Zeitschrift für die gesamte Versicherungswissenschaft, 114 (2025) 2: 227 – 269
https://doi.org/10.3790/zverswiss.2025.1459202

Bodily Injury Claims in German Non-Life Insurance – a Detailed Case Study

Marcel Wiedemann and *Andreas Herzog**

Abstract

In our paper we present a detailed case study of bodily injury claims in German non-life motor insurance. Based on a sample of approximately 2,800 claims of a large German non-life portfolio (with an extensive scope of attributes), we shall analyse the influence of the attributes derived in Wiedemann and John 2021. Our results can form the basis of the development of individual claims models.

Zusammenfassung

In unserem Artikel diskutieren wir die Ergebnisse einer detaillierten Fallstudie zu Personenschäden in der deutschen Kraftfahrtversicherung. Auf der Basis einer Stichprobe von ca. 2.800 Schäden eines großen deutschen Kraftfahrtportfolios (mit einer umfangreichen Zahl an Schadenattributen) untersuchen wir den Einfluss der in Wiedemann und John 2021 abgeleiteten Schadenattribute. Unsere Ergebnisse können als Basis zur Entwicklung von Einzelschadenmodellen dienen.

JEL classification: G22

Keywords: Bodily injury claims, non-life insurance, case study, individual claims models

Prof. Marcel Wiedemann
Hochschule Esslingen, University of Applied Sciences, Esslingen
Email: Marcel.Wiedemann@hs-esslingen.de

Andreas Herzog
HUK-COBURG Versicherungsgruppe, Coburg
Email: Andreas.Herzog@huk-coburg.de

* This research was supported by the HUK-COBURG insurance group (Germany). Both authors have links with the HUK-COBURG insurance group (consultancy, employment).

Duncker & Humblot · Berlin

1. Introduction

In recent years, individual claims models have come into fashion amongst non-life actuaries (see for instance Wüthrich 2018). They can provide a very powerful tool to many practical problems they are currently facing (see for instance Wiedemann and John 2018 & 2021) and avoid many of the shortcomings of traditional aggregated methods.

The main focus has so far been on specific mathematical models for individual claims modelling. We believe, however, that more attention needs to be given to the required data for modelling specific lines of business. To our knowledge, there has yet not been any research or case study focusing on this aspect. We strongly believe, that the main focus must be on obtaining statistically relevant data first, rather than focusing on more complex mathematical models (potentially based on insufficient data).

Over the last years, we have deducted a large project focusing on the development of individual claims models for bodily injury claims for a large German non-life motor insurance portfolio. Since the provisions for the corresponding claims usually form the largest share of all non-life provisions, they are of particular interest to non-life actuaries. Our project consisted of the following steps:

1. Data collection of the attributes described in Wiedemann and John 2021 for approx. 2,800 claims
2. A detailed analysis of the collected data to derive a set of relevant attributes for specific claims components
3. Development of individual claims models (modelling total payments and case reserves) for all claims components based on the derived relevant attributes
4. Adaption of claims systems so that relevant attributes can be collected systematically and IT-based by claims handlers. As a result, the information is directly accessible by all relevant departments (actuarial department, claims department, risk management, controlling, etc.)
5. Implementation of individual claims models into the claims system to propose case reserves for claims handlers automatically
6. Implementation of validation cycle for constant improvement of models

The aim of our paper is a detailed discussion of step 1 and 2 (case study). It is based on Wiedemann and John 2021 where bodily injury claims in German motor liability insurance were discussed from an actuarial point of view and (on that basis) attributes which seem relevant for claims modelling were derived. It remains to be shown, however, whether these attributes are actually statistically relevant. The discussion in Wiedemann and John 2021 is mainly based on Küppersbusch and Höher 2016, an excellent, detailed and extensive source on the matter of bodily injury claims in German non-life insurance.

Our paper is organised as follows. In Section 2, we discuss motivational aspects of our project and derive our understanding and definition of individual claims modelling. In Section 3, we discuss practical aspects of collecting a large data sample (step 1 from the list above). In Section 4, we discuss the attributes of our sample and present the results of our case study for the most relevant claims components (step 2 from the list above). We are planning to address step 3 in a future publication. In Section 5, we present some insights showing that the data discussed in Section 4 is truly helpful for a better understanding and modelling of bodily injury claims.

The results presented in this paper might be relevant for all German non-life actuaries dealing with bodily injury claims (for instance in motor insurance or general liability). The compensation of bodily injury claims derives from regulations of the German Civil Code (*Bürgerliches Gesetzbuch, BGB*) and is not part of contract terms of insurance companies (although the maximal limit of compensation is usually restricted by contract terms[1]). However, company specific aspects may affect the influence of the attributes discussed in Section 4 on claims costs (for instance, approaches to claims handling might be active or passive or legal aspects might be interpreted differently).

As a result of our project, we came to the firm conclusion that the developed individual claims models provide a significant improvement compared to standard methods on aggregated (triangle) data and open up further fields of application (such as claims steering, reinsurance optimisation, etc.). For this reason, we have extended individual claims modelling step by step and are currently covering almost all non-life lines of business (comprehensive motor insurance, home contents insurance, home insurance, general liability insurance, accident insurance, etc.).

The authors express their gratitude to the claims department of the HUK-COBURG insurance group. The results presented in this paper would not have been possible without their support and encouragement.

[1] In German motor liability insurance, the minimal limit of compensation for insurance contracts (as determined by law, *Pflichtversicherungsgesetz*) is currently at 7.5 million €.

2. Individual claims modelling – motivation & definition

The motivation of our project was not purely academic but stems from practical problems and questions with which actuarial departments might regularly be confronted. Typical problems might be the following:

- For reinsurance optimisation, one needs to understand the characteristics (esp. increases in costs) of large/major claims (which are usually bodily injuries claims).
- For optimising asset management, one needs a better understanding of future cashflows.
- Claims departments ask for benchmarks for case reserves to be used by claims handlers.
- Claims departments seeking actuarial advice in developing and assessing claims steering strategies.[2]
- Modelling automatic upfront payments after reporting of claims.
- The impact of changes in claims handling processes on claims payments and reserves need to be assessed.
- Legal changes and their effects on claims reserves need to be assessed[3].
- Regular forecasts show significant changes in claims payments and reserves compared to previous years with no obvious explanation.
- …

It is very hard – if not impossible – to answer the above questions based on just aggregated data since only very specific parts of the claims portfolio might be affected. From our experience, most analysis on aggregated data (even if segmented further) will be very complex and time-consuming but only lead to unsatisfactory and sometimes even self-contradicting results. A rapidly changing market environment, however, demands swift and high-quality answers to practical problems. It became clear to us that only by expanding our modelling approach to individual claims we could answer questions as the ones raised above comprehensively and, hence, generate significant company-wide benefits.

We define (the process of) individual claims modelling as follows:

1. It applies to reported claims which are not yet fully settled (although individual components might be settled, see later).

[2] If, for example, "severe" bodily injury claims can be detected at an early stage, one might consider specialist medical treatment to aid the healing process.

[3] For instance, the effect of the introduction of damages for pain and suffering for surviving dependants in Germany in 2017 needed to be assessed.

2. Consider a claim with reporting time t. Reaching any time t + n (with n ≥ 0 in years) the following is modelled for the considered claim:
   I. Expected ultimate claims costs U_{t+n} (per individual claims component) at current modelling time t + n.
   II. Expected (annual) future cash flows $C_{t+n} = (P_{t+n,1}, P_{t+n,2}, P_{t+n,3}, \ldots)$ (per individual claims component) at current modelling time t + n with $P_{t+n,i}$ denoting the expected individual payment in year i (after current modelling time t + n).

The chosen yearly time steps of part 2 for U_{t+n} as well as C_{t+n} seems natural but might, of course, be adjusted (depending on the characteristics of the underlying line of business, or if applications demand finer or coarser modelling).

Our approach also focusses on modelling future cashflows. This is a very essential point: It gives an expectation (reference point) for the future development of particular claims. To address some of the aforementioned problems, one is basically looking for claims deviating substantially from their expected behavior (which can be achieved by comparing expected and actual payments – one of the most relevant aspects in practice). Moreover, cashflows are needed for aspects like Solvency II, Asset Liability Management, etc. (for instance for discounting).

The idea behind this modelling approach is the following: Estimates for individual claims will need to be updated regularly (annually, as defined above) as more and more information is acquired. One must, however, try to ensure that updates are not overly volatile. We believe that this can be achieved in the following manner. At the time of reporting of a claim, normally no payment data is available. Hence, U_t and C_t (n = 0) must entirely be based on other claims attributes. In case of bodily injury claims, the attributes discussed in Section 4 of this paper might be used (for instance injury, age of claimant, etc.). The estimates U_t and C_t (n = 0) – if based on statistically relevant attributes – will give a robust initial assessment for the corresponding claim (without using any payment data). They are robust in the way that the attributes – once known – do normally not change during the course of claims handling (as for instance injury, age, occupation, etc. are fixed at the accident event). It is important to note, however, that at the time of reporting some attributes may not be known at all and, for others, there might only be an indication. In case of bodily injury claims for instance, at the time of reporting, injury, occupation, wage, etc. of the claimant might not be known, so claims handlers will have to make assumptions to assess the claim. Through communication with claimants, relatives, health insurance providers, etc., claims attributes can eventually be updated appropriately.

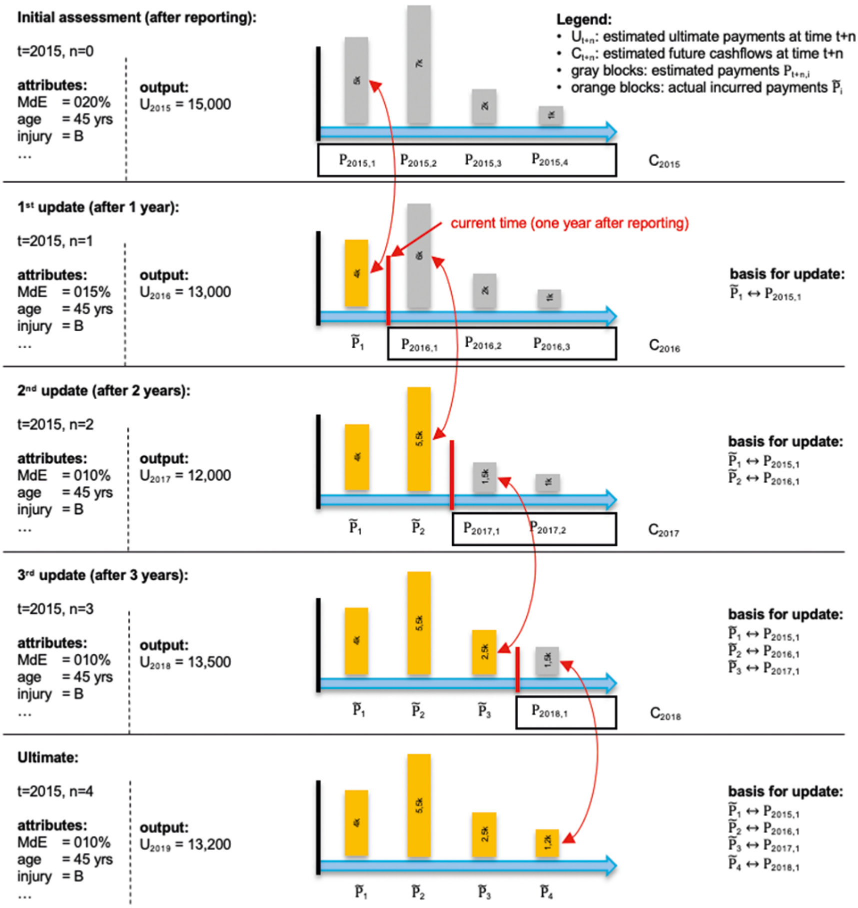

Fig. 1: A schematic example for individual claims modelling

During the course of claims handling, additional payment data will be acquired and for U_{t+n} as well as C_{t+n} (with $n > 0$) actual incurred payments can be taken into account. This follows the idea that the characteristics of claims which are ultimately "more expensive" or "less expensive" than their initial assessment will be reflected in actual payments being "higher" or "lower" than expected payments with increasing time. Figure 1 demonstrates this idea (for a description of the attributes see Section 4):

Meaningful updates for U_{t+n} as well as C_{t+n} (with $n > 0$) could for example be based on assessing the differences

$$\tilde{P}_{n-i} - P_{t+n-i-1,1} \,(i = 0, \ldots, n-1)$$

between the expected payments and the actually incurred payments $\tilde{P}_{n-i}$ of the considered claim for previous years (to make modelling more robust, one may consider differences of several years).

The process just defined reflects our vision of individual claims modelling driven by practical application. One might, of course, think that achieving the described level of detail is completely hopeless. We believe, however, that there is enough inherent structure in claims portfolios to actually make it work. The key for this is sufficient and "good" data to learn from.

The main content of this paper is a detailed discussion of our data sample as the main input for approaching U_t and C_t ($n = 0$) for bodily injury claims. We have so far achieved modelling initial assessments U_t ($n = 0$) (for almost all non-life lines of business) and to some extend part U_{t+n} ($n > 0$) (updates of ultimate claims costs after 1 year for certain claims components of certain lines of business). These models are also implemented in our claims systems as a tool for claims handlers. The data sample of our empirical study does, however, also contain detailed cashflow-data for all claims components. Preliminary analysis shows that the attributes discussed in this paper are also relevant for modelling cashflows. Moreover, our ideas for an updating process for U_{t+n} as well as C_{t+n} (with $n > 0$) (as described above) are also motivated by insights from our data sample. Both is currently still work in progress, yielding interesting areas for actuarial research.

