

# Chapter 10

## Three worked ELS examples

### 10.1 Introduction

In this chapter we will attempt to illustrate how the methods that we discussed in the previous chapters are applied in practice. We have chosen three examples to do so, which exhibit spread over the type of methods, substantive legal domain, and type of research question.

The first example deals with social security law, which falls under administrative law. In the example we showcase a qualitative analysis of the experiences of welfare recipients who have been forced to take part in programmes that are aimed to reduce – as it is euphemistically called – their ‘distance to the labour market’. These mandatory work programmes are intended to reduce dependence on welfare and to facilitate a return to paid employment. The research questions focus on whether the welfare recipients involved do experience these programmes as meaningful (or meaningless), and whether and why the programmes indeed – again in the experience of the persons involved – facilitate their return to the labour market. These questions as such belong within the third pillar of the trias ELSica that we introduced in section 1.3, the impact of laws.<sup>1</sup>

The second example deals with private law. In this example we illustrate the various steps in systematic case law analysis, focusing on a question that has long remained unanswered, namely whether parties who offer their apologies to victims, for instance victims of medical malpractice, thereby also admit liability in a legal sense, and thereby would weaken their position. The study neatly lays out the sampling procedure of court rulings, the doctrinal analysis of rulings and arguments used, combining this relatively standard analysis with qualitative and quantitative methods to arrive at an answer to the research questions. This study is best positioned somewhere within the first pillar of the trias, where questions are asked about assumptions in law.

The third example deals with international criminal law. This example uses simple descriptive quantitative analysis methods as well as a relatively advanced multivariate analysis to investigate whether sentences imposed by two international tribunals, the

---

<sup>1</sup>I am much indebted to Anja Eleveld who supplied much of the text for this section.

International Criminal Tribunal for the former Yugoslavia (ICTY) and the International Criminal Tribunal for Rwanda (ICTR), are predictable, or – as had been posited by a number of legal scholars – random. The study included all rulings at first instance and on appeal that had been handed down at both courts at the time the study was conducted. The research questions dealt with here are best positioned within the second pillar of the trias ELSica, the operations of the legal system.

## 10.2 Experiences with mandatory work programmes: qualitative analysis

In the Netherlands, as in many other welfare states, social security law contains regulations that compel welfare recipients to ‘use a facility offered by the municipality, including social activation, aimed at the transition to paid work’ (translated text from the Netherlands Participation Act, art. 9(1)(b)). This means that those who are receiving welfare can be forced to enrol in programmes that are intended to encourage a transition to the labour market.<sup>2</sup> That there can a certain degree of force is also illustrated by the fact that municipalities are obliged to cut benefits by 100% for at least one month if a recipient of social assistance benefits refuses to comply with the law.

Since the mid-1990s (the previous versions of) this regulation has been the legal basis for the obligation imposed in the Netherlands on welfare recipients to enrol in a Mandatory Work Programme (MWP). Despite the non-voluntary nature of these programmes, international quantitative studies using numerical measurement instruments have shown that MWP participants report relatively high levels of well-being (Carter & Whitworth, 2017; Knabe, Schöb, & Wiemann, 2017; Wulfgramm, 2011). Eleveld (2021) aimed to investigate this apparent paradox.

Instead of measuring well-being using one-dimensional linear measures, her study aimed to arrive at a more multifaceted and contextualized understanding of the effects of these programmes on welfare recipients’ well-being against the legal obligation to participate in MWPs. Related to this, the study aimed to explore to what extent the assumption that these legal regulations will lead to welfare recipients’ transition to (regular) work corresponds with MWP participants’ lived experiences. Since the researcher was primarily interested in the manner in which the forced nature of work impacts MWP participants’ well-being, the study primarily focused on MWP participants’ meaningful (or meaningless) work experiences. As such, this study aimed to answer the following main research questions:

- (1) *in what manner do welfare recipients participating in an MWP experience and interpret their work as meaningful or meaningless?*
- (2) *how are recipients’ interpretations and experiences related to their (subjective) prospects of transition to regular work?*

---

<sup>2</sup>This aim is elaborated in the so-called Memorie van Toelichting to the Participation Act (Kamerstukken II 2013/14 33801, No. 3, pp. 27–40), as well as in Memories van Toelichting to its predecessors Algemene bijstandswet and Wet Werk en Bijstand (Kamerstukken II 1991/92, 22545, No. 3, pp. 68–69; Kamerstukken II 1996/97, 25122, No. 3, pp. 25–26; Kamerstukken II 2002/03, 28870, No. 3, pp. 39–40).

### 10.2.1 Data collection

The study was conducted in three municipalities in the Netherlands to ensure spread over the Netherlands and urbanization: two medium-sized municipalities were chosen with just over 60,000 inhabitants and one large municipality with over 300,000 inhabitants. In all three municipalities, MWP were in force. In each municipality, the researcher visited five to seven MWPs, again to ensure spread over MWPs. The MWPs for the interviewees generally entailed unskilled work such as simple production activities, call centre work, maintaining municipal green spaces, postal delivery, canteen work, serving coffee to people in a nursing home or simple administrative work. The official goal of most of these MWPs was to develop and keep up basic working skills, such as arriving at work on time, obeying an employer's instructions and so on, their work attitudes and, if possible, to further explore their work preferences.

Eleveld conducted in-depth interviews with on average two to three recipients per MWP. Participants were mostly introduced to the researcher by supervision staff at the programmes, although Eleveld also recruited some through snowballing via earlier interviewees. All in all, a total of 42 recipients of social assistance benefits who were participating in an MWP were interviewed (with approximately equal numbers per municipality). In most MWPs, recipients worked for between 20 and 32 hours a week for a minimum of three months – although often for longer. Five respondents had been participating in one or more MWPs for periods of over two years. A total of 24 women and 18 men were interviewed. Most respondents were 40 years old or over, and one-quarter had a migrant background. The interviews were conducted in 2017 and 2018. The interviewees gave informed consent to participate in the research.

Interviews lasted on average one to one-and-a-half hours. The interview format was topical, with interview topics including participants' evaluation of the MWP, their views on prospects of a paid job, relationships with colleagues and supervisors, working conditions, and the role of sanctions. The format of the interview was free-flowing, with ample opportunity for interviewees to broach new subjects and add information. Interviews were conducted by the researcher. All interviews were recorded and later transcribed, leading to several hundreds of pages of transcribed material to be analysed. For more on the data collection, see Eleveld (2021) and Dermine & Eleveld (2021).

### 10.2.2 Analytic methods: coding and discovering patterns

Given the large amount of textified data available, the interview transcripts were analysed using a software package, in this case ATLAS.ti. Coding and structuring the data took place over four steps, which are detailed below, after which the research questions were answered using the patterns that the coding and structuring steps had revealed.

