

# Detecting Smartphone Insurance Fraud

Björn Angtoft and Olof Hallqvist

DIVISION OF INNOVATION ENGINEERING | DEPARTMENT OF DESIGN SCIENCES  
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY  
2019

MASTER THESIS

hedvig



# Detecting Smartphone Insurance Fraud

A study of the data that detects fraudulent smartphone  
insurance claims

Björn Angtoft and Olof Hallqvist



**LUND**  
UNIVERSITY

# Detecting Smartphone Insurance Fraud

A study of the data that detects fraudulent smartphone insurance claims

Copyright © 2019 Björn Angtoft and Olof Hallqvist

*Published by*

Department of Design Sciences  
Faculty of Engineering LTH, Lund University  
P.O. Box 118, SE-221 00 Lund, Sweden

Subject: Innovation Engineering (INTM01)  
Division: Innovation Engineering  
Supervisor: Emil Åkesson  
Co-supervisor: John Ardelius at Hedvig AB  
Examiner: Lars Bengtsson

# Abstract

The vast majority of Swedes are in possession of a smartphone, typically covered via one's household insurance policy and smartphones are included in upwards of 50 % of household insurance claims (Hedvig, 2019). Some estimates say that upwards of 40 % of all smartphone claims are fraudulent. Payments to fraudsters falls on honest policy holders pay for with higher premiums. There are several vulnerabilities in the current measures of Insurance Fraud Detection (IFD) on smartphones regarding the indicators that invoke suspicion and the economic incentives for investigation. The purpose of this report was to identify viable Smartphone Insurance Fraud Indicators (SIFIs) to be used for IFD, along with suggestions for further improving the fraud detection capability in areas of data acquisition and investments in analytical tools. A triangulation methodology was employed, which extended to interviews with 12 practitioners of IFD, attendance at 2 international insurance conferences and a review of 20 published academic papers and articles on insurance fraud. The primary result of this report was a compilation of 51 distinct SIFIs. By contrasting and comparing the findings from the three method areas, several examples of divergences between theory and practice were identified. The results can be used to expand and revise existing sets of indicators, as well as prioritising investments in new analytical tools. Although the research is focused on smartphone claims, the results also have the potential to be apply to other common belongings such as laptops, and tablets.

**Keywords:** insurance, insurance fraud, insurance fraud detection, smartphone, insurance fraud indicators

# Sammanfattning

De flesta svenskar äger en smartphone, en ägodel som normalt täcks via hemförsäkring. Smartphones ingår i uppemot 50 % av alla skador som rapporteras till försäkringsbolaget Hedvig (Hedvig, 2019). Vissa experter rapporterar att så mycket som 40 % av alla smartphoneärenden som anmäls till försäkringsbolagen är bedrägliga. Den ersättning som betalas ut till bedragarna subventioneras via övriga försäkringstagares premier. Det finns flera sårbarheter i den nuvarande utredningsmetodik som syftar till att identifiera bedrägliga försäkringsärenden (IFD) som rör smartphones gällande indikationerna som väcker misstänksamhet och de ekonomiska incitamenten för att initiera en utredning. Syftet med denna rapport var att identifiera användbara indikatorer för smartphonebedrägerier (SIFIs), samt att föreslå ytterligare förbättringar inom IFD på områden som datainsamling och investeringar i analytiska verktyg. En trianguleringsmetodik användes som inbegrep intervjuer med 12 praktiserare av IFD, deltagande på 2 internationella försäkringskonferenser och en literaturgenomgång av 20 stycken publicerade akademiska artiklar om försäkringsbedrägeri. Det primära resultatet av arbetet var en sammanställning av 51 olika SIFIs. Genom att kontrastera och jämföra resultaten från de tre undersökningsområdena kunde flera exempel på divergens mellan teori och praktik identifieras. Resultatet kan användas för att utöka och omvärdera befintliga uppsättningar av indikatorer och prioritera investeringar i nya analytiska verktyg. Denna rapport fokuserar på smartphoneskadorna, men delar av resultatet är potentiellt applicerbart även på andra vanliga tillhörigheter såsom datorer och läsplattor.

**Nyckelord:** försäkring, försäkringsbedrägeri, bedrägeridetektion, smartphone, indikatorer för försäkringsbedrägeri

# Acknowledgments

This report has been produced at the division of Innovation Engineering of the department of Design Sciences at the Faculty of Engineering at Lund University during the spring semester of 2019. The authors have collaborated with the insurance start-up Hedvig AB in Stockholm, whose representatives have been supportive in guiding the work. As part of the data collection, two conferences have been attended: *Insurance Innovators Counter Fraud 2019* in London, United Kingdom (18th of March 2019) and *FT Live Insurance Innovation Summit 2019* in New York City, USA (11th of April 2019). We want to extend our sincere gratitude toward Emil Åkesson, Lars Bengtsson, Carl-Johan Asplund, John Ardelius and everyone else who have contributed towards making this master's thesis possible.

Lund, June 2019

Björn Angtoft and Olof Hallqvist

# Table of contents

1 Introduction .....	8
2 Background .....	9
2.1 Definition and scale of insurance fraud.....	9
2.2 Current fraud investigation methods and its issues .....	10
2.3 Smartphone insurance fraud.....	12
2.4 Purpose and research question.....	13
3 Method.....	15
3.1 High-level approach .....	15
3.2 Practitioners.....	16
3.3 International insurance experts.....	17
3.4 Interviews round 2.....	19
3.5 Literature review .....	22
4 Results .....	25
4.1 Compilation of SIFI.....	25
4.2 Elaboration of SIFI compilation.....	30
4.3 Factors improving IFD capacity.....	47
4.4 Analysis of SIFI.....	51
5 Discussion and conclusions.....	56
5.1 Discussion of results.....	56
5.2 Conclusions .....	59
References .....	60
Appendix A Interviews .....	63

# 1 Introduction

The smartphone is one of the most abundant and frequently used possession among modern consumers. 90 % of Swedes over the age of 12 own a smartphone (Internetstiftelsen, 2018). Just like TVs, laptops, jewellery and other possessions that consumers care about, smartphones are typically covered via one's household insurance policy. Reportedly, smartphones are claimed for in upwards of 50 % of household insurance claims (Hedvig, 2019). How many of these claims that are fraudulent is not possible to know for certain, but some insurers have estimated that it amounts to 40 % of all smartphone claims (Gray, 2012).