3. Data sample – background & practical aspects

We were in the comfortable situation that some attributes for bodily injury claims were already available in our claims system (or supporting systems), so they could be obtained automatically. These attributes contained the following claimant information:

- reduction in earning capacity (MdE[4]) – taken from medical assessments (if available) or assessed by claims handlers
- description of injury (in some cases ICD[5]-codes)
- age, gender, wage, occupation, family status (if relevant), each at the date of accident
- health insurance coverage scheme at the date of accident

[4] MdE – Minderung der Erwerbsfähigkeit.

[5] ICD – International Statistical Classification of Diseases and Related Health Problems.

The description of injury has been used for a very coarse grouping of claims into different injury classes (see later). We believe, however, that our current grouping is too coarse. Although it provides valuable information for individual claims modelling, there is still room for further improvement. Text analysis or word mining might yield further insights which could improve models significantly. In addition, OCR-methods might be used to extract further information from medical documents. We are yet to explore this further.

We did not use the information from ICD-codes for modelling since they were only available in certain cases. Moreover, from an actuarial point of view, the codes from the so-called Weller-Database[6] seem more appropriate for modelling.

The remaining problem was to obtain the relevant payment data for all claims. Claims systems will usually offer claims handlers the possibility – as it was the case with our system – to segment payment data depending on payees. For instance, a payment settling a health insurance bill will usually be entered as "payment to health insurance provider". However, such a payment may comprise different claims components, for instance costs for medical treatment (constituting medical treatment) and costs for sick pay (constituting loss of earnings). To get a thorough understanding of how to untangle payment data (and where to find it – in the example just described, this can only be done by an analysis of the respective bill) remained the main task. In our case, almost all payment data needed to be extracted directly from paper files. Files for bodily injury claims may, however, be quite extensive (in some cases the information was spread over up to 10 binders, with almost 300 single payments in one case). Hence, finding the correct information in itself was quite a difficult task. Detailed practical advice and guidance on how to collect relevant information for payments and claimants can by found in Wiedemann and Herzog 2025a.

From our experience, it is very important to have a sound understanding of the data one wants to collect and where to find it. As many people were involved in collecting the data (due to the large number of cases in our study), we had to make sure that all of them were working in a similar fashion (for instance, invoice documents should be interpreted similarly). Standardised spreadsheets and instructions were very essential. Moreover, it was also very important to get the claims department involved and to work in close connection (since they are the experts on this matter). In our case, their expertise and support were essential for the success of our project.

Since personal injury claims involve numerous personal details, it is important to comply with data protection regulations (in particular the protection of

[6] See for instance: https://www.fsa.de/produkte/weller-tabelle/ [last download 26th March 2024].

personal data) when taking a sample. In particular, when dealing with health-related data, very stringent data protection requirements will need to be satisfied. This was also an important aspect for us. Using adequate anonymisation and clustering (e. g., age groups, wage groups, injury groups, etc.), we did ensure the protection of personal data.

Our sample consists of three subsets:

- Subset I: 899 claims (leg injuries)
- Subset II: 1,876 claims (all injuries)
- Subset III: 189 claims (nursing care claims)

The motivation of the subsets was as follows. Gathering the relevant data from large paper files is not straightforward and requires a very standardised arrangement of these files (so the same information can always be found in the same place), otherwise collecting data would be far too time consuming. At the beginning, it was not entirely clear to us whether the desired data could actually be retrieved in a consistent manner. So, subset I was basically a trial sample, followed by a detailed analysis. The aim was to investigate whether the collected data leads to a meaningful understanding of claims components and whether modelling is possible at all. For this reason, we tried to focus on injuries that are relatively frequent and rather homogeneous, which led to considering leg injuries. Once this was successfully done, we expanded our sample, generating subset II. After analysing subset I and II, we quickly realised that we were not able to properly understand and model nursing care costs. This was mainly because these claims are rare, not well represented in our sample and, more importantly, further attributes needed to be collected. Nursing care claims are, however, important to understand due to the fact that they can be very expensive and, hence, their provisions form a significant share of all provisions for bodily injury claims. For this reason, a third sample was taken, focusing entirely on nursing care costs. The analysis of this sample (subset III) is not part of this paper.

4. Data sample – details

4.1 Overview

In this subsection, we give an overview of the data collected in subset I & II (in total approx. 2,800 claims). The collected information is based on the discussion in Wiedemann and John 2021, where attributes which seem relevant for modelling from an actuarial point of view were derived for each claims component. Our data sample comprises the following information:

- Claimant information:
  - age (in years)[7]
  - gender
  - (gross) monthly wage[8,9,10] (in Euro €)
  - occupation (grouped: worker, employee, civil servant, self-employed, pupil/teenager, student, househusband/housewife, retired, unemployed, unable to work, n/a)
  - family status: marital status, number and age of children
  - health insurance coverage scheme (grouped: statutory health insurance (GKV[11]), employers liability insurance (BG[12]), accident insurance[13], private health insurance (PKV[14]), insurance scheme for civil servants (Beihilfe), other coverage schemes[15], no coverage & n/a)[16]
  - type of injury (grouped: hip (A), leg (B), neck (C), arm/shoulder (D), head (E), paraplegia (F), traumatic brain injury (G), mental illness (H), other & n/a (I))
  - reduction in earning capacity (MdE) – taken from medical assessments (if available) or assessed by claims handlers
- Payment data (individual payments in € with date and amount) for the following claims components:
  - pain and suffering
  - medical treatment
  - loss of earnings
  - nursing care costs
  - additional needs (excl. nursing care costs)
  - maintenance

[7] Appropriately binned for further analysis.

[8] Appropriately binned for further analysis.

[9] Taking into account total pension insurance contributions (not only claimants share), if applicable.

[10] The wage information of our sample serves as a basis for the assessment of damages for loss of earnings (together with a career projection). It might, for instance, be the wage at the time of the accident (if applicable) or an estimate (for instance, if claimants are pupils, students, unemployed, etc. with no known wage).

[11] GKV – Gesetzliche Krankenversicherung.

[12] BG – Berufsgenossenschaften.

[13] In the following, we shall not distinguish between employers liability insurance and accident insurance and abbreviate both by BG.

[14] PKV – Private Krankenversicherung.

[15] For instance foreign coverage schemes.

[16] “Other coverage schemes” and “no coverage & n/a” will later be grouped as “other”.

- burial
- other payments (physical/material damage, legal costs, …)

Moreover, we also collected information about the accident (date, description, …) itself. These aspects will be omitted in this paper as they did not lead to significant further insights.

Note that reduction in earning capacity (MdE) can either be taken from medical assessments (if available or demanded by claims handlers) or assessed by (experienced) claims handlers themselves on the basis of the injuries caused by the accident. As for any assessment, there will be a (reasonable) range of acceptable values.

It is important to mention that, in general, not all of the claimant information described above will be available immediately after the reporting of a claim. For some attributes, there might be a (considerable) delay (for instance, in the case that claimants are in hospital over long periods of time and cannot be contacted) to get all information needed for assessment. In such cases, claims handlers might start with assumptions (based on known information, experience, etc.) followed by updates later on. Because of this, but also due to changes in the circumstances of claims (for instance, healing processes worse than expected), attributes might change over time (for instance, injury and MdE). A sound understanding of the process of claims handling and especially at which time attributes are available (assumptions, updates) and how they might change over time is an essential point for modelling.

For our sample, we only considered claims where all components are either fully settled or where only fixed annuity payments remain (which mainly applies to damages for loss of earnings where fixed annuities might still be paid with all other remaining components being fully settled).

The time span of reporting years of our sample reaches from 2000 until 2014, so obviously inflation will have to be taken into account when interpreting the results of our sample. Moreover, for further modelling, all data should obviously be adjusted for inflation. However, to our knowledge, so far there has not been a detailed analysis of inflationary effects affecting a motor insurance portfolio, let alone bodily injury claims. In Wiedemann and John 2021, relevant indices which determine inflation of each component were discussed and we shall also try to analyse whether these indices are actually relevant. This, however, is a very difficult task and our findings can only serve as a starting point. On the basis of our data, a more detailed understanding of inflationary effects on bodily injury claims is unfortunately not possible (since this needs a larger sample and more detailed injury data).

In the following subsections, we shall analyse total payments (sum of individual payments per claim) for the following claims components:

- pain and suffering
- medical treatment
- loss of earnings
- additional needs (excl. nursing care)

In each subsection, we will only consider claims with non-zero total payments for the respective claims component. For a detailed analysis, we usually need to restrict our dataset in order to avoid sparse cells (leading to unmeaningful results). Hence, our findings will only apply to the attribute ranges considered. We shall mainly focus on the results derived from our data sample and not go into any details of the presented attributes, their characteristics or their connections. For further details and background information, we refer the reader to Küppersbusch and Höher 2016 or Wiedemann and John 2021 (for a summary from an actuarial point of view).

As mentioned before, nursing care costs are excluded since it required an extra data sample (the insights gained from this sample can be found in Wiedemann and Herzog 2025b). Maintenance is excluded too, since cases are rare and claims handling follows standardised methods and formulae in most cases (depending on family status etc.). Burial cost and other payments are also excluded from our analysis due to their insignificant magnitude (compared to overall payments).

All results are based on subset I & II of our data sample. For reasons of disclosure, the actual magnitude of total payments will be omitted in the following. Instead, the vertical axes in each block of charts are normalised similarly. This, however, does not provide any restriction to detecting potentially relevant attributes and their influence.

As a result of our project, we decided to adapt our claims systems, so relevant attributes (as discussed above) can be collected systematically and IT-based by claims handlers. Hence, our database increases automatically and continuously over time (without any additional effort). However, due to the fact that the settlement period for bodily injury claims spans over many years, we are yet to increase our database significantly.

In this context it is also important to focus on data quality aspects. It is essential that claims attributes are entered as early and as correctly as possible as well as updated as promptly as possible by claims handlers. Especially final updates of attributes at the closing of claims are very essential from an actuarial point of view (but might initially be less important for claims handlers, since their work is completed).

4.2 Pain and suffering

The charts of Figure 2 give an overview of total payments for pain and suffering for the attributes MdE, age, and type of injury (whiskers of box-plots show 5 % and 95 % quantile).

The strong dependence of total payments on MdE is already discernible (independent of age and injury, see also the charts of Figure 3). The dependence on age, however, seems more delicate and requires further analysis. In order to avoid unmeaningful results caused by sparse data cells, we shall focus on the following data for further analysis (the charts of Figure 3 show mean total payments and number of claims for the respective attributes):

- MdE: 0 %, …, 50 %
- Injury: B, C, D

The charts of Figure 3 underline the strong dependence on MdE, independent of age and injury. A slight effect for ages above 60 (esp. for higher MdE groups) might be detected, as expected. One must, however, bear in mind that the data for the MdE groups above 30 % in our sample is rather small (esp. for the age group 21 – 40). Injury does seem to play a roll, as suggested by the last chart of Figure 3. As a result, individual claims models should at least be based on MdE and injury.

In Wiedemann and John 2021, it was mentioned that the consumer price index (CPI) for Germany seems a relevant index for inflation of damages for pain and suffering. During the relevant period of cases in our sample, the mean inflation was around 1.5 % per year, so rather moderate and more or less constant. An analysis of the inflation in our data is, of course, rather difficult since differing attributes of claims need to be taken into account, which will require detailed modelling. However, as seen above, a rough idea might be derived from considering the development of total payments for individual MdE groups over time (for different injuries), since MdE is the most relevant attribute. For this, one needs to keep in mind that other aspects (like injury and severity) do play an important role and, hence, there might still be significant volatility in each MdE group. Moreover, since the settlement period of a claim might span over several years with multiple payments, inflation may also affect the settlement period of claims. However, there does not seem to be a consistent legally motivated view on how the settlement period is affect by inflation (or if it is affected at all). From an actuarial point of view, we shall consider inflation to only be a reporting year effect not affecting the settlement period for damages for pain and suffering.

The charts of Figure 4 show the index of total payments for pain and suffering (with base year 2000) for the MdE groups 10 %, 20 %, 30 % and the average index of those three indices (per reporting year). For this, we further restrict to the age groups 20 – 60. We also exclude the reporting years 2012 and later due to the low number of claims in those years.