#### *Step 1: Descriptive (first-level) coding*

In a first round, all parts of the text that directly or indirectly referred to the participants' valuation of MWP work were marked by the authors. These texts were then re-read and re-coded, predominantly using gerunds as primary codes to describe what was going on in the text (Charmaz 2012), labelled as 'Thinking along with the work programme's management', 'Feeling depressed at home', 'Rediscovering oneself', 'Learning the Dutch language', 'Feeling insecure', 'Complaining that time passes slowly', 'Enjoying

contacts with colleagues'. These first-level codes (along with subsequent second-level codes) and the original text fragments are given in Table 10.1.

### *Step 2: Focused (second-level) coding*

In a next step, Eleveld combined the numerous primary codes<sup>3</sup> with similar codes, where possible simultaneously re-wording in more abstract terms (i.e. second-order categories) the descriptive codes. This entailed adding a second, interpretative layer to the relatively 'flat' first-layer codes. For example, 'Thinking along with the management' was given the second-order code 'Experiencing autonomy', attaching interpretation to the literal, first-level code. As another example, the primary code 'Rediscovering oneself' was given the second-level code 'Personal development – being yourself and knowing yourself'. While the label of this second-order code was still quite close to the primary code, in other instances the researcher actually took more liberty when formulating a second-order code. To give an example, the primary code 'Feeling depressed at home' was re-coded as 'Experiencing a more structured life', with the author motivating this through the participant clarifying his statement that working in an MWP established a basic rhythm in his life and, as such, pulled him out of a destructive situation. The example shows how coding does not amount to simply categorizing the answers, but can already entail interpreting (and as such analysing) the data.

It should be noted that, after the interviews had been read and re-read, some quotations could in fact be linked immediately to interpretive or 'second-order' interpretive codes. Here, the first-level coding step could be skipped, because the meaning of the text fragment was already clear, and a second-level code could be given right away. Examples of these secondary order codes are 'Feeling useful/contributing', 'Experiencing forced work', 'Gaining self-esteem', 'Feeling being treated inferior' and 'Lacking a prospect of future work'. These second-level codes that were found without a first step, and the original texts are given in Table 10.2.

During this process, some second-order codes were also merged. For example, the second-order code 'Feeling bored' was merged with a second-order code 'Learning nothing', as it turned out that quotations were often linked to both these codes. Re-reading such double-coded quotations (fragments of text that have been attached multiple codes) revealed that boredom was closely related to the lack of opportunities for self-development. Also, the second-order codes 'Gaining self-esteem' and 'Feeling respected' were merged, as re-reading the double-coded quotations indicated that having self-esteem was often preconditioned upon being respected. As such, merging second-order codes also constituted an analytic step.

### *Step 3: Axial coding*

In a next, third step, a large number of second-order codes were grouped together in broader umbrella-type categories such as 'Meaningful work experiences' or 'Meaningless work experiences'. This so-called 'axial coding step', in which dimensions or axes are identified along which the second-order codes can be placed or ranked is another analytical step. Here the researcher seeks to combine or even rank the second-order

<sup>3</sup>We use 'primary code', 'first-level code' and 'first-order code' interchangeably, as we do 'secondary order code', 'second-level code' and 'second-order code'.

**Table 10.1:** First- and second-level codes and quotations

First-level code	Second-level code	Quotation
Thinking along with management	Autonomy at work	"I really like that it has yet to be set up. So that you get to think about a lot of things. It's informal"
Feeling depressed at home	Structured lives	"I couldn't bring myself to do anything. I had very heavy mental blocks"
Rediscovering oneself	Personal development-being and knowing yourself	"I also need to know where I stand in my life and that's really what it is about (...) to discovering yourself again because I was really stuck with myself"
Learning the Dutch language	Personal development skills	"Also, my language improved, because at home we talk our language, Croatian and in terms of (Dutch) language I had forgotten a lot"
Feeling insecure	Lack of supervision	"When she said I can't come anymore, I said: "what are you telling me. Then my heart started beating (...) I can't manage this". She said, ask if you need help, ask Karen or the girls at the reception, they know everything. The next day I stood there alone, and I thought where to start and what to do"
Complaining time passes slowly	Feeling bored	"I have to work too many hours. Time doesn't go so fast. Half a day would be better"
Enjoying contacts with colleagues	Building & enjoying social relationships	"It was a fun group. It was like having a warm blanket wrapped around you, like coming home. You were immediately included in the group"

**Table 10.2:** Second-level codes and quotations

Second-level code	Quotation
Useful/ contributing	“I have no problem with having to do something for my benefits. I don't mind that. Then I still feel that I am somewhat useful. When I was still at home, I started asking myself in the morning, why do I get out of bed? I might as well stay in bed. Then you start to feel useless”
Forced work	“I'm only here to avoid a financial sanction”
Gaining self-esteem	“They kept saying: “You are doing well”. I really liked that’ ,
Feeling being treated inferior	“Paper picking that was really the lowest point of my life I would say. I found it something I never thought I would have to do”
Lacking a prospect of future work	“It's all fun but at some point you think: “Yes, there's really no point to it all because no one will hire you”. And then you stop looking at it positively. In the beginning, you're bursting with energy and really want to do it. But at some point you think: “Yes, that's it, yes”” .

**Table 10.3:** Grouping second-order codes related to meaningful work experiences

Subset	Second-order code
Conditions	Structured lives
	Autonomy at work/agency
Dimensions	Personal development skills
	Personal development being and knowing yourself
	Useful/contribution
	Building and enjoying social relationships
Consequences	Gaining self-esteem
	Getting social recognition

codes. What distinguishes this step from the previous steps can best be explained as follows. During the first steps, the aim is to categorize and interpret the text fragments: first using descriptive codes (codes that simply describe or summarize what is said), then using interpretative codes (codes that add meaning to what is said). In quantitative terms, one could frame these as univariate steps. In the third step, the axial coding, the researcher attempts to see how the second-order codes relate, what associations there are between them, or whether some kind of underlying or manifest structure can be discerned. Here the analysis is – one could say – multivariate (or multi-category).

In the analysis of MWP, Eleveld did just that. She first grouped second-level codes into larger sets, or clusters of codes, such as ‘Meaningful work experiences’. In doing so, she searched for codes that cluster in the sense that they refer to the same themes or general topics. After having done so, she also investigated how these codes related. Eleveld for instance discovered how the codes in the category ‘Meaningful work experiences’ could be distinguished into three subsets, differentiating between *conditions* for meaningful work experiences, aspects or *dimensions* of meaningful work experiences, and *consequences* of meaningful work experiences (see Table 10.3).