Since smartphone claims are typically of relatively low value, it can be difficult to motivate assigning them full-scale fraud investigations, meaning fraudulent claims can easily pass undetected. This calls for tools and procedures that make the detection process more cost-efficient. A commonly used approach in IFD is to identify a set of fraud indicators that can be used to support the manual work by adjusters and investigators, or as input variables for automated processes using data mining algorithms to identify suspicious claims. With this report aim to identify a diversified set of fraud indicators viable for smartphone IFD, referenced throughout the report as SIFIs.

Fraud can occur in many different parts of the interaction between the policy holder and the insurance company, not just during claiming. However, this report is concerned only with fraudulent claims, specifically concerning smartphones covered by household insurance.

The report is made in collaboration with the Swedish insurance start-up Hedvig AB. The scope and level of analysis has been set accordingly, with Sweden as the focal point and sources primarily from Europe and the U.S as reference points.



## 2 Background

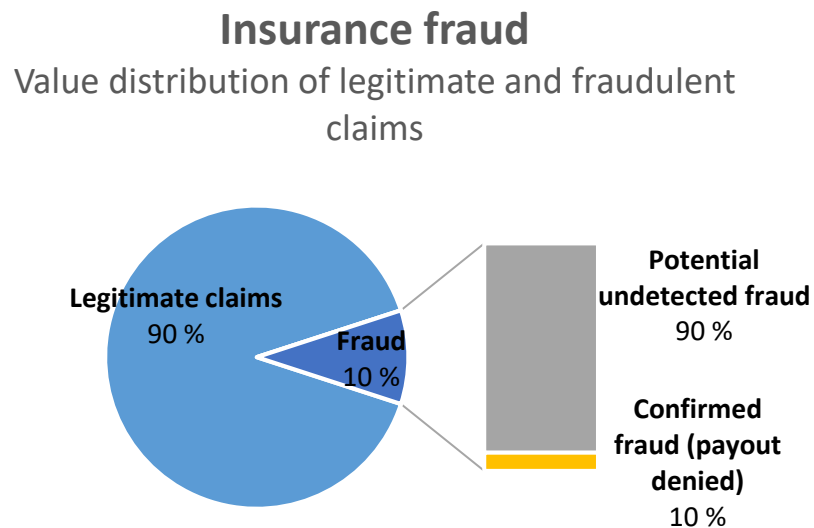
### 2.1 Definition and scale of insurance fraud

The definition of fraud that is used throughout this report comes from (Viaene & Dedene, 2004). In short, the definition of insurance fraud states that for an insurance claim to be considered fraudulent, the presence of at least the following are required:

- Material misrepresentation (in the form of concealment, falsification or lie)
- Intent to deceive
- Aim of gaining an unauthorized benefit.

In addition to this general definition, there are various types of insurance fraud. For example, fraud can be committed by a new customer submitting false details in the insurance application, or by an existing policy holder making a deceitful insurance claim in order to receive an illegitimate compensation. As previously mentioned, this report only deals with fraudulent insurance claims, specifically concerning smartphones covered by household insurance. It is also helpful to distinguish between soft fraud and hard fraud. Soft fraud occurs when a claimant exaggerates the damage of an otherwise legitimate claim to receive a higher payment than entitled to, which is a known tactic to compensate for the deductible. Hard fraud is made with the sole purpose of swindling the insurance company and means that the claim itself is fabricated by the claimant or orchestrated by several colluding parts. The definitions are gathered from (Viaene & Dedene, 2004). Svensk försäkring and Larmtjänst are two industry associations that support the Swedish insurance industry by e.g. producing statistics and spreading knowledge about new trends in frauds. Since fraud is an act of deceit and a crime which can be difficult to prove, grasping the full scale of the problem is difficult. Larmtjänst (2019) reports that 3 million insurance claims were made in Sweden in 2018, resulting in ~60 billion SEK in compensation being paid out to policy holders. 7462 claims (2 ‰) were investigated for fraud, leading to 480 million SEK in pay-out denials. These cases are known as “confirmed fraud”, but the value of the actual amount of fraudulent insurance claims remains unknown. In 1994, it was proposed that 5 - 10 percent (of all compensation paid out) was a reasonable estimation for the total value of insurance fraud in Sweden (Persson & Bongenhielm, 1998). This figure is still widely used by practitioners today and is also referred to by Larmtjänst (2019) when estimating the annual scale of insurance fraud in Sweden to a total of 3 - 6 billion SEK. Most cases of suspected insurance fraud in Sweden occur within home

insurance (51 %), with motor insurance claims coming in second at 40 % of claims that get denied after investigation (Larntjänst, 2019).



**Figure 2.1 Value of the potentially undetected fraud to the confirmed. The figure has been developed using data from (Larntjänst, 2019)**

## 2.2 Current fraud investigation methods and its issues

In Sweden, it is the responsibility of the insurance companies to investigate insurance fraud due to the insurance policy being under civil law, which means that the parties included in the contract themselves must prove that they are compliant with the contract. The purpose of the investigation is therefore not primarily to report suspected fraudsters to the police, but rather to deny them payment for invalid or false claims (Korsell, et al., 2015).