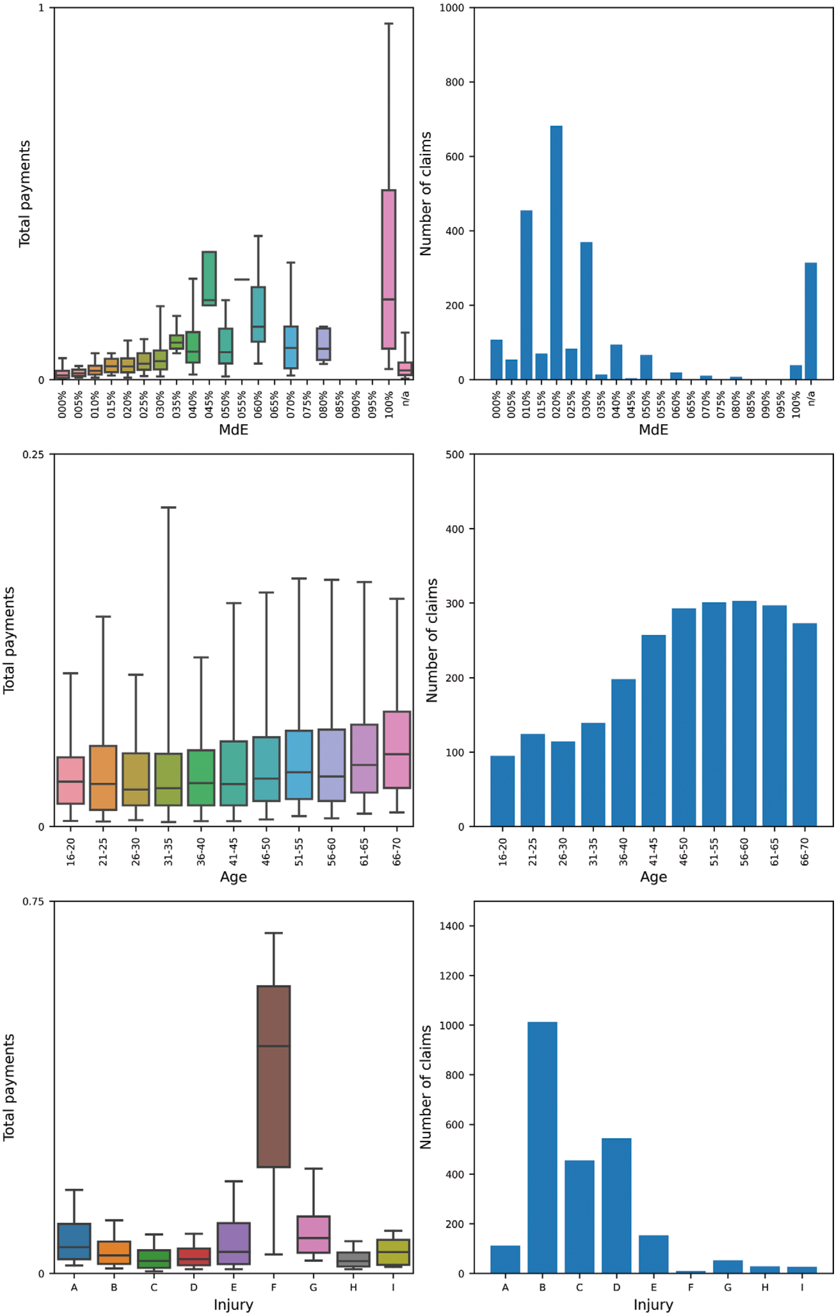

Fig. 2: Pain and suffering – overview

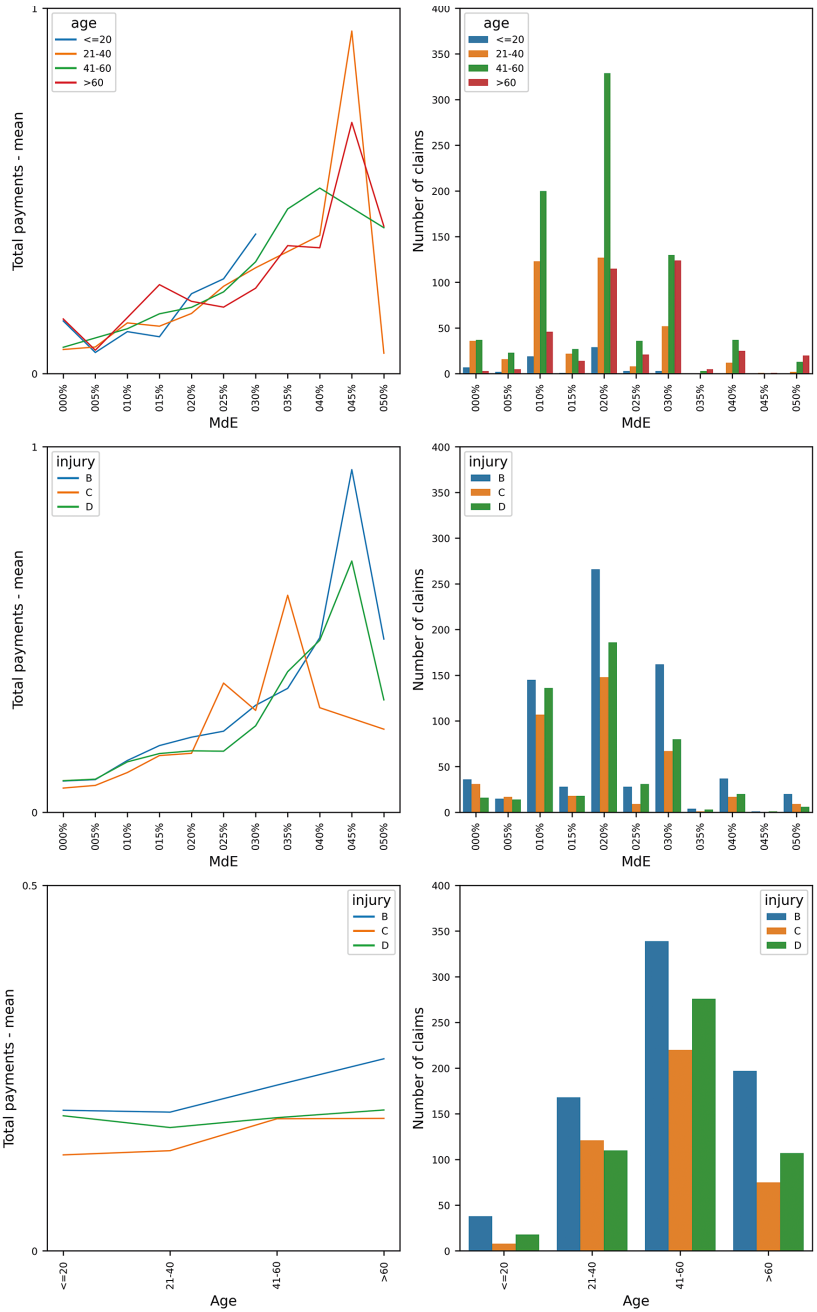

Fig. 3: Pain and suffering – details

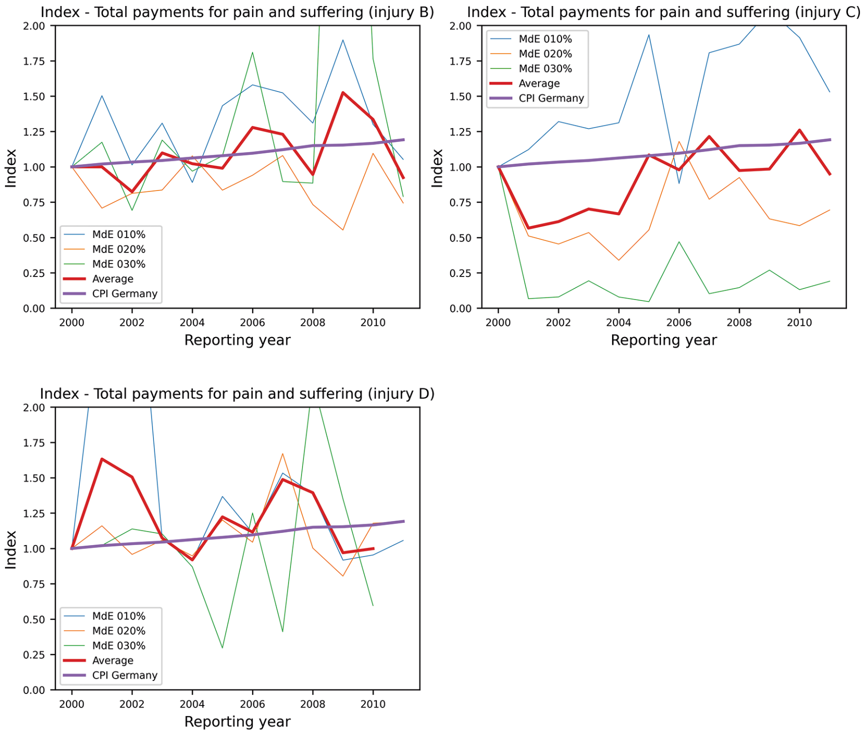

Fig. 4: Pain and suffering – index of total payments

The volatility is, as expected, very high and this analysis can only serve as a motivation (note that injury B forms the largest injury group). We may, however, derive that the average inflation in our sample does not seem inconsistent with the change of CPI for Germany. Hence, for further modelling, total payments might indeed be adjusted on the basis of CPI.

This, of course, also means that the underlying data of our analysis needs to be adjusted for inflation in a similar manner. However, since we are only interested in a qualitative (and not quantitative) understanding of relevant attributes and their influence (in this paper), we omit presenting the corresponding charts since they show similar effects to the ones presented above (due to the fact that inflationary effects seem rather moderate as shown above).

4.3 Medical treatment

The charts of Figure 5 give an overview of total payments for medical treatment for the attributes MdE, age, type of injury, and scheme of coverage (whiskers of box-plots show 5 % and 95 % quantile).

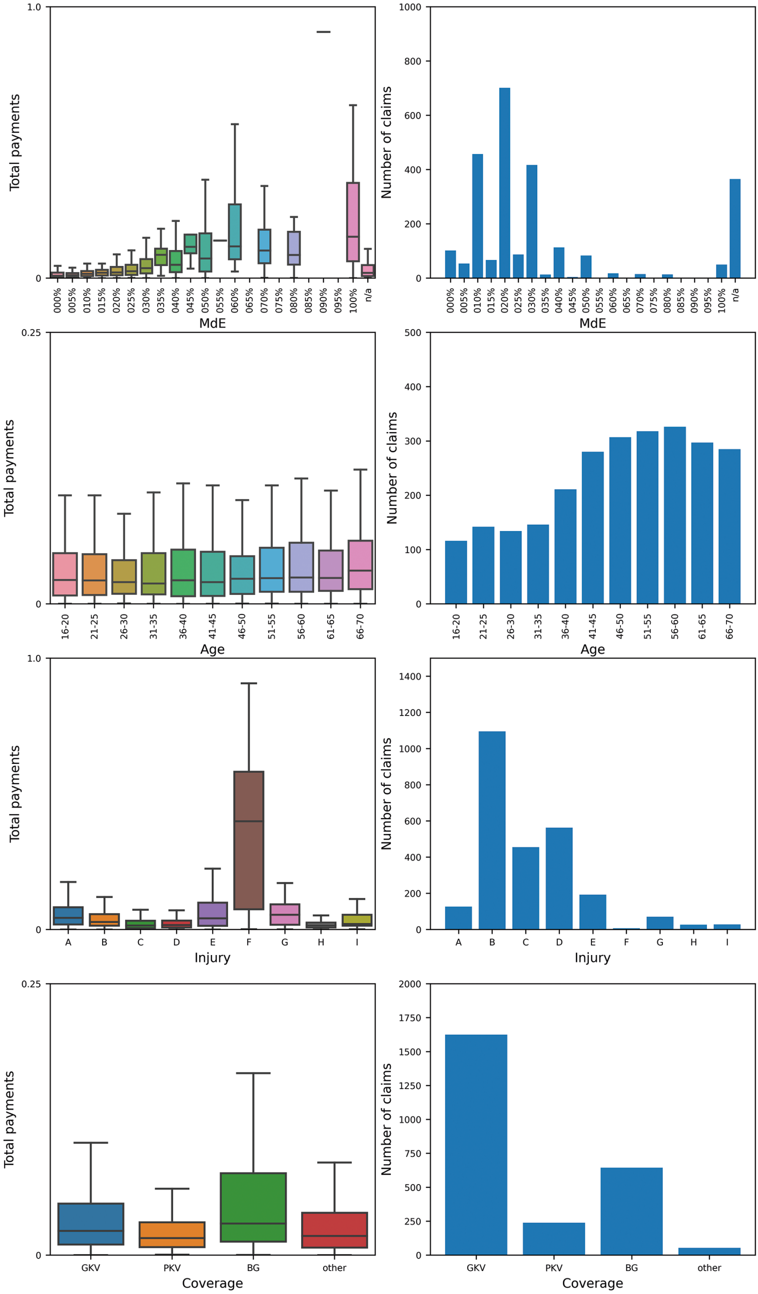

Fig. 5: Medical treatment – overview

As described in Wiedemann and John 2021, MdE might be used as a proxy for the severity of injuries. The importance of this attribute is underlined by our data.

In order to avoid unmeaningful results caused by sparse data cells, we shall focus on the following data for further analysis (the charts of Figure 6 & 7 show mean total payments and number of claims for the respective attributes):

- MdE: 0 %, …, 50 %
- Injury: B, C, D
- Coverage scheme: GKV, BG

The charts of Figure 6 & 7 underline the strong dependence on MdE. Overall, age only seems to have a slight effect for the considered MdE groups. The relevance of injury and coverage scheme is, of course, obvious (for details see Küppersbusch and Höher 2016) and highly underlined by our data. As a result, individual claims models should at least be based on MdE, injury, and coverage scheme.

In Wiedemann and John 2021, it was mentioned that the development of base rates (of the DRG system) is a relevant indication for the inflation of payments for medical treatment. Figure 8 shows the annual changes of the corresponding base rates since 2006.