To illustrate findings, qualitative researchers often use exemplars. Exemplars are specific examples in the dataset that illustrate the findings (Tracy, 2020; see also section 7.1.2). To present her data, Eleveld also used a number of exemplars, one of these being a participant named Jeroen. He was a brilliant exemplar as he illustrated vividly how the four identified dimensions of meaningful work were related. He said:

*“What I like about this work? First of all, the camaraderie and the atmosphere. I’m happy to be working again. I am among people. I like to help customers. They come to me asking questions and sometimes I am able to help them and sometimes not. I feel that I am meaningful for society. I’m not only sitting at home, alone, a loner. I am someone. I can be myself again. I can do things, have a drink with colleagues. That is fun, you are connecting with new people. And the best thing of all is that you can simply be human and be yourself.”*

**Table 10.4:** Grouping second-order codes related to meaningless work experiences

Subset	Second-order code
Extrinsic meaningless	Bad working conditions
	No or insufficient 'payment'
	Forced to work
Intrinsic meaningless	Boring/learning nothing
	Lack of supervision
	No prospects of future work
	Feeling being treated as inferior
	Not being recognized by others

In this process of axial coding, it is possible to search for various kinds of relationships between codes. And this is also what Eleveld did. In addition to her cluster of meaningful work experiences, she had also identified a cluster of *meaningless* work experiences. Within this subset, she further distinguished between *extrinsic* meaningless work experiences, experiences to do with how the work was organized, and *intrinsic* meaningless work experiences, that is experiences to do with what according to the interviewees the work entailed in practice (see Table 10.4).

#### *Step 4: Discovering patterns*

In a final step, Eleveld analysed the patterns in the data, that is, whether some respondents had particular profiles of answering patterns. Eleveld had found that MWP work could be experienced both as meaningful and as meaningless, across various dimensions. She had also found that some interviewees experienced their work as exclusively meaningful or exclusively meaningless, but that other participants had mentioned both meaningful and meaningless work experiences. Based on an additional analysis of the data (not reported in detail here) Eleveld had also concluded that the assumption that MWP work would lead to a transition to paid work did not meet with most participants' lived experiences. On the contrary, quite a lot of them complained about the lack of future work prospects. Eleveld wanted to look for patterns in the dataset in order to understand why and how MWP participants experienced their work as meaningless and/or meaningful and how these experiences were related to their complaints about the lack of future work prospects, their personal characteristics and the characteristics of the particular MWP they worked in.

To this end, the respondents were grouped according to firstly their meaningful and meaningless work dimensions, and secondly the extent to which they complained about the lack of future work prospects. Based on these two principles, participants were allocated to five different categories. First, one group of participants was formed who referred to one of the meaningful work dimensions, without referring to any of the intrinsic meaningless work dimensions, one could say a 'positive' group. Next, from those participants who referred to both meaningful and meaningless work dimensions, first a group was formed that complained about the lack of future work prospects and



second a group where participants did not utter such complaints. This aspect (whether or not participants were positive or negative about future work prospects) was of course a key variable, as the programmes had been designed to improve work prospects. The remaining participants were divided into a group that participated in the two MWP that were – instead of supporting them to move to paid work – primarily used to punish welfare recipients (a salient characteristic as this would likely also have an association with respondents' lived experiences and future outlook) and a group comprising the remaining participants in one of the other MWPs.

Eleveld subsequently explored what characterized and distinguished the participants in each of these groups, apart from the principles that were used to form them in the first place. To this end she explored whether the categories could be connected to: 1) one or more specific meaningful or meaningless work dimensions; 2) extrinsic meaningless work dimensions (e.g. working conditions, the experience of forced work) or the nature of work; 3) a specific municipality (remember that the study was conducted in three municipalities); 4) personal characteristics (e.g. sex, age, migrant background, years of regular work experience, engagement in volunteer work, a combination of work and care); or 5) any other second-order code that had many quotations (e.g. treatment by the municipality; work identification; framing of work, work ethos; experiencing/using autonomy within work). In this step, the quotations were re-read to explore which 'emotions' were characteristic of these group members. The procedure that was employed is clearly iterative, and illustrates the fluid and flexible analysis modes that are employed in qualitative research.

Table 10.5 gives the patterns that were found. First, while participants in the above-mentioned groups 1, 2 and 3 all referred to some of the meaningful work dimensions, participants in groups 1 and 3 specifically emphasized that their meaningful work experiences were related to their personal development and to gaining self-esteem and social recognition. Participants in groups 1 and 3 were also clearly more engaged in volunteer work and the combination of work and care activities, compared to the participants in any of the other groups. These participants also experienced more autonomy compared to participants in any of the other groups (despite the forced character of the MWP), either because of the (empathic) way they were treated by the municipality or because of the freedom they experienced in the MWP. All in all, these participants seemed to be less bothered by their (lack of) prospects for paid work: other issues such as self-development and volunteer work were more important to them.

Participants in these groups were also, compared to the participants in the other groups, relatively often partly incapacitated for work. In addition while participants in group 1 were relatively old (group 1), participants in group 3 had a fairly long work history. Many participants in these groups did not envisage a quick transition to paid work because of their work disability and/or old age. For them the MWP work offered rather a good alternative to sitting at home. However, participants in group 1 were generally happier with their work compared to participants in group 3 who were more critical of the MWPs but seemed to accept their situation under current conditions and sought to make the best out of it. These participants also tended to frame their MWP work as regular work.

Members of group 2 seemed to be more ambitious when it came to the transition to paid work. These participants also seemed to endorse a stronger work ethos compared

to members of groups 1 and 3. However, many of them were disappointed because of the lack of skills development and the lack of clear job prospects.

Members of group 4 (who participated in the disciplinary programmes) and (to a lesser extent) those in group 5 complained about the forced nature of the work, which also influenced the way they experienced their work – as exclusively meaningless. Yet in this group the boring nature of the work (instead of the lack of future work prospects) stood out. This probably also explains why members of group 4 tended to be more resigned about their situation compared to members of group 5.

Members of group 5 were dissatisfied and sometimes even angry with their situation. Like participants in group 3, they complained about the ways they were treated by the municipality. However, in contrast to members of group 3, they were also disappointed about the lack of job prospects and the insufficient ‘payments’ for the work done.

### 10.2.3 Answering the research questions

Based on the analysis in the first three steps, Eleveld concluded that the assumption that MWP work would lead to a transition to paid work did not match most participants’ lived experiences. In addition, she drew the following (broad) conclusions. First, legal regulations that compel welfare recipients to participate in an MWP imply for many of them that they are forced to perform work that they experience as meaningless – either because the work is boring, or because they do not see much prospect of a transition to paid work, or they get insufficient supervision, or they feel that they are treated as inferior or are not socially recognized in any other way.

However, second, the analysis indicated that these regulations may as well – seemingly paradoxically – increase participants’ sense of agency and autonomy, especially when respondents had earlier felt they had lost control over their lives. In these instances, participation in MWPs may (also) give rise to meaningful work experiences along various dimensions (feeling useful, building and enjoying social relationships, personal development) and result in an enhanced sense of self-esteem.

Third, the analysis indicates that welfare recipients tend to experience MWP work as more meaningful when the legal regulations that compel them to participate in an MWP are implemented in less restrictive ways or when the work permits them to enjoy some autonomy and where they are less focused on transition to paid work. Under these conditions welfare recipients may focus more, through the MWP, on the development of their inner selves and appreciate their MWP participation, or simply accept their MWP participation for lack of better alternatives. However, where MWP participants expect that participation in an MWP will result in a transition to paid work, initial acceptance or enjoyment may turn to disappointment or anger, especially in the event of strict municipal implementation.

