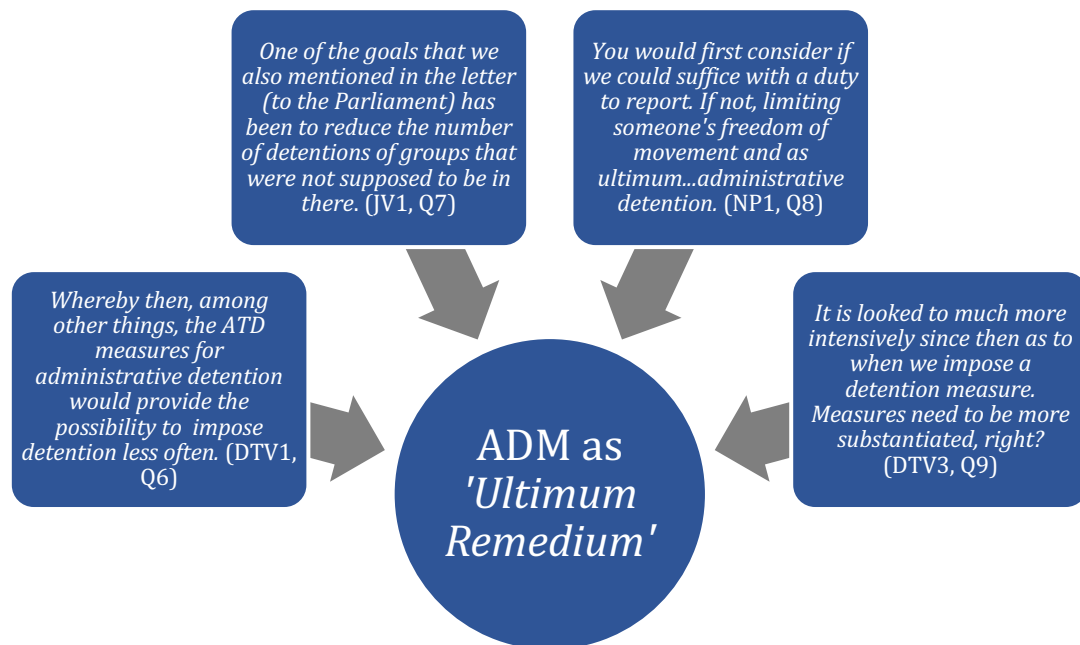


Figure 11. Statements made regarding the use of administrative detention as an '*ultimum remedium*' in practice

Secondary rationales

The information from the respondents confirmed the other aspects of the rationale, such as the intended cost-efficiency advantage of the ATD and less-coercive measures being seen as providing a more dignified and humane treatment. However, the rationale that these measures could meet the same practical policy objective of supervision to effectuate return as the detention measure was given a more nuanced and critical view from the respondents. With some stating that, although it is difficult to provide any hard evidence, the idea behind the ATD measures according to them is not that these are just as or comparable in meeting the practical policy objective. Rather, these measures do provide more of a risk of individuals absconding or leaving to an unknown destination. Most respondents agreed that the idea behind the provision of such supervisory measures is that the other rationales for their inclusion should outweigh these potential risks. In addition, some respondents expressed their regret that the original focus on the sustainable return as a rationale has lost its importance, mainly due to the recent political developments from the contextual factors (see Chapter 4.1.1). However, the idea that the alternative measures can be used to provide supervision during the return process in the same vein as the ADM is still applicable.

Definition and interpretation of the concept of alternatives to detention in practice

The most surprising revelation from the interviews concerned the interpretation of the basic concept of ‘alternatives to detention’ and the rationale behind it. Opinions differed as to whether these supervisory measures function as a literal alternative to detention, in the sense that these are imposed as a replacement for an administrative detention measure. Some argued that it is not that black and white and that some of these measures are in fact used as an actual alternative. However, a reported problem that was recognized is that it is difficult to gain insight into the precise motivations and considerations of the executive civil servants that are involved in the choice for a measure. More significantly, some state that these measures were perhaps introduced at the beginning as alternatives to detention, but only because there were no reasonable alternative options besides detention before the program. Rather, these “alternatives” are now included with the administrative detention as an option range for supervisory measures. It is therefore not the case of imposing a detention measure OR an alternative. As one respondent explained:

We do not interpret the definition in such a way, because detention and any other measure are just as well all included within the range of options for imposing supervisory measures. And when we are confronted with an irregular-staying migrant in practice we, together with the police, would just determine the measure that is best applicable in that situation [...] and only when we notice that these supervisory measures are non-applicable, for example,

because we expect that someone would evade supervision, only then would we arrive at a determination of the “ultimum” supervisory means of administrative detention. (DTV1, Q10)

Thus, this relates to the overarching rationale of the program, in which the supervisory measures are not imposed in a ‘OR this, OR that’ manner, but rather in an ‘IF this, THEN that’ individual assessment. As another respondent states: *“The policy was meant to be that there were structural possibilities for stakeholders to impose an alternative if detention was non-applicable” (JV1, Q11).* Taken together, these results indicate that the use of the term ‘alternatives to detention’ no longer covers these measures. Instead, these measures are an integral part of a full range of options of supervisory measures, including the administrative detention measure.

4.2.3 Input in practice

A prevailing view from several respondents has been that the legislative changes did not have a significant impact on the approach of many of the involved stakeholders. When asked about the changes with the duty to report measure, one respondent stated: *“Legally nothing, as it was already in place before. However, it is now applied more often as an alternative or lighter measure for administrative detention”* (NP1, Q12).

The program theory & logic model described that input was given to provide a basis for the individual assessment of the application of the measures. A recourse unmentioned in that chapter was that a program was also started to improve the professionalization of the R&DS. One aspect of which was the introduction of a form of case-management (although respondents do point out that some NGO's, such as the International Coalition on Detention, might disagree on it being a form of case-management). The R&DS adapted the approach by the method of 'werken in Gedwongen Kader' (working in detention centers) to the characteristics of the Dutch immigration process. Moreover, the departure supervisors were also better informed on all the alternatives, which was also crucial in the other involved government agencies.

Another unmentioned input that was mentioned in an interview concerned the financial input as a result of a part of the recalculated budget of the Custodial Institutions Agency. Financial resources became available after discussions recognized that the Custodial Institutions Agency does not require as many resources, as the detention capacity would be decreased. A sum of €10 million from the available budget was redirected towards strengthening the R&DS, building a new location, and setting up the government fund of €1 million per year for subsidizing return-related projects by NGO's.

The implementation of the program did not make it necessary to overhaul the process around the data-systems that are used for the registration of relevant information around these measures. Registration of the application of freedom-restricting and administrative detention measure was already in place. New so-called 'registration obligations' were implemented for new supervisory measures, whereby registration of the duty to report and Bail were added as intake reasons.

As for the input provided by the pilots as mentioned in the program theory, although these are recognized as providing the groundwork for the further implementation of the supervisory measures, some issues were brought up with the effectuation of the pilots. These issues were already mentioned in the program theory, including the strict pre-selection for inclusion and the small size of the sample. The respondent from Amnesty International also expressed these concerns: *“Yes, regarding those pilots [...] a couple of things not enough is known about these [...] and it was too small. Only a tiny group was able to participate in it”* (AI1, Q13). This coincided with the document review of the program theory. As the document review also showed the same

absence of detailed information on the results from these pilots for the present study, as only a one-page summary of the output and outcomes was found.

4.2.4 Activities as implemented in practice

The activities as described in the program theory & logic model consisted of the individual assessment for the choice of a measure, a description of the options as intended in the program, as well as the re-assessment of the applied measure to an individual during the return process. One reoccurring theme through the interviews has been the low application rate of the bail measure. The low rate was also expected in the implementation of the program, mainly due to the findings from the pilots that most individuals did not meet the required conditions. Although this reason was also given in the interviews, three additional causes were also provided by the respondents that could explain the low rate. A first given cause was mentioned by one respondent, who thought that there were some signals that the registration of the application of the measure was sometimes inaccurate. Thus, the application numbers of the bail measure could be an inadequate representation of the use in practice. That inaccuracy was explained as coming from the fact that these kinds of processes can be incredibly complicated, which could cause a non-registration or a faulty registration of some aspects. The respondent explicitly mentioned that this was a preliminary indication, and the respondent was unsure if this was actually the case. Despite the possible inaccuracy of the data, information from other stakeholders confirm that the application remains low, especially compared to other measures. Another cause for this is that the use of a bail is more difficult than the application of a duty to report, while these are both considered a relatively "light" measure on the level of restriction on someone's freedom of movement. Commenting on the application of the bail measure, one respondent explained:

If you have to choose between a bail, which requires a pile of paperwork and were you also need to ask money from the irregular-staying migrant, that needs to be set apart, after four weeks you have to declare forfeited as he has not departed. Or you impose a duty to report, with return supervision from the R&DS, which falls into your daily practice. You will see that a bail is not always offered as a possibility. (JV1, Q14)

A final given cause is also that the bail is less popular as an advised measure by involved lawyers or NGO's. Although the measure was clearly described and communicated as an option by the Ministry, it can be concluded that the measure remains an unpopular measure for all those involved. This is not to say that it does not have a purpose. Most respondents that commented on the application of the bail measure do recognize that it can be useful, though only for an insufficient number of cases from the population.

Regarding the duty to report, the main issue expressed by the respondents was that although it is described as one measure, it can take many different forms. In fact, some respondents stated that it is difficult to consider the duty to report as one single measure. Several dimensions exist on which a duty to report can vary: (1) The duration of application (as there is no pre-defined maximum duration), (2) the frequency of the reporting requirement, (3) registration of the irregular-staying migrant reporting to the FRC, COA or police, and (4) with or without supervision by the R&DS. So, the implementation of the duty to report is not unambiguous and is done more by a flexible approach. Another measure that is different in practice is the expulsion order. Due to the implementation of the European return decision within 28 days, the expulsion order is often only expressed in words when an irregular-staying migrant is encountered, and the police want to re-affirm their obligation to depart.

The maximum duration of 12 weeks of being able to remain in an FRC with a freedom-restricting measure was already found to be a guideline which must be kept only in principle, while exceptions can be made. The respondents provided further information on that guideline, stating that:

The choice for a maximum stay of 12 weeks was indeed chosen when the FRC was established, most likely due to the fact that you could appeal and that would take around three months before a decision was made then [...] so we are still guessing precisely why. Those 12 weeks have stayed. It could be that irregular-staying migrants need more time, which is assessed by the departure supervisor. (DTV2, Q15)

A prevailing view among the respondents was that a freedom-restricting measure in the FRC has the clear advantage over measures, such as the duty to report and bail, that there is more supervision of the irregular-staying migrant. The involved departure supervisor, therefore, has more contact with an individual, thus increasing the chances for adequate progress to effectuate return. Even more so, many NGO's have an office within an FRC, thus increasing the opportunities for these to be involved in the process. All in all, this results in shorter communication lines between all involved organizations and the irregular-staying migrant. NGO's expressed their concerns as found in the documents for the program theory beforehand about the FRC, as some of them regard it as a form of detention instead. The respondent of Amnesty International shared some of these concerns, as long-term stays in an FRC can have a similar adverse effect on someone's wellbeing as detention. However, the respondent agreed that the circumstances in a FRC are already very much improved compared to the administrative detention centers. The implementation of a subsidy for projects by NGO's is seen as a well-regarded option, with some projects being a hit-or-miss but overall providing positive results.

A final noticeable observation from the interview-data has been the implementation of the FL and CFF, although both were not considered as real separate alternative measures. Still, these provisions to take into account the vulnerabilities surrounding children with or without their families were implemented for these specific populations. Especially the implementation of the CFF has been positively mentioned by many of the respondents as an example of how to accommodate children within the return process.

Individual assessment before and during application of a supervisory measure

Another critical part of the activities has been the individual assessment before and possible re-assessments after the application of a measure. Generally speaking, most respondents agreed that the individual assessment is implemented to a great extent. It is especially noticed with the duty to report, FRC, and detention measures that the demands for the assessment have increased as to meet the legal requirements following the implementation of the program. However, that assessment is difficult to understand or follow for outside organizations. One concern expressed by the respondent from Amnesty International was that it is difficult to see what kind of assessment was made in a case as they argue that the assessment is mostly based on interpretation. These organizations would like to see some form of standardization or clarification of the process as to enable them to verify the assessment process when needed. Especially relevant for Amnesty International were the unclear criteria for when a person is considered a vulnerable person, which require a different approach as described in the program theory (Chapter 4.1). The unclear wording of when someone meets the criteria would be improved with the law that is under consideration, although it is the opinion of Amnesty that some additional criteria would be needed to clarify better when someone is considered a 'vulnerable person.' The view from the respondent of the R&DS differed in regard to the general need for a more standardized format for an assessment. Although it is recognized and understood that such organizations would like to know more about the process, each case has unique characteristics that demand extensive knowledge and experience from those involved in the assessment. That complexity also makes it difficult to use a standardized process for the assessment or to gather and provide clear and accessible data on the consideration that was made. One could argue that the positive change, whereby an individual assessment for each case can be made, limits the possibility to provide insightful information on the considerations that are the basis of that assessment.

A final relevant finding from the interview data involved the re-assessment of the application of a supervisory measure. As the program now includes a stepwise approach to the application, this makes it easier for a departure supervisor to adjust or reconsider the choice of a supervisory measure. Attributed to the increased number of suitable options, but also because the

process of application has become more transparent for the relevant officials as to why a certain measure was applied.

Structure of the supervisory measures

An overview can be made of the structure of the supervisory measures in practice, based on the comparative information and the logic model (Figure 10). Figure 12 presents the structure as divided into four groups of supervisory measures. The first group consists of the measures designated as 'procedural measures' by the present study. Both the return decision with the possible informal use of an expulsion order and the taking of documents are mostly used as a procedural measure, instead of these being a real measure as to provide supervision. The measures seen as the least restrictive, actual measures for supervision are the duty to report and bail options from the second group. With the duty to report applied in far more cases than the application of a bail. The third group contains the fairly restrictive measure of a freedom-restricting measure, which is almost always combined with a stay at the FRC. In addition, the FL is a specific accommodation for families with underage children or unaccompanied minors. Though not a real supervisory measure, the FL together with the CFF should be seen as part of the accommodations as to provide supervision on those that fall under the category of the vulnerable target group. Lastly, the most restrictive measure is the ADM as an '*ultimum remedium*.' The CFF is an accommodation to adequately detain families with underage children or unaccompanied minors, thus being a highly restrictive measure for providing supervision on those from the vulnerable target group together with the FL as a less-coercive type of accommodation. The subsidy of NGO's was not regarded as an actual supervisory measure by the stakeholders, in accordance with the document analysis, thus that measure is not incorporated in the overview.

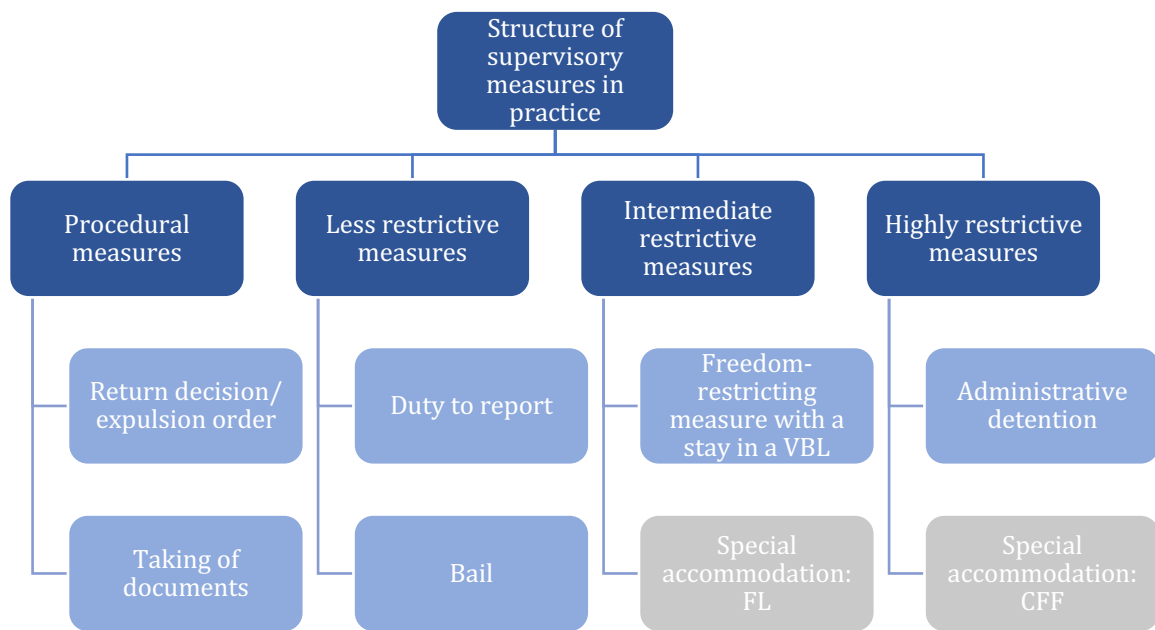


Figure 12. Overview of the structure of the supervisory measures in practice

4.2.5 Output of the adapted policy program in practice

The program theory identified three categories of outputs from the documents and initial talks with the stakeholders. Further interviews confirmed that the three broad categories of humanitarian, practical, and efficiency related outputs were accurate categorizations. However, some of the aspects of these categories were perceived differently in practice.

Practical output

The central theme from the respondents concerning the outputs has been the focus on the practical condition output. The government agencies involved in the return process all register relevant data on the application of the measure for which these agencies are responsible. The most important output regarding the application is the inflow and outflow registration of an irregular-staying migrant. An irregular-staying migrant that is imposed a duty to report is registered by the police on the date of its application, along with their other characteristics. Whenever the application of the measure ends, the end-date is recorded together with the reason for the decision as the outflow. It can thus be the case that there are several registrations of inflow and outflow within the return process of specific irregular-staying migrant, as they can be imposed with several measures during the return process (see also Figure 13).