An analysis of the inflation in our data is (as in the case of total payments for pain and suffering), of course, rather difficult since differing attributes of claims need to be taken into account, which will require detailed modelling. However, a rough idea might be derived from considering the development of total payments for individual MdE groups over time (for different injuries), since MdE is the most relevant attribute. One needs to keep in mind, however, that other aspects (like injury and severity) do play an important role and, hence, there might still be significant volatility in each MdE group. Moreover, since the settlement period of a claim might span over multiple years with multiple payments, inflation will, of course, also affect the settlement period of claims. Hence, an analysis can only be based on claims with short settlement periods (1 – 2 years), which will usually be the case with minor or medium injuries (lower MdE groups).

The charts of Figure 9 show the index of total payments of medical treatment (on the basis of 2006) for the MdE groups 10 %, 20 %, 30 % and the average index of those three indices. Due to the introduction of the DRG system in 2004, earlier reporting years are also excluded. Moreover, we exclude the coverage scheme BG and the reporting years 2012 and later due to the low number of claims in those years.

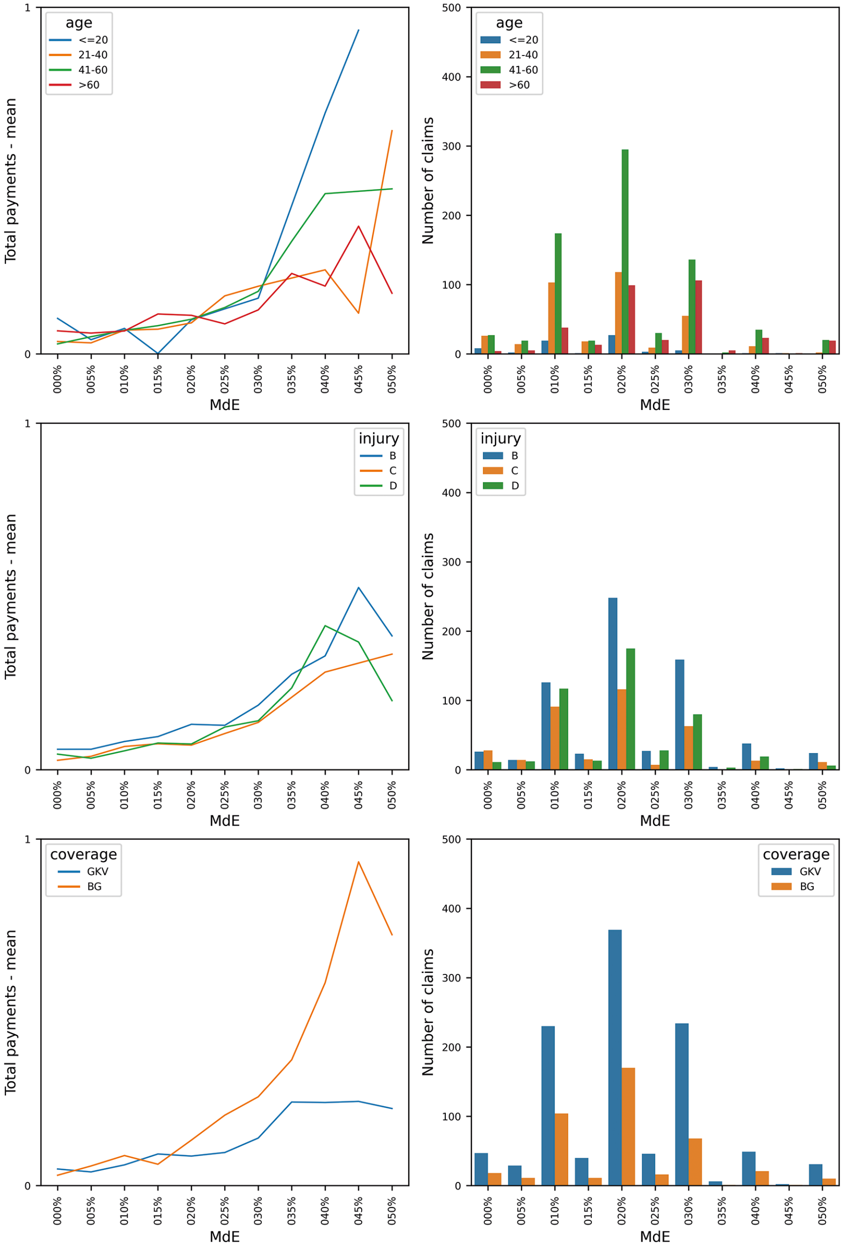

Fig. 6: Medical treatment – details (I)

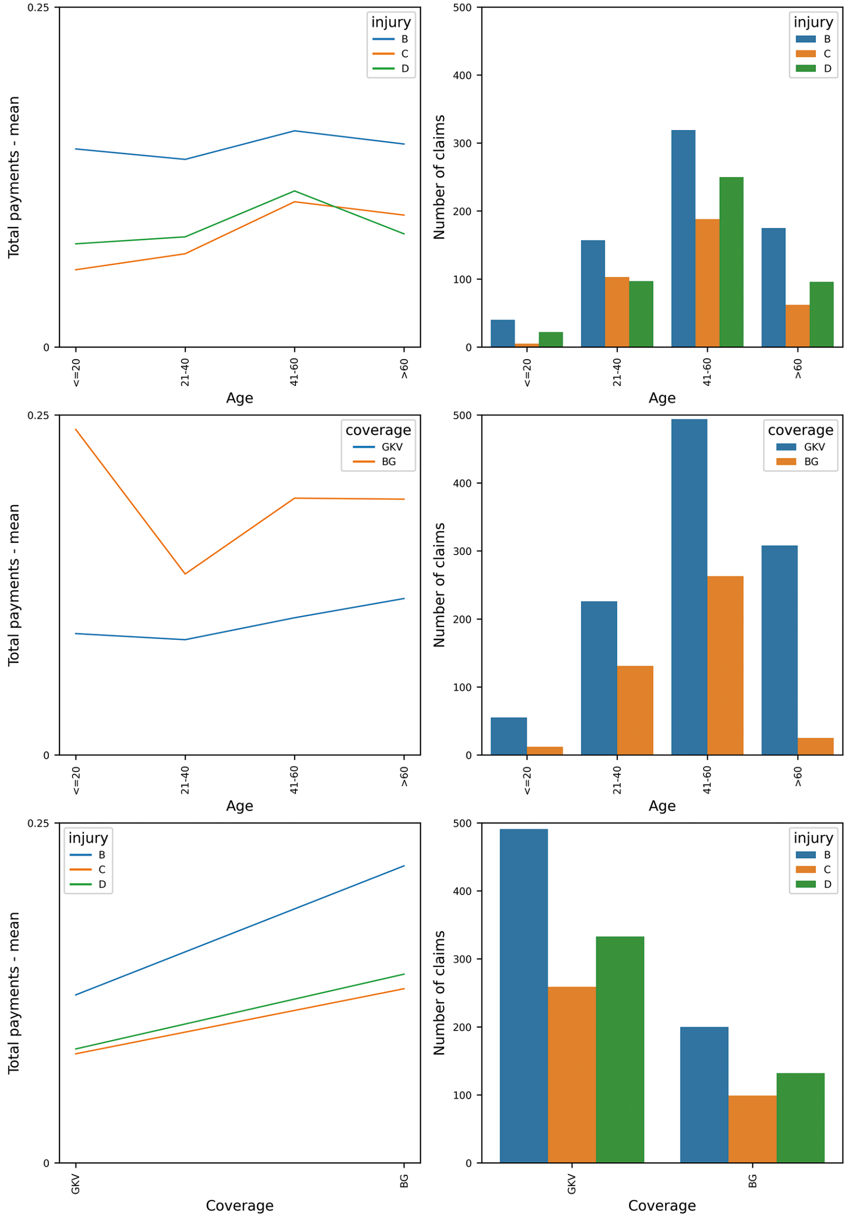

Fig. 7: Medical treatment – details (II)

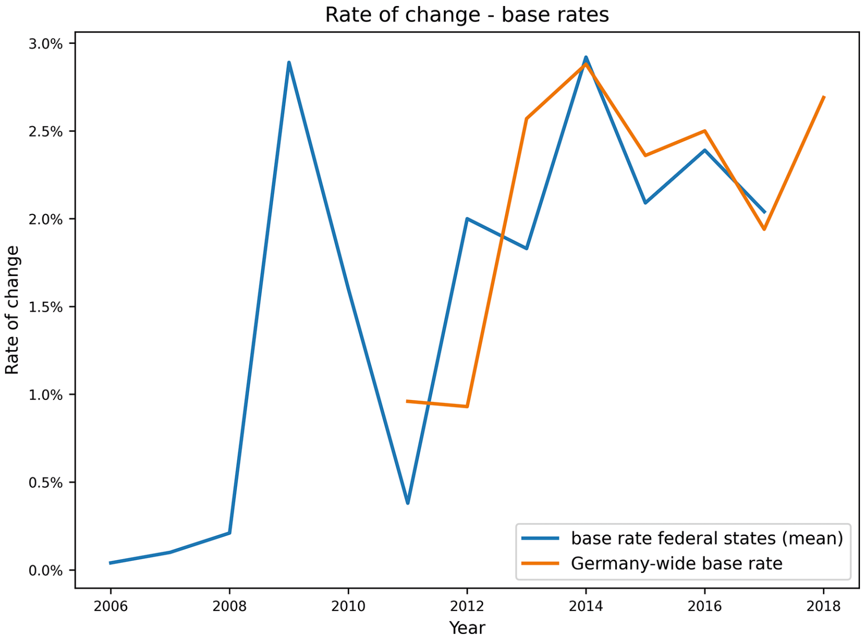

Fig. 8: Rate of change – base rates

The volatility is, as expected, very high and this analysis can only serve as a motivation. We may, however, derive that the average inflation in our sample does not seem inconsistent with the development of base rates for injury B and C. The development of the index for injury D does, however, show significant deviations (mainly due to the initial jump from 2006 to 2007) which may be caused by a change in the severity of claims. Analysing this further, however, would need more detailed data. Nevertheless, for further modelling, we suggest to adjust total payments on the basis of (Germany-wide) base rates for the coverage schemes GKV and PKV (they constitute the major share of all claims).

As in the case of total payments for pain and suffering, we omit presenting results based on inflation adjusted data since they show similar effects to the ones presented above (due to the fact that inflationary effects seem rather moderate as shown above).

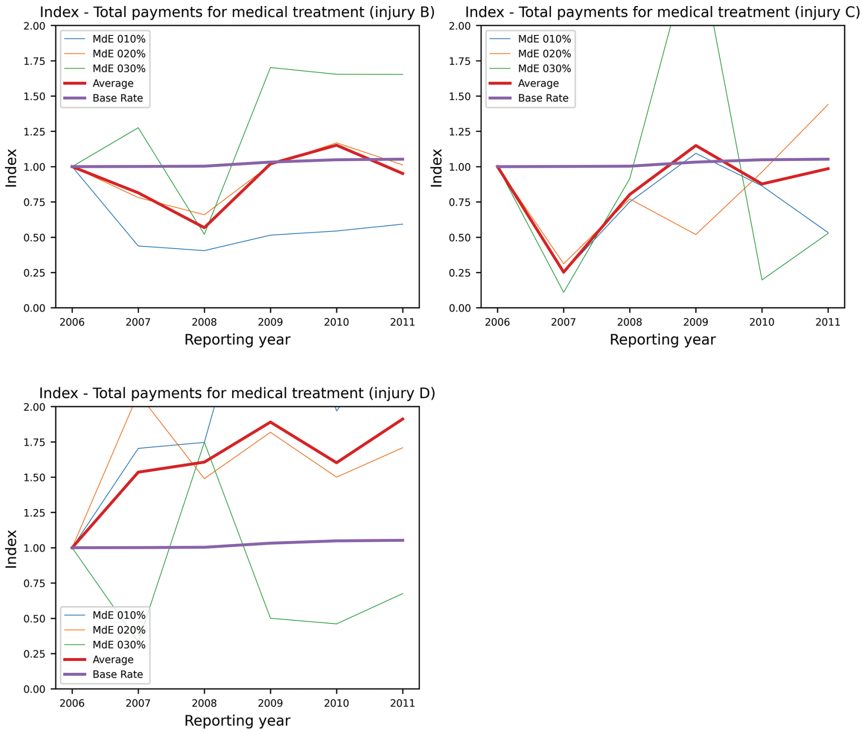

Fig. 9: Medical treatment – index of total payments

4.4 Loss of earnings

4.4.1 Housekeeping

The charts of Figure 10 give an overview of total payments for housekeeping for the attributes MdE, age, and family status (whiskers of box-plots show 5 % and 95 % quantile).

As described in Wiedemann and John 2021, MdE might be used as a proxy for impairment in housekeeping. The importance of this attribute is underlined by our data.