In sum, Eleveld concludes that quite a number of welfare recipients do not experience and interpret their work as meaningful. It is striking that many respondents also mentioned both meaningful as well as meaningless aspects of their MWP experiences. The legal regulations that compel welfare recipients to work in an MWP lead often to meaningless work experiences which negatively impact well-being. These regulations may, however, also give rise to meaningful work experiences and, as a

**Table 10.5:** Identifying patterns

Group (N)	Criteria	Main characteristics	Salient emotion
1 (N=10)	Exclusively referring to meaningful work dimensions	Focus on development inner self Positive about treatment by municipality Best option under current conditions Relatively old age Partly incapacitated	Content/ happy
2 (N=6)	Referring to both meaningful and meaningless work dimensions, including the lack of prospects of paid work	Ambitious, focus on future paid work Work is important	Disappointed
3 (N=11)	Referring to both meaningful and meaningless work dimensions, but not the lack of prospects of paid work	Ambiguous, valuing development inner self Enjoying autonomy within work Experiencing work negatively in all dimensions Best option under current conditions Partly incapacitated	Content/ acceptance angry
4 (N=6)	Exclusively referring to meaningless work dimensions and working in a disciplinary MWP	Work is experienced as forced work Forced character colours work experiences Feeling bored at work	Resigned
5 (N=9)	Exclusively referring to meaningless work dimensions, but not working in a disciplinary MWP	Focus on future paid work and developing skills Complaining about the bad treatment by the municipality and (to lesser extent) experiencing work as forced work	Disappointed/ angry

result, in increased well-being, especially when legal regulations are implemented in a non-restrictive way, leaving more autonomy to welfare recipients and where MWP participants do not expect that their work will lead to paid work transitions.

The analysis in addition shows that welfare recipients' experiences with MWPs depend on recipients' background and personal characteristics, and on the manner in which these programmes are implemented. The qualitative analysis showed that while programmes were designed to ease and in fact at times force welfare recipients' transition to the labour market, they appear to be appreciated most, and contribute most to well-being, when they are seen to be employed not for these purposes. This shows that recipients' interpretations and experiences are linked to their (subjective) prospects of transition to regular work, but in a sense in a reverse manner than anticipated: it is particularly the unambitious, those who do it in a sense 'on the side' and who do not have many expectations, who gain the most.

Regardless, one perhaps unfortunate conclusion from the research appears to be that the envisaged stimulus for transition to the labour market does not appear to materialize. Those employed in the programmes find them satisfying as a pastime that is better than sitting alone at home, provides activities that improve their self-image, and makes them feel useful. The programme does not appear to be perceived as an instrument for social mobility. The punitive aspects of the programmes, when experienced as such, also do not appear to serve as a push in the direction of the labour market, but to incur incapacitating emotions, such as resignation. Those who are very much focused on re-entry into the labour market are mainly disappointed.

Obviously, one question that remains is to what extent the group that was interviewed (mainly referred to the researcher by supervisory staff at the MWPs) is a special group, and whether these findings can be generalized to the entire group of welfare recipients who partake in such programmes. Given other information gathered by Eleveld from observation and from interviews with supervising staff and civil servants, it is unlikely that welfare recipients who benefited from the MWPs in the sense that they quickly found work were not sampled, as such successful transitions rarely occurred. While Eleveld had suspected that perhaps staff who referred interviewees to her might have selected relatively positive ones, this also turned out to be not the case: many referred interviewees were in fact quite negative about the programmes.

The example shows how qualitative data can be coded and combined to reveal non-linear, intricate patterns in the data, employing iterative methods. Sometimes particular choices are made by the researcher – who is immersed in and eminently familiar with the data – in a manner that would not always be replicable by other researchers. Regardless, it becomes eminently clear from the analysis that MWPs cannot be reduced to being simply 'effective' or 'not effective', and that the experiences of participants cannot be reduced to high or low scores on a one-dimensional well-being scale. Experiences are multi-layered and variegated, and depend on personal characteristics, expectations and programmatic context. Qualitative analysis methods are suited par excellence to discovering such patterns.

### 10.3 Apologies: systematic case law analysis

In daily life, people may suffer harm. Passengers may become hurt in traffic accidents, medical practitioners may make mistakes that disable patients, professional sports trainers may employ harsh training methods, girls raised in youth care institutions often bear the scars of sexual abuse. It happens that victims of traffic accidents, medical malpractice, sexual violence in youth care institutions, or inappropriate/transgressive behaviour seek compensation, financial or otherwise. Some of these cases evolve into disputes and are then taken to court.

Research has shown that victims generally do not just seek monetary compensation. They foremost seek acknowledgement that what happened was wrong, and recognition of the harm they suffered. Also, many victims seek an apology. In fact, it is known that if apologies are offered to a damaged party, that this speeds up the settlement of procedures, reduces the likelihood that the damaged party will enter formal legal proceedings, and contributes overall to the well-being of the recipient as well as the person offering the apology, and to a feeling of restoration after the wrongs that were done (Robbenolt, 2003, 2006; Wijntjens, 2020).

However, in the practice of compensation for wrongs suffered, many still presume that apologies entail a risk. The basis for this widespread belief, which is also prominent in the legal academic literature, is based on a number of exemplary or notorious cases, mostly similar cases that are referred to by subsequent authors, or simply cases that a legal scholar happens to know of (Wijntjens, 2020). Offering apologies is labelled by some as ‘legally dangerous’ (Farmer, 2015), as an apologetic statement may be admissible evidence at trial to establish liability or to prove some other element of an offence. Also, it has been noted that insurance companies may instruct the insured to be reticent in offering apologies, and to speak only summarily and with great care on what happened, with mention made of lawyers even ordering their clients to remain silent (Cohen, 1999). As a way out of this manifestly undesirable situation, so-called ‘apology laws’ were introduced in the USA in 1986, intended to encourage injurers to apologize, as these laws ensure that at least some types of apologies cannot be used against them in litigation. Since then, these laws have also spread throughout the common law world (Australia, Canada, England and Wales, Scotland, Ireland and Hong Kong; see Helmreich, 2012, Vandebussche, 2018).

For the Netherlands, Wijntjens (2020) conducted a large scale study to determine to what extent this fear – of apologies amounting to an admission of liability in legal proceedings – has an empirical basis in legal practice. To investigate this, she employed systematic case law analysis. Systematic case law analysis differs from conventional legal analysis – in which issues are presented in one case or a small group of exceptional or weighty cases – in that it performs a so-called content analysis (see section 7.3) on a large group of similarly weighted cases to find overall patterns. As such, it is set up to prove a point not out of one author’s rhetorical power but because the patterns that are found in case law have been uncovered through systematic and transparent analysis of the content of rulings, in such a way that the data collection, data analysis and findings are reproducible (see also section 2.2).