When a damage occurs and the policy holder makes a claim, he or she will file a claim either by a telephone call, a web form or, as introduced by recent entrants to the insurance market, via text or voice with a chatbot. A description of the investigation process was included in (Korsell, et al., 2015), focusing on two main control functions: the claims adjuster and the investigation unit (or insurance inspector). Below follows a short summary of the general investigation process as described in (Korsell, et al., 2015).

The claims adjuster makes the first assessment of the claim to see if, and to what extent, it is covered in the policy. If the claim is legitimate and trustworthy, the adjuster pays out the amount that the policy specifies. If the adjuster has good reason

to believe the claim is fraudulent, the claim will be denied and in some cases the policy will be terminated. It is explained how the control functions have a set of informal and formal criteria respective when assessing claims for fraud. The informal criteria are judgements based and the formal criteria concerns hard rules for decision making. The criteria cover e.g. how the claimant acts in communication with the adjuster, the claims history and the customer type.

The adjuster is judging the claimant's behaviour when reporting a claim on:

- **Preciseness:** If the claimant is very detailed in the descriptions, this signals that the claim report was made up in advance.
- **Ignorance:** If the claimant cannot describe what has happened or what is missing or if more and more items are added to the claim, this signals soft fraud.
- **Stress:** If the claimant seems stressed, this signals fright of getting caught.
- **Pervasiveness:** If the claimant seems very eager to reach a final decision of whether the claim will be paid out or not, this can signal that the claimant wants to rush the process so that the adjuster might miss to check something.

Further informal judgements are described in (Korsell, et al., 2015), which follow below. The claimant is judged on *claims history*, which is described as the main indicator of a fraudulent person. If there is some suspicion of fraud, the adjuster will look the claimant up in GSR, where a brief description of all adjusted claims of all claimants in Sweden are stored; many claims at several different insurers can indicate a fraudulent claimant. The adjuster will also check the personal finances of the claimant to assess whether there is economy for buying expensive items or if there is much debts. A *plausibility assessment* is made based on the claim and the credibility of the claimant. On top of that an *ethical judgement* is made to morally grade the claim and what the company risks are in case of an incorrect payment; smaller claims are thought not to harm the company as much and hence are less thoroughly assessed. The *insurance policy* is examined, looking at the time between signing the policy and filing the claim, and how much premium has been paid by the policy holder; it is generally viewed suspicious with claims reported shortly after the policy is written. (Korsell, et al., 2015) describes how the adjuster will, if working for a major insurance company, forward suspicious claims to fraud investigators. These, in turn, have their own set of informal criteria. *Well documented cases* are shown more interest from the investigators, which can include e.g. receipts or pictures that can be technically analysed for falsification or manipulation. The *claimant's age* is regarded as significant as most detected fraudsters are younger males. The investigators will also try to meet the claimant in person to judge the willingness to complete the process and if the story seems reliable.

One of the central conclusions from (Korsell, et al., 2015) is that the warning flags used by the insurance companies, such as claims history and atypical behaviour when reporting, has not been chosen due to what characterizes a fraudulent person.

Rather, the reason they are used is that they are easily communicated within the insurance companies and since they simplify the allocation of work assignments. Many parameters that are examined are examples of selection biases, which evolve from finding something where you look. For example, as younger males are in the majority of the confirmed fraudsters, the adjuster might scrutinize younger males for every type of claim, which would increase the chances of detecting more frauds from younger males. Further, *Intuition* is one of the most prevalent reasons for deeper investigation. Adjusters with long experience allegedly develop a so-called “police gaze”, enabling them to identify anomalies in a claim. This means that the insurers are dependent on the experience of the single adjuster to detect most frauds. Hence, the knowledge of fraud detection lies not within the organizations, but within the individual employees, and thus disappear when these knowledgeable employees stop working at the company. A big issue is that the large focus on historic claims behaviour has the consequence that the claimants look “normal” for years before the insurers suspect anything.

Relying on *intuition* is recognized also by the swiss insurance company Zurich. In a reference guide to claims investigation, the questions encouraged to be examined by the adjuster include if the claimant has *financial problems*, is too *aggressive* or *nice*, is *too familiar with the insurance procedure*, *readily accepts reduced claim* or has a *history of multiple claims*. On top of this, the adjuster is foremost recommended to trust the gut feeling: “*Do you feel something is not right with the claim? Do you believe what you are being told?*”. (Zurich Municipal, u.d.)

## 2.3 Smartphone insurance fraud

*“...close examination of internal components can show that a phone supposedly dropped down a lavatory – one of the commonest alibis – was in fact damaged earlier than alleged”*  
*“People say they drove over it, but when you look at it you see there are hammer marks” (Gray, 2012)*

Although there are special types of insurance policies tailored specifically to smartphones, most people are covered via their home insurance. In Sweden, 90 % of the population owns a mobile phone, (*Källa: Svenskarna och internet 2018*). 97 % of people living in Sweden are covered via home insurance (Svensk Försäkring, 2019). The public data available do not specify the amount of smartphone claims being made in Sweden. However, smartphone claims are often included under what is known as *allriskförsäkring*, a subcategory of home insurance. *Allriskförsäkring* covers e.g. damaged smartphones, cameras or lost wallets. In total, the claims categorized as “allrisk” are by far the most common type of claim. In 2017, some 305 000 allrisk claims were made, amounting to 1 billion SEK paid out in total. It should be noted that theft is categorized separately. As a category it amounts to

significantly fewer claims, but about the same amount as for allrisk claims are paid out in total (Svensk Försäkring, 2019). The share of allrisk and theft claims that are smartphone claims remain unknown. However, the insurance company Hedvig report that as much as 50% of their processed claims are smartphone claims. How many of these claims that involve fraud remain unknown, and as of today the company's database does not include enough data to support meaningful estimates.