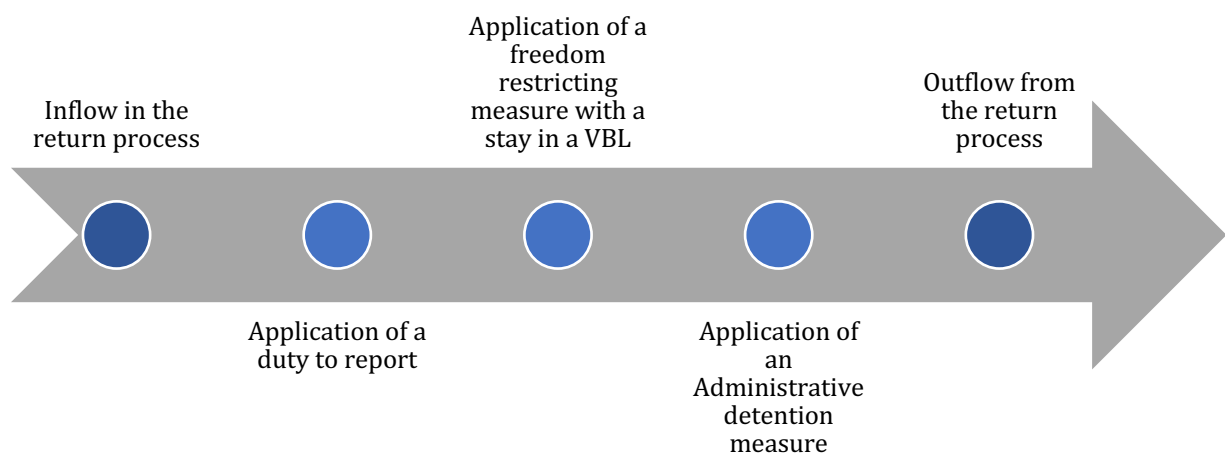


Figure 13. Hypothetical process of an irregular-staying migrant in the return process

The example above is a very straightforward example, one that is sadly not reflected in practice. Some of these issues are described in the discussion (see also Chapter 5). However, one of the

problems that emerged from the interviews that are of interest for this section entails the difference in the intended output of the supervisory measures that were implemented as an addition to the detention on the compliance rate. The program theory described that one could consider the output of no departure or a demonstrable confirmed departure of an irregular-staying migrant as an indicator for the case resolution rate. Interpreting the output of the other form of departure, leave for an unknown destination (or 'Met Onbekende Bestemming' (MOB) registration in the Netherlands), becomes more complicated as it is unknown if the irregular-staying migrant has left the Dutch territory or not. Furthermore, the outcome of a return process is also registered as a MOB if, for example, an irregular-staying migrant absconds from a duty to report. The MOB output thus makes a further interpretation of the related outcomes more difficult.

Efficiency output

As for the efficiency outputs, the findings indicated that the duration of the measure as an output should be moved from the efficiency category to the practical category of the outputs (see also Figure 10). The duration of the application of a measure as an output is relevant for both the effectiveness and necessity outcomes of the program. Though, the interview data indicated that the duration has more to do with the other practical outputs. A concern that was expressed by one of the respondents was about the view that the implementation of the measures had resulted in an extended duration of the return process. In some cases, these measures could function as to extend the process of someone who does not want to return, enabling them to stay longer in the Netherlands. Only when sufficient grounds are given is it possible to start a forced return procedure and impose an administrative detention measure. As expressed by a respondent:

If I look at it that way, but I have become pretty cynical about it, then I would think that it just provides an elongation of a stay of an irregular-staying migrant in the Netherlands and that it eventually leads to more detention. Because it has become a means to stretch the duration (of the return process), as we can't enforce (the return). (JV1, Q16)

Humanitarian output

As for the humanitarian output, a common concern shared by the respondents was how any such output would show itself in practice. Most respondents considered it a fundamental underlying assumption that less-restrictive measures pose less of an adverse effect on the wellbeing of an irregular-staying migrant. A possible problem for the implementation of the program is therefore if that assumption is enough, or if an effort should be made to clarify what kind of output one

would expect when considering the effect of the program on the circumstances of the irregular-staying migrants.

4.2.6 Outcomes & objectives

A number of issues were identified in the interview data regarding effective program implementation. These findings on the outcomes and objectives are bundled together in the following part.

One recurrent view from the respondents had been that, although a critical intended outcome (especially as a part of the rationale behind the program), the efficiency related outcome has become more of an underlying outcome in the background. More emphasis is put on the outcomes on the circumstances of the irregular-staying migrant and the practical results.

Most importantly, the influence of recent political developments will likely influence the interpretation of the practical outcome of the supervisory measures for the foreseeable future (see also Chapter 4.2.1). Several respondents explained that one could interpret the outcomes from the output of MOB in two different ways. The general point of view was that these measures should contribute to the general aim of the return process to, simply put, effectuate more return. This included both the more preferable demonstrable departure output as well as the, although less preferred, MOB output. For example, one respondent said:

As long as the irregular-staying migrant has an obligation to leave the Netherlands you could say; when someone goes MOB [...] that someone has stayed behind illegally, or you could say that someone has met his obligation to depart until it is proven otherwise. (NP1, Q17)

The first “negative” view would be that someone has stayed illegally in the Netherlands or somewhere else in the EU. The second more “positive” view from the government perspective is that it is still a form of return, until proven otherwise. The comments from the respondents indicate that previously, the more “positive” view as an interpretation of the output and outcome for the supervisory measures. However, a shift in the priorities of the current government has been noticed in practice. With the causes for the shift being attributed to recent political developments, as well as a rise in the number of MOB’s as an outflow output (see also Chapter 5.1.1). The new coalition government has expressed their intention to make more use of forced return and that demonstrable departure should be the priority. The intention has already been expressed in practice, as a new assignment letter for the R&DS included the provision that more work should go towards effectuating demonstrable departure. The risk for absconding is perceived as far greater with the application of the alternative measures. Thus, the new intentions would probably predominately interpret the MOB output in the “negative” view. And, those

maintaining the previous interpretation would regard the MOB output as more of a part of the overall output of the return rate.

A similar effect of the recent political developments can be expected as to the impact of the program. The individual assessment will likely incorporate higher demands for effectuating demonstrable return, making a choice for an alternative less attractive. Although one could argue that this is an adverse change, the respondents differ in their perceptions on the issue. The respondent from Amnesty International lamented the increased prioritization of demonstrable and forced return. Other respondents also expressed that it is unfortunate that the intention has moved away from sustainable return to forced and demonstrable return. However, other respondents argued that the application of an administrative detention measure has become too complicated, even when it is deemed necessary. In that way, the balance has been tipped towards the preference on alternatives to much. Administrative detention still has a function and is necessary for those that are unwilling to cooperate or fail to meet the other conditions. So, the hope is that by readjusting the intention to some extent towards forced returns and the application of administrative detention, a better balance is found between the consideration and application of the supervisory measures as a whole.

The following chapter comprises of a discussion of the further aspects of the present study on the basis of the results.

5. Discussion

The discussion comprises of the last three aspects of the study and the related research questions (see Chapter 3.1& 3.7). The first section of the program plausibility & data availability consists of the summative aspect of the discussion (see Chapter 5.1). The second section of program utility consists of the formative aspect (see Chapter 5.2). The third section gives several evaluation options (see Chapter 5.3). Incorporated quotes from the interviews are described in the same way as in the results (see Chapter 4.2).

5.1 Program Plausibility & Data Availability: Summative Analysis

The following part of the discussion is aimed at providing a preliminary assessment of the expected effects from the program. In addition, the second part of the section describes the data that is still needed to enable further evaluations. This section is part of the summative fourth step of the EA-model by Wholey (2015a) as used for the present study (see also Chapter 3.5.4).

5.1.1 Program Plausibility

An initial set of outputs and outcomes was formulated for the present study in the program theory & logic model (see also Chapter 4.1, Program Theory). These were compared with the interview data, as to see how the intended implementation corresponds to the implementation in practice. On the basis of that comparison, several outcomes are presented in Table 7. The outcomes are divided into the three outcome categories, proportionality, effectiveness, and necessity. Each of the included outcomes is described with the inclusion of relevant indicators, data descriptions, and initial assessment which will be discussed further.

Table 7. Overview Program Plausibility

<i>objective</i>	<i>Outcome</i>	<i>Indicator</i>	<i>Description</i>	<i>Preliminary Evaluation</i>
Necessity objective	Cost reduction	The average costs per day for each measure	The total organizational costs for the Ministry, related to the application of a measure per day, within the return process	General assumption that the application of an 'alternative' measure is more cost-efficient, although difficult to substantiate
	Simplification	Ease of application and uniform use of the same supervisory measures by each organization	The simplification and unification of the application of the supervisory measures in the return process	General indication of mixed results
Effectiveness objective	Application	Number of applications	The number of application of supervisory measures	Definite agreement that the 'alternative' measures are applied more
	Successful cases	The case resolution rate	The rate to which a measure is imposed when the return process ends and what kind of outflow is then registered	General concern that 'alternative' measures more often lead to unsatisfactory case resolutions, although difficult to determine the causal relation
	Compliance	Compliance rates	The rate to which irregular-staying migrants abscond during the application of a specific measure	General concern that the application of an 'alternative' measures increases the risk of absconding, although difficult to determine the causal relation
	Duration	Duration of the application of a measure	The duration to which a measure is imposed on an irregular-staying migrant during the return process	Some concerns that the 'alternative' measures are imposed for a longer consecutive time
Proportionality objective	Treatment of irregular-staying migrants	Well-being of the irregular-staying migrants	Physical and mental health effects of supervisory measure on irregular-staying migrants	General assumption of definite improvement, although difficult to substantiate
		Perception of treatment	Perception of treatment of irregular-staying migrants by relevant NGO's & Government agencies	General assumption of definite improvement, although only preliminary conclusions can be made
	Correct application of supervisory measures	Review of assessment	Disagreement with the imposed supervisory measure by the irregular-staying migrant	Indication that the number of objections is significantly lower for 'alternative' supervisory measures compared to the administrative detention measure

Necessity objective

As mentioned in the program reality section of the results, the outcomes that were included from the necessity category were found to be of less importance to most of the respondents (see also Chapter 4.2.6). There was a general assumption that the newly implemented or adjusted measures were cheaper to impose than the ADM. Be that as it may, that assumption was based on previous pilots and a resulting comparison between the average costs per day. One respondent

argued that the duration of these measures could be higher, as described concerning the duration outcome. If that were true, the overall costs of an ATD become higher than is previously assumed. Furthermore, the same respondent also added that:

However, I do think that you need to include all the related costs of an alternative (regarding a cost-effectiveness calculation). So, also, what does a departure supervisor cost, as he has to manage the process, and what does these cost for providing shelter and care [. . .] However, you can still think that the measures remain cheaper. (JV1, Q20)

Based on that statement and the views from the other respondents, it is most likely still the case that these measures require less financial resources. Though, if new calculations were done which include the aspects as mentioned earlier, the difference between the alternative measures and administrative detention on overall costs and the average per-day costs would probably be smaller.

The outcome of the increased simplicity and uniformity of the policy surrounding the supervisory measures was difficult to ascertain from the documents or respondents. These were organizational changes, designed to streamline the process of applying the supervisory measures. Commenting on the simplicity outcome, one of the respondents said:

The simplicity is not just related to time, but in that, you would also know that others within your organization are also applying it (the alternative supervisory measures), that you can use the same forms [. . .] and the simplicity is also that you could do a bit of profile development. (JV1, Q21)

The aspect of profile development was explained as being able to provide some structure in the application of the choice of a measure. Though it was also explicitly mentioned that in no form this would be an actual form of profiling, but that one would know that a family would go to a family center for example. Unfortunately, these intended outcomes of the program were considered unsuccessful by the respondent. As it was reported by the respondent that:

And that was also (the intention) with simplifying the process of application. That they would know; these are the established measures that we have. Those you can use but it never really got off the ground. That when you have characteristic A and characteristic B, you can apply measure C (with due consideration of the individual case of course). That, unfortunately, never happened. (JV1, Q22)

One reason that the intended outcome has not been seen in practice was that it would necessitate that the R&DS would get a leading role in the application of the ATD measures. Unfortunately, the R&DS is not legally authorized to apply these measures, as that is part of the role of the police as the organization with the competence/authorization to use enforcing measures. Another given

reason is that although some competence has been given to the R&DS, as they now have the authorization to apply the detention measure, some problems still exist. The difference in the priorities of the organizations of the National Police and the Royal Marechaussee compared to the R&DS can put a strain on the possibilities for intensive cooperation that is often necessary for the complicated return process. The respondents attribute the difference in priorities to the fact that the R&DS only needs to concern itself with the return process, while the National Police and Royal Marechaussee have many other priorities to consider.

Effectiveness objective

The outcomes of the effectiveness outcome were difficult to assess on their results, as the main difficulty with the outcomes was that the outputs were difficult to interpret. Many of the outputs are affected by other influences besides the application of a supervisory measure, and as such, it is difficult to determine the causal relationship with high certainty. For example, several of the respondents indicated that the number of applications of administrative detention has decreased in recent years and the number of other supervisory measures had risen. One could, therefore, assume that these two trends are related. However other factors were possibly also of influence, and current statistical systems cannot assess the different relationships to the necessary extent. One respondent reported that a factor for the decreased application of ADM had been the reorganization of the Dutch National Police. The reorganization was already mentioned as an additional contextual factor, which caused a decrease in the number of object-controls for locations where irregular-staying migrants illegally stayed. These kinds of factors do not show their direct presence in the statistical data and thus require extensive awareness of the contextual factors.

Similar issues present themselves with the successful case rate outcome and compliance outcome. The majority of respondents agreed that the alternative measures pose more of a risk of absconding supervision and being less effective as to effectuate return. The latter was mainly explained by the lack of pressure that can be exerted by a government agency during the application of a duty to report or bail. The freedom-restricting measure was more positively regarded as it provided more options to influence the progress of the return process. In addition, one respondent also felt that these supervisory measures are used as to extend the return process as long as possible and to postpone an eventual departure. Tough, these indications could not be substantiated by actual statistical analysis. One issue was the registration of MOB, as described in the program theory (see also Chapter 4.1.5). That registration can have multiple meanings, thereby making any accurate interpretation of the compliance rate or success rate of the measures improbable. Here too, many other influences could interfere with any interpretation of the outcomes as with the number of applications. The increase in the rise of MOB registrations is also

attributed to the characteristics of the irregular-staying migrants, such as their nationality, age and if they fall under the Dublin-regulation. Unfortunately, the current analytical database KMI+ of the Ministry of Justice only collects aggregate data on the supervisory measures. Thus, any data is not linked to a specific case/irregular-staying migrant, thereby making any analysis that could assess the causal relationship difficult.

A final significant complication is the fact that there is not "one" duty to report measure as previously stated in the comparison of the activities (see also Chapter 4.2.4). Therefore, respondents found it difficult to determine the results of the duty to report measure.

Proportionality objective

The proportionality category of the program, as formulated in the program theory, involved the expectation that the policy program would result in the possibility that less-coercive measures can be applied that would still ensure the primary function of these measures to provide adequate supervision. Consequently, it is a general assumption of the respondents that the "lighter" measures have less of an impact on the mental and physical health of irregular-staying migrants as compared with the administrative detention measure. Though, they find it hard to substantiate these views, as most of them agree that any evidence is difficult to find. However, the interview data provided a preliminary and positive assessment of the perception of an NGO and relevant government agencies towards the improved circumstances of irregular-staying migrants. Significant steps are made towards the provision of administrative detention as an '*ultimum remedium*' with the addition and improvements of the other supervisory measures. Though, some things can still be improved. Amnesty International especially expressed that advancements are still necessary as to improve the overall regime of the supervisory measures, to provide better guidance besides supervision and that the duration of the application of the freedom-restricting order and administrative detention measure should be shorter. Providing better guidance of the irregular-staying migrants within the return process is also something other respondents provided as a concern. With some feeling that there were too many instances in which someone is imposed with a return decision in 28 days, duty to report or bail without additional guidance. As expressed by one respondent:

A duty to report on its own is ineffective. Because why is the FRC so effective? As you have to report there, you get return supervision from the R&DS, and you have a place to stay. So that together ensures that you feel safe, you have your basic needs, and there is a form of control.
(JV1, Q18)

Another outcome of the proportionality category is the level of correct application of the supervisory measures. The measures are supposed to be individually assessed for each case,

where several conditions are considered that are the foundation for a particular measure. An additional indicator for the proportionality outcome is thus if the application of the measure is in fact deemed proportional by the judiciary. An irregular-staying migrant has the right to request a review by a judge on the application of a supervisory measure. These reviews have seen more critical demands by the judges for the government agencies to provide more substantiated considerations of the assessment according to several of the respondents. In addition, the respondent from the legal department of the INS stated that the number of objections on the application of an alternative measure was far lower than with the administrative detention measure. The comment below illustrates these findings:

You do not often encounter any appeals against [...] the (alternative) supervisory measures in the "process representation field [...] for example, at a freedom-restricting location such as Vlagtwedde [...] and against a duty to report [...] those we do not perceive at all, as there are seldom any procedures against those kinds of decisions. And these lighter measures than detention lead much less often to disagreement with the irregular-staying migrant themselves. (IND1, Q19)

The lower number is of course also influenced by the fundamental difference in the willingness to cooperate between irregular-staying migrants imposed with an ADM or ATD. Even more interesting, a fair percentage of the objections for an 'alternative' measure was filed by the irregular-staying migrant to be imposed with an ATD instead of ADM or nothing at all. Especially the stay in an FRC was often requested, as to provide shelter and their basic needs. Objections in these cases were often filed as the R&DS was of the opinion that the irregular-staying migrant was uncooperative with the return process, and thus did not meet the necessary conditions for such a measure.

5.1.2 Data Availability

The previous section (Chapter 5.1.1) described the preliminary results on the outcomes, based on the comparison between the program theory and the program in practice. The following section is aimed to provide a more in-depth consideration of the measures and data on which the preliminary evaluation was based, while also mentioning those that are still needed. The section is divided into the three outcome categories, presenting an overview and considerations for the data availability of each.

Necessity data

The data needed for the necessity/effectiveness outcomes are based on the financial and bureaucratic resources required for the application of supervisory measures. The financial costs associated with the application of the supervisory measures are not measured continuously, as is also the case for the bureaucratic resources. Most of the data that is available on the financial costs were presented in pilot studies, in which broad estimates were made for the supervisory measures and their costs in total and per day. The pilot 'Placed in detention' (In bewaringstelling) by the R&DS was reported to include estimations on the costs of imposing ADM and ATD measures. The internal report on the alternatives to detention also include estimations on the costs of the applications. Although, all these estimations did not incorporate the expected longer application duration of the alternatives or the new situation. Limited information on the availability of the data necessary to determine the improved simplicity of the process was provided by the respondents or the documents, except the statement of one of the respondents that the implementation did not come through. The given reason was that this was due to the inability to solidify the leading role of the R&DS in the return process. Possible data sources should, therefore, be sought in the perceptions on the issue by relevant employees of the involved government agencies.

Effectiveness data

Almost all of the data measurement for the practical outcomes are based on numerical registrations of the application of the supervisory measures, and the inflow/outflow registration of the measure. These registrations are initially made by the involved government agency responsible for the application of a measure. Several statistical systems are used across these agencies, all collecting other aspects relevant for their purpose in the return process. For example, the following comment illustrates the data that is registered with the FRC:

Well, we (The R&DS at the FRC) have our own systems, right? And the most important one in the system is the departure file. And everything the departure supervisor does is registered. And the departure file is the most important, as it lists the entire history of the specific irregular-staying migrant. (DTV2, Q24)

There are several possible ways in which the imposed measure for an irregular-staying migrant can end. It could be that the migrant has left, either demonstrable or non-demonstrable. Additionally, the irregular-staying migrant can also be imposed another measure, at which the migrant is registered in the outflow of the previous measure and registered in the inflow of the other measure. Thus, it is entirely possible that an irregular-staying migrant is imposed with several measures in sequence. To make it even more complicated, irregular-staying migrants can also file a new application for asylum. These individuals would then be replaced in the asylum procedure. In sum then, many different types of outflow outputs are registered across the government agencies. With these registrations often being put in different systems for each of the government agencies. And so, the available data is there, but only difficult to aggregate.