In order to avoid unmeaningful results caused by sparse data cells, we shall focus on the following data for further analysis (the charts of Figure 11 show mean total payments and number of claims for the respective attributes):

- MdE: 0 %, …, 50 %
- Family status: single, married, divorced

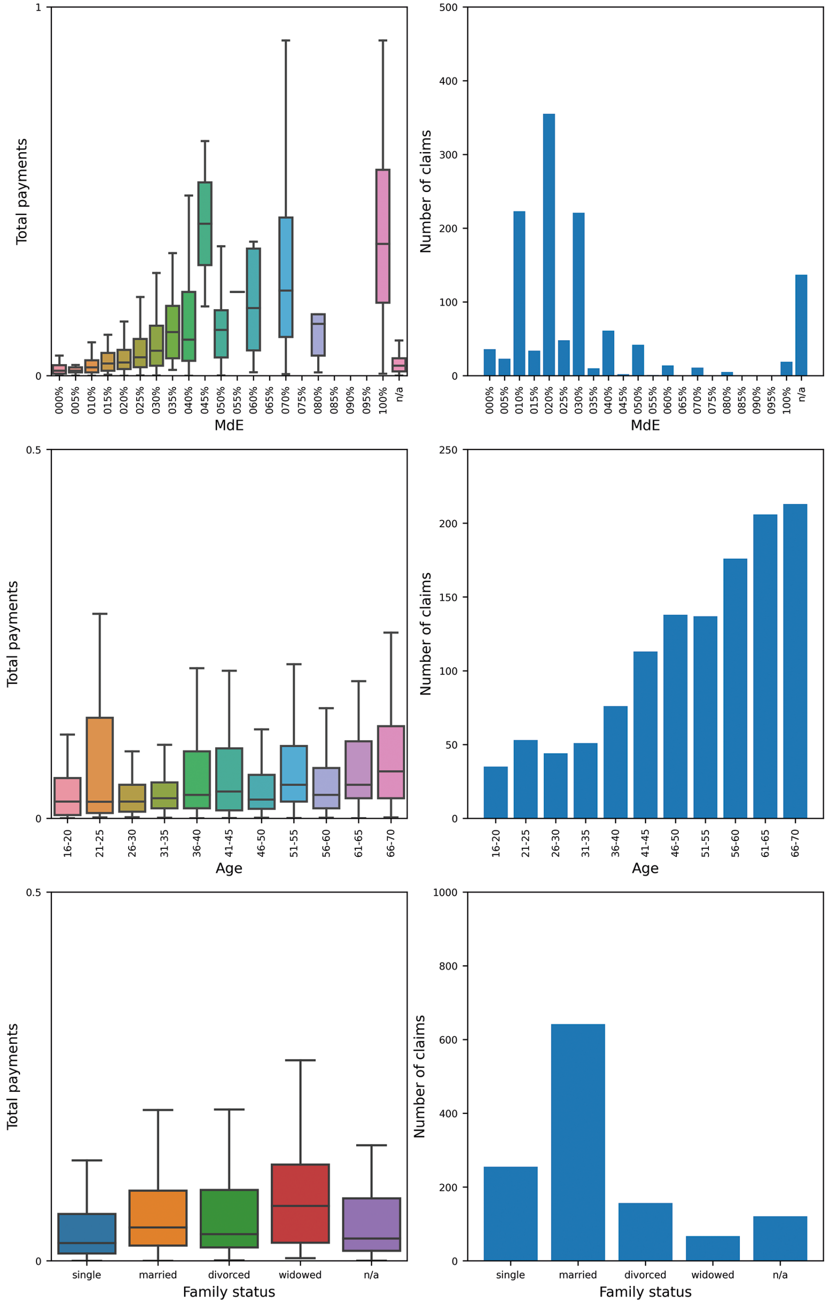

Fig. 10: Housekeeping – overview

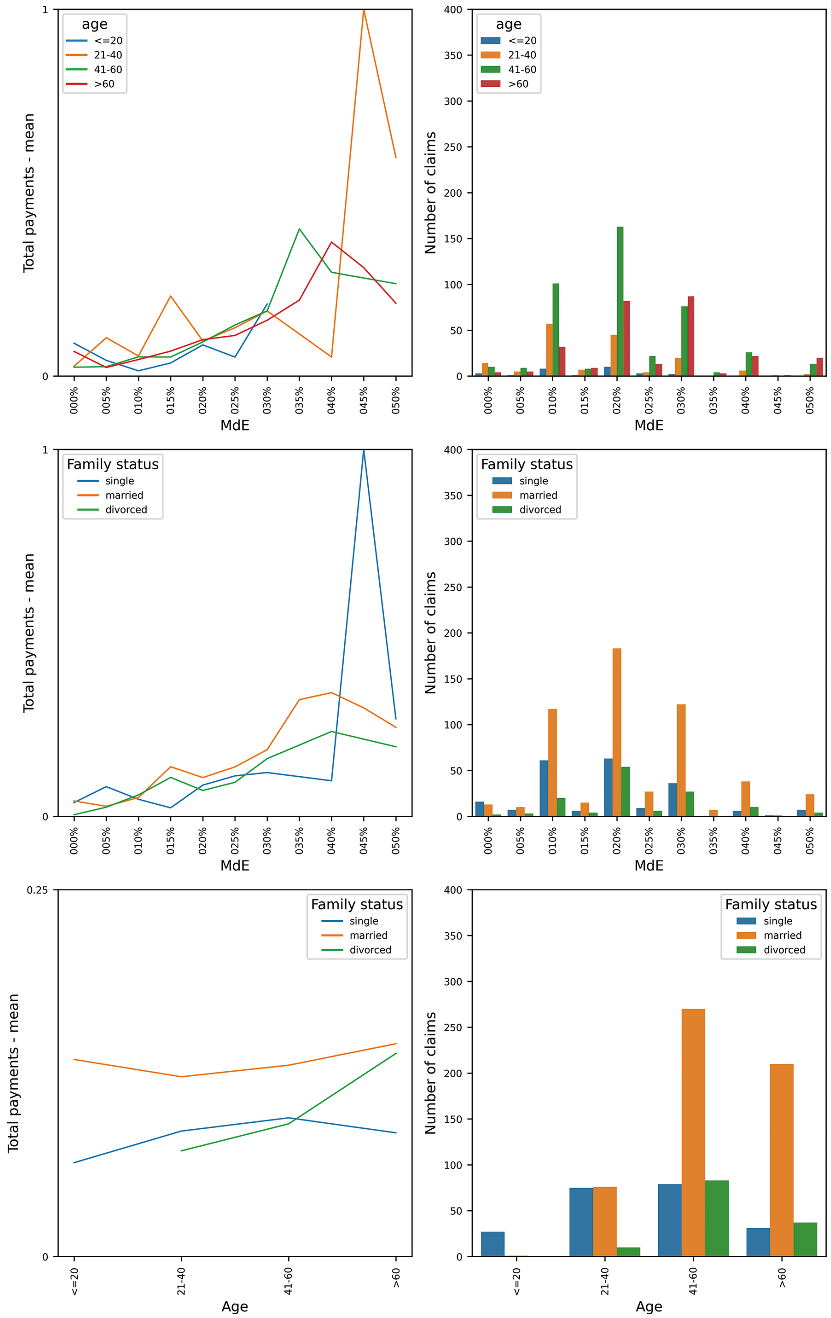

Fig. 11: Housekeeping – details

The charts of Figure 11 underline the strong dependence on MdE. The relevance of family status is also underlined by our data (for details see Küppersbusch and Höher 2016). The situation with age is more delicate since there might be opposing effects. Compensations might be restricted to the end-age of 75 (for details see Küppersbusch and Höher 2016), entailing falling damages with increasing age. However, impairments (due to the accident) might be higher for older age groups, hence, yielding an opposing effect. As a result, individual claims models should at least be based on MdE and family status.

Due to the smaller magnitude of payments and the lower number of cases, an analysis of inflationary effects is omitted.

4.4.2 Loss of earnings not including housekeeping

Analysing loss of earning claims (not including housekeeping) is more delicate due to the fact that payments will strongly depend on the age and wage of the claimant. In order to be able to compare claims, both effects must be taken into account. In this situation, a risk measure motivated by the following example turns out to be appropriate.

Let us consider a claimant earning 1,000 € per month; if – as a result of the accident – the claimant is permanently incapacitated, 1,000 € per month will need to be compensated (together with potential pension insurance contributions[17]) until retirement age (potential career changes and wage increases will need to be taken into account as well), constituting the worst case cashflow. The actual cashflow, however, might be very different. In the example shown in Figure 12, the claimant is able to work 50 % after three months.

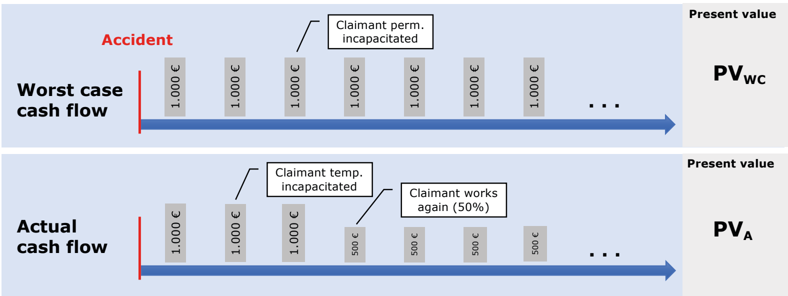

Fig. 12: A schematic example for payments for loss of earnings

[17] One might take potential pension insurance claims into account by increasing gross wages appropriately (since contributions are a fixed share of gross wages in most cases, which might, however, change slightly over time). The effect on RLE will be a rescaling. As mentioned before, the gross wages of our sample already take this into account.

Hence, one might consider the ratio of the present values of both cashflows, defining the following risk measure which we call risk loss of earnings (RLE):

$$\mathrm{RLE} = \frac{\mathrm{PV_A}}{\mathrm{PV_{WC}}}$$

In order to calculate RLE, the wage must be known. The wage data of our sample will either be the known wage at the time of the accident or – if no wage is known (for instance in case claimants are pupils, students, unemployed, etc.) – an estimate for assessing damages for loss of earnings. Cases with estimated wages are of minor significance as shown in the last chart of Figure 14.

In theory, RLE should be between 0 and 1. In practice, however, there might be exceptional cases. Obviously, once RLE, age, and wage are known, $\mathrm{PV_A}$ can be calculated. Moreover, with this approach, inflation in wages will automatically be taken into account. In order to analyse damages for loss of earnings (not including housekeeping), we shall focus on RLE. This approach is different to all other claims components, where our direct focus is always on total payments.

For simplicity, our calculations of present values are based on an interest rate of zero percent. We have based the worst cashflow on gross wages (of our sample) and a fixed retirement age of 67. Potential future wage increases will not be taken into account in the calculation of $\mathrm{PV_{WC}}$. This will suffice to determine relevant attributes and their influence on RLE. The actual magnitude of RLE will, however, depend on the interest rate together with a projection of career and wages. Hence, both will need due attention for modelling in practice (esp. since lump sum settlements might play an important role). One might also consider changing interest rates over time (for instance based on financial market data).

The charts of Figure 13 & 14 give an overview of RLE for the attributes MdE, age, scheme of coverage, occupation, and wage (whiskers of box-plots show 5 % and 95 % quantile). In case of unknown wage or age above 67, RLE is not defined. The wage information is binned into intervals with the following convention: the interval (0 €, 500 €] is denoted by "<=500", the interval (500 €, 1,000 €] is denoted by "<=1000", etc.

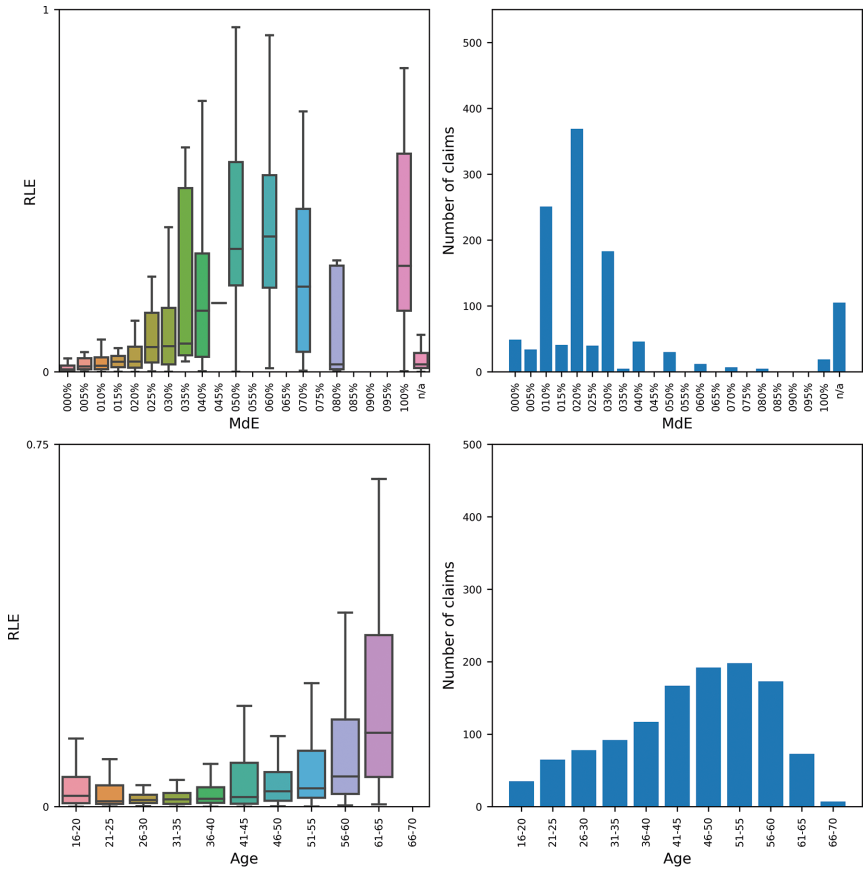

Fig. 13: RLE – overview (I)

One can already see a strong dependence on MdE, age, and occupation, as expected. As discussed in Küppersbusch and Höher 2016, assessing damages in case of self-employed, unemployed, children, pupils, etc. is very difficult since forecasting career and earnings contains a high level of uncertainty. In the case of employees and labourers, forecasts are not as difficult since they may be based on the existing work history (esp. earnings). Moreover, as discussed in Küppersbusch and Höher 2016 or Wiedemann and John 2021, the separation between physical and non-physical work is important. In our case, this is reflected by the occupation groups "labourer" and "employee".