Systematic case law analysis, which we did rank under empirical legal studies, is therefore in a sense a particular kind of legal-empirical study, namely a study where

the data are always formed by jurisprudence texts. In general, in systematic case law analysis, a representative or purposive set of rulings is sampled, the text in the dossiers or rulings coded, and the codes and relations between them analysed. In that sense, systematic case law analysis is a mixture of qualitative and quantitative analysis: qualitative methods are generally used to code and organize fragments of text, and quantitative methods may be used next to assess – for instance – how often certain evidence is regarded as admissible, or whether there is a relation between the outcome of sex discrimination cases and the gender composition of judges’ panels.

In the following we will describe Wijntjens’ (2020) systematic case law analysis that focused on the role of apologies in legal proceedings. In our example, we will focus on the manner in which she attempted to answer her research question, namely to what extent and in what manner apologies play a role in civil and medico-disciplinary legal proceedings.

### 10.3.1 Data collection

Wijntjens selected court rulings from the database of court rulings that are published online by the Netherlands Council for the Judiciary (*rechtspraak.nl*). This database contains a large number of court rulings, but not all proclaimed court rulings. For instance, all rulings by the Supreme Court are published, all rulings by the Intellectual Property Division of the Civil Sector of the District Court of The Hague, and all rulings in (attempted) homicide cases. For other types of rulings it is so far unclear how choices on publication are made. This means that it is unclear how the sample of rulings published is formed. We may not therefore assume Wijntjens’ is a probability sample (see section 3.2 and section 3.3.1), and that means also that we are unsure whether the findings are generalizable. In fact, the sample cannot be regarded as more than a convenience sample.

As Wijntjens was interested in civil court rulings and rulings by disciplinary law, she first selected all rulings that were present in the database up to October 2018, the time she started data collection. To do so, she first selected all rulings present in the database under the header ‘civil law’: a first set already comprising a hefty 140,056 rulings!

Next, she selected all published medical disciplinary law rulings via an online disciplinary rulings database published by the Netherlands government. All rulings by the Dutch Disciplinary Committees Health Care (*Tuchtcollege voor de Gezondheidszorg*) are published since 2010. This resulted, at the time data were collected, in 8,283 rulings. This means that the sample she drew from this database in fact constitutes – in a statistical sense – a population (see also section 3.1), namely the entire population of all rulings from 2010 up to the moment data collection began.

To narrow down the set of documents to be studied, Wijntjens next conducted two further steps. First, she automatically searched within all digital 140,056 civil plus 8,283 disciplinary rulings for the keywords \*excuse, \*excuses, \*regret, \*apology, \*apologies, \*sorry and \*repent. These keywords had been chosen on the basis of a study of jurisprudence on apologies, and on the basis of an earlier study by Van Dijk

(2017) on ordered apologies.<sup>4</sup> This reduced the set of rulings to 3,452 civil court and 872 disciplinary law rulings, and after removing some double hits to 3,117 and 821 rulings respectively. Wijntjens now had a sample of almost 4,000 rulings in which at least one of the keywords appeared.

Next, Wijntjens scanned all rulings, and judged for every case whether it was indeed relevant for her study. She decided that that was the case when it emerged from the ruling that one or more parties had offered apologies or when it was on the other hand explicitly mentioned that apologies had not been offered. It is clear that working through these almost 4,000 rulings is a massive task. Wijntjens did so using Excel.

After scanning the large set, a smaller set of rulings remained. For the civil court rulings 1,254 remained: in 915 it was written that apologies had been offered, and in 358 it had been noted that no apologies had been given (in 19 cases the text of the ruling noted both that apologies had and had not been given). For the disciplinary court cases, 644 rulings remained, of which apologies had been offered in 563 cases and no apologies had been offered in 96 cases (here in 15 cases it had been noted that apologies both had and had not been offered). This left all in all roughly 2,000 rulings.

As the example clearly shows, the case law analysis entailed trudging through a massive amount of textified material, out of which a smaller and smaller number of cases remain that are relevant for the study at hand. In the case of Wijntjens' study, the initial amount of almost 150,000 rulings was reduced to a little under 1,900 cases, a reduction by 98.72%, or, conversely, went through a funnel with just 1.28% of the original rulings remaining for coding and analysis.

Wijntjens subsequently trimmed her dataset further. She selected, within the set of cases, those cases in which apologies were not simply mentioned somewhere in the text, but actually played a role in the case. Again a smaller subset of cases contained explicit mention of the apologies made or not made that were functional in the case (this was so for 570 cases).

### 10.3.2 Analytic methods: coding

Next, cases had to be read and the text coded. Wijntjens enlisted support for this task. A group of Master's students aided Wijntjens in the reading and coding process. Wijntjens developed a codebook and coding scheme. To further reliability of coding, Wijntjens selected 75 rulings that were coded by herself as well as by the Master's students. In an iterative process of comparing codes, improving their codebook and re-coding new rulings, they arrived at a final codebook and final coding scheme that were clear and workable and led to consistent codes.

To assess reliability, Wijntjens then selected another 20 rulings and had these coded independently by herself and a research assistant. The interrater reliability, computed as percentage agreement between Wijntjens and the research assistant, varied between 90% and 100%, with corresponding values for the reliability measure Cohen's kappa (see section 2.6.1) between 0.62 (insufficient to reasonable agreement) and 1.00 (perfect agreement).

---

<sup>4</sup>This study by Van Dijck focused on so-called ordered apologies, cases where the injured party claims an apology, or the court orders one. Wijntjens' study in fact also focused on ordered apologies, but we do not discuss her findings on that subtopic here.

The coding scheme was a mixture of inductive and deductive coding or, in qualitative methodology-speak, emic and etic coding. Deductive codes had been taken from the academic literature and jurisprudence analysis, and were quite coarse. Inductive codes emerged during the coding process, such as apologies offered or refused, which the judge employed to judge the intentions of parties.

### **10.3.3 Analytic methods: classification and finding patterns**

Wijntjens conducted several analyses on her data. In the first of these, she investigated the argumentative schemes that the judges used to arrive at their rulings. She assessed what role the fact that apologies had or had not been offered played in judges' argumentations. The analysis that she conducted is essentially a doctrinal analysis, assessing whether apologies played a subordinate role, a conjunct role, or a decisive role in assessing the evidence on which the conclusion about the case would be based that the judge reached.

According to her codebook, apologies played a subordinate role if they were mentioned only as a support to other, stronger evidence as indicated by the ruling. Wijntjens gives an example of a client who requested dissolution of a contract with a contractor who had agreed to improve the client's website and webshop. The judge had to decide whether the period for granting the contract had been exceeded by the contractor. The court ruled that this was not the case. In support of this opinion, it was mentioned in the ruling that this was substantiated by the fact that the client had apologized for delays he had caused.

Apologies played a secondary role if they were mentioned in conjunction with other evidence, such as in a case where a group of shareholders had asserted that a company had misled them. The court ruled that it could indeed be doubted whether the optimism broadcast by the company was justified. In the ruling, first of all, reference was made to a statement made by the then chairman of the board in an informative shareholders' meeting in which he apologized for misleading investors. Secondly, it was mentioned that even at the height of the financial crisis much remained unclear about the company's portfolio. The two arguments in conjunction formed the evidence.