Luckily, steps were made to centralize the exchange of information between the involved stakeholders. These developments resulted in the database system KMI+, which retrieves periodical data sets from each of the associated government agencies and combines these into a central data warehouse. Data is, for example, gathered on the application numbers of each of the alternatives, the numbers of inflow and outflow, the type of inflow and outflow, plus additional relevant categories. The agencies initially supplied the information on a monthly basis, while recent developments made it possible to supply the data on a weekly basis. And so, KMI+ is a potential treasure-trove of potential data for further evaluative purposes. However, respondents stated that KMI+ in its present form is unable to provide enough data to determine the causal relationships between the data. This mainly has to do with the fact that it is based on aggregated data, while it does not provide usable data on an individual case level. As commented by a respondent concerning the issue:

No, it is an aggregated database system. So, it is not based on V-numbers (the unique number by which an irregular-staying migrant is identified). We do provide the V-numbers, but the use of the datasheets is by non-recognizable V-numbers for the relevant government agencies so to speak. Though, one could look at the underlying V-numbers. That is not the primary goal of the collection of the database. (DTV1, Q25)

Thus, readily collectible information for each case is difficult to acquire at this moment. However, respondents do mention that developments were made to make such data available in the central KMI database.

Proportional data

As mentioned in the previous section, the proportionality outcomes were mostly based on the perceptions of the stakeholders from both the government agencies and Amnesty International. These are more qualitative measurements, gathered by conducting a low number of interviews. If a more accurate evaluation would be deemed necessary, data from three possible other perspectives are needed. First, more data needs to be gathered on the perceptions from both relevant government agencies and additional NGO's. These could be collected by conducting further qualitative interviews specifically on the proportionality of the program, or by conducting questionnaires with pre-formulated questions and answer possibilities. The advantage of the first measurement is that it provides more in-depth knowledge on the topic, while the latter can be spread to more relevant stakeholders and could provide more substantive and clear information. The data from the questionnaires are also more easily analyzed with the use of more statistical methods. Additionally, one respondent suggested that further interviews could include the perceptions of the assistant public officers and other stakeholders that are involved in the assessment of the choice for a supervisory measure. These perceptions would provide further insight into the considerations that were made, which conditions were prioritized, and how the category of vulnerable groups are assessed as previously expressed as a concern by Amnesty International (see also Chapter 4.2.4).

Additional statistical data can be obtained from the legal affairs department of the INS on the number of filed objections on all the supervisory measures, including ADM. These can be compared as to provide further evidence for the perception of the respondent of the INS that the number of objections is far lower for the other supervisory measures compared with the ADM. Though, the respondent did warn that its registration software INDIGO is not very usable as to run queries on the data.

Data availability on the perception of the irregular-staying migrants themselves or the effects of applying a supervisory measure on their mental and physical health would be fairly complicated according to the respondents. The first issue would be to get a population that is as “neutral” as one can be as an irregular-staying migrant in the return process. The following comment illustrates these difficulties:

Look, we (The R&DS) already tried to do a limited “customer-satisfaction” study by interviewing irregular-staying migrants. Like, ‘how do you consider the supervision you get during your return process? However, it depends to a great extent on the setting in which someone is approached. Someone who is placed in detention or a restrictive measure will react differently, in general, then someone with a duty to report. (DTV1, Q23)

Retrieving relevant data after the return process is also tricky, as these people have departed the Netherlands, left for an unknown destination or moved to another immigration process. Several respondents agree that as of now, not a lot is known about what happens to irregular-staying migrants after the ending of a return process. With some regarding it as a knowledge gap or black-hole for both government agencies or other stakeholders. One study in 2013 'From treatment to departure: An investigation into the operation of immigration detention' [Van bejegening tot vertrek: Een onderzoek naar de werking van vreemdelingenbewaring] by Alphen, Molleman, Leerkes, & Hoek (2013) did conduct surveys with irregular-staying migrants that were placed in detention. Though the study did incorporate some aspects of the physical and mental health influences of the measure, the aim of the study was more directed towards the contributions of involved personnel. An important step that the study incorporated was to conduct the surveys with irregular-staying migrants with a diverse range of their individual characteristics, such as age, nationality, and more. Thereby limiting the possibility of any other factors that could influence their perceptions. Still, many different options of surveys and questionnaires exist that could be used to measure the mental and physical health perceptions of individuals, which could be adapted to evaluate the aspect for irregular-staying migrants imposed with any of the supervisory measures.

5.2 Program Utility: Formative analysis

This section comprises of the formative aspect of the study (see also Chapter 3.1). It gives the recommendations that are needed for the implemented program to improve the causal links between the program components and, to increase the possibility for the use of further evaluations of the program. This section is based on the fifth step of the EA-model as is used in the present study, combining the information from the previous chapters (see also Chapter 3.5.5).

5.2.1 Recommendations regarding program components

A necessary first change involves the causal links between the program components. The return process has a highly complicated structure, with many involved agencies that differ in their approach, priorities and the tracking & recording of data. One of the intended outcomes of the implementation of the program was to increase the simplicity and uniformity of the supervisory measures. As mentioned in the previous sub-chapter (Chapter 5.1.1), that has not been the case in practice. The question would thus be, what can be changed to enable the activities of the program to result in the outcome?

A first recommendation is to further increase the role of the R&DS as the central hub of the return system. Although the organization has already been established as the central implementing organization for the return policy, some further improvements would be necessary regarding the supervisory measures. As previously mentioned (see Chapter 4.2.1), the R&DS is still reliant on other organizations to impose the duty to report, bail or the taking of documents measure. The applications of these measures are the responsibility of the National Police. Recent reorganizations have shown that the priority of such an organization has to shift to other areas. As the R&DS already has obtained the ability to impose the administrative detention measure, one could argue that enabling the R&DS to impose the other supervisory measures as well already provides further centralization. The benefit of which is that one central organization is responsible for the application of the supervisory measures, while the other organizations could still function as to assist in effectuating the supervision. This would also aid in the centralization of the data registration, as further evaluation would benefit significantly if one organization is the central actor within the return process of each case. As seen in Figure 13, an irregular-staying migrant now “moves” from the registration of each of the involved organizations. These moves are registered as outputs by the R&DS in contrast to the aggregated registrations made in KMI+. Though, the precise reasoning behind the move to another measure is not captured in the registrations. And therefore, only the numbers applications and their ratio can be assessed as of now. Thus, to enable the possibility to make any determinations on the causal relations between the statistical data, it is; (A) Of added value if the R&DS is responsible for the application of all the

measures; and (B) necessary to include the procurement of the reason for a switch to another measure in current databases.

Another recommendation is to establish a common terminology, as the used terms are often interpreted differently by the involved organizations. As one respondent states:

The terminology is often another problematic issue. It would be nice if we could all agree on what is meant with a forced return, voluntary return, Removal. So that we mean the same thing, from the Council of State to the Aliens Police Department. (IND1, Q27)

A recommendation that is already worked on by those implementing the data warehouse KMI+, is to get the organizations who supply the data sets to use the same terms.

5.2.2 Recommendations regarding the collection of data

A second recommended change would be to enable one-off, or more preferably, repeatable follow-up collections of information on those that have departed either demonstrable or (non-) demonstrable. Information that can be obtained from irregular-staying migrants who fall into the category of demonstrable departure could be less biased and also provide a better overview of the whole return process. As mentioned earlier, previous attempts to analyze the perceptions of irregular-staying migrants were attempted with individuals who were still in the return process under one of the supervisory measures. And this increases the chance of getting socially desirable answers from the irregular-staying migrant, as they may feel that this could affect their return process if they cooperate. Coming into contact with people who fall in the (non-) demonstrable return category is likely difficult or impossible. However, improved information exchanges between EU countries, together with the further analysis of the MOB output, could provide further analysis of these kinds of returns in relation to the supervisory measures.

5.2.3 Recommendations regarding the terminology of ATD

Another, and possibly most important, of a recommendation would be that the supervisory measures that were implemented as the “alternatives to detention” are to be regarded by all relevant stakeholders as part of the range of supervisory measures, including the administrative detention measure. A current issue is that the other measures are regarded as real replacements for detention. And while that was initially the case in the earliest stages of the program, that is not the case anymore. As commented upon by a respondent: *“The policy was meant as to provide structural possibilities for all involved stakeholders to impose an alternative when detention is not in question”* (JV1, Q27).

Figure 14 presents a broad interpretation of the common perceptions of the ‘alternative’ supervisory measures from before, after, and a misrepresentation of some, all based on the

comparison between the interviews and program theory. According to the respondents, the situation before the policy was as is broadly described by part A of Figure 14. After the assessment of the case, it was most often a choice between a freedom-restricting measure with a possible stay in an FRC or administrative detention. A duty to report was also already implemented but in a less structural and limited manner. The expulsion order and taking of documents were more of a procedural measure. A common misperception is that the program intended to let the measures function as a replacement of the administrative detention measure as presented by part B of the figure, while this is not the case. The current and eventual intention of the implementation was to provide a step-wise range of options that can be chosen from, depending on the conditions and characteristics of each irregular-staying migrant (Part C, of Figure 14). Each of these options has their own level of restrictiveness, conditions, and characteristics. While these all serve the same purpose, as to provide supervision of the irregular-staying migrant in the return process. And thus, a pre-selection is made for each of the measures as to who is imposed with one. Especially for vulnerable target groups, a more critical assessment of a measure is made. While special options are provided for families with underage children and unaccompanied minors to stay in the family locations when they have not left after the legal departure term of 28 days, and a closed family center for enforcing the return of those that can be effectuated within two weeks. Although there is some disagreement if these are alternatives or not, it is best to see these as separate provisions for providing supervision that is less of a strain on the children involved.

It is thus crucial for the Ministry of Justice & Security and the other involved government agencies to emphasize that the 'alternative' measures are now part of a package of options, including the administrative detention measure. With each measure having its own function as a supervisory mean in the return process.

5.2.4 Recommendations regarding maximum application duration

A final recommendation that can be made is to implement some form of end term for the application of the duty to report measure. Several of the respondents stated that they find that the measure is difficult to enforce if there is no pre-stated end term of the measure, making it difficult to exact control on the return process and follow the developments made by the irregular-staying migrant necessary to depart.

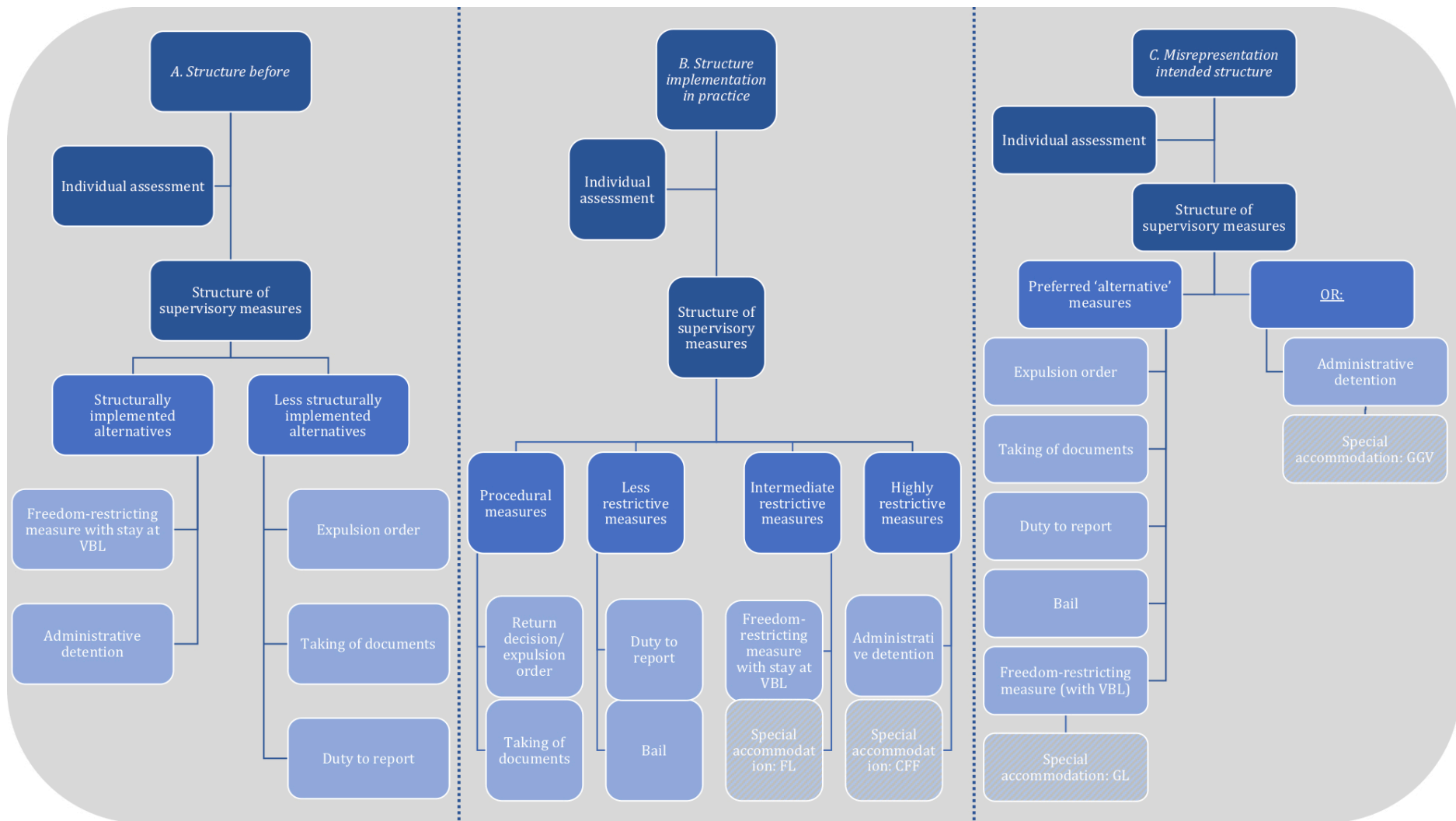


Figure 14. Overview of the different structures of the supervisory measures

5.3 Program Evaluation

The following section presents several evaluation options that can be used by the Department of the Directorate-General for Migration as the suggestions for further evaluation. Several options are presented in Table 8, while detailed descriptions can be found in the Appendix (see Appendix D). The section begins by describing the ‘status quo’ option, in which no evaluation is undertaken. And thus, that option explains what the benefits are of any further evaluation.

5.3.1 The status quo: Perception based supervision

We didn't have such a clear overview of all the unique measures that were applied before all of this. So that; what was imposed and by whom? And does it work or not? What leads to a return? And I haven't found a single research by which you see that on a large scale. It always targets one measure, that worked (only) in a pilot project. (JV1, Q28)

The quote of the respondent above highlights one of the issues that surround the alternative measures. The less-coercive measures are, in general, perceived by governments, NGO's, advocacy groups, international organizations, and the scientific field as more cost-effective, more humane and sufficiently effective in comparison to the administrative detention measure. However, most of the information is based on pilot projects, targeting one measure with (most often) a preselected limited target group. This is entirely understandable, as these kinds of pilots can be controlled more easily from the influences that make an evaluation in practice highly complicated. However, this also has the consequence that the implementation of these measures is most often based on the presumption of the benefits and consequences of these measures. Especially in the return process, as previous literature has mainly addressed the use of such measures in the entry/asylum process. Based on several, mostly qualitative, studies and common knowledge, it is generally presumed that irregular-staying migrants are treated more humane with the use of the alternatives. Several studies have calculated the cost differences between detention and other measures, with (often broad) estimations indicating that the alternatives cost less. Based on the pilots, it is presumed that the alternatives lead to an increase in the risk for absconding and non-departure. Though, the shared view is that these are sufficiently efficient while providing the additional benefits on the previously mentioned aspects.

If unevaluated, the program would likely continue without any major crises. Effectuating return is a highly complicated process, with many external influences that can change the outcome. It is, therefore, necessary for the involved government agencies to incorporate a high level of flexibility and reliance on the assessments made by the involved officials. The individual assessment is dependent mainly on the experience, knowledge, and perception of the departure supervisors, assistant public officers, and other executive officials.

However, further evaluation can provide an additional source of information on which the application of the supervisory measures is based. Evaluation of the proportionality of the program can assess the progress that is made in the treatment of the irregular-staying migrants. The resulting data can be used to show that the implementation of the program was worth it and that the Ministry of Justice & Security has adequately reacted to the necessity to change the return policy. The data can also be used to enable a better application of the supervisory measures, as it could provide broad indications for which measures offer the lowest impact on the well-being of the irregular-staying migrants. And also, to enable a more precise assessment and prediction of the effects of a measure on an individual's well-being.

Evaluation of the efficiency/necessity of the program could confirm the claim that the program leads to a more cost-effective return policy. Similar to the proportionality evaluation, it can also provide additional information that makes the application of the measures more precise. When there are accurate estimations of the costs per day on the use of supervisory measures as they are implemented in practice, these can be combined with the information on the proportionality and effectiveness. Thus, a full determination or categorization can be made as to the total and 'per day' costs of each measure in the application. And that can be incorporated into the individual assessment. It should be stated that a prevailing opinion of the respondents has been that the cost-efficiency is of lesser importance than the other aspects.

And finally, the effectiveness evaluation is the one with the most diverse range of opinions from the respondents. The number of applications and the result of the return process are aspects that form the basis of the effectiveness assessment of the related governmental organizations. It is also an important part of many of the pilot-and other scientific studies. However, it has become apparent that the possibilities to go further in the analysis of these practical results is difficult to achieve on an extensive and comparative scale. It was, therefore, the opinion of some of the respondents that more effort should be placed on especially the proportionality evaluation. As an effectiveness evaluation requires significant resources, centralization of data and terminology, and a constant overview of all external factors, these practical limits do not weigh up to the possible advantage of the evaluation by some. Especially, because recent studies have already provided some indication that the measures themselves do not impact the result to a great extent. With the previously mentioned external factors having more of an influence on the result. However, other respondents provided several manners in which these issues can be bypassed, while also providing the view that further development of KMI+ could potentially offer the data for determining the causal link between the data. As is the same with the other two evaluations, the advantage of such an evaluation is to provide substantive evidence for the implementation of the program. In addition, the information on the effectiveness of the supervisory measures can be combined with the others to provide a more precise application of the measures. Knowing which

measure results in a more effective return by which category assists in the considerations made for the choice of a measure by an assessment of the individual cases. It is thus, as already mentioned several times to emphasize, not to enable a form of profiling of the cases (with all its negative associations). But further evaluation should provide a structural background of tangible information that can be used as a filter for the selection of appropriate measure, with the final choice depending on the case-management by the involved executive officials that are responsible for imposing the measure.