First of all, we shall analyse the wage-independence of RLE. For this, we restrict our data to (the charts of Figure 15 show mean RLE and number of claims for the respective attributes):

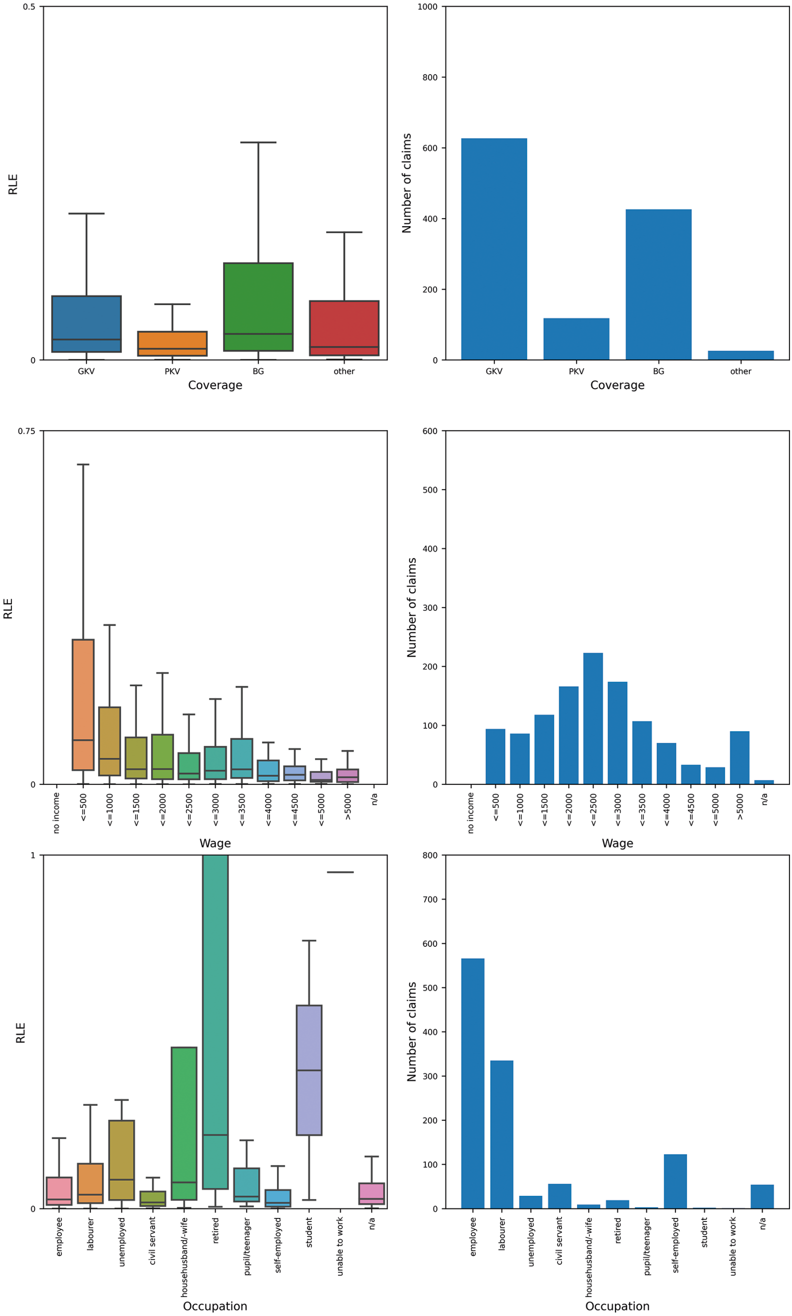

Fig. 14: RLE – overview (II)

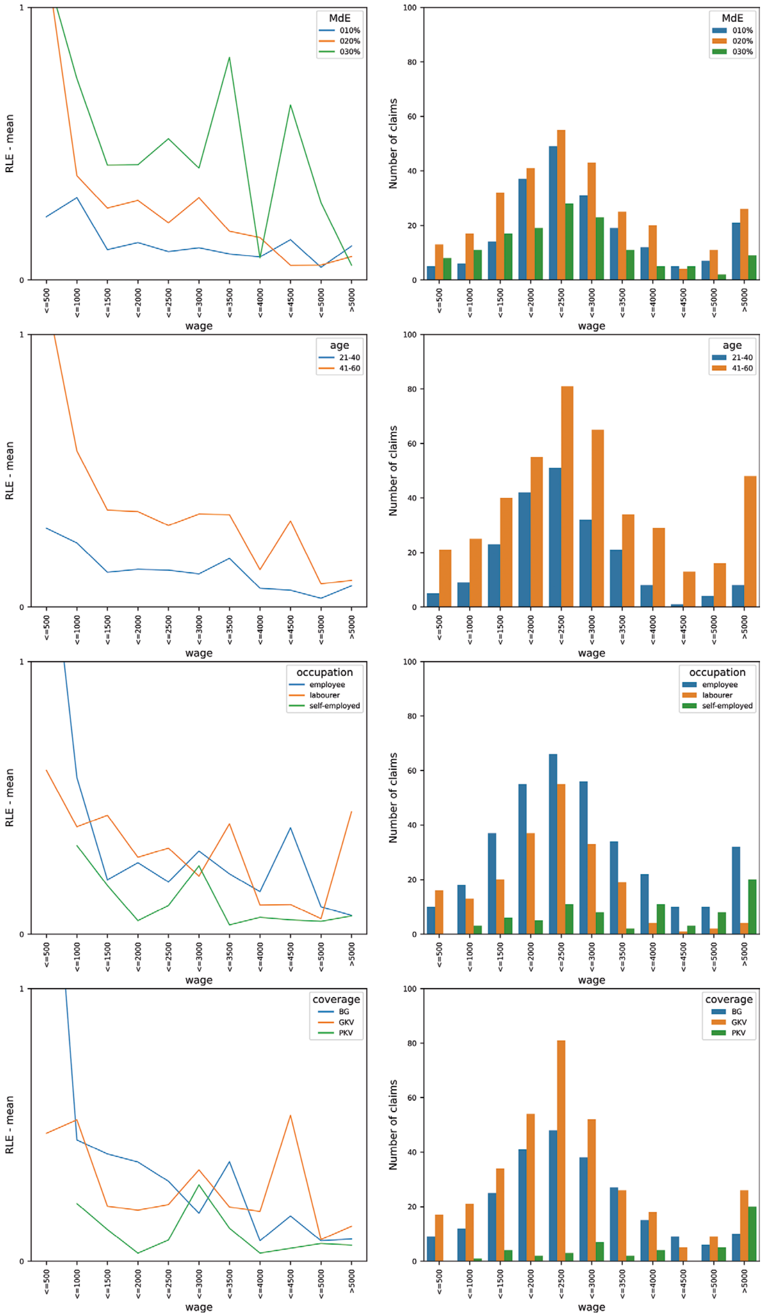

Fig. 15: RLE – analysis of wage-independence

- MdE: 10 %, 20 %, 30 %[18]
- Age: 21 – 60
- Coverage scheme: GKV, PKV, BG
- Occupation: employee, labourer, self-employed

The charts of Figure 15 suggest that RLE is indeed sufficiently independent of wage and justifies the use of RLE as an appropriate risk measure. The anomalies shown for low wage groups might be caused by career projections which are not based on current low wages (for instance in case of claimants undergoing professional training at the time of the accident with significant projected future wage increases).

In order to avoid unmeaningful results caused by sparse data cells, we shall focus on the following data for further analysis (the charts of Figure 16 & 17 show mean RLE and number of claims for the respective attributes):

- MdE: 0 %, …, 50 %
- Age: 21 – 60
- Coverage scheme: GKV, PKV, BG
- Occupation: employee, labourer, self-employed

To summarise the charts of Figure 16 & 17, the attributes MdE, age, coverage scheme, and occupation seem highly relevant for modelling RLE. Their respective influence is as expected (for details, see Küppersbusch and Höher 2016). As a result, modelling loss of earnings claims should be based on RLE as a risk measure. Total payments can be modelled using RLE together with age and wage (which are known or need to be estimated for a given claim). Since inflation in wages will automatically be taken into account, this approach is particularly appealing. It is, however, important to monitor the changes of RLE over time (for example for different occupational groups).

4.5 Additional needs (not including nursing care costs)

The charts of Figure 18 give an overview of total payments for additional needs for the attributes MdE, age, and type of injury (whiskers of box-plots show 5 % and 95 % quantile).

Cases of paraplegia clearly stand out. As described in Wiedemann and John 2021, MdE might be used as a proxy for the severity of injuries.

---

[18] Restriction to the biggest MdE groups, to avoid sparse cells. This will only affect the first chart shown in Figure 15.

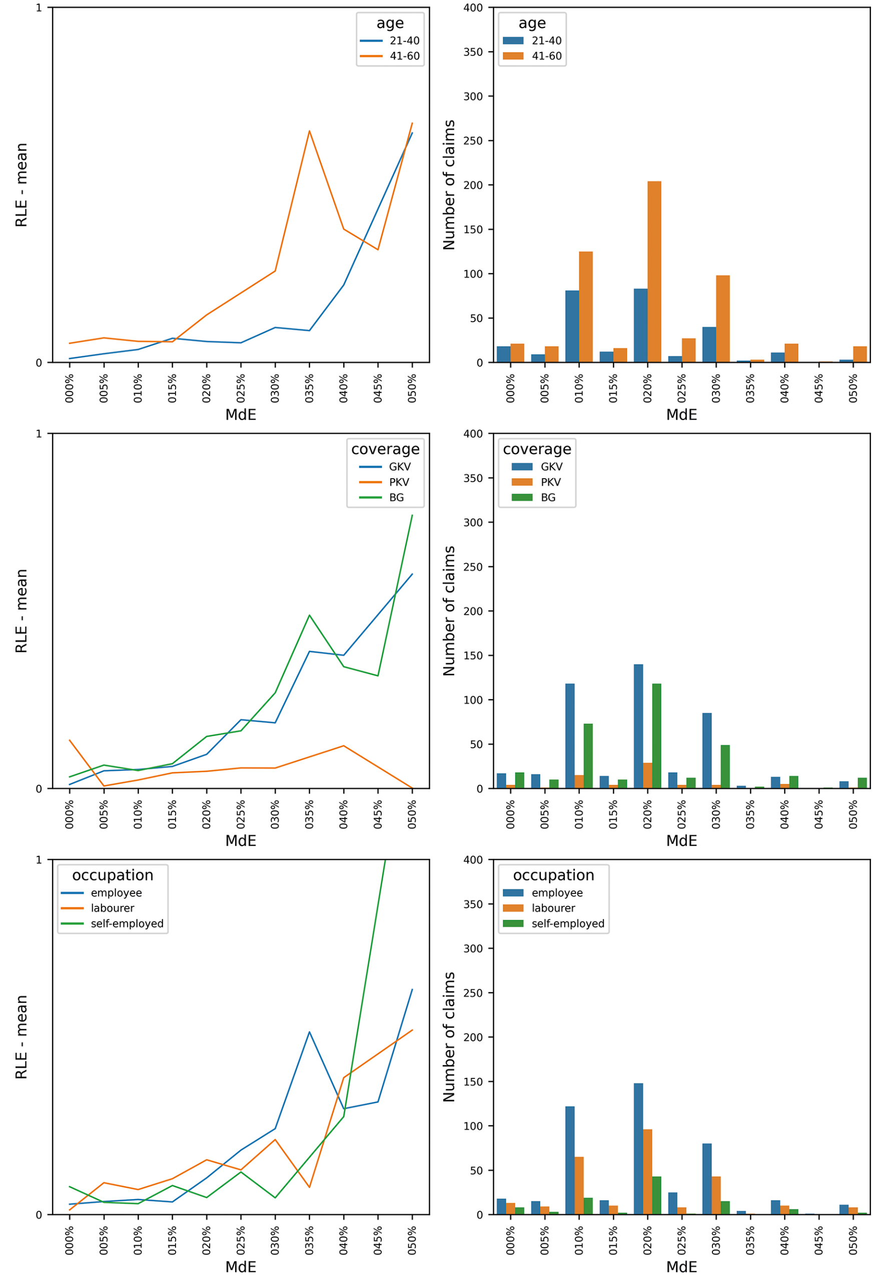

Fig. 16: RLE – details (I)