Wijntjens found very few rulings where apologies were decisive in the ruling. Out of all 570 coded and rulings analysed, she found seven judgments in which the court considered that the apologies of the person causing the damage were regarded as constituting an acknowledgement of liability. This amounts to 1.2%.

Next, Wijntjens delved deeper in the patterns in the data. She outlined additional roles that apologies could play, roles that had emerged abundantly during the emic coding. For that, she selected – leaving out the seven cases in which it had been ruled that apologies had constituted an acknowledgement of liability – those rulings in which the fact that apologies had (or had not) been offered were explicitly mentioned in the ruling itself, the decision reached. For that, she zoomed in on a subset of 160 cases, in which this occurred. Table 10.6 illustrates how strongly the data were funnelled in the entire research process.

Wijntjens found that, in 111 of these 160 cases, apologies had been offered, and in 49 cases it had been noted in the ruling that apologies had not been offered. Next, she investigated what impact apologies had for the person who had offered apologies



**Table 10.6:** Data reduction in apologies study

Inclusion criteria	Selection method	Resulting N
all published civil court rulings on <i>rechtspraak.nl</i> as well as medical disciplinary law rulings by the Dutch Disciplinary Committees Health Care	automated	148,294
search terms *excuse, *excuses, *regret, *apology, *apologies, *sorry, *repent	automated	3,938
apologies offered, not offered or claimed	scanning	1,898
apologies functional in case	reading	570
apologies mentioned in ruling	coding	160

**Table 10.7:** Apologies versus outcome for apologizing party

Apologies	positive effect (N = 58)	no effect (N = 53)
No apologies	no effect (N = 16)	negative effect (N = 33)

on the outcome of the case. Where apologies had been offered, in 58 rulings it could be shown that these were to the advantage of the person who had apologized: there were instances where a sanction was moderated, where less damages had to be paid, where the judge ruled that the apology indicated good intentions, which supported offering ‘second chances’. In 53 rulings the apology did not have a registered effect: examples are situations where the case was deemed too serious, or where the apology was perceived as insincere. These tallies show how in the investigated rulings apologies turn out to be beneficial in more than half of cases.

Looking at the cases where it was mentioned that no apologies had been offered, the picture was different. In 16 out of these 49 cases, the lack of an apology did not have an effect on the ruling, mostly because the judge had ruled that the case did not call for an apology. In more than two-thirds of cases, however, the absence of an apology was regarded as an aggravating factor, translating into for instance higher sums for pains and damages.

The second analysis is summarized in Table 10.7. While Wijntjens herself did not compute any statistics for these results, the frequencies in the table show how the two variables (apologies and favourableness of outcome) are clearly associated. In fact, the odds ratio for this table equals 2.257 (significant at the 5% level), which can be translated by saying that withholding an apology increases the risk of a negative outcome by 126%.

### 10.3.4 Answering the research questions

Wijntjens’ study debunked – for the Netherlands – a persistent belief. Compiling all available evidence, her study first showed that if apologies had been given, they were mentioned in only a fraction of rulings. Her study also showed that in such cases apologies generally played no more than a supportive role in judges’ reasonings. Apologies almost never played a dominant, decisive evidentiary role. Secondly, her analysis showed how apologies in more than half of cases worked out well for the apologizing party, resulting in more positive or less negative outcomes. Importantly, a failure to apologize incurred a sizeable increased risk of negative consequences. All in all, it appears as if apologizing bears little risk, and indeed may bring positive side-effects, with not apologizing a risky strategy as it may backfire.

One could doubt whether the findings – the sample of civil court rulings was as we said a convenience sample – are generalizable, in other words whether it was a particular kind of ruling that was published in the online database of rulings, and whether in unpublished cases apologies would play a different role. While this cannot be ruled out, a similar pattern as in the civil cases was found for the disciplinary rulings – which do constitute a population in a statistical sense. A second issue is that, as Wijntjens remarks, apologies may have an impact that cannot be read from the rulings: it may be that judges are influenced by the fact that one party has apologized but that this influence is not expressed in writing and therefore does not play a noticeable role in the ruling. Last, it may be that apologies have been offered, but that no note is made by the clerk of this: these cases are then wrongly classified as cases without an apology. Obviously, apologizing and not apologizing are not actions by parties that are randomly distributed over cases: we should therefore be cautious not to interpret the findings overly causally.

Regardless of these methodological issues, it does appear as if, given the beneficial effects that apologies can have in the aftermath of wrongs, there is little risk in expressing them, and it may be more risky not to do so. From the point of view of the injured party, apologies are generally welcomed. And, as Van Dijck (2017) showed, even apologies that are ordered, and are therefore not spontaneous and may even be insincere, may have a legal purpose.

## 10.4 International sentencing: multivariate analysis

Our third example focuses on international sentencing. Many of the institutions that deal with perpetrators of international crimes, such as international criminal tribunals and the International Criminal Court (ICC), are relatively new. Most of them therefore do not have much jurisprudence to fall back on; they do not have a legal tradition, specific legal guidance or even quantified sentencing guidelines. This raises a number of issues that relate back to fundamental principles of justice. One important such issue is consistency: similar cases should be treated similarly. Defendants with similar profiles in terms of crimes committed, role in the criminal enterprise, similar circumstances should receive comparable sentences.

As Hola (2012) discusses, consistency of sentencing should be evident at two levels: the level of the ‘approach’ to sentence determination (i.e. reasoning) and the level of the outcome of sentencing. As ‘approach’ Hola defines the manner in which sentences are determined, decisions being based on common standards or general underlying principles, uniformly applied to the facts of each case. At the second level, consistency in outcome, Hola states that comparable cases should receive comparable sentences.

The first level calls for research that investigates the manner in which standards are referred to and actually used. This would be socio-legal research in a sense, research that scrutinizes legal decisions, in which texts are scanned for active or passive reference to principles, in a sort of systematic case law analysis, and comparison of the application of such principles across cases and jurisdictions or tribunals is investigated. The second level calls for – as we will show below – fairly technical, quantitative ana-

lytic methods that are able to assess to what extent sentences are actually consistently determined, and therefore predictable.

The literature on consistency in sentencing is surprisingly sparse. What is out there has mostly focused on Anglo-Saxon countries, mainly in the context of sentencing guidelines (for example Ostrom et al., 2008), with the study on the Netherlands by Wermink et al., 2015 one exception. A few studies have focused on sentencing in international criminal law, amongst which Meernik & King (2003) on sentence lengths in general at the ICTY, Meernik (2014) on the impact of plea bargaining on sentence length, and King, Meernik, & Kelly (2016) on gendered effects in sexual violence cases.