5.3.2 Overview of the evaluation options

Several evaluation options were formulated on the basis of the information retrieved by the present study. The evaluation categories divided these options, so included options are the proportionality, effectiveness, and necessity of the program. Therefore, the following section does not present separate versions of an integral plan, but each option describes one approach or aspect to evaluate. It is thus possible for the department of the directorate general for migration to “pick and choose” or adept the evaluation to their preferred approach.

Each option provides information on the original four aspects as advocated by Wholey (2015a) in addition to two added aspects with the present study: (1) the Purpose, (2) the Data collection, (2) the Data availability, (3) Data method, (4) Estimated resources, and (5) Usability (see also Chapter 3.5.6). Table 8 presents an overview of these options, while the full in-depth descriptions can be found in the Appendix (see Appendix D). It should be noted that the overview and descriptions of the options are estimations and suggestions and need further elaboration or adapted for actual evaluation.

Necessity options

Two options are listed in Table 8 for the evaluation of the necessity of the implemented program. Similar to the proportional options, these options evaluate different aspects of the outcome. Based on the comparison of the interview data and the documents, the financial outcome was a relevant and important part of the rationale for the implementation of the program (see Table 11, Option 1). However, the view of most of the respondents is that the actual outcome on that aspect no longer has the priority and attention of the Directorate-General, nor the political domain. It is thus determined that the first option is of low importance for being incorporated in an evaluation in the near-future, based on the current situation and influence of contextual factors (see also Chapter 4.2.1). The second option was deemed more important, as it was one of the leading rationales for the implementation of the program (see Table 12, Option 2). The respondents have stated that organizational limitations inhibited improving the return policy in that regard. Further

evaluation of the current situation could provide other possibilities for reaching this objective, while also being easier to effectuate such an evaluation.

Effectiveness options

Two options are included for the effectiveness evaluation of the implemented program, a more realistic simple evaluation (Table 13, Option 1) and a rigorous evaluation (Table 14, Option 2). In contrast to the options from the evaluation of the proportionality, the second option is a continual and more in-depth addition to the first option. Much of what is proposed in the first option is already being done to a greater or lesser extent. The number of applications and outflows of the supervisory measures, as well as the type of outflows and the overall outflow of the return process, are already registered in the data warehouse KMI+. However, the first option wants to gather that data over multiple years and analyze these as part of a study specifically on the supervisory measures of the implemented program. The reason that it is still relatively complicated, as seen is that one missing variable of added value is missing. The present study has not been able to find any evidence that KMI+ includes data on the (average) duration of the application of each of the supervisory measures (including ADM). Further evaluation should include that variable, as the duration of application is an important factor in the determination of the basic effectiveness of the implemented program. The complicated second option has the potential to provide an incredibly compelling result if it can be accomplished. And not only for the situation in the Netherlands will it be interesting, since there is also an international need for more comparative studies of the practical results of the (alternative) supervisory measures in practice. However, the second option is determined to be of secondary importance by the present study. The respondents are of the opinion that the chance is high that the differences in the results on the effectivity are insignificantly low for the “alternative” measures as compared with ADM. In addition, they would argue that the results on the proportional, and to a lesser extent the necessity, aspects are of more value. The general discourse is most wonder if the possible results of such an evaluation are worth the complicated and resource intensive effort of implementing such an evaluation. One way to simplify the option is to only include the data from the freedom-restricting measure with a stay in the FRC & FL and the ADM. These measures have more data already available, or their characteristics make it easier to obtain additional data compared to measures such as the duty to report and bail measures (see also Chapter 5.1.2).

Proportionality options

As can be seen from Table 8, four evaluation options are presented from the data retrieved for the present study regarding the proportionality of the implemented program. These options all incorporate separate aspects of the proportionality, although all are aimed at evaluating the just application of supervisory measures. In addition, these also incorporate the analysis of the effects of the application of these measures on the treatment and well-being of the irregular-staying migrants. As such, a high priority is given to incorporate one or multiple options into an evaluation, as the proportionality of the implemented program should be the most critical outcome to clarify according to the respondents. The four evaluation options consider the individual assessment (Table 15, Option 1), the opinions from all involved stakeholders (Table 16, Option 2), the effects on someone's physical and mental health (Table 17, Option 3), and the improved communication of the supervisory departures (Table 18, Option 4). A balance was struck between the priorities and the complexities related to evaluating these aspects. Option three includes actual interviewing or administering of questionnaires with irregular-staying migrants that are (or were) applied with a supervisory measure. Such an approach is fairly complicated and resource-intensive to accomplish. The second option can also include interviews with irregular-staying migrants but on a smaller scale. The estimation of the complexity and resource intensity was based on the average between including and excluding respondents from that group. In addition, it would be advised to limit the scope of the third option to the health effects of the irregular-staying migrants. Including other aspects (e.g., treatment, general view of the Dutch return process) would increase the risk of certain biased responses from the interviews or questionnaires. The increased risk is more likely when the irregular-staying migrants think that the answers they give have any influence on their circumstances in the return process.

Table 8. Overview of the Evaluation Options

<i>Option</i>	<i>Purpose</i>	<i>Data collection</i>	<i>Data availability</i>	<i>Data analysis</i>	<i>Resources</i>	<i>Utility</i>
Necessity evaluation						
(1) Costs vs. Costs (Table 11)	Evaluation of the total costs and costs per day for the measures in practice	All (estimated) associated costs of applying each supervisory measure	(p): Previous studies with cost-estimations (a): Data from option 5 Estimation of actual costs of measures in practice	Estimation/calculation of the total costs for the number of applications Estimation/calculation of the average total costs Estimation/calculation of the average costs per day	Access to financial data One or multiple researcher(s), preferably with a financial background Time: 6-9 months	<i>P:</i> ● <i>C:</i> ● ● ● ● ● <i>RI:</i> ● ● ●
(2) Possibilities for Simplicity (Table 12)	Evaluation of the possibilities to streamline, simplify, and centralize the application of supervisory measures	Perceptions from practical, policy, and judicial fields on the possibilities to 'simplify the process'	(p): Some preliminary indications from the present study (a): In-depth or expended data on the specific aspect	Qualitative analysis based on multiple interviews with respondents OR Quantitative analysis of questionnaires completed by respondents	Respondents Questionnaires One researcher Time: 6-9 months	<i>P:</i> ● ● ● <i>C:</i> ● ● ● <i>RI:</i> ● ●

Effectiveness evaluation						
(1) The Basics (Table 13)	Evaluation of the basic variables of application, duration and outflow rates of the supervisory measures	The number of applications and resulting outflow, duration of applications and overall return processes, and type of outflow of all the supervisory measures	(p): Aggregated monthly data on the number of applications, outflows, and type of outflow in KMI+ Preliminary indications of contextual influences by the present study (a): Aggregated <u>weekly</u> data is underway Aggregated data on the average duration of application of the measures and return processes	Basic quantitative analysis over multiple years (as far as 2013) In addition: a limited qualitative overview of all other influences that need to be incorporated into the interpretation	Access to KMI+ data warehouse Basic statistical software Stakeholders with a good overview of the influences Researcher, with experience of statistical analysis Time: 3-6 months	P: ● ● ● C: ● ● ● RI: ● ●
(2) A Step into the Unknown (Table 14)	In-depth analysis beyond option 1 on the influence of the supervisory measures on the outflow	Data on multiple variables on V-number (individual case level)	(p): Aggregated data available in KMI+ Individual case level data available for several of the variables (a): Structural operationalized registration of indicators for well-being Most data registered in underlying datasets for KMI+ or in databases from the government agencies	Complex quantitative analysis: Binary Logit Model Discrete Time Logit Model	Access to KMI+ data warehouse Access to databases from government agencies Complicated statistical software Powerful hardware for the calculations Multiple researcher(s), with extensive knowledge and experience with sophisticated statistical analysis Time: 12 months	P: ● ● C: ● ● ● ● ● RI: ● ● ● ● ●

Proportional evaluation						
(1) Assessing the Assessors (Table 15)	Evaluation of the individual assessment	Considerations for the application of a measure	(p): Databases involved organizations as well as aggregated data KMI+ (a): Opinions/views executive officials	Basic quantitative analysis of database information Qualitative analysis of interviews OR Quantitative analysis of questionnaires	Access to database Respondents Questionnaires One researcher Time: 6-9 months	P: ● ● ● ● ● C: ● ● ● RI: ● ● ●
(2) Substantiating the Perceptions (Table 16)	Evaluation of the treatment of irregular-staying migrants	Perceptions from stakeholders (Government, NGO, Irregular- staying migrant)	(p): First indications present study (a): In-depth or expended data on the specific aspect	Qualitative analysis of interviews OR Quantitative analysis of questionnaires	respondents questionnaires one or multiple researcher(s) Time: 6-9 months	P: ● ● ● ● C: ● ● RI: ● ●
(3) Limiting the Consequences (Table 17)	Evaluation of the effect of a supervisory measure on the well-being	Perceptions of those that experienced the application of a measure	(p): Limited findings from previous studies (a): Accurate data on the actual measures used in practice in the Netherlands	Preferably: quantitative analysis of questionnaires Alternative: qualitative analysis of interviews	Available & cooperative respondents (scientific) Questionnaires Multiple researchers/eval uators Time: >12 months	P: ● ● ● C: ● ● ● ● ● RI: ● ● ● ● ●
(4) Professionalizing of Communication (Table 18)	Evaluation of the adapted case- management program by the R&DS	Perceptions of departure supervisors trained with the methodology in practice	(p): Implemented program by R&DS (p) (a): Follow-up data on the results of the program	Quantitative analysis of questionnaires In addition: a qualitative analysis of a small number of interviews	Respondents Questionnaires One or a pair of researcher(s) Time: 6-9 months	P: ● ● ● ● ● C: ● ● RI: ● ●

Note. Adapted from Wholey (2015a). Data availability: (p)= present, (a)= absent). Utility: (P) =Purpose, (C)= Complexity, (RI)= Resource Intensity.

6. Conclusions

The conclusions comprise of a description of the main findings (Chapter 6.1), limitations (Chapter 6.2), implications (Chapter 6.3), and recommendations (Chapter 6.4). The given implications and recommendations have been made in general and mostly concerned the scientific or societal field, as the discussion already included the practical implications and suggestions for the specific program under study.

6.1 Main Findings

The present study analyzed the adapted policy program of supervisory measures in the return process of the Netherlands from 2013. The intention of the supervisory measures policy program in the Netherlands was to improve the application of supervisory measures for those irregular migrants in the return process. The rationale was that the policy program would lead to a better individual assessment and choice of application of a measure, by providing more measures and implement these in a structured approach. The central research question that guided the current research has been:

To what extent is the Directorate-General for Migration of the Netherlands ready to evaluate the adapted policy program of the use of supervisory measures in the return process?

The central research question comprised of several underlying research questions, each directed towards one of the five aspects of the research design of this study. The study has incorporated an Evaluability Assessment Model by Wholey (2015a) to determine if the adapted policy program is ready to be evaluated. A program theory and logic model were formulated as a first aspect that described the intended components of the program and how these relate to each other. After which, a comparison was made between the proposed program and the actual implementation in practice by interviewing several stakeholders by the second aspect. These two aspects formed the results of the present study. The discussion consisted of the summative assessment on the likeliness that the components of the adapted policy program would lead to the desired outcomes and objectives. And, the formative analysis which provided the necessary recommendations for enabling further evaluation. A final aspect concerned the presentation of several evaluation options. The most significant findings of these aspects for preparing the adapted policy program further towards its full evaluation can be formulated as follows:

- **Main summative finding:** The comparison from the *results* shows that in practice shows that the assessment of the effectiveness of the program is mostly done based on the assumptions by the stakeholders and limited quantitative data from the pilots, different

data systems and the aggregated data from the data warehouse KMI+. The *summative discussion* presents the relevant objectives, outcomes, and outcomes which should be included in the further evaluation, as well as the necessary data process for such an evaluation.

- **Main formative finding:** The *results* show that the previously named “alternatives to detention” should now be regarded as integral parts of the range of supervisory measures. Each of these measures (including ADM) has its own characteristics, conditions, and purposes. The *formative discussion* recommends that the Ministry of Justice & Security clarifies that these measures are not intended as actual replacements of ADM but rather function as “less-coercive measures.”

These findings resulted in the presentation of several evaluation options for evaluating the necessity, effectiveness, and proportionality of the adapted policy program of supervisory measures in the return process of the Dutch immigration policy.

6.2 Main Limitations

The present study incorporated a framework of an Evaluability Assessment Model by Wholey (2015a). Some limitations were found concerning the use of an EA and the adoption of this specific framework. The adapted framework is one of several different variations with their own characteristics and approach (see also Chapter 2.2). This increases the flexibility of the EA method, as a variation can be incorporated with the best “fit.” Furthermore, most models are described by a broad framework, which increases the flexibility to adapt such a model to the program even more. Some adaptations of the EA-model were made for the present study, by including the framework for the interview-guides from Trevisan & Walser (Smith, 1989). Furthermore, the present study included two additional aspects for the evaluation options, as the original four aspects were deemed too limited to provide a decent overview of each of the options.

Another limitation of the EA approach is to find the appropriate balance between the scale of the EA and how deep the analysis of the specific aspects can go. The study had a broad scope, considering many different aspects of an evaluation of the supervisory measures as part of the policy program. An EA is intended to be a quick form of pre-evaluation, thus necessitating that any broad-scoped EA remains shallow on the depth of the analysis of each of the aspects. Any actual evaluation can limit their scope and provide in-depth analysis of one of the aspects of the policy program. Thus, the present study tried to balance the depth of the analysis to give a general overview of all the aspects of the policy program, providing in-depth information when necessary. Other EA-models, such as the one by Smith (1989), give more attention to delineating the scope of the EA. Still, the model by Wholey (2015a) was included as it provided a better combination between the summative and formative aspects of an EA.

The model was adapted to the specific characteristics of the immigration policy field in the Netherlands and the policy program under consideration. Therefore, any generalization of the findings to other supervisory measures policy programs or the broader immigration policy field from other countries remains limited. Studies that involve the policy programs from other countries in the EU can more easily use these findings, as the 2008/115/EC Return Directive ensured an increase of similarities in the return policies of the EU member states that fall under the directive (EUR-Lex, 2008)..

Other limitations were more related to the process of implementing the EA-model. The methodology includes some of the steps that were made to increase these aspects, such as: the use of the use of semi-structured interview guides, a codebook based on concept-driven & data-driven coding, and the use of interview techniques that increase the validity (D'Ostie-Racine et al., 2013; Gibbs, 2007).

The present study conducted a limited number of interviews, with stakeholders that were seen as the most important to obtain information from (see also Chapter 3.2). In addition, an interview was conducted with someone from Amnesty International to provide a balanced view of the policy program.

One of the interviews had a corrupted recording of the second part, due to some technical issues. The problem was quickly recognized afterward. A memo was drafted after the interview based on the notes and recollection of the interviewer after recognizing the problem an hour after the interview. The memo was sent to the respondent for verification.

Lastly, the interviews were conducted in the native language of the interviewer and respondents. Any quotes used in the results were translated from Dutch to English while trying to remain as close as possible to the original meaning. However, it is possible that some translation issues are present in these quotes. The original Dutch quotes are, therefore, listed in the Appendix (see Appendix C). With the English and Dutch version of each quote assigned a specific number.

6.3 Main Implications

Despite these limitations, the findings of the present study still provide some interesting implications for both the scientific and societal field beside those already mentioned in the discussion. Evaluability Assessments are already made across a vast range of different policy fields. So far, scientific research on the supervisory measures has concentrated on the investigation of, and advocating for, the implementation of less-coercive measures or the “alternatives to detention.” However, the scientific analysis of the implementation of these measures in practice has only recently begun (Bruycker et al., 2015; Chmelickova, 2006; Directorate-General for Migration and Home Affairs, 2013; EU Commission, 2014; Field & Edwards, 2006). The complexity and lack of structural data gathering with the implementation of these measures have made it difficult to accurately determine the efforts that have been made in the EU. The present study provides a specific consideration of the readiness of the related adapted policy program in the Netherlands. However, these findings could be useful for other researchers that wish to conduct evaluative research for other cases related to supervisory measures, the return process, or immigration policies. Recent situations such as the Migration Crisis of 2015 have shown the necessity of well-implemented immigration policies across the EU. The ‘alternative’ supervisory measures provide many advantages over the administrative detention measure as additional and preferred options for providing supervision in the return process. Luckily, many steps have been made by the Netherlands and other EU member states to implement and improve the use of supervisory measures in the return process. Though, little is still known about the effects of the implementation in practice. The findings of the present study are meant to aid the process of gathering more substantial information on the implementation of the measures in practice. Thus, hopefully leading to a contribution of the present study to increasing and improving the use of supervisory measures in the return process of the Netherlands and beyond.

6.4 Main Recommendations

The discussion of this study already provided some recommendations and presented several evaluation options, though specifically for the Directorate-General for Migration and the policy under consideration (see also Chapter 5.2). A general recommendation regarding the evaluation of supervisory measures in the return process would be to focus on the evaluation of these measures across Europe. The 2008/115/EC Return Directive was an important influence on the implementation of the policy program in the Netherlands. That influence is likely to be found in other EU member states. Considering the evaluation from a European viewpoint should include two aspects. A first aspect is that more needs to be done to share information on the return process between different EU member states. Information on irregular migrants is often spread across several countries that were involved in the immigration process of the individual. Accurate information on, for example, non-demonstrable departure necessitates the ability to find additional information on an irregular immigrant from the other EU Member States. A second recommendation would be to draft general guidelines for monitoring and evaluating the use of the range of supervisory measures including ADM. As immigration policy is increasingly being influenced by EU-level decision making, clear and universal concepts, indicators and measurements could increase the possibilities for comparative research on the use of the measures.

The scientific field surrounding immigration policies has grown dramatically over the recent decades as immigration has become a higher priority as a policy domain with its own unique issues and complexities. Some progress has been made by recent studies, comparing the return policies of different countries or conducting small-scale or pilot studies on the application of alternative measures. Still, it is difficult to say what the influence of these kinds of measures has on the well-being of irregular migrants, the return rates, or the cost-efficiency of these measures in practice. Although difficult to achieve, these aspects should be studied by scholars as to determine the precise causal relationships. Input from social (psychological, sociological), economical, and statistical areas of expertise need to be incorporated into the study of the immigration policy field to be able to make these kinds of contributions.