There are mainly three sorts of smartphone claims: *theft*, *loss* or *damage*. As mentioned, there are no statistics available from Svensk Försäkring on discrete possession level, but here is reason to believe it is a growing issue. According to the U.K. based fraud prevention organisation Cifas, the number of fraudulent household insurance claims increased by 52% between 2017-2018 (Cifas, 2019). In 2012, the British insurance company Assurant reported to have observed a 100-fold increase in fraudulent iPhone claims in two years' time (Assurant, 2012).

## 2.4 Purpose and research question

Lost and damaged smartphones account for up to 50 % of insurance claims at Hedvig AB and amount to a significant share of the aggregated compensation paid out by the company and occupy many labour hours. Payments to fraudsters cause higher premiums for honest customers and by increase the number of detected smartphone frauds, premiums can decrease. Smartphone claims are often characterized by low value relative to other types of insurance claims, why full-scale fraud investigations can typically not be economically motivated, and many suspicious claims are instead readily paid. Currently, intuition is often the base for claims adjusters to choose to investigate a claim further, which leaves space for personal bias to govern the decisions and makes the assessment heavily dependent on individual's knowledge rather than formalized data criteria. Further, there is a high pressure on the claims adjuster to assess claims quickly in the pursuit of satisfying customer expectations.

The selection of claims to investigate needs to be preceded by a correct assessment of the plausibility of fraud based on correct assumptions of the fraudster profile as well as the possibility of the investigation to reach a conclusion in order to use resources efficiently. It is thus critical to obtain and understand the kind of data which is indicative of smartphone insurance fraud as well as the level of evidence adequate to deny payment. Producing a dataset of indicators of smartphone fraud would increase the ability of the insurer to make correct decisions. It would also make the knowledge part of the organization rather than dependent on the individual adjusters and investigators, increasing the long-term capacity of smartphone IFD. Improving the capacity of quicker assessments would decrease the cost of investigation and could increase the number of detected frauds. The purpose of this master's thesis is to improve the smartphone IFD performance by producing a

dataset of indicators and suggesting key areas for increasing the smartphone IFD capacity. The result could help the practitioners to reprioritise their efforts when assessing claims in the future. This is summarized with the following research questions:

- Primary research question: *What data detects smartphone insurance fraud?*
- Secondary research question: *How can the capacity of smartphone insurance fraud detection be improved?*

# 3 Method

## 3.1 High-level approach

To fulfil the purpose and answer the research questions, three types of information sources have been approached:

1. *Swedish and international practitioners from the insurance industry;*
2. *Internationally acknowledged experts within IFD and general insurance;*
3. *Published research on IFD in the form of academic literature.*

The practitioners were approached in order to describe the current methods used in the assessment of smartphone claims, fraud detection and to identify associated flaws. The IFD experts were approached to point to modern best-practices for successful IFD procedures. To obtain evidence-based fraud indicators for smartphone IFD, published academic research was also included to obtain research data. Employing different methods or data sources is common practice for increasing validity of the results and is often called triangulation (Bryman & Bell, 2015). The practitioners were interviewed, while the experts' views were derived from seminars at an international industry conference. The literature was studied in the context of a literature review.



**Figure 3.1 Triangulation by consulting different sources of information is one way to increase validity of results.**

## 3.2 Practitioners

Practitioners were approached on two occasions using interviews. The purpose of the first interview round was to describe different aspects of IFD on a higher level in terms of general practices in fraud assessment, motives for current investigation procedure and identifying possible flawed areas. The second round of interviews focused on smartphone fraud in particular and what measures were used against this type of fraud specifically. All data were kept in folders separating them by round and whether they were Swedish or international respondents.

### 3.2.1 Interviews round 1

For the interviews in round 1, one fundamental interview guide was used, consisting of questions on the topic of the overall issues of insurance fraud. The interviews were conducted using a semi-structured approach, leaving space for the practitioners to elaborate on subjects tied to their respective area of expertise. As recommended by Bryman and Bell (2015), the guide accommodated some elasticity, allowing for questions to be reformulated depending on the participant. The questions covered industry practices, legal and ethical considerations, investigation procedures, different approaches to claims processing, fraud indicative data and trends within insurance fraud.

The following procedure was used when collecting data from interviews of round 1:

1. Snowball sampling
2. Interviewing
3. Transcribing
4. Coding transcript to concepts
5. Comparing with existing concepts

Step 1 and 2 were performed until the final sample of participants had been reached. Steps 3-5 were then performed for every recorded interview. *Snowball sampling* is a technique for choosing participants to include in research and can be described as follows: a first seed is selected and interviewed, after which the participants are asked if they can recommend 2-3 others to interviewed, who will be in turn interviewed and asked for further recommendations (Goodman, 1961). The reason for employing the technique was partly to be flexible in the choice of whom to approach, while still ensuring that the participants were of interest to the research, and also to get a wide variety of practitioners in the sample. The first seed was a practitioner working as Head of Claims, referred to as S1. From there, we got in touch with a private investigator, S3, and so on. For a list of all individuals interviewed as well as the interview guide itself, please refer to Appendix 1. The



participants have been anonymized in the report and are referred to as an alias consisting of a letter and a number. The aliases are also found in Appendix 1.

The interviews were then transcribed and coded. Coding is the process of labelling theoretically relevant data, after breaking it down to smaller constituents. coding was continuously reevaluated and reformulated as new patterns emerged. Two ways to code were employed, inspired by the coding used for grounded theory according to Bryman and Bell (2015):

1. *Open coding* is the process of categorizing data into concepts by describing the data in slightly more general terms than presented. The open coding generates granular codes, which then are used to create more and more general concepts by grouping elements of the same subject under a common label;
2. *Axial coding* is a procedure where the data is reassembled after the open coding, to form new connections between concepts;

Using this approach, general subject of IFD were identified and general suggestions for improvements gathered. The result from this round was used to enable the level of specificity in the second round of interviews which focused on smartphone fraud detection. All interviews in round 1 were recorded and transcribed exactly as the participants expressed themselves. A memo containing the thought process of how the codes were constructed in the analysis of round 1 was kept throughout that part of the research to be able to see the development of our findings.