Most studies have investigated predictability using multiple regression models (see section 9.2) and expressed predictability in  $R^2$ , the percentage of variance in sentence length that can be explained by predictor variables, variables in the model that are supposed to impact sentence length (such as number of counts, severity of crime, aggravating and mitigating circumstances and cooperation with the prosecution). If  $R^2$  is for instance 0.50, it is said that the model is able to explain 50% of the variance of sentence length (see section 9.2.4). Most studies found  $R^2$  to hover around 60%.<sup>5</sup>

Hola studied both consistency in approach (2012, chapter 2) and consistency in outcome (Hola, Bijleveld, & Smeulders, 2009) at the two international tribunals that have – still – generated the largest number of sentences in recent history: the International Criminal Tribunal for Rwanda (ICTR) and the International Criminal Tribunal for the former Yugoslavia (ICTY). It is this second study that we will focus on here. Hola, with her co-authors, studied consistency in outcome. Hola investigated the impact of a number of legally relevant factors, such as mode of individual responsibility, mitigating factors, guilty plea and the like, on sentencing outcome. She wanted to know, for instance, whether superior responsibility led to higher sentence lengths, and whether those accused of graver crimes or those with higher rank would receive longer sentences.

### 10.4.1 Data collection

Hola focused on the ICTY. The ICTY website has case information sheets for all cases prosecuted at the ICTY, which contain information on the indictment, the trial and the judgments. The same goes for many other international tribunals and for rulings at the International Criminal Court (ICC). Data on international sentencing are therefore generally freely available (although they may contain massive amounts of text, see for instance Chlevickaite (2022), who in her assessment of factors contributing to credibility and reliability analysed 93 ICTY, ICTR, and ICC trial judgments from 1996 to 2019, which totalled an estimated 50,000 pages).

Hola studied all 63 cases that had by the time the analysis was conducted been handed down, whether by the Trial Chamber if the case had been dealt with only at first instance, or by the Appeals Chamber if it had been decided on appeal. While these

---

<sup>5</sup>For instance, Ostrom et al. (2008) reported that  $R^2$  equalled 67.2% in Michigan, 55.4% in Virginia for assault and 49.3% for burglary, and a whopping 86.1% in Minnesota – which the authors qualify as extraordinary and attribute to very strict sentencing guidelines. Wermink et al. (2015) reported an  $R^2$  of 57.4% for sentences for adults in the Netherlands.

**Table 10.8:** Sentence length in years, by offence, offender rank, guilty plea and mode of individual responsibility

	trial			appeal		
	average sentence (SD)	median sentence	N	average sentence (SD)	median sentence	N
Average	15.3 (10.0)	14.0	60	14.6 (8.7)	14.0	56
Crimes a/humanity	17.2 (10.2)	17.0	47	16.2 (8.9)	15.0	43
War crimes	8.4 (5.4)	7.0	13	9.1 (4.5)	9.0	13
High ranking	17.2 (10.2)	17.0	47	16.2 (8.9)	15.0	43
Middle ranking	13.5 (9.6)	10.0	25	13.3 (8.5)	10.0	24
Low ranking	13.9 (6.5)	15.0	23	14.1 (6.1)	15.0	20
No guilty plea	15.9 (10.6)	15.0	40	15.1 (8.9)	15.0	36
Guilty plea	14.1 (8.9)	12.5	20	13.6 (8.3)	12.5	20
Superiors	11.8 (10.6)	7.8	18	10.7 (6.3)	9.0	15
Perpetrators	17.5 (8.0)	17.5	20	17.3 (7.7)	18.0	18
JCE	16.3 (9.8)	15.0	22	14.6 (8.4)	13.0	19
Planners	18.5 (9.2)	18.5	2	18.5 (9.2)	18.5	2
Instigators	19.8 (15.4)	25.0	3	20.7 (12.1)	25.0	3
Orderers	21.8 (12.8)	20.0	8	16.8 (7.8)	18.0	9
Aiders	14.1 (8.4)	13.0	17	16.2 (8.5)	15.0	18

Source: Hola et al., 2009

63 cases comprised at data collection a population of cases in a statistical sense, for practical purposes it is a small sample. This limits possibilities for analysis, as we will see below.

#### 10.4.2 Analytic methods: bivariate data analysis

Hola and her co-authors first gave descriptives of these cases. They calculated that the average number of counts for which a perpetrator was convicted equalled 4.6 (the number in first instance had been slightly higher at 5.1). Perpetrators were most often convicted for persecution (35 in total). Next, Hola described sentence length in relation to a number of legally relevant factors. Table 10.8 lists average sentence lengths (in years), as well as sentence lengths broken down by a number of categories of offenders, and by whether or not a defendant had pleaded guilty. The table gives not only means, but also median values, as the distribution of sentence length was skewed (see section 6.3.1). The table shows that sentences on appeal were on average somewhat shorter. Also, we see that crimes against humanity were sentenced more harshly, and that those who did not plead guilty received longer sentences as well. These findings are what one would expect.

There were some unexpected findings too, however: while high-ranking offenders in general received the longest sentences, the sentences for middle-ranking offenders were surprisingly low, and those of the low-ranking offenders higher again – and in fact (if looking at the median on appeal) just as high as those of the high-ranking offenders. This finding is surprising as one would expect that those with greatest responsibility for the crimes would receive the longest sentences, and those with the least responsibility, lowest in the hierarchy of culpability, correspondingly the lowest sentences.

Hola further investigated this issue by looking at the more specific modes of individual responsibility underlying an offender's conviction. She categorized these modes into Superiors, Perpetrators, Joint Criminal Enterprise (JCE), Planners, Instigators, Orderers and Aiders. Sentence lengths for these groups are also given in Table 10.8.

This categorization revealed a similar surprising finding, namely that those with superior responsibility received the lowest sentences on average, lower even than the aiders. Hola and her co-authors explain this by suggesting that superior responsibility entails the passive behaviour of a superior, such as a failure to punish or prevent subordinates from committing crimes. The omissions of superiors are perhaps perceived to be less serious than active participation in criminal activities, less serious than the activities of the executioners, those with 'blood on their hands'. However, on a bivariate level, Hola identified several other counterintuitive sentencing patterns at the ICTY, for instance that aiders and abettors (that is, facilitators of crime with arguably a very limited role in the crimes) received on appeal sentences almost as high as orderers (who are arguably more culpable due to their decisive influence on hands-on perpetrators). Thus, sentence length is likely not only a function of a perpetrator's role, but also of his or her role in conjunction with other aspects of the crime, such as cruelty or the number of counts.

This led to the conclusion that it does not make sense to investigate the effect of individual responsibility bivariately. As individual responsibility may be *confounded* (see section 5.1) with other characteristics of the cases and perpetrators, such as hands-on cruelty, one may end up with hard-to-interpret findings like the ones just discussed. Individual responsibility may in fact be confounded with various additional factors; for instance, it may also be that aiders often have in general been convicted of less serious crimes. If aiders receive on average a shorter sentence, is that because they were aiders – or because they were responsible for less serious crimes? Or do aiders receive shorter sentences because they often entered a guilty plea?