References

- Adams, W. C. (2015). Conducting Semi-Structured Interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 492–505). New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch19>
- Advisory Committee on Alien Affairs. (2013). *Aliens' detention or a less intrusive measure?* Den Haag: ACVZ.
- Alphen, B. van, Molleman, T., Leerkes, A., & Hoek, J. van. (2013). *From treatment to departure: An investigation into the operation of immigration detention [Van bejegening tot vertrek: een onderzoek naar de werking van vreemdelingenbewaring]*. Den Haag; Boom Lemma uitgevers.
- Amnesty International. (2011). *Detention of immigrants in the Netherlands: It should and could be done differently [Vreemdelingendetentie in Nederland: Het moet en kan anders]*. Amsterdam: Amnesty International.
- ATLAS.ti Scientific Software Development GmbH. (2016). ATLAS.ti (Version 8.0). Berlin: Scientific Software Development GmbH.
- Basilien-Gainche, M.-L. (2015). Immigration Detention under the Return Directive: The Cjeu Shadowed Lights. *European Journal of Migration and Law*, 17(1), 104–126. <https://doi.org/10.1163/15718166-12342073>
- Bloomfield, A. (2016). Alternatives to Detention at a Crossroads: Humanisation or Criminalisation? *Refugee Survey Quarterly*, 35(1), 29–46. <https://doi.org/10.1093/rsq/hdv018>
- Bloor, M., & Wood, F. (2006). *Keywords in Qualitative Methods*. London: SAGE Publications Ltd. <https://doi.org/10.4135/9781849209403>
- Broeders, D., & Leerkes, A. (2010). A Case of Mixed Motives?: Formal and Informal Functions of Administrative Immigration Detention. *British Journal of Criminology*, 50(5), 830–850. <https://doi.org/10.1093/bjc/azq035>
- Bruycker, P. de, Bloomfield, A., Tsourdi, E., & Petin, J. (2015). Alternatives to Immigration and Asylum Detention in the EU - Time for Implementation. Retrieved from <http://odysseus-network.eu/wp-content/uploads/2015/02/FINAL-REPORT-Alternatives-to-detention-in-the-EU.pdf>
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24(4), 409–429. <https://doi.org/10.1037/h0027982>
- Centers for Disease Control and Prevention. (1999). Framework for program evaluation in public health. *MMWR. Recommendations and Reports: Morbidity and Mortality Weekly Report*. 48(RR-11), 1–40.

- Chmelickova, N. (2006). Survey on Alternatives to Detention of Asylum Seekers in EU Member States. The regional coalition 2006. Retrieved from <https://idcoalition.org/wp-content/uploads/2008/12/survey-on-alternatives-to-detention-of-asylum-seekers-in-eu-member-states.pdf>
- Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, 24(2), 119–128. [https://doi.org/10.1016/S0149-7189\(01\)00003-9](https://doi.org/10.1016/S0149-7189(01)00003-9)
- Costello, & Kaytaz. (2013). Building Empirical Research into Alternatives to Detention: Perceptions of Asylum-Seekers and Refugees in Toronto and Geneva. Retrieved from <http://www.refworld.org/docid/51a6fec84.html>
- Cronbach, L. J., & Shapiro, K. (1982). *Designing evaluations of educational and social programs* (1st ed). San Francisco: Jossey-Bass.
- Directorate for Migration Policy. (2011). *Internal report Alternatives for Detention*. The Hague: Directorate-General for Immigration.
- Directorate-General for Immigration. (2013). Appendix: results of the pilots in light of alternative administrative detention. The Hague: Ministry of Justice & Security.
- Directorate-General for Migration and Home Affairs. (2013). *Evaluation on the application of the Return Directive (2008/115/EC)* (Final report No. DR-02-13-466-EN-N). Luxembourg: Publications Office of the European Union.
- D'Ostie-Racine, L., Dagenais, C., & Ridde, V. (2013). An evaluability assessment of a West Africa based Non-Governmental Organization's (NGO) progressive evaluation strategy. *Evaluation and Program Planning*, 36(1), 71–79. <https://doi.org/10.1016/j.evalprogplan.2012.07.002>
- Dutch House of Representatives. (2015). *Legislation concerning the return of irregular-staying migrants and administrative detention (Law 'Return and administrative detention')*. Retrieved April 23, 2018, from <https://www.tweedekamer.nl/kamerstukken/wetsvoorstellen/detail>
- Edwards, J. A., & Lampert, M. D. (Eds.). (1993). *Talking data: transcription and coding in discourse research*. New Jersey: Lawrence Erlbaum Associates.
- Esaiasson, P., Persson, M., Gilljam, M., & Lindholm, T. (2016). Reconsidering the Role of Procedures for Decision Acceptance. Retrieved from /core/journals/british-journal-of-political-science/article/div-classtitlereconsidering-the-role-of-procedures-for-decision-acceptancediv/A39F7FB7B150F7D229E936E421E1F4DC
- EU Commission. (2014). Communication from the Commission to the Council and the European Parliament on EU Return Policy. Brussel: EU Commission.

- EU Commission. (2016a). *Irregular migrant*. Retrieved January 24, 2018, from https://ec.europa.eu/home-affairs/content/irregular-migrant-0_en
- EU Commission. (2016b). *Third-country national found to be illegally present*. Retrieved January 24, 2018, from https://ec.europa.eu/home-affairs/content/third-country-national-found-be-illegally-present_en
- EUR-Lex. (2008). *Directive 2008/115/EC of the European Parliament and of the Council of 16 December 2008 on common standards and procedures in Member States for returning illegally staying third-country nationals*. Retrieved February 26, 2017, from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:348:0098:0107:en:PDF>
- EUR-Lex (2013a). Directive 2013/32/EU of the European Parliament and of the Council of 26 June 2013 on common procedures for granting and withdrawing international protection. Retrieved from <http://data.europa.eu/eli/dir/2013/32/oj/eng>
- EUR-Lex (2013b). Directive 2013/33/EU of the European Parliament and of the Council of 26 June 2013 laying down standards for the reception of applicants for international protection. (2013). Retrieved from <http://data.europa.eu/eli/dir/2013/33/oj/eng>
- European Migration Network. (2016). Guidelines for monitoring and evaluation of AVR(R) programmes. Retrieved from http://emn.ie/files/p_20160129124148guidelines_for_monitoring_and_evaluation_final_jan2016.pdf
- Field, O., & Edwards, A. (2006). Alternatives to Detention of Asylum Seekers and Refugees. Retrieved from <http://www.refworld.org/docid/4472e8b84.html>
- Finckenauer, J. O., Margaryan, S., & Sullivan, M. L. (2005). Evaluability Assessment in Juvenile Justice: A Case Example. *Youth Violence and Juvenile Justice*, 3(3), 265–275. <https://doi.org/10.1177/1541204005276267>
- Fiske, L. (2016). *Human rights, refugee protest and immigration detention*. London: Palgrave Macmillan.
- Gibbs, G. (2007). *Analyzing Qualitative Data*. London: SAGE Publications, Ltd. <https://doi.org/10.4135/9781849208574>
- Global Detention Project. (2015). *The Uncounted: The Detention of Migrants and Asylum Seekers in Europe*. Retrieved February 23, 2017, from <https://www.globaldetentionproject.org/the-uncounted-the-detention-of-migrants-and-asylum-seekers-in-europe>
- Global Detention Project. (2016). *Netherlands Immigration Detention Profile*. Retrieved February 17, 2017 from <https://www.globaldetentionproject.org/countries/europe/netherlands>

- Goodrick, D., & Rogers, P. J. (2015). Qualitative Data Analysis. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 561–595). New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch22>
- Hare, J., & Guetterman, T. (2014). Evaluability Assessment: Clarifying Organizational Support and Data Availability. *Journal of MultiDisciplinary Evaluation*, 10(23), 9–25.
- Klaver, J. F., Telli, S., & Witvliet, M. (2015). *Regional plan - Return of irregular-staying migrants with a duty to return. An examination on interventions that stimulate voluntary return [Regioplan - Terugkeer vertrekplichtige vreemdelingen. Een verkenning van interventies om zelfstandige terugkeer te stimuleren]*. Retrieved February 8, 2017, from http://www.regioplan.nl/publicaties/rapporten/terugkeer_vertrekplichtige_vreemdeling_en
- Leerkes, A., Galloway, M., & Kromhout, M. (2011). To return, or not? Determinants of return intentions and return attitudes among irregular-staying migrants who have exhausted all legal remedies [Terug of niet? . Determinanten van terugkeerintenties en -attitudes onder (bijna) uitgeprocedeerde asielmigranten]. *Mens & Maatschappij*, 86(2), 122–157.
- Leers, G. B. M. (2011). Letter to chamber of representatives: Alternatives to detention. The Hague: Ministry of Justice & Security. Retrieved from <https://zoek.officielebekendmakingen.nl/kst-19637-1483.html>
- Mangiaracina, A. (2016). The Long Route Towards a Widespread European Culture of Alternatives to Immigration Detention. *European Journal of Migration and Law*, 18(2), 177–200. <https://doi.org/10.1163/15718166-12342098>
- Marsh, K., Venkatachalam, M., & Samanta, K. (2012). An economic analysis of alternatives to long-term detention: Final Report. London: Matrix Evidence. Retrieved from <http://detentionaction.org.uk/wordpress/wp-content/uploads/2011/10/Matrix-Detention-Action-Economic-Analysis-0912.pdf>
- McLaughlin, J. A., & Jordan, G. B. (2015). Using Logic Models. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 62–87). New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch3>
- Ministry of Justice & Security. (2013). *Teeven takes measures after the inspection report on Dolmatov- news item - [Teeven neemt maatregelen na Inspectierapport over Dolmatov - Nieuwsbericht -]*. Retrieved March 17, 2017, from <https://www.rijksoverheid.nl/actueel/nieuws/2013/04/12/teeven-neemt-maatregelen-na-inspectierapport-over-dolmatov>

- Ministry of Justice & Security. (2016a). *January to June 2016 report from the organizations cooperating in the immigration process*. Retrieved January 22, 2017, from <https://www.rijksoverheid.nl/documenten/rapporten/2016/11/08/rapportage-vreemdelingenketen-periode-januari-juni-2016>
- Ministry of Justice & Security. (2016b). *January to December 2015 report from the organization cooperating in the immigration process*. Retrieved February 8, 2017, from <https://www.rijksoverheid.nl/documenten/rapporten/2016/03/18/rapportage-vreemdelingenketen>
- Ministry of the Interior and Kingdom Relations. (1992). *Common law on administrative legislation*. Retrieved February 22, 2017, from <http://wetten.overheid.nl/BWBR0005537/2018-04-11>
- Ministry of the Interior and Kingdom Relations. (1992). *Common law on administrative legislation*. Retrieved February 22, 2017, from <http://wetten.overheid.nl/BWBR0005537/2018-04-11>
- Ministry of the Interior and Kingdom Relations. (2000a). *Aliens Act 2000*. Retrieved February 22, 2017, from <http://wetten.overheid.nl/BWBR0011823/2017-01-01>
- Ministry of the Interior and Kingdom Relations. (2000b). *Aliens circular 2000 (A)*. Retrieved November 9, 2017, from <http://wetten.overheid.nl/BWBR0012287/2017-10-01>
- Ministry of the Interior and Kingdom Relations. (2000c). *Aliens decree 2000*. Retrieved November 9, 2017, from <http://wetten.overheid.nl/BWBR0011825/2017-10-01>
- Ministry of the Interior and Kingdom Relations. (2014). *Subsidy scheme supporting voluntary return*. Retrieved November 9, 2017, from <http://wetten.overheid.nl/BWBR0034899/2014-03-13>
- Mulder, L. (2012). Adaptations of the Aliens Act 2000. Retrieved from <https://zoek.officielebekendmakingen.nl/stcrt-2012-26248.html>
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). Planning and Designing Useful Evaluations. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 7–35). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch1>
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3 ed). Thousand Oaks: Sage Publications.
- R&DS. (2015). Directory for Repatriation and Departure. Retrieved from https://www.dienstterugkeerenvertrek.nl/binaries/Leidraad%20Terugkeer%20en%20Vereeniging%20terugkeerenvertrek_tcm49-145473.pdf
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: a systematic approach* (7th ed). Thousand Oaks, CA: Sage.

- Rutte, M., Van Haersma Buma, S., Pechtold, A., & Segers, G.-J. (2017). Coalition agreement "Trust in the future." Retrieved from <https://www.kabinetsformatie2017.nl/documenten/publicaties/2017/10/10/regeerakkoord-vertrouwen-in-de-toekomst>
- Sampson, R., & Mitchell, G. (2013). Global Trends in Immigration Detention and Alternatives to Detention: Practical, Political and Symbolic Rationales. *Journal on Migration and Human Security*, 1(3), 97–121. <https://doi.org/10.14240/jmhs.v1i3.14>
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (Eds.). (1991). *Foundations of program evaluation: theories of practice* (Reprinted). Newbury Park: Sage Publications.
- Smith, M. F. (1989). *Evaluability Assessment: A Practical Approach*. Dordrecht: Springer Netherlands. Retrieved from <http://public.eblib.com/choice/publicfullrecord.aspx?p=3105592>
- Smith, M. F. (1990). Evaluability assessment: Reflections on the process. *Evaluation and Program Planning*, 13(4), 359–364. [https://doi.org/10.1016/0149-7189\(90\)90020-W](https://doi.org/10.1016/0149-7189(90)90020-W)
- Teeven, F. (2013). Letter to chamber of representatives: Commitments to Reports and Advice on immigration policy. Parliamentary papers. Retrieved from <https://www.rijksoverheid.nl/documenten/kamerstukken/2013/09/13/brief-tweede-kamer-toezeggingen-op-rapporten-en-adviezen-vreemdelingenbeleid>
- The European Migration Network. (2014). *The use of detention and alternatives to detention in the context of immigration policies: EMN Synthesis Report*. Retrieved February 26, 2017, from http://emn.ie/cat_publication_detail.jsp?clog=1&itemID=2745&t=6
- Trevisan, M. S. (2007). Evaluability Assessment From 1986 to 2006. *American Journal of Evaluation*, 28(3), 290–303. <https://doi.org/10.1177/1098214007304589>
- Trevisan, M. S., & Walser, T. M. (2015). *Evaluability Assessment: Improving Evaluation Quality and Use*. Thousand Oaks: SAGE Publications, Inc. <https://doi.org/10.4135/9781483384634>
- Trevisan, M. S., & Yi Min, H. (2003). Evaluability assessment: a primer. *Practical Assessment, Research & Evaluation*, 8(20).
- Welch, M., & Schuster, L. (2005). Detention of asylum seekers in the US, UK, France, Germany, and Italy: A critical view of the globalizing culture of control. *Criminology and Criminal Justice*, 5(4), 331–355. <https://doi.org/10.1177/1466802505057715>
- Wholey, J. S. (1979). *Evaluation--promise and performance*. Washington: Urban Institute.
- Wholey, J. S. (2015a). Exploratory Evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 88–107). New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch4>

- Wholey, J. S. (2015b). Use of Evaluation in Government. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 798–815). New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch30>
- Zimmerman, M. A., & Holden, D. J. (Eds.). (2009). *A practical guide to program evaluation planning: theory and case examples* (New ed). Thousand Oaks: SAGE Productions.

Appendix.

A. Example of an interview guide

Vragenlijst onderzoek alternatieve toezichtmaatregelen voor detentie

Naam: David Kuiper

Functie:

Datum afname: 01-12-2017

Inleiding:

Graag zou ik willen beginnen met een korte inleiding voor het interview. Voor het ministerie van Justitie en Veiligheid, specifiek de Directie Regie Migratie Keten, Directie Migratiebeleid en de Dienst Terugkeer en Vertrek (DT&V) voer ik een onderzoek uit naar de mogelijkheden voor een evaluatie van de implementatie van alternatieve toezichtmaatregelen voor detentie. Het gaat hierbij om het aanbod van toezichtsmaatregelen voor terugkeer na de pilots van 2012 en de daaropvolgende verankering van de toezichtsmaatregelen in het terugkeerbeleid. Het interview zal zich richten op de toepassing van toezichtsmaatregelen binnen het terugkeerproces.

Het interview zal tussen de 30 tot 60 minuten duren. Deelname is vrijwillig en kan op elk moment tijdens het onderzoek worden ingetrokken. U hoeft geen vragen te beantwoorden waarbij u zich ongemakkelijk voelt. Graag zou ik willen vragen of het mogelijk is dat dit gesprek word opgenomen. Voor het onderzoek is het nodig dat het gesprek omgezet kan worden naar tekst voor de verdere analyse. U zal geanonimiseerd worden en niet met naam in de resultaten voorkomen. De uitgeschreven tekst zal eerst aan u aangeboden worden ter controle. Na deze controle zal de opname van het gesprek worden verwijderd. De uitgewerkte tekst van het interview zal voor maximaal 3 jaar bewaard worden door de student in verband met de eisen van de Radboud Universiteit. Een kopie van deze verklaring zal naar u worden toegestuurd.

Heeft u nog vragen voordat we beginnen?

Datum:

Datum:

Handtekening Geïnterviewde:

Handtekening student:

1. Algemeen

- a. Functie
 - i. Wat houdt uw functie in?
 - ii. Welke ervaringen heeft u binnen uw huidige en/of eerder functie(s) met administratieve detentie en de alternatieve toezichtmaatregelen?

2. Algemeen beeld beleid

- a. Het onderzoek betreft de verbeterde verankering van alternatieve maatregelen voor detentie bij migranten die moeten terugkeren specifiek na de uitvoering van de pilots in 2011 en de kaderbrief in 2013.
 - i. Kunt u mij uitleggen wat uw beeld is van het beleid omtrent deze alternatieve toezichtmaatregelen?
 - ii. Wat is de doelgroep voor de alternatieve maatregelen? voor wie is dit bestemd?

3. Specifieke onderdelen

- a. Context
 - i. Wat was volgens u de aanleiding voor het aanpassen van het terugkeerbeleid omtrent de alternatieve toezichtmaatregelen?
 - 1. Welke problemen probeert men hiermee op te lossen?
 - 2. Klopt dit met uw eigen beeld van het probleem?
- b. Rationale
 - i. Wat is de intentie/gedachtegang achter de invoering van dit beleid?
 - 1. En hoe ziet u dit? En ziet u dit zelf anders?
- c. Objectives
 - i. Wat zijn volgens u de belangrijkste doelen van het beleid achter de alternatieve toezichtmaatregelen?
 - 1. Is er nog iets veranderd aan deze doelen volgens u sinds de veranderingen in 2013?
 - 2. Zou het beleid volgens u nog met andere doelstellingen rekening moeten houden?
- d. Input
 - i. Welke middelen zijn er aangewend en hoe zijn deze ingezet voor de toepassing van de alternatieve toezichtmaatregelen?
 - 1. Is er nog iets veranderd aan de aanwending en inzet van deze middelen sinds 2013?
 - 2. In hoeverre heeft e migratiecrisis invloed gehad op de aanwending van middelen?
 - 3. Ontbreken er nog mogelijke middelen volgens u?
- e. Activities
 - i. Welke maatregelen behoren volgens u tot de alternatieve toezichtmaatregelen voor detentie binnen het huidige terugkeerbeleid?
 - 1. Hoe is de implementatie van deze maatregelen verlopen?
 - 2. Zijn er nog veranderingen opgetreden aan de maatregelen gedurende het implementatieproces
- f. Output
 - i. Wat voor verwachte output zouden deze maatregelen opleveren
 - 1. Wat voor onderdelen bestaat deze output uit?
 - 2. In hoeverre is dit volgens u nu ook gebeurd?
- g. Impacts
 - i. Welke verwachte impact zou de verbeterde implementatie van de alternatieve toezichtmaatregelen hebben op het huidige terugkeerbeleid?
 - 1. Welke positieve en negatieve impact?