### 3.3 International insurance experts

International IFD experts and general insurance experts were included in the report to describe the state-of-the-art knowledge in the industry. The data gathering in this part included attending two industry conferences, which covered topics such as transformative innovations, customer relationship, big data management and advanced analytics to prevent fraud. We have considered the speakers at these conferences to be *experts* in this report, as they have been invited by the organizers due to being at the forefront of their fields of expertise. Attending the conferences was also a way to access samples from a highly capable population within IFD at once, as the attendee list could be used to target interview objects.

Name	Organizer	Date
<b>Counter Fraud 2019</b>	Insurance Innovators	2019-03-19
<b>Insurance Innovation Summit</b>	Financial Times Live	2019-04-20

**Table 3.2 Attended conferences**

### 3.3.1 Counter fraud 2019, London (UK)

At the *Counter Fraud 2019* conference 19<sup>th</sup> of March 2019, 204 people attended representing 177 start-ups, regulators, third-party solution providers and incumbents within P&C, who all had the objective of developing new fraud detection methods. Mainly European companies were present. The speakers held positions ranging from *associate director of fraud, head of fraud intelligence and strategic development, principal data scientist, head of fraud* and *head of P&C analytics*. The perspectives of represented organizations covered the whole value chain of insurance in aspects of fraud prevention and detection, data management, regulation and enforcement in Europe, which made this conference a valuable data source for this report. This conference was attended to gather suggestions for innovative general strategies for increasing the fraud detection capabilities, to attain considerations regarding data privacy and regulation matters and collect SIFI that could be included in our results.

Two seminar series were held in parallel at the conference; one focusing on *Customer Experience* and one focusing on *AI applications for insurance fraud*, of which the latter was attended. This was chosen due to the focus on data and analytics, which would be more relevant to the report. The seminars consisted of lectures and panel discussions and the insights we gathered were taken down as notes and coded to the subjects of most interest to the purpose of this report. The titles of the seminars can be seen in table 3.3 below<sup>1</sup>.

---

<sup>1</sup> For a more elaborated agenda, please refer to the organiser's website <https://marketforcelive.com/insurance-innovators/events/counter-fraud/#agenda>

SESSION	TITLE	SUBJECTS
1	A tech revolution in counter fraud – grasping the opportunities	<ul style="list-style-type: none"> <li>• Rejuvenating the fight against fraud</li> <li>• Tackling new customer demands with agile leadership</li> <li>• Leveraging next generation data: the evolution of fraud processing and fraud prevention</li> <li>• Optimising new tech in counter fraud: a holistic approach</li> </ul>
2	AI and advanced analytics: new weapons in the fight against fraud	<ul style="list-style-type: none"> <li>• Exploring the transformative potential of AI in counter fraud</li> <li>• Harnessing advanced analytics to defeat crime</li> <li>• Optimising the use of AI and advanced analytics in counter fraud: grasping the opportunities</li> </ul>
3	Fighting fraud in the new data landscape	<ul style="list-style-type: none"> <li>• Understanding the potential of new data sources</li> <li>• Optimising data strategy and management</li> </ul>
4	Future trends in fraud	<ul style="list-style-type: none"> <li>• Motor</li> <li>• Cyber</li> <li>• On-demand insurance</li> <li>• Insights from beyond insurance: broadening the conversation</li> </ul>

**Table 3.3 Sessions of the AI applications for insurance fraud**

### 3.4 Interviews round 2

The purpose of the second round of interviews was to describe current practice for detecting smartphone insurance fraud, including what data is used. A new interview guide was developed for this purpose based on the results from the first round of interviews and questions were formed to allow for description of fraud tactics specific to smartphone claims. For the second round, several of the practitioners from the first round were approached again, but participants of round 1 that did not

work with fraud detection in practice were excluded. This time the sample was expanded, again using the snowball sampling technique to include professional Swedish IFD practitioners.

Further, an attendee list from the Insurance Innovators conference in London, Counter Fraud 2019, was used to expand the sample. Out of the 204 attendees, 104 were selected to be included on the basis of their work title, which needed to show that the person worked with IFD on a professional level. Five of these accepted to participate. A semi-structured interview guide was used to leave space for elaboration of the replies and related follow-up questions. To be able to compare the replies better, less elasticity was allowed regarding the main questions than in interview round 1. The first question regarded the estimation of the share of fraudulent claims of the total amount of smartphone claims filed to them. This was posed to see how diverse the views were among the practitioners of the urgency of the problem and also functioned as a good starting point of the interview. The remaining questions were posed to gather what data is currently used for smartphone IFD and illuminating the approaches for detection from different angles. The questions were fairly open in the way they were posed to not infer ideas from the interviewer on the respondents. An exception from this was the formulation of the second question, where some data was suggested. The reason for this was that those data points had been described in the report from (Korsell, et al., 2015) to be general, which was validated in the responses from interviews of round 1, and could therefore be used as a way in which we showed the respondents that we were knowledgeable of the process to some degree. The interviews were conducted over the phone and recorded, but not transcribed. Instead notes were taken down throughout the interview of key findings and to ensure validity of the notes, participants were consulted after the interview and asked for approval or if they would have added, changed or removed any of the notes. This is called response validation and is employed to increase the dependability of the results (Bryman & Bell, 2015). The quotes were then compared on each question between the responses to see if there was consensus or if opposite views were exposed. The guide is found in Appendix 1.