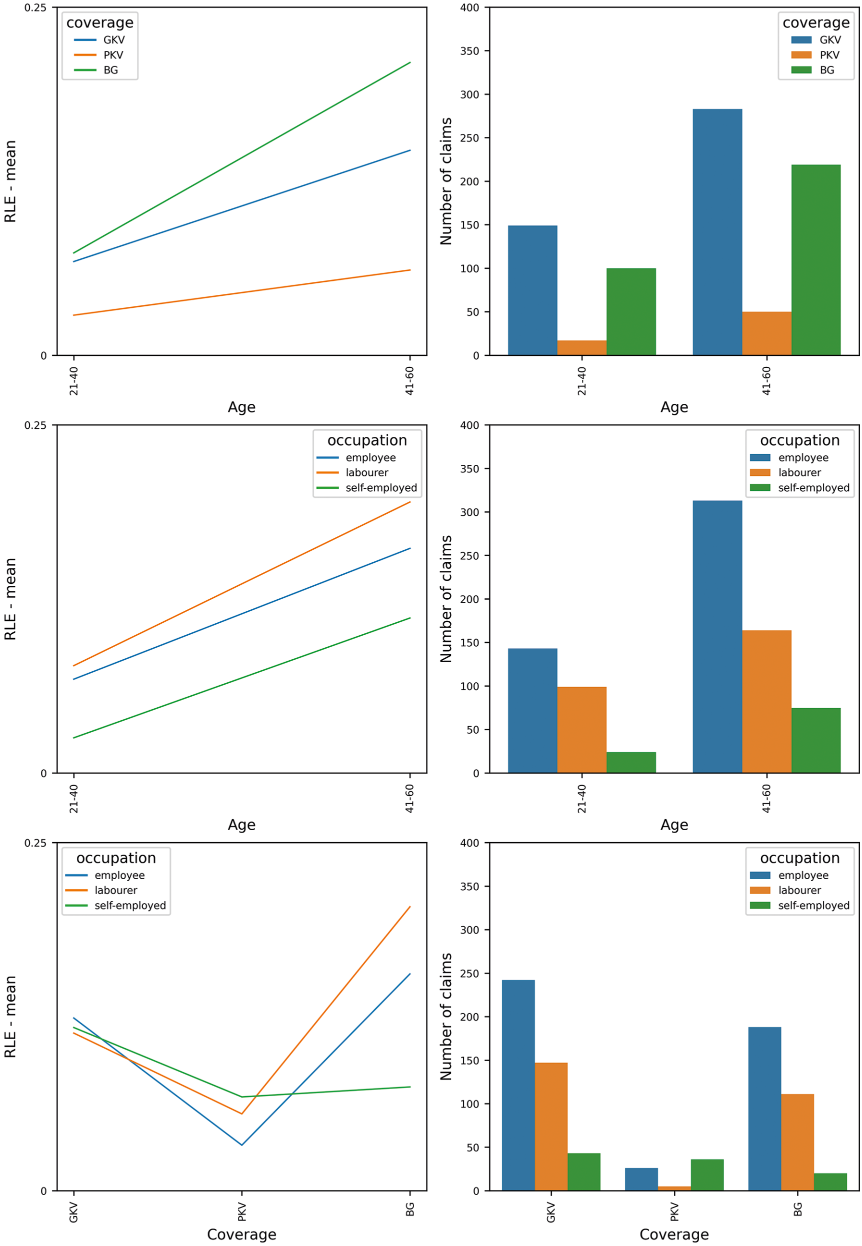

Fig. 17: RLE – details (II)

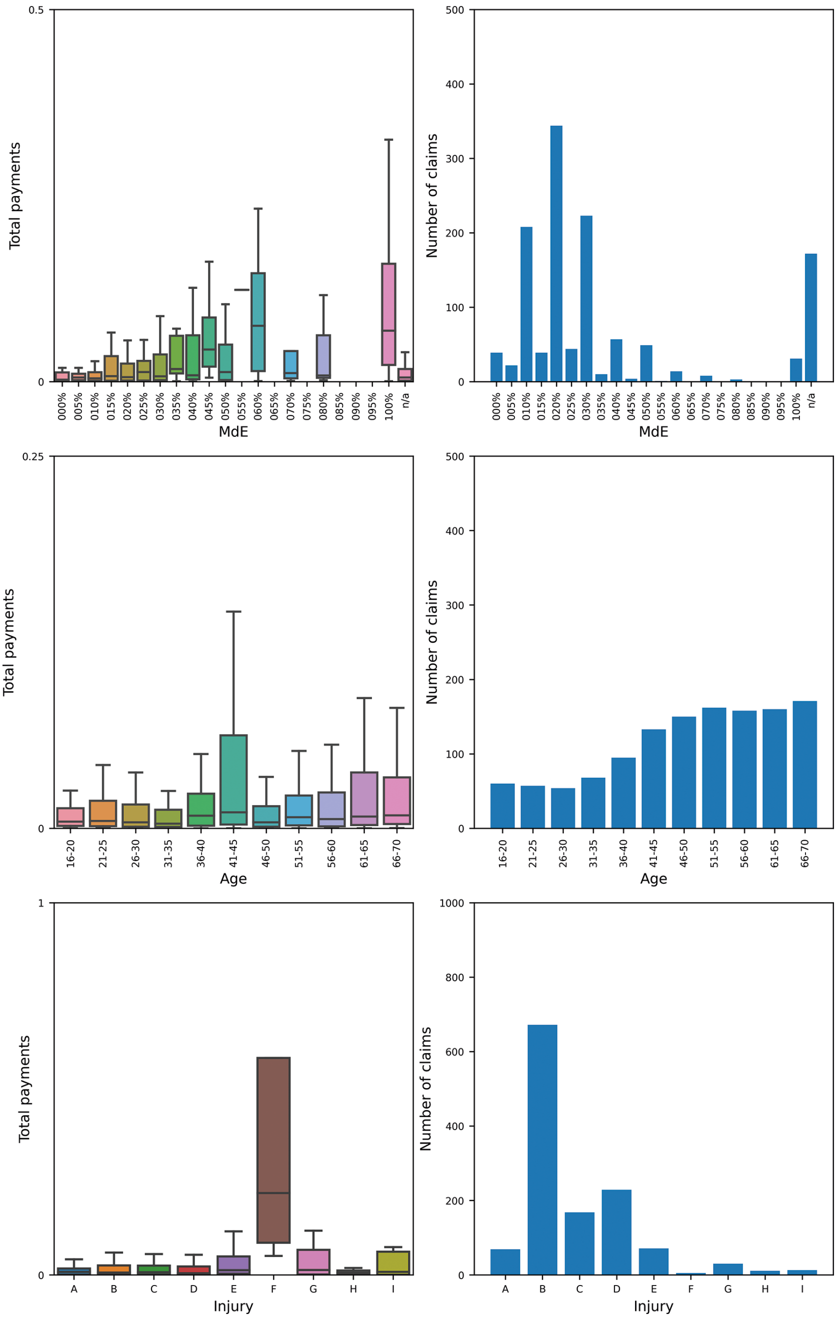

Fig. 18: Additional needs (not including nursing care costs) – overview

In order to avoid unmeaningful results caused by sparse data cells, we shall focus on the following data for further analysis (the charts of Figure 19 show mean total payments and number of claims for the respective attributes):

- MdE: 0 %, …, 50 %
- Injury: B, C, D

As described in Küppersbusch and Höher 2016, the care situation of the claimant (outpatient, inpatient, nursing care) might also be important. This information, however, was not available to us. The charts of Figure 19 underline the dependence on MdE (for the age groups 41 – 60 and > 60), age, and injury. As a result, individual claims models should at least be based on MdE, age, and injury. One must bear in mind, however, that damages in this case are of a smaller magnitude and, hence, a flat rate model might suffice in practice.

Due to the smaller magnitude of payments and the lower number of cases, an analysis of inflationary effects is omitted.

5. Improving predictions with individual claims models – some motivational aspects

In Wiedemann and John 2021, relevant attributes which determine the magnitude of total payments for bodily injury claims were derived from a detailed analysis of the (legal) regulations for compensation (see also Küppersbusch and Höher 2016). Together with the results of the last section of the paper in hand, we get a sound overall picture of relevant attributes which should be used as a modelling basis. From this point of view, they are clearly the ones to consider when working towards individual claims modelling of bodily injury claims (in German non-life motor insurance). Moreover, this approach reveals the relevant real-world attributes from claims handling, so modelling can be directly connected to real-world effects. In this way, the results of actuarial reserving are directly accessible, interpretable, and, hence, usable by other departments (for instance claims department).

Table 1 summarises the findings of the last section, describing the modelling basis for different claims components (most relevant attributes in bold face).

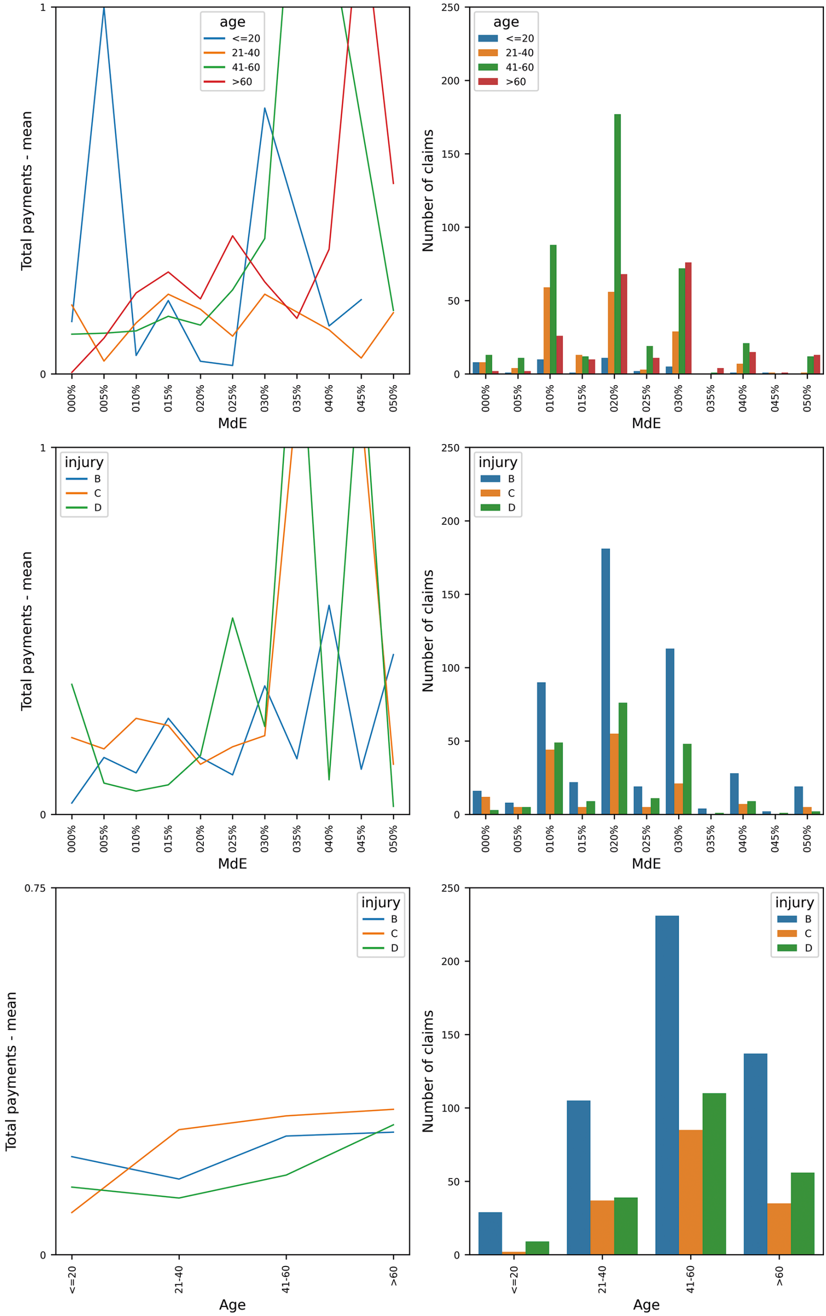

Fig. 19: Additional needs (not including nursing care costs) – details

Table 1

Relevant attributes for modelling of claims components

Claims component	Relevant attributes for modelling
Pain and suffering	**MdE**, injury, age
Medical treatment	**MdE**, injury, coverage scheme
Loss of earnings–housekeeping	**MdE**, family status
Loss of earning not including housekeeping (risk loss of earnings)	**MdE, age, wage**[19], coverage scheme, occupation
Additional needs	**MdE**, injury, age

As discussed in Section 2, our motivation for developing individual claims models stems mainly from practical problems which could not be approached by using aggregated methods. At this point, however, it remains an open question whether the insights from our data sample are really helpful for achieving individual claims models yielding improved predictions of claims reserves for bodily injury claims (compared to using just aggregated data). A comprehensive answer to this will always require detailed modelling using the attributes presented above. This, however, is out of the scope of this paper. Moreover, assessing the "quality" of an individual claims model is not a purely mathematical task. Models must be challenged on a daily basis by actually putting them into operation (in our case, they assist claims handlers on a daily basis). Furthermore, the implementation of feedback cycles (from claims handlers) as well as monitoring and controlling cycles is absolutely essential as the basis for further improvement. Nevertheless, in the remainder of this section, we aim to present some motivational aspects (based on our data sample), showing why the attributes discussed (in Section 4) are really helpful for actuarial reserving.

For further analysis, we focus on the most relevant components pain and suffering, medical treatment, and loss of earnings (not including housekeeping) and consider the following subset of our data sample (as these are the largest cells):

- Reporting years: 2000 – 2012
- MdE: 10 %, 20 %, 30 %
- Age: 21 – 40, 41 – 60
- RLE: in the interval [0,1] (to exclude extreme outliers)

This subset contains of just over 1,200 claims (approx. 45 % of our data sample). Since our following arguments are of purely motivational character, we do not adjust our data for inflationary effects. Moreover, to keep arguments sim-

[19] Wage as a necessary attribute for transition to total payments.

ple and transparent, we mainly focus on the effects of MdE since it is one of the most relevant attributes.