2. En hoe ziet u dit?

4. Data-analyse

- a. In hoeverre is het mogelijk om bij alle instanties en individuen binnen de keten een evaluatie uit te voeren?
- b. Uit mijn eigen analyse van onder andere de kaderbrief over de alternatieven voor detentie uit 2013 als voornaamste leidraad kom ik uit op vier doelen voor de verbeterde implementatie van de alternatieven voor detentie. Namelijk het gebruik van minder restrictieve maatregelen, eenvoudiger zijn in de uitvoering, goedkoper zijn in gebruik en een beter of minimaal vergelijkbaar aantal terugkeerders.
 - i. In hoeverre kloppen deze doelen met uw eigen beeld van de doelen voor de alternatieven?
 1. Wat ontbreekt of klopt er niet?
 - ii. Wat voor indicatoren zijn er volgens u waarmee dit geëvalueerd kan worden?
- c. Welke data is er volgens u nodig om een evaluatie te doen met betrekking tot de doelen van de alternatieve toezichtmaatregelen?
 - i. Welke data wordt daarvan al verzameld door het ministerie?
 1. Wat betreft de mindere mate van restrictie op de bewegingsvrijheid? *(Uit eerdere gesprekken bleek dat dit kan worden gezien als de verantwoordelijkheid voor de uitzetprocedure die bij de vreemdeling komt te liggen in plaats van bij de "overheid". Hierbij worden er handvatten en informatie aangereikt aan de vreemdelingen kwestie zodat diegene zelf de benodigde stappen kan zetten.)*
 2. Wat betreft de eenvoudigheid van uitvoering?
 3. Wat betreft de kosten?
 4. Wat betreft het vergelijkbare of betere terugkeer aantallen?
 - ii. Welke data is er nog nodig? En per maatregel? -> documentinname, meldplicht, VBL & GL
- d. In hoeverre is het mogelijk om bestaande data en documenten te gebruiken voor verdere evaluatie?
- e. Zijn er nog factoren (middelen, procedures, politieke) die de mogelijkheid tot evaluatie kunnen beïnvloeden?
- f. In hoeverre is het mogelijk om iemand tijdens het proces tot uitzetting te volgen?
 - i. Welke gegevens zijn hierbij belangrijk?
 - ii. Welke ketenpartners zijn hierbij belangrijk?

Dit waren alle vragen. Heeft u zelf nog vragen en/of opmerkingen?

Ik wil u bedanken voor uw tijd, wanneer u wilt kunt u op de hoogte worden gehouden van het onderzoek.

B. Codebook for Data analysis

Table 9. Mixed codebook for data analysis

Theme	Description	Category	Description	Code	Code-component	Description
A: Program intent	Concept-driven: Describes the structure of the policy. It shows the intent of the policy and that the objectives, indicators, and priorities are clear.	A 1: Program context	Concept-driven: Provides information about the context in which a program operates. Informs evaluability readiness, options and increase evaluation feasibility	A 1 1: Contextual factors	A 1 1 1: Antecedent influences	Concept-driven: The basic characteristics that are present during the start of the program
					A 1 1 2: Mediating influences	Concept-driven: Influences that become apparent during the implementation of the program
		A 2: Program perspectives	Concept-driven: Perceptions of the program and its components as well as perceptions of how the program works and should work	A 2 1: Related program logics	A 2 1 1: Rationale	Concept-driven: The mission or idea behind the need to implement the program
					A 2 1 2: Objectives	Concept-driven: The stated goals or objectives the program intends to accomplish
					A 2 1 3: Situation before	Data-driven: What the situation was before the program
				2 2: Program structure	A 2 2 1: Input	Concept-driven: The necessary resources to support the program
					A 2 2 2: Activities	Concept-driven: The services that the program implements to produce the necessary program outputs
					A 2 2 3: Outputs	Concept-driven: The products of the services provided by the Activities
				A 2 3: Outcome structure	A 2 3 1: Outcomes	Concept-driven: The expected short-term and intermediate effects as the results from the activities and outputs
					A 2 3 2: Expected impact	Concept-driven: The extent to which the activities and outputs have resulted in a long-term change

B: Program plausibility	Concept-driven: Lists the objectives which should be measured and how these are supposed to be evaluated.	B 1: Program implementation	Concept-driven: Determining the extent to which program implementation aligns with program theory & logic model	<div> <div>B 1 1 1: General implementation of the program</div> <div>B 1 1: General implementation</div> <div>B 1 1 2: Perception of the implementation</div> <div>B 1 1 3: Changes to implementation</div> <div>B 1 1 1: Implementation of the specific components</div> <div>B 1 2: Implementation of components</div> <div>B 1 2 2: Perception of specific implementation</div> <div>B 1 2 3: Changes to the implementation of specific components</div> </div>			Concept-driven: Are the program and its specific components implemented as intended in the program theory. Where does it differ if it is not the case?
		B 2: Research logic	Concept-driven: The inclusion of findings on the research logic of the program theory	<div> <div>B 2 1: New reports relevant to the program theory</div> <div>B 2 2: Older or similar evaluations</div> </div>			Data-driven: Code for new documents/ literature mentioned Data-driven: Code for older documents/ literature mentioned

C: Data accessibility	Concept-driven: Analysis the policy in its practical implementation and to uncover the feasible measures and data for evaluation.	C 1: Data needs	Concept-driven: Statements on which evidence/data is needed to determine the objectives of the program	C 1 1: Indicators	Data-driven: Indicators that can be used for the objectives mentioned in the interview
				C 1 2: measures	Data-driven: Data that can be used to measures the objectives mentioned in the interview
		C 2: Existing data	Concept-driven: Statements on the sources for the data that are already in place	C 2 1: Current state	Data-driven: Current state of sources of data that can be used to measures the objectives mentioned in the interview
				C 2 2: Developments	Data-driven: developments with the data source mentioned in the interview
		C 3: Data collection	Concept-driven: The accessibility of data collection	C 3 1: Sources of data	Data-driven: The mentioning of a potential source of data
				C 3 2: stakeholders	Data-driven: The mentioning of involved stakeholders and statements on the cooperation
				C 3 2 1: Which C 3 2 2: Cooperation	
		C 4: Data issues	Data-driven: Difficulties in data collection	C 4 1: problems C 4 2: solutions	Data-driven: The mentioning of potential problems and related solutions
D: Utility of evaluation	Concept-driven: Presents the policy or its parts of which the objectives can be measured and evaluated. This is meant to assess the sensibility of implementing an evaluation.	D 1: Usefulness	Concept-driven: Is it sensible to implement an evaluation	D 1 1: Evaluation possible D 1 2: What is in place D 1 3: What needs to change	Data-driven: If an evaluation is possible, what is already in place for evaluation, and what needs to change for an evaluation
		D 2: Difficulties/ barriers	Concept-driven: Are there any, resource, procedural or political factors that could be barriers to evaluation	D 2 1: Which barriers D 2 2: Solution	Data-driven: Which barriers and which solutions

E: Other	Data-driven: Includes other data-driven codes.	E 1: Job description	Data-driven: Explanation of their Job and experiences with the ATD measures
		E 2: ATD experience	

Note. Adapted from Centers for Disease Control and Prevention (1999), D'Ostie-Racine et al. (2013), McLaughlin & Jordan (2015), Smith (1990), Trevisan & Walser (2015), Wholey (1979, 2015a), Zimmerman & Holden (2009)

C. Quotes for the Results

Table 10. Overview of the quotes used in the Results

<i>Respondent</i>	<i>Code of the Quote</i>	<i>Original 'Dutch' version</i>	<i>English translation</i>
NP1	Q1	Denk ook deels Europese wetgeving die toch steeds meer invloed heeft op...ook op het vreemdelingenbeleid nationaal.	I would argue that European legislation more and more has an influence on migration policies at a national level as well
DTV2	Q2	En we zien ook dat vreemdelingen uit bepaalde landen hè, daar wordt wel snel op gereageerd.	We also see a fast response to irregular-staying migrants from certain countries.
IND1	Q3	Duurzame terugkeer betekent natuurlijk dat mensen weer bereid en in staat zijn om te integreren in de samenleving waar ze ooit vandaan zijn gekomen.	Sustainable return, of course, means that people should be willing and able to re-integrate into the society of which they came from.
JV1	Q4	Ja dat was toen ook een hele belangrijke, nu hoor je daar weinig meer over...maar duurzame terugkeer, dat iemand teruggaat met middelen en mogelijkheden [. . .] dat niet alleen, maar dat hij ook met opgeheven hoofd terugkwam. En dat hoor je nu helemaal niet meer.	Yes, that was a very important one then (regarding sustainable return). Nowadays you don't hear as much about it but sustainable return means that someone returns with some means and possibilities [. . .] not only that, but it is also so an individual can return with some dignity. And that is something you don't hear about anymore
NP1	Q5	Wat je wel ziet met de alternatieven binnen de politie... waar voorheen eigenlijk een inbewaringstelling in vreemdelingenbewaring toch redelijk gebruikelijk was, zie je wel een hele kentering he. Waar eigenlijk gezegd werd; 'vreemdelingenbewaring tenzij het niet mogelijk is'. Wordt nu veel meer gekeken naar welke alternatieven hebben we. Of welke maatregelen kunnen we opleggen, welk lichter maatregel kunnen we opleggen. En inbewaringstelling/vreemdelingenbewaring dus echt als ultimum remedium wordt toegepast.	What you see now with the alternatives and their use within the police, where previously an administrative detention measure was imposed fairly regularly, this has changed significantly. Where previously one would say; 'administrative detention unless it is not applicable,' nowadays we look much more towards the possible alternatives that we have. Or which measures we can impose, which lighter measures we can impose. And administrative detention can thus really be applied as an ultimum remedium.
DTV1	Q6	Waarbij toen onder andere de genoemde alternatieven voor bewaring een mogelijkheid zou moeten kunnen bieden om minder vaak de bewaringsmaatregel toe te passen	Whereby then, among other things, the ATD measures for administrative detention would provide the possibility to impose detention less often
JV1	Q7	Een van de doelen die we in de brief ook noemden was inderdaad het terugdringen van bewaring van groepen die daar niet in hoorde.	One of the goals that we also mentioned in the letter (to the Parliament) has been to reduce the number of detentions of groups that were not supposed to be in there.
NP1	Q8	Je gaat eerst kijken, kunnen we volstaan met de meldplicht. Zo niet, beperken bewegingsvrijheid en uiteindelijk als ultiem...vreemdelingenbewaring.	You would first consider if we could suffice with a duty to report. If not, limiting someone's freedom of

			movement and as ultimatum administrative detention.
DTV3	Q9	Want sindsdien wordt toch wel veel zwaarder gekeken naar, wanneer leggen we die maatregel van bewaring op. Hè, maatregelen moeten beter onderbouwd worden hè?	It is looked too much more intensively since then as to when we impose a detention measure. Measures need to be more substantiated, right?
DTV1	Q10	We kennen het begrip als zodanig niet, omdat in het palet toezichtmaatregelen valt bewaring en elke andere vallen daar gewoon onder. En de praktijk is dat we worden geconfronteerd met een illegale vreemdeling bijvoorbeeld, dan kijken we samen met de politie welk toezichtmaatregel het beste van toepassing is [...] en pas als we merken dat al die toezichtmaatregelen niet van toepassing zijn, bijvoorbeeld omdat we denken dat die zich gaat onttrekken aan toezicht dan pas komen we pas bij het uiterste toezichtmiddel in bewaringstelling.	We do not interpret the definition in such a way, because detention and any other measure are just as well all included within the range of options for imposing supervisory measures. And when we are confronted with an irregular-staying in practice we, together with the police, would just determine the measure that is best applicable in that situation [...] And only when we notice that these supervisory measures are non-applicable, for example, because we expect that someone would evade supervision, only then would we arrive at a determination of the "ultimatum" supervisory means of administrative detention.
JV1	Q11	Het beleid was bedoeld dat er een structureel mogelijkheden waren voor zowel ketenpartners om een alternatief op te leggen wanneer er geen bewaring aan de orde was.	The policy was meant to be that there were structural possibilities for stakeholders to impose an alternative if detention was non-applicable
NP1	Q12	Juridisch eigenlijk niks. Want die was er voorheen ook al. Maar dat die dus nu veel meer toegepast wordt als alternatief of lichter middel voor vreemdelingenbewaring.	Legally nothing, as it was also in place beforehand. However, it is now applied more often as an alternative or lighter measure for administrative detention.
AI1	Q13	Ja ik heb eigenlijk die pilots [...] een paar dingen...en er is te weinig over beken [...] het is te klein geweest. Het is een hele kleine groep geweest die eraan mocht meedoen.	Yes, regarding those pilots [...] a couple of things not enough is known about these [...] and it was too small. Only a tiny group was able to participate in it.
JV1	Q14	Als je moet kiezen tussen een borgsom waarbij je een hele berg papierwerk voor hebt, om ook nog eens geld te moeten vragen van de vreemdeling. Dat weer apart moet zetten, dat na 4 weken dat het verbeurd moet verklaren want hij is niet vertrokken. Of je doet een meldplicht, met terugkeer ondersteuning van de DT&V. Wat gewoon in jouw dagelijkse praktijk is, dan zie je dat er een borgsom niet altijd wordt aangeboden.	If you have to choose between a bail, which requires a mountain of paperwork and were you also need to ask money from the irregular-staying, that needs to be set apart, after four weeks you have to declare forfeited as he has not departed. Or you impose a duty to report, with return supervision from the R&DS, which falls into your daily practice. You will see that a bail is not always offered as a possibility.
DTV2	Q15	Bij de oprichting van de VBL is er inderdaad gekozen voor maximaal 12 weken. Dat heeft waarschijnlijk toentertijd te maken gehad met het feit dat je nog in beroep kan gaan en meestal het ongeveer drie maanden duurt	The choice for a maximum stay of 12 weeks was indeed chosen when the FRC was established, most likely due to the fact that you could appeal and that would take around three months before a decision was made then [...]

		voordat je een uitspraak krijgt [. . .] daar gissen we nog een beetje naar. Die twaalf weken is wel nog steeds erin gebleven. Het kan zo zijn dat vreemdelingen langer de tijd nodig hebben. En dan maakt de regievoerder een afweging.	so we are still guessing precisely why. Those 12 weeks have stayed. It could be that irregular-staying migrants need more time and the departure supervisor then makes an assessment.
JV1	Q16	Als ik er zo naar kijken maar ja ik ben daar redelijk cynisch nu in dan denk ik dat het juist zorgt voor een verlenging van het verblijf van de vreemdeling in Nederland. En dat het uiteindelijk leidt tot meer bewaring. Omdat het is nu een rekmiddel geworden. Want we kunnen het niet afdwingen.	If I look at it that way... but I have become pretty cynical about it, then I would think that it just provides an elongation of a stay of an irregular-staying in the Netherlands and that it eventually leads to more detention. Because it has become a means to stretch the duration (of the return process), as we can't enforce (the return).
NP1	Q17	Zolang als de vreemdeling de verplichting heeft zelf Nederland te verlaten, kun je zeggen ja als iemand MOB is gegaan [. . .] kun je zeggen dan is iemand illegaal achtergebleven of je kunt ook zeggen dan heeft diegene aan zijn vertrekverplichting voldaan, tot het tegendeel blijkt.	As long as the irregular-staying has an obligation to leave the Netherlands you could say; when someone goes MOB [. . .] that someone has stayed behind illegally, or you could say that someone has met his obligation to depart, until it is proven otherwise.
JV1	Q18	Alleen de meldplicht is niet effectief. Want waarom is de VBL effectief? Omdat je je daar en moet melden en je hebt terugkeer begeleiding van de DT&V en je hebt onderdak. Dus dat bij elkaar zorgt ervoor dat jij je veilig voelt, je hebt je basisbehoeften, er wordt controle...	A duty to report on its own is ineffective. Because why is the FRC so effective? As you have to report there, you get return supervision from the R&DS, and you have a place to stay. So that together ensures that you feel safe, you have your basic needs, and there is a form of control.
INS	Q19	En nou...als in procesvertegenwoordigings-land maak je dus lang zo vaak niet mee dat er beroepen worden ingesteld vreemdelingenwet-maatregelen [. . .] bijvoorbeeld op een...eh...gezinslocatie of op een vrijheidbeperkende locatie zoals in Vlagtwedde [. . .] En tegen meldplichten [. . .] dat onttrekt zich eigenlijk al helemaal aan onze waarneming, omdat daar zelden procedures tegen dat soort besluiten worden gevoerd. En die lichtere middelen dan bewaring, die leiden veel minder vaak tot ontstemming bij de betrokkenen zelf	You don't often encounter any appeals against [. . .] the (alternative) supervisory measures in the "process representation field [. . .] for example at a freedom-restricting location such as Vlagtwedde [. . .] and against a duty to report [. . .] those we do not perceive at all, as there are seldom any procedures against those kinds of decisions. And these lighter measures than detention lead much less often to disagreement with the irregular-staying migrant themselves.
JV1	Q20	Maar ik denk 'je moet wel alle kosten voor een alternatief meenemen'. Dus ook, wat kost een regievoerder... want hij moet het wel in beheer hebben en wat kost het aan opvang en onderdak [. . .] maar nog steeds kan je denken dat het goedkoper is.	But I do think that you need to include all the related costs of an alternative (regarding a cost-effectiveness calculation). So, in addition, what does a departure supervisor cost, as he has to manage the process, and what does these cost for providing shelter and care [. . .] However, you can still think that the measures remain cheaper.