### 10.4.3 Analytic methods: multivariate data analysis

To disentangle all such confounding, the authors chose to do a multivariate regression analysis. In a multivariate regression analysis, one investigates the effect of each independent variable on an outcome variable, and the regression weights express the impact of the respective variables *given the other variables in the model* (see section 9.2.6). Thus, one would be able in a regression analysis to investigate the effect of being an aider *over and above* the effect of variables present in the model. The model would give the net effect of being an aider, net of whether someone had shown excessive cruelty, had confessed to his crimes, had committed war crimes, etc. The regression coefficient reflects what a variable (being an aider) adds to the prediction of the dependent variable

**Table 10.9:** Regression analysis of ICTY sentence lengths in years

	b	$\beta$	p-value
High rank	10.9	0.5	<.001
Middle rank	2.2	0.1	<.231
# Guilty counts	0.7	0.4	<.001
# Mitigating factors	-0.6	-0.2	<.025
Conviction for crimes against humanity	6.1	0.3	<.001
Instigation	7.7	0.2	<.032
Perpetration	3.4	0.2	<.079
Superior responsibility	-3.5	-0.2	<.058

---

$R^2 = .653$   
 $R^2_{\text{adj}} = .602$   
 $N = 63$

Source: Hola et al., 2009

sentence length given all other factors (type of crime, mitigating factors, guilty plea, etc.) in the model. This is the reason why regression analysis is a suitable technique for the question at hand, namely what the impact is of one variable on the dependent variable, controlling for any distorting effects of other factors.

Table 10.9 gives the results of the regression analysis, which was conducted using SPSS. We give only the unstandardized regression weights (see section 9.2.2 and section 9.2.6), which enable us to translate the impact of the successive independent variables on sentence length in years.

In a first step to build a final model, 14 predictors had been incorporated.<sup>6</sup> Even though the authors had wanted to include more predictors, the small sample size did not allow this (and 14 predictors was actually stretching it given the 63 cases). Of these 14 variables, a number were not significant (such as, perhaps remarkably, whether the defendant had pleaded guilty or not), and these were removed and the analysis re-run with a smaller set of variables, which is the final set that we will discuss here. See also 'Model building' in section 9.1.2.

However, before we discuss the results, we note that the authors had had to solve for the fact that in two instances (Stakič and Galič), life sentences had been handed down. With 'life' as the sentence length, there is no sentence length to predict ('life' being indeterminate), so that the value for the dependent variable is technically speaking

<sup>6</sup>One variable was added as a 'dummy variable': the variable 'perpetrator rank' had three categories and was not interval level (at most it may be considered ordinal level; see section 6.2.1). This variable was therefore included as two dichotomous variables, one indicating whether a perpetrator had high rank (1) or not (0), and a second one indicating whether a perpetrator had middle rank (1) or not (0). Such 'dummy coding' enables to include non-interval variables in the model. Significance is then assessed over the various 'splits' of the variable (here 'high rank' versus not, and 'middle rank' versus not).

missing. For this analysis, they therefore first re-coded life sentences to the expected duration of imprisonment given the convicted's age; for Stakič this led to an imputed value of 30 years, for Galič a value of 20 years. One could argue that these values are in a sense arbitrary and do not reflect that life imprisonment is intended to express a qualitatively differently sentence – much heavier – than a determinate one. In that sense, the imputed value of 20 years for Galič may be seen as too low given that for instance the average sentence of orderers at trial was already 21.8 years. The authors ran several other models, in which different imputed values for these life sentences were used. The interpretation however remained substantively the same. Given these results, the authors conclude that the findings are *robust* against different imputations of values for the life sentences. Obviously, as there were just two life sentences for which values had to be imputed against 61 determinate sentences, the results were not likely to be affected much.

The variables in Table 10.9 in the final model were all significant at the  $p < 0.10$  level, and therefore retained. This may appear a puzzling statement as the attentive reader will see that this is not the case for the  $p$ -value for the dummy variable 'Middle rank'. It was noted above that 'High rank' and 'Middle rank' were two dummy variables added to assess the effect of the nominal variable rank of the perpetrator. We see that the dummy variable that measured whether a perpetrator was of high versus low rank *did* add significantly to the prediction: a perpetrator of high rank was predicted to be handed down a sentence that was 10.9 years longer than a perpetrator of low rank. Perpetrators of middle rank, however, did not receive significantly longer sentences than those of low rank (the difference is indeed much smaller at 2.2 years).

After rank, instigation contributes importantly to sentence length: instigators received on average a 7.7-year longer predicted sentence than perpetrators who had not been classified such. Next, those convicted for crimes against humanity are predicted to receive a 6.1-year longer sentence, with the reference category here being a conviction for war crimes (at the time of data collection no conviction for genocide had yet occurred). In the model, perpetrators are handed down 3.4 years more than non-perpetrators, the former likely being the perpetrators with 'blood on their hands' as Hola labelled them. And, mirroring the preceding finding, superior responsibility predicts a 3.5-year shorter sentence. This counterintuitive pattern had already been observed in the bivariate tabulations, but the regression analysis shows that this is not due to any confounding with the other independent variables in the model as multiple regression gives the impact of each variable net of the other variables in the model.

Lastly, we see how mitigating circumstances and guilty counts almost balance each other out: each additional mitigating circumstance subtracts 0.6 years (or approximately 7 months) from the predicted sentence, and each additional guilty count adds 0.7 years (or approximately 8.5 months) to the predicted sentence. Said differently, if we were to have two perpetrators with exactly the same scores on all the variables, and if perpetrator A had 3 guilty counts and perpetrator B had 5 guilty counts, perpetrator B would be predicted to be handed down a 1.4-year longer sentence.

We are each time speaking about predicted sentences, because not all sentences will have been predicted perfectly by the model. The fit of the final model is .653, meaning that 65% of the variation in sentence length can be predicted from the seven variables included. The fit of the model is thus quite reasonable. We also give the  $R^2_{\text{adj}}$ , the fit



measure that penalizes the inclusion of (too) many variables in the model given its  $N$  (see section 9.2.4). The sizeable drop from an  $R^2$  of .653 to an  $R^2_{\text{adj}}$  of .602 shows that the model may indeed have been slightly overloaded.

#### **10.4.4 Answering the research questions**

All in all, the multiple regression analysis leads to a number of conclusions. Firstly, the analysis shows that the fit values of this model are adequate; they are in fact comparable to the fit values of ‘national’ analyses on sentencing consistency. International sentencing at the ICTY and ICTR therefore does not appear less consistent than regular national sentencing, which is a striking finding given its infancy, limited jurisprudence and the lack of sentencing guidelines. Second, the analysis showed that higher-ranking offenders did indeed receive longer sentences: in the model, being a high-ranking offender adds 10.9 years to sentence length.

The analysis also showed that the puzzling finding that those with superior responsibility received lower sentences is a true effect and not due to any confounding with the variables in the model: perpetrators with superior responsibility on average did receive a 3.5-year lower sentence, net of all other characteristics of their cases. After rank, it is those who were actively perpetrating the international crimes and the instigators who received the harshest punishments. Last, as expected, those convicted for graver crimes (crimes against humanity versus war crimes) could expect longer sentences.