Any attempt to improve reserving of bodily injury claims (and, hence, also to achieve individual claims models) should always put injuries and especially the resulting impairment at the heart of modelling since all claims components are directly linked to them. It is clear that the magnitude of total payments (as well as the handling period) for bodily injury claims will strongly depend on the impairment due to the injures caused by the accident. As mentioned before, our injury data is unfortunately not detailed enough. However, the attribute MdE (which describes the general impairment of earning capacity as a result of the accident event and might also be used as a proxy for the severity of injuries) turns out to be highly relevant as a result of our case study and should, hence, give significant advantage for making predictions.

Figure 20 shows the shares of MdE groups in our considered subset per reporting year.

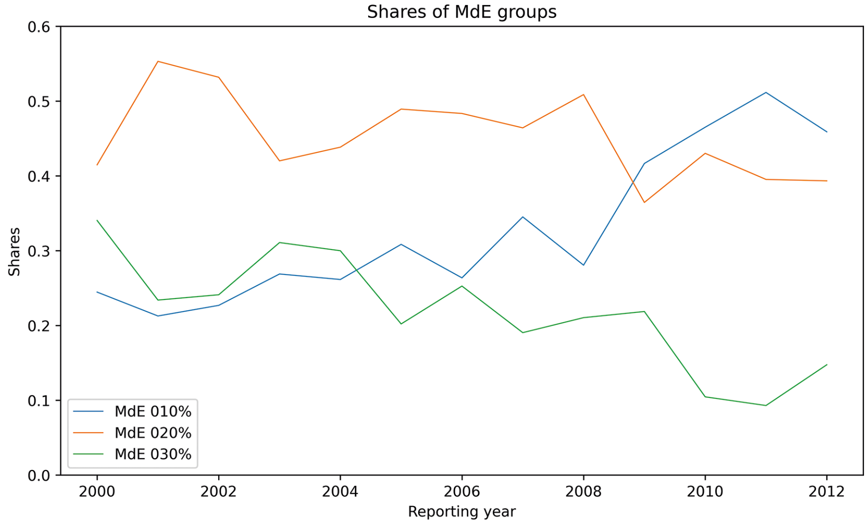

Fig. 20: Shares of MdE groups per reporting year

As we can see, the shares of MdE groups are volatile over time which should apply to any portfolio of bodily injury claims. We remark that the sharp drop for MdE group 30 % in recent reporting years is a consequence of our data sample: As reporting years increase, the settlement period of claims in our sample decreases since we only consider (essentially) settled claims (see also Section 4.1). Claims with shorter settlement periods, however, tend to be of smaller magnitude. This is also an important aspect when working with aggregated data: Aggregated payment data of early development years (esp. first development year) will be "dominated" by payments for claims of smaller magnitude. Moreover, one of the main problems when making predictions based on aggregated data is the fact that for a given reporting year, the composition of claims (shares of magnitudes of claims, especially the share of "major" losses) is not known and not taken into account. This usually complicates calculations as it is unclear whether previous reporting years are comparable at all (note that a differing composition of magnitudes of claims will also result in differing cashflow patterns, since bodily injury claims of larger magnitude are usually of longer tail). Significant volatility in the shares of the most relevant attribute values of Table 1 (as in the case of MdE shown Figure 20) over time will be an obstacle for aggregated methods in general, as this introduces artificial volatility in aggregated data (claims triangles) by superposing cashflows with differing patterns (which may be understood once the distribution of attribute levels is known).

Figure 21 shows the following:

- First row: mean total payments for the components mentioned above (as well as their sum) per reporting year for different MdE groups[20]
- Second row: mean total payments for different MdE groups per reporting year for each component

As we can see, different MdE groups have very different levels of mean total payments for different claims components (note that the MdE group 30 % is significantly smaller, resulting in higher volatility). Especially for high MdEs (usually corresponding to major losses), levels of total payments will be significantly higher (see also Section 4) and their share (of all claims per reporting year) will be volatile over time as well. The total number of claims with high MdEs in our sample is, however, rather small making a detailed analysis impossible. As this is an important aspect for modelling; further relevant data for this critical aspect was generated with subset III of our data sample (nursing care claims) since MdEs of this subset are usually large (mainly 100 %).

The charts of Figure 21 also suggest that there seems to be a reasonable stability (for mean payments of pain and suffering and medical treatment) for different MdE groups over the period of considered reporting years which is also an essential point for modelling.

Although MdE also has a significant effect on payments for loss of earnings, it is hard to interpret the third column since many other attributes (esp. age and wage) will play an important role as well. Note that the drop in mean total payments for more recent reporting years is also a consequence of the fact that our

[20] Only considering claims with non-zero payments for the respective component.

Fig. 21: Mean total payments for different MdE-groups

data sample consists of (essentially) settled claims (resulting in claims with shorter settlement periods in more recent years and, hence, shorter periods for loss of earnings). Passing to RLE, however, we get a similar picture as for pain and suffering and medical treatment, especially very different RLE-levels for different MdE groups as well as a reasonable stability over time as shown in the charts of Figure 22 (note that the MdE group 30 % is significantly smaller resulting in higher volatility). In this case, however, the effect of age (as seen in Section 4) must be factored in as well (as it also has a considerable effect). Figure 22 shows the following:

- First row: mean RLE per reporting year for different age groups
- Second row: shares of age groups per reporting year
- Third row: mean wage of age groups per reporting year[21]

[21] Only considering cases with known wage.

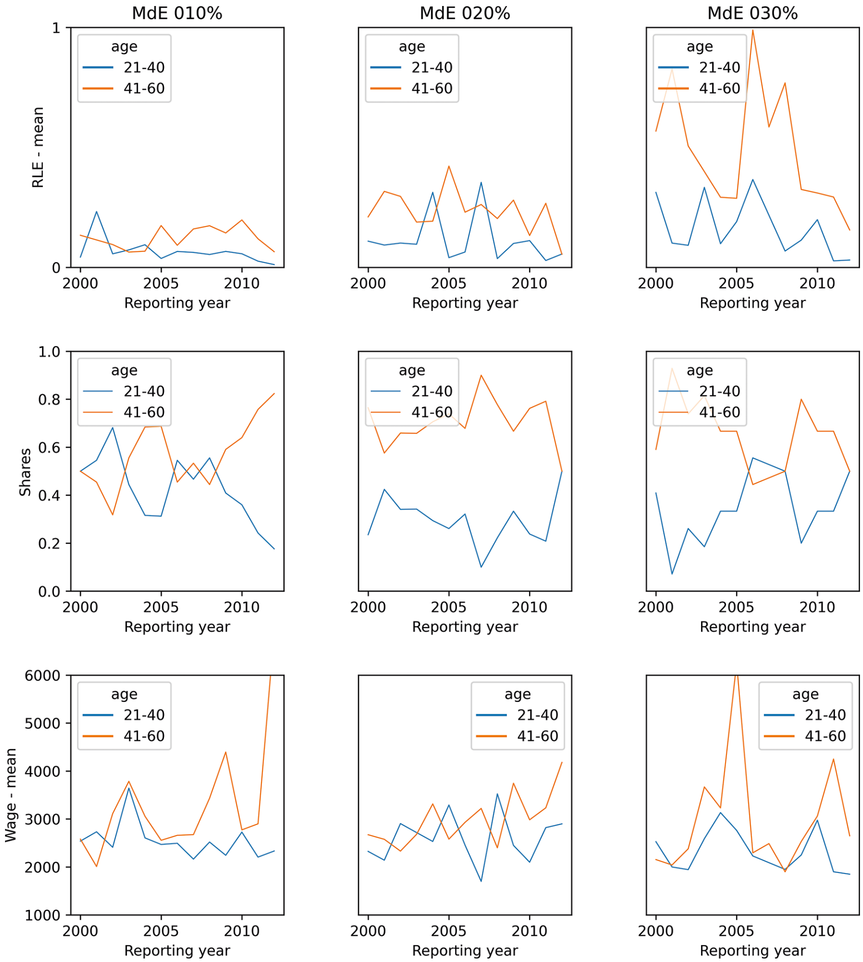

Fig. 22: Mean RLE, age shares, and mean wages for different MdE-groups

The shares of the considered age groups per reporting year are also very volatile which should also apply to any other portfolio of bodily injury claims. Additional volatility will be brought into play by differing wage distributions per reporting year (the charts of Figure 22 only show the average wage per reporting year), directly affecting the magnitude of total payments for loss of earnings. Both aspects can be taken out of the equation by using the corresponding attributes for making predictions based on modelling RLE.

This preliminary analysis already suggests that there is plenty of "structure" in bodily injuries claims data and that already bringing just the attributes MdE and age (for loss of earnings) into play leads to a significantly better understanding of the magnitudes of claims amounts and achieve individual claims modelling. The most simplistic approach would be just using mean total payments/RLE's per MdE and possibly per age (for loss of earnings) from the past – giving the relevant magnitudes of claims – together with the respective current attribute levels, resulting in better estimates for more recent reporting years (compared to estimates on aggregated data) because only "relatively little" or insufficient payment data is known (keep in mind that bodily injury claims are usually long tail claims). Moreover, one must keep in mind that further attributes (as shown in Table 1) will also play a role, yielding a path to further improvements. As a result, we are proposing to use the attributes from Table 1 for modelling initial ultimate claims cost U_t and cashflows C_t ($n = 0$, as described in Section 2) for individual claims without the use of any incurred payment data.

A vital second step will be the analysis of cashflow patterns to model run-off periods and, hence, U_{t+n} and C_{t+n} ($n > 0$) covering claims with reporting years further in the past (for these claims, the actual incurred payment data is substantially greater and, therefore, more representative for the ultimate losses of the claims). Preliminary analysis shows that the attributes in Table 1 are also relevant for understanding these aspects. This is currently still research in progress and we believe it opens interesting paths for actuarial research. We believe that our suggested approach will also lead to significantly better results (compared to estimates based on aggregated data) in this case since, in particular, different magnitudes of claims are taken into account.

One might, of course, argue that our elaborations only demonstrate that one simply needs to subdivide aggregated data slightly further and there is no need for individual claims modelling. However, subdividing by MdE, age, wage, etc. will inevitably be too cumbersome (too many triangles to consider) and automatically lead to working on the level of individual claims.

We find it also very important to point out that the development of individual claims models cannot be done by actuarial departments on their own and it goes well beyond actuarial and mathematical models and techniques. As described in John and Wiedemann 2018, a strong collaboration with claims departments and their expertise is essential. Especially expert input for modelling and calibrating is absolutely essential. They are "living" models which need strong ties to the operational world.

6. Summary

Modern non-life actuarial departments will be confronted with numerous questions concerning claims and their settlement (payments, reserves) on a regular basis (for instance in connection with profit projections, risk management, reinsurance optimisation, Asset Liability Management, claims steering, etc.). On the basis of mostly aggregated data, however, answering detailed questions is rather difficult, if not impossible. This was the starting point of our research into personal injury claims with the aim of getting a better understanding of their characteristics and, hence, generating added value for all stakeholders involved.

The first step of our project was the case study described in this paper. It has been shown that the attributes discussed in Wiedemann and John 2021 can indeed be obtained in practice, and, moreover, that these attributes are indeed relevant for modelling the respective claims components.

Our results pose the question of the applicability of traditional aggregated reserving methods (based on triangle data) to personal injury claims since none of the information described is taken into account explicitly. It is important to understand that the composition of claims components (pain and suffering, medical treatment, etc.) as well as their individual attributes (MdE, age, etc.) and settlement patterns will differ significantly from year to year (also due to inhomogeneity because of the relatively low number of personal injury claims). It is therefore hard to imagine how aggregating claims will yield homogenous triangles.

Hence, we propose to move forward towards a more detailed reserving of personal injury claims on the level of individual claims. However, it is clear that this can only be achieved by starting with sufficiently detailed and standardised data. As already mentioned in Section 1, on the basis of our results, we have developed individual claims models for all claims components of personal injury claims and we are planning to discuss these models as part of a future publication.

Moreover, we believe that individual claims modelling will open up many new and important areas of actuarial research which are highly relevant in practice: modelling cashflows on the level of individual claims, modelling injuries and MdE based on type of accident, actuarial models for claims steering, etc. The basis for such attempts is detailed data as presented in this paper.

References

Amberger, M./*John*, D./*Wiedemann*, M.: Aufbruch in ein neues Zeitalter. Versicher. Wirtsch. 05, 68 – 74 (2018).

John, D./*Wiedemann*, M.: Actuarial Data Analytics – Der Weg zur Einzelschadenreservierung. Aktuar 04, 226 – 230 (2018).

Küppersbusch, G./*Höher*, H. O.: Ersatzansprüche bei Personenschaden, 13th ed. C.H. Beck, München (2020).

Wiedemann, M./*John*, D.: A practitioners approach to individual claims models for bodily injury claims in german non-life insurance. ZVersWiss 110, 225 – 254 (2021), https://doi.org/10.1007/s12297-021-00511-2.

Wiedemann, M./*Herzog*, A.: Fachliche Aspekte zur Erhebung einer Datenbasis für die Modellierung von Personenschäden. DAV Journal 01/2025, 352 – 361 (2025a).

Wiedemann, M./*Herzog*, A.: Nursing care costs for bodily injury claims in German non-life insurance – a detailed case study. To appear in ZVersWiss (2025b).

Wüthrich, M.: Machine learning in individual claims reserving. Scand. Actuar. Journal 06, 465 – 480 (2018).