JV1	Q21	Niet alleen qua tijd, de eenvoud zit zowel in de tijd maar ook in dat je weet dat binnen jouw organisatie andere het ook toepassen, dat je dezelfde formulieren kan gebruiken [...] en de eenvoud zit er ook in dat als jij...dat je een beetje profiel ontwikkeling kan gaan doen.	The simplicity is not just related to time, but in that, you would also know that others within your organization are also applying it (the alternative supervisory measures), that you can use the same forms [...] and the simplicity is also that you could do a bit of profile development.
JV1	Q22	En dat was ook met wat makkelijker maken voor de uitvoering dat ze weten, dit zijn de vaste alternatieven die we hebben. Die kan je gebruiken en het is nooit echt van de grond gekomen. Dat al je kenmerk A en B hebt, dan pas je maatregel C toe. Dat is jammer genoeg niet gekomen.	And that was also (the intention) with simplifying the process of application. That they would know; these are the established measures that we have. Those you can use but it never really got off the ground. That when you have characteristic A and characteristic B, you can apply measure C (with due consideration of the individual case of course). That, unfortunately, never happened.
DTV1	Q23	Kijk, we hebben al eens een keer een beperkt, als DT&V, geprobeerd een beperkt eh... klanttevredenheid onderzoek te doen. Door vreemdelingen te interviewen hè? 'Hoe vindt je dat je begeleid wordt in je terugkeertraject'. Maar dan ligt het er net aan in welke setting je iemand benadert. Iemand die in bewaring zit of een beperkte maatregel is opgelegd zal anders reageren, in zijn algemeenheid, dan iemand met een meldplicht.	Look, we (The R&DS) already tried to do a limited "customer-satisfaction" study by interviewing irregular-staying migrants. Like... 'how do you consider the supervision you get during your return process? But it depends to a great extent on the setting in which someone is approached in. Someone who is placed in detention or a restrictive measure will react differently, in general, then someone with a duty to report.
DTV2	Q24	Nou we hebben natuurlijk onze eigen systemen hè? En ehm... het belangrijkste in het systeem is het vertrek plan. En alles wat de regievoerder doet, wordt vastgelegd. En het vertrekplan is het allerbelangrijkste in ons systeem. Want daarin staat dus eh... de hele geschiedenis van deze vreemdeling.	Well, we (The R&DS at the FRC) have our own systems, right? And the most important one in the system is the departure file. And everything the departure supervisor does, is registered. And the departure file is the most important, as it lists the entire history of the specific irregular-staying migrant
DTV1	Q25	Nee, het is een geaggregeerd databasesysteem. Maar niet op V-nummer, dat is althans niet de bedoeling. We reiken wel als ketenpartners V-nummers aan, maar het gebruik van de datasheets zeg maar dat is voor de keten partners is dat op niet herkenbaar V-nummer laat ik het zo zeggen. Kijk, er kan wel gekeken worden naar de V-nummers onderliggend. Dat is niet primair het doel van de database verzameling.	No, it is an aggregated database system. So, it is not based on V-numbers (the unique number by which an irregular-staying migrant is registered). We do provide the V-numbers, but the use of the datasheets is by non-recognizable V-numbers for the relevant government agencies so to speak. Though, one could look at the underlying V-numbers. That is not the primary goal of the collection of the database.
IND1	Q26	Die terminologie dat is nog wel eens een lastige kwestie dat zou prettig zijn als we het is een keertje met ze allemaal erover eens zouden zijn wat we verstaan onder gedwongen terugkeer, vrijwillige terugkeer, uitzetting. Dus van	The terminology is often another problematic issue. It would be nice if we could all agree on what is meant with a forced return, voluntary return, Removal. So that we mean the

		Raad van State tot en met Vreemdelingenpolitie, dat we allemaal hetzelfde bedoelen.	same thing, from the Council of State to the Aliens Police Department.
JV1	Q27	Het beleid was bedoeld dat er een structureel mogelijkheden waren voor zowel ketenpartners om een alternatief op te leggen wanneer er geen bewaring aan de orde was.	The policy was meant as to provide structural possibilities for all involved stakeholders to impose an alternative when detention is not in question.
JV1	Q28	Hiervoor hadden we niet zo'n goed overzicht voor unieke maatregelen die werden opgelegd. Dus dat, Wat wordt er nu opgelegd en door wie? En werkt het of niet? Wat leidt nou tot terugkeer? En ik heb geen enkel onderzoek kunnen vinden waarbij je dat op grote schaal ziet. Het is altijd gericht op 1 maatregel, die in een pilotproject daarvan werkte.	We didn't have such a clear overview of all the unique measures that were applied before all of this. So that; what was imposed and by whom? And does it work or not? What leads to a return? And I haven't found a single research by which you see that on a large scale. It always targets one measure, that worked (only) in a pilot project.

D. Evaluation Options

Table 11. Necessity evaluation option 1

Costs vs. Costs			
Purpose	One of the initial rationales for the implementation of the program under consideration of the present study was that the use of the 'alternative' measures would lead to a more cost-efficient return policy. ADM is regarded as a costly measure to impose, requiring a high level of resources to effectuate. The assumption is that the application of the other supervisory measures requires fewer resources to effectuate, thus being more cost-efficient. These assumptions are mostly based on previous information from pilots and small-scale studies. The purpose of the first option is to evaluate the current costs of the range of supervisory measures and to see if these assumptions are right.		
Data Collection	All associated costs of applying a supervisory measure need to be determined and collected to estimate the costs for each of the supervisory measures. In addition, basic statistical data from the first option of the effectiveness evaluation should be collected. Possible related costs are also provided by the guidelines for monitoring and evaluation of AVR(R) programs by the EU commission, which lists several aspects of the costs for a similar program as the supervisory measures (European Migration Network, 2016).		
Data Availability	Data that is currently available:		Data that is still absent:
	<ul style="list-style-type: none"> - Previous studies and pilots that include an estimation of the costs involved with the application of the supervisory measures: (Directorate for Migration Policy, 2011; Marsh, Venkatachalam, & Samanta, 2012) - Most of the statistical data form option 1 of the effectiveness evaluations 		<ul style="list-style-type: none"> - Estimations of the actual costs of the application of a measure in practice - Statistical data on the average duration of application per supervisory measure
Data analysis	Financial estimation/calculation: <ul style="list-style-type: none"> - Estimations/calculations of the total costs for the number of applications of each measure - Estimations/calculations of the average total cost per supervisory measure - Estimation/calculations of the average costs per day per supervisory measure 		
Resources	The option requires any required access to available financial data. One or multiple evaluators are needed with a financial background. An estimation of the costs is estimated to take between 6 to 9 months		
Utility	Priority:	●	Complexity: ● ● ● ●
	Resource intensity: ● ● ●		
Utility	The option is of low priority, mainly because the respondents indicate that the priority has moved to other areas.		
	The evaluation will be relatively difficult to implement as an accurate estimation depends on finding all the related costs and how accurate these are themselves estimated.		
Utility	The evaluation requires a moderate level of resources. It requires access to the financial data and some time and effort from multiple evaluators.		

Table 12. Necessity evaluation option 2

Possibilities for Simplicity		
Purpose	The second option for an evaluation of the necessity/efficiency of the program considers the intent of the program to provide a more centralized, simple structure for the application of the measures. Incorporating a central role for the DT&V, as well as a more standardized use of the supervisory measures were meant to make the application of the measures easier. Unfortunately, as discussed in the results, the view is that these efforts have not seen all of their intended effects in practice. This evaluation would be aimed to analyze further what obstacles were there that inhibited the implementation and what possibilities there are for further improvement.	
Data Collection	Data needs to be collected on the efforts, obstacles and implications that were apparent that influenced the implementation of the program as to limit the intended effects on this aspect. Therefore, the perceptions from a practical, policy and legal level need to be collected from those involved. The present study already provides some directions to which the data collection can be directed to; the mixed results for centralizing the role of the DT&V in the return process and the concerns expressed regarding the differences in the priorities of organizations involved.	
Data Availability	Data that is currently available: - Some preliminary indications from the present study	Data that is still absent: - In-depth data on the specific aspect of simplicity
Data analysis	- Qualitative analysis of the perceptions of the respondents; Interviews can be conducted with respondents, aimed to provide comprehensive information on the intended effect of a more streamlined, centralized application of the supervisory measures by the implemented program. These can be analyzed by transcribing, coding and further interpretation. Such an approach would be more straightforward, although the consequence is that interviews are a less accurate method for gathering health-related information	
Resources	Not too many resources are required for the option. Willing and cooperative respondents need to be found, as well as a researcher/evaluator with some experience with conducting interviews and analyzing these accordingly. The option is estimated to take somewhere between 6 to 9 months.	
Utility	Priority: ● ● ● The evaluation option is of moderate priority. While possibly not as noticeable as other aspects of the implemented program, the simplicity aspect was more important than would be gathered from the documents. Several respondents stated that improvements were made, although more progress was unfortunately not achieved.	Complexity: ● ● Assessing the aspect would be pretty feasible, as it involves a qualitative approach with few extra requirements. Resource intensity: ● ● The evaluation requires a relatively low level of resources, as it can be done in 6-9 months, with relatively few necessary resources and one researcher/evaluator.

Table 13. Effectiveness evaluation option 1

The Basics		
Purpose	The Ministry of Justice & Security is already tracking basic statistical outputs in the aggregated data warehouse KMI+. A basic evaluation option is aimed at providing a comparative overview of these statistical outputs over several years, as far back as is gathered by KMI+ till 2013. Comparisons can be made of the differences in rates of application (inflow), inflow type, outflow, outflow type and the average duration of the applications and the overall return process need to be analyzed with the incorporation of any contextual influences. Possible aspects that this option could examine are; if the 'alternatives to detention' as implemented with the program show any increased application; if the use of ADM has gone down over recent years (specifically if a drop in application can be seen after the implementation of the program); if it is true that the average duration of the return process has increased after implementation of the program.	
Data Collection	Several variables need to be included in the data collection: the number of applications of a measure (inflow), the type of inflow, the outflow, type of outflow, and average duration of application. Preferably, as many data points in time should be included to provide a more accurate overview of the variances in the numbers.	
Data Availability	Data that is currently available:	Data that is still absent:
	<ul style="list-style-type: none"> - Aggregated monthly data on the number of applications (inflow), outflows, types of inflow and outflow in KMI+ - Some preliminary indications of contextual influences by the present study - See also the periodical reports from the organizations cooperating in the immigration process (e.g., (Ministry of Justice & Security, 2016b)) 	<ul style="list-style-type: none"> - Aggregated data on the average duration of the application of the measures - Aggregated data on the average length of a return process - Aggregated data is soon to be handed over by all involved government agencies on a weekly basis
Data analysis	<ul style="list-style-type: none"> - Quantitative analysis of statistical data: A basic statistical analysis can be done to analyze the possible variances in the data. - In addition: A limited number of interviews can be conducted as to provide a qualitative overview of all other influences that need to be incorporated into the interpretation 	
Resources	Conducting a basic statistical evaluation requires the access to the KMI+ data warehouse, basic statistical software, stakeholders that can describe possible contextual influences, a researcher with a basic understanding of statistical analysis and should take around 3 to 6 months.	
Utility	Priority: ● ● ●	Complexity: ● ● ●
	The evaluation option is of medium priority, mainly because the reporting of the statistical data is already done on a periodical basis. Further evaluation provides a comparative overview of the statistical data from these reports, with the addition of the durational variables. It thus is aimed at giving an overview of the numbers over recent years.	<p>The evaluation will be relatively easy to implement. Most of the data is already present in the KMI+ data warehouse. The only issue would be to include the durational variables, and how difficult it would be to retrieve these from KMI+ or the underlying data. In addition, some measures, such as the duty to report, are difficult to gather accurate data on for the average duration of application.</p> <p>The evaluation requires a relatively low level of resources, as it can be done in 3-6 months, with relatively few necessary resources and one or a pair of researchers/evaluators.</p>

Table 14. Effectiveness evaluation option 2

A Step into the Unknown		
Purpose	The second option for an evaluation of the effectiveness of the program builds upon the first option, as it aims to evaluate the causal relationship between the application of the supervisory measures and the outflow output of these measures and the overall outflow of the return process. Such an evaluation tries to link the practical output with the effectiveness outcomes. Although challenging, such an evaluation could provide the evidence for the possible influence these measures have on the outflow of demonstrable and non-demonstrable return. Moreover, it can be used as an example of the possibilities that further developments of KMI+ as the data warehouse for the immigration process can provide. With individual case (V-number) level analysis enabling more precise analyses, and together with a weekly handing over of the datasets thus making it useful to measure relevant changes in the data warehouse faster and more accurate.	
Data Collection	Several additional variables need to be included in the data collection in addition to the variables from effectiveness option 1. Most of all, the data needs to be collected on an individual case level, instead of as aggregated data. The following variables were obtained from the present study as relevant to collect: <ul style="list-style-type: none">- Case indicators (V-numbers + number of return procedure)- Age- Gender- Country of origin- Variable for each measure (Expressed in duration of application)- Inflow type- Outflow type (Dependent variable)- Unaccompanied minor (Expressed dichotomous)- Part of a family (Expressed dichotomous)- Indication of physical or mental health- Considered part of the vulnerable target group	
Data Availability	Data that is currently available: <ul style="list-style-type: none">- Aggregated monthly data on several of the variables available in KMI+- Possible individual case level data available on some of the variables- For the variables that influence return; see also the study by (Leerkes et al., 2011)	Data that is still absent: <ul style="list-style-type: none">- Structural and operationalized registration of indicators for well-being- Most data on the variables are registered in underlying datasets for KMI+ or in databases from the specific government agencies
Data analysis	<ul style="list-style-type: none">- Quantitative analysis of statistical data: To evaluate the effects of the supervisory measures, a sophisticated form of statistical modeling is necessary. Commonly used models such as a regression model are not possible with these variables. Possible other statistical models that can be used are:<ul style="list-style-type: none">o Binary Logit Model: A model that can be used to determine the likeliness that an event happens with a change in an independent variable. For example, the odds of a demonstrable return increase/decrease X times for every month someone is imposed with a duty to report. The advantage of the binary logit model is that it can also include more categories of the dependent variable in an extended Ordered Logit or Multinomial Logit Model. Though, the use of an extended model further increases the complexity.o Discrete Time Logit: A model that is used to determine the time it takes for a particular event to occur, and the influence on the time by any independent variables. For example, the time it takes for an irregular migrant to leave demonstrable and the importance of variables such as age, gender or country of origin. The advantage of this model is that it allows the analysis in regard to the time in the return process. Though, it has the problem of <i>censoring</i> as selecting a time-period automatically means that in some cases it is unknown when these became at risk of experiencing the event and actually experienced the event.	
Resources	The option requires extensive resources, as access to the data warehouse KMI+ is needed, in addition to further access to the underlying datasets and databases from the involved government agencies. Furthermore, the statistical models require complicated statistical software (SPSS, R) and powerful hardware to process the calculations. Multiple evaluators/researchers are needed that have extensive knowledge and experience with sophisticated statistical analysis. As the data gathering and analysis is very complicated, such an evaluation would require around 12 months to complete.	
Utility	Priority: ● ● The evaluation option is of low priority, mainly because the respondents indicate that other evaluations have more preference. Furthermore, the respondents question if the possible results are worth the effort.	Complexity: ● ● ● ● ● The evaluation will be difficult to implement. Though most of the data is located somewhere in KMI+, underlying datasets or databases of the government agencies, the data needs to be gathered, operationalized/transformed and put in one statistical file for further analysis. In addition, some variables are yet to be found in any database. It is thus complicated to determine the causal relationships between the variables accurately.
	Resource intensity: ● ● ● ● ● The evaluation requires extensive amounts resources. It requires broad access to most of the data that is available, the right means for the statistical analysis, and a lot of time and effort from multiple evaluators.	

Table 15. Proportional evaluation option 1

Assessing the Assessors			
Purpose	The purpose of the first evaluation option is to analyze the individual assessment step of the policy program. The individual assessment is meant to function as a form of case management on several conditions for each case. The evaluation would determine which considerations are made by the involved executive officials, which conditions are prioritized, and if any improvements can be made. Ultimately, the evaluation should aim to provide more substantive evidence for the indication from the present study that the individual assessment has improved under the current program. Thus, showing that the function of the administrative detention measure as an “ultimum remedium” has indeed been achieved and the other supervisory measures are viable options considered in the individual assessment.		
Data Collection	Data that needs to be collected is aimed at the considerations that are made for the application of a supervisory measure. It thus involves the data collection on the input of the implemented program directed towards an improved individual assessment, the actual individual assessment and the output related to the individual assessment.		
Data Availability	Data that is currently available:	Data that is still absent:	
	<ul style="list-style-type: none">- Aggregated data on the inflow numbers for the supervisory measures, as well as the inflow type registrations- Registrations of the application of a measure in the databases of responsible government agencies- Limited record of the consideration made for the application- INDIGO database of the INS can be used to retrieve data on any filed objections and the judicial reviews	<ul style="list-style-type: none">- Opinions and reflections by executive officials on the considerations within an individual assessment- Opinions and reflections by executive officials on any possible improvements & recommendations	
Data analysis	<ul style="list-style-type: none">- Basic quantitative analysis of database information: Basic statistical analysis of the data retrieved from the databases should provide some insight into the differences in the numbers of each inflow type for each supervisory measure- Quantitative OR qualitative analysis of the perceptions of the respondents: The primary analysis depends either on the need for more expanded (quantitative) or more in-depth (qualitative) evaluation of the individual assessment. Interviews can be conducted with respondents, that provide comprehensive information on the considerations that are important in making the assessment. These can be analyzed by transcribing, coding and further interpretation.- Questionnaires could be conducted that provide a more expanded overview of the considerations made, based on predetermined aspects that would be important to ask the respondents. These can be analyzed using statistical methods.- In addition: One could also conduct a limited number of interviews that form the basis for the quantitative questionnaires		
Resources	This option requires brief access to the databases of all relevant government agencies, as well as access to KMI+. Furthermore, a group of respondents needs to be found that are willing and able to cooperate with the study. The evaluation option probably requires one researcher, with some additional guidance. Due to a possible combination of both data gathering and analysis of information from the databases as well as from respondents, it is estimated that the evaluation requires between 6 to 9 months for completion.		
Utility	Priority:	• • • • •	Complexity: • • •
	Resource intensity:	• • •	
	The evaluation of the individual assessment is regarded as a high priority for the implemented program. Mainly as it is the primary impact of the program as to enable a range of options that can be applied depending on an assessment of each case.		
	The evaluation is likely to be of average complexity, as neither the analysis of the databases nor the interviews require a difficult method. However, as both forms are incorporated, the combined use could lead to some issues.		
	The evaluation requires an average level of resources, with a single researcher, average duration and a combination of the different sources of information.		