### **3.4.1 Insurance Innovation Summit 2019, New York City (US)**

FT Live is the global conference and event division at the Financial Times Group with worldwide arrangements in a wide variety of industry sectors. The Insurance Innovation Summit 2019 in New York was attended by companies within property & casualty insurance, life insurance, reinsurance, credit rating, tech consulting and technology supplying. The speakers held positions such as CEO, CFO, CIO, CTO, CMO and other senior-level executives.

The conference was chosen as close to a complete spectrum of insurance sectors was represented with speakers on executive level. Although there was no particular

focus on fraud detection, the subject was tangent in several seminars. On the main stage, different aspects of innovation in the insurance industry was the subject. During a “break-out session”, two talks were held simultaneously about ecosystems and customer experience of which the latter was attended. The form of the seminars was lecture or panel discussion, an oversight of the sessions can be viewed in the table below. Insights that were of value to this report were taken down as notes and coded to the subject that was of most interest to us. This mainly contributed to the question of how to improve IFD capabilities. Table 3.4 below describes the agenda<sup>2</sup>:

SESSION	TITLE	SUBJECT
1	Separating the signal from the noise	Broad trends affecting the industry
2	A better way of doing business in the digital age	Digitalizing the organization
3	Profound cultural change and a new skills agenda	Required skills in the workforce
4	Shifting the business model to prevention	Using telemetrics to let policy holders affect their premiums
5	Financial wellness: what will motivate people to take action?	Engaging employees in their personal finances
6	Regulatory insights for the path ahead	Complying to financial regulations
7	How to invest in frictionless customer experience	Digitizing to meet customer expectations
8	View from the front lines	Prospects of insurtech companies
9	Riding the wave of artificial intelligence	Opportunities for continued innovation
10	The Holy Grail of big data	Regulatory and societal challenges of big data acquisition

**Table 3.4 Agenda of the FT Live Insurance Innovations Summit 2019**

---

<sup>2</sup> For a more elaborated agenda, please refer to the organiser’s website <https://live.ft.com/Events/2019/FT-Insurance-Innovation-Summit-2019>

### 3.5 Literature review

The literature review was the final step of the triangulation process. The review was performed in a systematic way with the use of a strict study protocol. However, the approach should not be mistaken for that of a comprehensive systematic review, typically employed in larger studies. The purpose of the review was to compile results from a variety of published scientific articles that would be relevant specifically for the detection of smartphone insurance fraud. The study protocol details the method of analysis; how searching was made, how the articles were assessed and what inclusion/exclusion criteria were used.

In addition to database searching, snowballing was used to identify relevant articles. The snowballing method is recommended by e.g. (Jane Webster, 2002) and (Wohlin, 2014). (Jane Webster, 2002) propose snowballing as the main method to find relevant literature and recommends both backward snowballing and forward snowballing. Backward snowballing means using the reference list of screened articles to identify new articles. Forward snowballing means identifying new articles by searching for articles citing the screened article. The snowballing approach requires a starting set of articles, here named TIER1. The articles belonging to TIER1 were identified using the search engine Google Scholar. The choice of Google Scholar was made in order to mitigate selection bias that favours certain publishers or universities (Wohlin, 2016). It is not evident what criteria should be fulfilled before the starting set (TIER1) can be considered complete. (Wohlin, 2014) recommends that the starting set should be diverse with regards to publishers, years and authors, and that it should be drawn from searches using keywords from the research questions as well as different wording and synonyms. Because of time constraints, it was expected that more papers would surface than there would time to assess. In this situation, (Wohlin, 2014) suggests settling with a number of relevant and highly cited articles. The study protocol (3.5.1) describes this procedure in more detail. Once TIER1 had been established, snowballing commenced. This included both backward and forward snowballing. Papers that surfaced from snowballing were tested using TIER2 inclusion criteria. Should a paper qualify for TIER2, it was screened similarly to the articles belonging to TIER1. 21 articles were screened before knowledge saturation was considered to have been reached. TIER2 represents all articles that surfaced from snowballing and that qualified from screening.

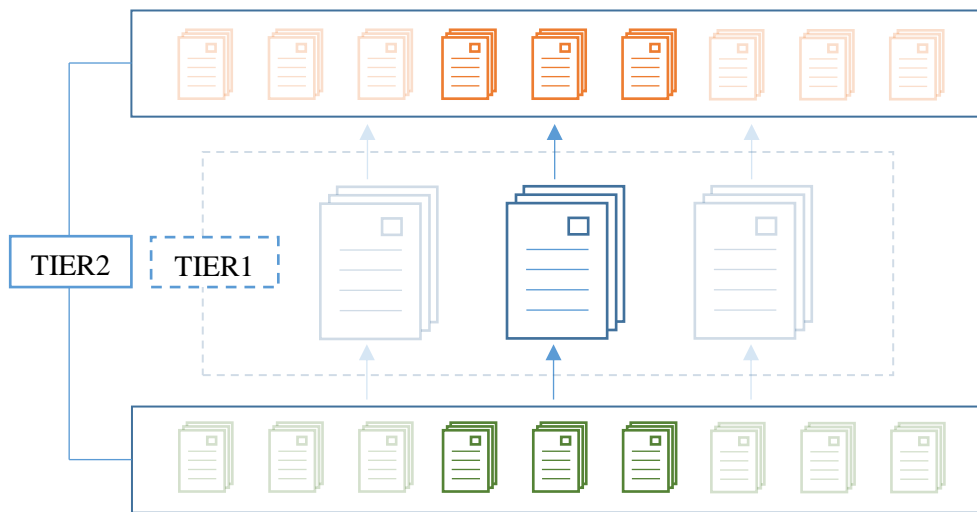


Figure 3.5 Depiction of literature review

### 3.5.1 Study protocol

The review process can be broken down into four steps.