Table 16. Proportional evaluation option 2

Substantiating the Perceptions			
Purpose	The second evaluation option incorporates a practical approach to a broad analysis of the treatment of irregular migrants under the implemented program and if improvements have been made. The third option is specifically aimed at the physical and mental health effects of the supervisory measure, while this option takes a broader approach that includes a qualitative view of all involved stakeholders. As such, it consists of both aspects of the third and fourth option. This option can be seen as a sort of ‘satisfaction research.’ In which the perceptions of all involved stakeholders are gathered and analyzed, as to see if the implemented program has made the treatment of the irregular migrants better in general.		
Data Collection	Data needs to be collected from three different perceptions. The first perceptions that need to be collected are from the relevant government agencies involved in the return process from all layers of the agencies (from practice, policy, to judicial). The perceptions of the NGO's function as a more critical counterbalance. The third perceptions that should be included, if possible, are of those who have been in the return process and were imposed with a supervisory measure. Specifically, data needs to be collected on the perceived improvements, (unforeseen) consequences, and recommendations on the treatment of irregular migrants. The present study already contains some indications and relevant stakeholders that can be used as a starting point.		
Data Availability	Data that is currently available:		Data that is still absent:
	<ul style="list-style-type: none">- The present study provides some first indications- The present study provides some stakeholders that are relevant		<ul style="list-style-type: none">- In-depth data on the specific aspect from the first two groups.- Expanded data of more relevant stakeholders from the first two groups than with the present study- In-depth or expanded data on the perceptions from irregular migrants that have been imposed with a supervisory measure in their return process. Possible respondents could be found in the group who got an asylum residence permit after a newly requested asylum admission was made while they were in the return process
	<ul style="list-style-type: none">- Qualitative analysis of the perceptions of the respondents: A qualitative analysis could be used for a more in-depth evaluation of the opinions. Interviews can be conducted with respondents, that provide comprehensive information on the improvements, (unforeseen) consequences, and recommendation on the treatment of irregular migrants. These can be analyzed by transcribing, coding and further interpretation		
Data analysis	OR <ul style="list-style-type: none">- Quantitative analysis of the perceptions of the respondents: Questionnaires could be conducted that provide a more expanded overview of the improvements, (unforeseen) consequences, and recommendation on the treatment of irregular migrants. These can be based on predetermined aspects that are important to ask the respondents, which can be analyzed using statistical methods		
Resources	The option requires a group of willing and cooperating respondents from all three aforementioned groups. Most difficult would be to find such respondents between irregular migrants. As mentioned in the data availability, a possibility would be to find respondents from a group who has acquired an admission after being put in the return process with a supervisory measure imposed. The evaluation option probably requires one or a pair of researchers/evaluators, with some additional guidance provided by the Directorate-General for Migration. Furthermore, it is estimated that the evaluation would need between 6 to 9 months to complete.		
Utility	Priority:	• • • •	Resource intensity: • • •
	The evaluation of the general treatment of irregular migrants is regarded as a high priority by the respondents. It is thus important to conduct some form of evaluation on the effect of the implemented program on the aspect.	Complexity: • • The evaluation is likely to be of average complexity, as the analysis of either the interviews or questionnaires is of average complexity. The only issue would be to construct a questionnaire (if used) that captures all the aspects while not being too broad.	The evaluation requires an average level of resources, with one or a pair of researchers/evaluators, an average duration and three sources of information. With the finding of respondents from the third group potentially being the most problematic part.

Table 17. Proportional evaluation option 3

Limiting the Consequences			
Purpose	It is generally accepted by both governments, NGO's and the scientific field that limiting the freedom of movement by imposing a supervisory measure could potentially affect someone's well-being. However, this has mostly been an assumption, with only limited studies providing substantive evidence for the effects for the application of the 'alternatives to detention.' The third evaluation option is to present a proposal for such an evaluation of the impact of a supervisory measure on the well-being of an irregular migrant. It is recognized that such an evaluation is incredibly difficult to effectuate on a small scale, let alone on a comparative, larger scale. The difficulty lies in finding enough individuals that are willing to cooperate, while also paying attention to the questions asked to limit the chance of any biased responses (as the individual might feel that responding with a socially desired answer could influence their return process). Still, it would be interesting to analyze the different effects on the mental and physical health of those individuals from the various measures.		
Data Collection	Data needs to be collected from irregular migrants that are or have been, imposed with a supervisory measure. A selection can be made to include individuals from specific measures. For example, it is easiest to collect data from those imposed with a freedom-restricting measure that stay in a FRC with those imposed with an ADM. Other measures such as the duty to report and bail are more challenging to collect data from. It is recommended to limit the evaluation solely to the effects on the physical and mental well-being of the irregular migrant, as to make the review more manageable.		
Data Availability	Data that is currently available:	Data that is still absent:	
	- Limited findings from previous (pilot/small) studies: (Costello & Kaytaz, 2013; Fiske, 2016)	- Accurate data on the actual measures used in practice in the Netherlands - The mental health effects - The Physical health effects	
Data analysis	- Preferably: quantitative analysis of questionnaires; to measure these effects as accurate as possible, scientifically based questionnaires need to be found that are aimed at measuring the perceived mental and physical health of individuals. These need to be adapted towards the specific characteristics and situation of the return process. Moreover, these questionnaires should be administered before, during and after the imposing of a measure to measure any differences accurately		
	OR - Alternative: qualitative analysis of the perceptions of the respondents; Interviews can be conducted with respondents, aimed to provide in-depth information on the physical and mental health of the irregular migrants. These can be analyzed by transcribing, coding and further interpretation. Such an approach would be more straightforward, although the consequence is that interviews are a less accurate method for gathering health-related information		
Resources	The option requires a group of willing and cooperating respondents from the population of the return process. Most difficult would be to be able to administer these questionnaires before, during and after the irregular migrant has been imposed with a measure. Easiest would be, just as with option two, to first only include respondents from the FRC and ADM. It is estimated that an evaluation of this scale would require multiple researchers/evaluators if it incorporates the use of questionnaires. Furthermore, such an evaluation cannot be done retroactively, but instead needs irregular migrants that are going to be in the return process. Thus, this would require an extended period of more than 12 months to be able to gather enough data.		
Utility	Priority:	● ● ●	Complexity: ● ● ● ● ●
	Resource intensity:	● ● ● ● ●	
Utility	The evaluation of the physical and mental health effects of limiting the freedom of movement by a supervisory measure is of average priority as can be obtained from the findings of the present study. Although important to know, the difficulties surrounding such an evaluation would make its implementation of questionable value. Even more so as most stakeholders already assume that less-coercive measures pose less of a risk on these aspects as common sense.		
	The evaluation is likely to be exceedingly difficult to effectuate. Necessary scientifically based questionnaires need to be obtained and adapted, the possibility to administer these questionnaires before, during and after an application of a measure should be made possible and overcoming any obstacles that could interfere with an accurate measurement of the effects need to be overcome.		
Utility	The evaluation requires a high level of resources as it takes much time, requires multiple researchers/evaluators, access to the population that is imposed with a measure and a demanding methodology.		

Table 18. Proportional evaluation option 4

Professionalizing of Communication						
Purpose	The departure supervisors from the DT&V work according to the methodology of ‘werken in gedwongen kader’ (working in detention centers) since its implementation in 2015. That methodology was adapted by the University of Applied Sciences of Utrecht to the needs of the DT&V. It is the intention that the method will be evaluated in the foreseeable future. That evaluation is included as an evaluation option in the present study, as it was one of the inputs given that assists in the communication and assessment of the application of the supervisory measures. The purpose of such an evaluation is to analyze the use of the methodology in the interactions between the departure supervisor and the irregular migrant. For example, it can investigate if the adaptation of the method was sufficient, or if any adjustments need to be made further.					
Data Collection	Data needs to be collected from departure supervisors that were trained in the use of the methodology in their communications with irregular migrants. Preferably, the data collection should include the perception of a varied group of supervisors. Both supervisors that were trained from the start and those that were already supervisors beforehand provide different takes on the methodology. In addition, some information should be collected from either the irregular migrants themselves or from NGO's to give a balanced view of the implementation.					
Data Availability	Data that is currently available:	Data that is still absent:				
	- Implemented methodology in its adapted form by the University of Applied Sciences of Utrecht	- Follow up data on the implementation of the program				
Data analysis	<ul style="list-style-type: none">- Quantitative analysis of questionnaires: The primary approach to the evaluation involves the use of quantitative analysis of surveys to measure the perceptions of the departure supervisors on the implementation of the methodology. These can be analyzed using statistical methods- In addition: A limited number of interviews can be conducted as to; (1), provide the basis for the quantitative questionnaires, (2) Include the perceptions of NGO's and irregular migrants, (3) provide some limited in-depth information on the implementation of the methodology					
Resources	The option requires a group of willing and cooperating departure supervisors as respondents, as well as a limited number of respondents from NGO's and irregular migrants. An evaluation can likely be done by one researcher, with additional guidance from the Directorate-General for Migration. It would probably take around 6 to 9 months to complete.					
Utility	Priority:	● ● ● ● ●	Complexity:	● ●	Resource intensity:	● ●
	This evaluation option is (most likely) already being planned or in the process of being implemented. The methodology has been implemented and worked with, so an evaluation would now be appropriate to determine if the implementation was successful and if it leads to the desired effects in the communication with irregular migrants. Thus, this evaluation option is of a high priority to be implemented further.		The evaluation will be fairly easy to implement. As the University of Applied Sciences made an adapted form of the methodology, any evaluation could use that as a framework for the analysis. The only issue would be to incorporate some of the perceptions from the irregular migrants themselves.		The evaluation requires a relatively low level of resources, as it can be done in 6-9 months, with relatively few necessary resources and one or a pair of researchers/evaluators.	

E. Executive Summary for the Department of the Directorate-General for Migration (in Dutch)

Executive summary

'Evaluability Assessment' van het aangepaste beleid omtrent de toepassing van (alternatieve) toezichtsmaatregelen in het terugkeerproces

Frank E. Hendriks, Radboud Universiteit Nijmegen 2018

Doel:

Het bepalen of de Directie-Regie Migratieketen klaar is om het aangepaste beleid voor de toepassing van alternatieve toezichtsmaatregelen te evalueren



In 2013 zijn er wijzigingen aangebracht in het beleid rondom de toepassing van toezichtsmaatregelen in het terugkeerproces. Op dit moment is er nog geen uitvoerige evaluatie gedaan naar de effecten hiervan. Een eerste aanzet hiertoe liet zien dat er behoefte was naar een voorbereidende analyse naar de mogelijkheden voor een evaluatie. Dit onderzoek heeft als doel gehad om dit uit te werken. Hierbij richtte het zich met name op het verduidelijken van de doelen, uitkomsten en benodigde data. Daarnaast diende het om de verschillende invalshoeken en aanpakken voor een dergelijke evaluatie inzichtelijk te maken.

Aanpak:

Toepassing van het Evaluability Assessment (EA) model door middel van een document analyse en het afnemen van semi-gestructureerde interviews

Evaluability Assessment:

EA is een vorm van inleidend evaluatief onderzoek dat gericht is op het bepalen of een organisatie klaar is voor het uitvoeren van een evaluatie van bepaald beleid of een programma. Dit onderzoek heeft een EA-model van Wholey (2015) toegepast, aangezien dat model een goede combinatie is tussen het summatieve aspect (welk effecten zouden er geëvalueerd moeten worden) en het formatieve aspect (wat kan er worden verbeterd om verdere evaluatie mogelijk te maken). De zes stappen van dit model kunt u vinden in de rechter kolom. Om de benodigde informatie te verkrijgen zijn er twee analyses toegepast:



Document analyse: Verschillende relevante documenten werden verzameld. Deze werden onder andere aangedragen door de werkgroep en andere betrokkenen. Informatie hieruit werd inzichtelijk gemaakt door gebruik te maken van het kwalitatieve software programma ATLAS.ti.



Interview analyse: Acht betrokkenen zijn geïnterviewd door middel van semi-gestructureerde vragenlijsten. Deze betrokkenen leverden een overzicht van de beleidsmatige en cijfermatige aspecten van het beleid. Informatie hieruit werd getranscribeerd en gecodeerd voor verdere analyse met behulp van ATLAS.ti.



EA-model van Wholey (2015)

Stap 1

Betrek beoogde betrokkenen en andere belangrijke ketenpartners

Stap 2

Verduidelijk/beschrijf het beleid

Stap 3

Analyseer het beleid zoals geïmplementeerd in de praktijk

Stap 4

Beoordeel de plausibiliteit van het beleid

Stap 5

Bereik overeenstemming over noodzakelijke wijzigingen in het beleid of de implementatie

Stap 6

Bereik overeenstemming over de richting van verdere evaluatie en het beoogde gebruik

Resultaten & Discussie:

De resultaten bestaan uit twee delen:

'Program Theory & Logic Model': Dit deel bestaat uit de uitwerking van de tweede stap. Op basis van de documenten en korte gesprekken met de werk groep en andere betrokkenen is een beschrijving gemaakt van het beleid op een aantal aspecten. Een 'Logic Model' is gemaakt om een overzicht te geven deze beschrijving (zie de bijlage).

'Program Reality': Dit deel bestaat uit de uitwerking van de derde stap. Een vergelijkende analyse is uitgevoerd tussen het beleid zoals bedoeld en het beleid zoals geïmplementeerd, op basis van het vorige deel en de interview data.

Deze resultaten leiden tot de volgende drie onderdelen van de discussie:

'Program Plausibility & Data Availability': Het summatieve aspect van deze studie. Dit deel bestaat uit het identificeren van de doelen, uitkomsten, indicatoren en beschikbare data voor verdere evaluatie. Dit deel is gebaseerd op de vierde stap van het EA-model.

'Program Utility': Het formatieve aspect van het onderzoek. Dit deel bestaat uit het geven van verschillende aanbevelingen om verdere evaluatie mogelijk te maken. Dit deel is gebaseerd op de vijfde stap van het EA-model.

'Program Evaluation': Het laatste aspect van deze studie. Dit deel bestaat uit het presenteren van verschillende evaluatie opties (incl. Een 'status quo' optie) die betrokkenen activeren en informeren over de verschillende mogelijkheden voor verdere evaluatie. Dit deel is gebaseerd op de zesde stap van het EA-model.

De bevindingen van de discussie worden hieronder samenvat.

'Program Plausibility & Data Availability': Doelen, uitkomsten, indicatoren, data

Doel	Uitkomst & Indicator	Data
1 Noodzakelijkheid: Terugkeerbeleid efficiënter maken in het gebruik van de benodigde middelen	1. <i>Kostenreductie:</i> Gemiddelde kosten per dag per maatregel 2. <i>Versimpeling:</i> Uniforme en eenvoudige toepassing van zelfde maatregelen door elke organisatie	1. <i>Kostenreductie:</i> <ul style="list-style-type: none"> • Totale kosten van toepassingen • Gemiddelde kosten totaal en per dag 2. <i>Versimpeling:</i> Vragenlijsten en/of interviews met uitvoerende ambtenaren
2 Effectiviteit: De toepassing van de toezichtsmaatregelen blijft ten minste even effectief	1. <i>Toepassingen:</i> Aantal toepassingen per maatregel 2. <i>Terugkeerpercentage:</i> Percentage van vertrek 3. <i>Ontwikingspercentage:</i> Ontwikking toezicht 4. <i>Duur maatregel:</i> Tijd van oplegging maatregel	<ul style="list-style-type: none"> • <i>Basis:</i> Geaggregeerde data op alle vier de aspecten • <i>Uitgebreid:</i> Data op case-niveau op alle vier de aspecten en andere variabelen (zie studie Tabel 14)
3 Proportionaliteit: Zorgen voor een humaner en waardiger terugkeerbeleid	1. <i>Gezondheid individu:</i> Mentale & fysieke gezondheid 2. <i>Perceptie behandeling:</i> Rechtmatigheid behandeling 3. <i>Beoordeling assessment:</i> Of de vreemdeling het eens is met de gekozen maatregel	1. <i>Gezondheid individu:</i> Vragenlijsten en/of interviews 2. <i>Perceptie behandeling:</i> Vragenlijsten en/of interviews 3. <i>Beoordeling assessment:</i> Database analyse, vragenlijsten en/of interviews

Aanbevelingen

- ⚙ Blijf werken aan het centraliseren van het terugkeerbeleid en de toepassing van de toezichtsmaatregelen. Zorg voor een uniforme structuur voor de oplegging van de maatregelen.
- ⚙ KMI+ biedt al uitgebreide mogelijkheden voor analyse. Voor verdere analyse van de toezichtsmaatregelen op een uitgebreide manier zou het ook mogelijk moeten zijn om data op case-niveau te kunnen gebruiken.
- ⚙ Het kan nog beter worden gecommuniceerd dat er geen 'alternatieven voor detentie' zijn in het terugkeerbeleid. Elke maatregel heeft zijn eigen voorwaarden, kenmerken en functie. De toezichtsmaatregelen bestaan dus uit een pakket van maatregelen waaronder administratieve detentie. Een betere term voor de andere maatregelen zou zijn 'minder-restrictieve maatregelen'.

Evaluatie opties

Evaluatie opties zijn gepresenteerd verdeeld over de drie verschillende doelen:

💡 Noodzakelijkheid

- *Kosten evaluatie*: Gericht of de evaluatie van de totale kosten van de maatregelen (zie studie, Tabel 11)
- *Evaluatie van simplificatie*: Analyse van potentie om het beleid verder te simplificeren (zie studie, Tabel 12)

💡 Effectiviteit

- *Basis evaluatie*: Gericht op een eenvoudige analyse van de effectiviteit (zie studie, Tabel 13)
- *Uitgebreide evaluatie*: Gericht op een uitgebreide analyse van de effectiviteit (zie studie, Tabel 14)

💡 Proportionaliteit

- *Assessment evaluatie*: Evaluatie van de afweging voor een bepaalde maatregel (zie studie, Tabel 15)
- *Perceptie van behandeling*: Analyse of het beleid tot een betere behandeling heeft geleid (zie studie, Tabel 16)
- *Gezondheid evaluatie*: Evaluatie van de effecten op de fysieke & mentale gezondheid (zie studie, Tabel 17)
- *Communicatie evaluatie*: Evaluatie van de nieuwe gespreksmethode bij de DT&V (zie studie, Tabel 18)

CONCLUSIES

Summatieve conclusie: De vergelijking van de resultaten laat zien dat de effectiviteit van het beleid nog voornamelijk wordt gedaan op basis van de assumpties en inzichten van relevante betrokkenen. Kwantitatieve data is moeilijk te gebruiken, aangezien er daarvoor eigenlijk geaggregeerde informatie nodig is op case-niveau. *Formatieve conclusie*: De resultaten laten zien dat de oorspronkelijke 'alternatieven voor detentie' nu onderdeel zijn van het totaal pakket van toezichtsmaatregelen. Deze moeten dan ook anders worden benoemd dan nu het geval is, zodat het duidelijk is dat deze maatregelen geen alternatief in letterlijke zin zijn. Elke maatregel heeft zijn eigen voorwaarden, kenmerken en functie.

Het beleid is op dit moment nog alleen te evalueren op basis van algemene assumpties en informatie. Aanbevelingen en opties uit deze studie kunnen worden gebruikt om verdere uitgebreide evaluatie mogelijk te maken.

LIMITATIES

- ✗ Vanwege praktische limitaties zijn er een beperkt aantal interviews afgenomen. Door zorgvuldige selectie is toch geprobeerd een representatief beeld te krijgen voor het onderzoek.

BRONNEN

- ➡ Wholey, J. S. (2015). Exploratory Evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 88-107). New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119171386.ch4>

Bijlage Logic Model: Overzicht beleid