1. Searching
2. Assessment of article abstract
3. Screening
4. Snowballing

#### 3.5.1.1 TIER1

The database Google Scholar was used to search for candidate articles for TIER1. Searching was conducted using the following explicit search strings: “Insurance fraud detection”; “Home insurance fraud detection”; “Home insurance fraud”; “Household insurance fraud detection”; “Household insurance fraud”; “Mobile insurance fraud detection”; “Mobile insurance fraud”; “Mobile phone insurance fraud detection”; “Mobile phone insurance fraud”; “Phone insurance fraud”; “Smartphone insurance fraud detection”; “Smartphone insurance fraud”.

Only articles published within the past 20 years were included, excluding any article published before 1999. Articles that did not include any of the keywords in their titles or subtitles were excluded. A first assessment was made using the information available in the article abstract. If the abstract was considered relevant to our research, the article qualified for the screening phase. The screening phase served to identify text segments on the topics of; 1) insurance fraud indicators and 2) information valuable for generating new smartphone insurance fraud indicators. All articles that did include such text segments qualified for TIER1.

### 3.5.1.2 *TIER2*

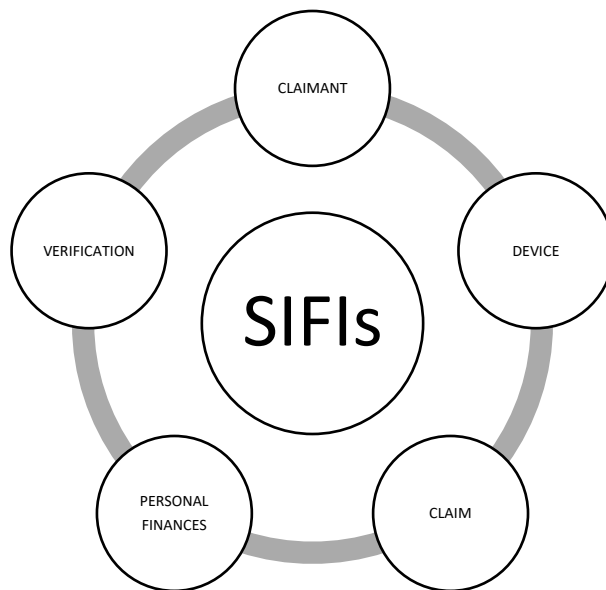
First, backward snowballing was performed on all text segments identified as relevant in the articles belonging to TIER1. If any of these segments contained a reference to a new article, that new article was screened for text segments including one of the two topics previously specified. Once this step had been performed for all relevant text segments of a TIER1 article, forward snowballing commenced, using the same screening approach. Articles that surfaced from the forward snowballing only qualified for TIER2 if they were published within the last 3 years (2016-2019), had at least 2 citations, and included text segments on at least one of the two subjects previously specified. This approach allowed access to recent articles that could be relevant to our research while at the same time making sure we did not spend too much time screening hundreds of articles. The key findings from the text segments identified as relevant from TIER1 and TIER2 were then synthesized (4.3) in order to extract their most relevant contributions to our research.



# 4 Results

## 4.1 Compilation of SIFI

Five data categories were defined in order to get an overview of data sources indicative of smartphone insurance fraud; claimant characteristics, financial situation, device data, claim data and verification data. The reason for dividing data into distinct categories is to make the data more understandable and easier to analyse. In addition to this categorization, data can be either “structured” or “unstructured”. The degree of structure is relevant since this determines what analytical methods can be used to process the data. In total, 51 fraud indicators were identified.



**Figure 4.4.1 The categories of SIFI**

#### 4.1.1.1 Claimant characteristics

CODE	FRAUD INDICATORS	CLARIFICATION
CC01	Time as customer <sup>c</sup>	Longer time as customer, lower risk of fraud
CC02	Number of previous insurance claims <sup>a, c</sup>	Fraudulent home insurance claims are often made by first-time claimants
CC03	Previous employee or temporary staff member <sup>c</sup>	Claimant has information that can be used for deceitful purposes
CC04	Claimant appears to be claims-wise <sup>a, c</sup>	Has proven to be highly indicative across different algorithms
CC05	Uncooperative <sup>a, c</sup>	If uncooperative, higher fraud risk
CC06	Socio-economic status of neighbourhood <sup>a</sup>	Lower status, higher economic incentives
CC07	E-mail and social media status <sup>b</sup>	“If the claimant has had a Gmail account for 5+ years and <200 LinkedIn contacts he/she is probably not a fraudster”
CC08	Difficult to contact - Avoids use of telephone/e-mail <sup>c</sup>	
CC09	Claimant has a criminal history <sup>c</sup>	
CC10	Readily accepts lower compensation <sup>c</sup>	Can indicate that getting a payout is more important than the size of it.
CC11	Claimant is eager that the claim is processed quickly <sup>a, c</sup>	
CC12	Claimant in age bracket 20-30 <sup>a, c</sup>	Most fraudsters are in this age bracket
CC13	Endurance of claimant <sup>a</sup>	Claimants with illegitimate claims tend to withdraw claim if repeatedly asked for more proof

<b>CC14</b>	Number of policy changes <sup>a, c</sup>	Many changes are indicative of systematic fraud behaviour.
<b>CC15</b>	Has claimant made similar claims with other insurers recently? <sup>a</sup>	Note: “recent” is not well-defined
<b>CC16</b>	Claimant is emotional and/or refers to children and family situation <sup>a</sup>	Signals that claimant attempts to invoke empathy
<b>CC17</b>	Network of claimant <sup>a, b</sup>	Claimant connected to other involved parties, such as the third-party repair shop
<b>CC18</b>	Claimant recently bought policy extension, such as <i>allrisk</i>	
<b>CC19</b>	Claimant checked the extent of coverage with insurer before claiming	This indicates that claimant is exploring the possibility of defrauding the insurer

**Figure 4.4.2 Source of indicator: a = practitioners, b = experts, c = literature**

#### 4.1.1.2 Financial situation

<b>CODE</b>	<b>FRAUD INDICATORS</b>	<b>CLARIFICATION</b>
<b>FS01</b>	Number of account overdrafts <sup>c</sup>	Data acquisition enabled by PSD2. Strained financial status increases incentives for fraud.
<b>FS02</b>	Unemployed <sup>c</sup>	Employer details obtained through Försäkringskassan (private employer) or Statens Pensionsverk (public employer). Strained financial status increases incentives for fraud
<b>FS03</b>	Low average account balance <sup>c</sup>	Strained financial status increases incentives for fraud. Data acquisition enabled by PSD2
<b>FS04</b>	Debts to Kronofogden <sup>a</sup>	Strained financial status increases incentives for fraud.

<b>FS05</b>	Mortgage status – amount and maturity <sup>a</sup>
<b>FS06</b>	Payment history with current insurer <sup>a</sup>
<b>FS07</b>	Mismatch between financial status & phone model Claimant with low financial status should probably not be expected to own the most expensive device.

**Figure 4.4.3 Source of indicator: a = practitioners, b = experts, c = literature**

#### 4.1.1.3 Device data

CODE	FRAUD INDICATORS	CLARIFICATION
<b>DD01</b>	Brand & Model <sup>c</sup>	Higher fraud incentive for a model with lower value depreciation
<b>DD02</b>	IMEI number <sup>a, c</sup>	Has the phone been previously claimed for at another insurer?
<b>DD03</b>	Time between purchase and claim <sup>a, c</sup>	
<b>DD04</b>	History of previous owners <sup>a</sup>	Several previous owners? Phone may have been used by colluding fraudsters.
<b>DD05</b>	Attempted selling object on site like e-bay prior to claim date <sup>a</sup>	
<b>DD06</b>	Value of the device <sup>a, c</sup>	Higher value, higher incentive to claim
<b>DD07</b>	Phone damage <sup>a</sup>	Does a technical examination yield results in line with story of claimant?
<b>DD08</b>	Phone blocked (or blacklisted) by mobile network operator <sup>a</sup>	
<b>DD09</b>	Same phone used for reporting the claim as the phone which is claimed for <sup>a</sup>	
<b>DD10</b>	Picture of phone packaging <sup>a</sup>	Could work as proof of ownership, if receipt is missing

<b>DD11</b>	Phone usage <sup>a</sup>	Mobile network operator can verify that the phone is used by the claimant
-------------	--------------------------	---

**Figure 4.4.4 Source of indicator: a = practitioners, b = experts, c = literature**

#### 4.1.1.4 Claim data

CODE	FRAUD INDICATORS	CLARIFICATION
<b>CD01</b>	Month of claim <sup>c</sup>	Study found 2x fraudulent phone claims in September compared to February.
<b>CD02</b>	Claim made <1 year after policy was underwritten <sup>c</sup>	Binary indicator.
<b>CD03</b>	Time between underwriting and claim <sup>a, c</sup>	Non-binary indicator. The sooner the claim is made, the more suspicious it is.
<b>CD04</b>	Time between claim and release of new smartphone model <sup>a, c</sup>	Phone claims increase significantly when new models are released.
<b>CD05</b>	Time between incident and claim <sup>c</sup>	The more time that has passed since the incident, the more suspicious
<b>CD06</b>	Type of claim <sup>c</sup>	<i>Accidents</i> are more common among fraudulent household insurance claims than <i>theft</i>
<b>CD07</b>	Claim in connection to holiday	More fraudulent claims tend to occur during holidays.

**Figure 4.4.5 Source of indicator: a = practitioners, b = experts, c = literature**

#### 4.1.1.5 Verification Data

CODE	FRAUD INDICATORS	CLARIFICATION
<b>VD01</b>	No presence of witnesses <sup>a, c</sup>	
<b>VD02</b>	Plausibility of explanation for accident <sup>a, c</sup>	

<b>VD03</b>	Claimant not willing to provide sworn statement <sup>c</sup>	Sworn statement = oral/written assertion of facts stated under oath.
<b>VD04</b>	No police report filed in the case of theft <sup>a</sup>	If a claim concerns theft, a police report is usually demanded
<b>VD05</b>	Bank statement, receipts, tickets, Swish statements, pictures on social media <sup>a</sup>	To verify events in claim report
<b>VD06</b>	Pictures on social media <sup>a</sup>	To verify events in claim report
<b>VD07</b>	Documents have been tampered with <sup>a</sup>	

Figure 4.4.6 Source of indicator: a = practitioners, b = experts, c = literature

## 4.2 Elaboration of SIFI compilation

### 4.2.1 Findings from practitioner interviews

<i>Alias</i>	<i>Nation of employment</i>	<i>Profession</i>	<i>Round of interview</i>
<i>S1</i>	Sweden	Head of Claims	1, 2
<i>S2</i>	Sweden	Head of Claims Adjusting	1, 2
<i>S3</i>	Sweden	Private Insurance Fraud Investigator	1, 2
<i>S4</i>	Sweden	Head of Insurance Fraud Investigation	1, 2
<i>S5</i>	Sweden	CEO of industry association	1
<i>S6</i>	Sweden	Administrator at industry association	1

<i>S7</i>	Sweden	Head of Insurance Fraud Investigation	2
<i>S8</i>	Sweden	Head of Insurance Fraud Investigation	2
<i>U1</i>	UK	Head of fraud	2
<i>W1</i>	Switzerland	Data Scientist	2
<i>F1</i>	France	Data Scientist	2
<i>N1</i>	The Netherlands	Business developer at insurance fraud detection solutions supplier	1

**Table 4.7 Participants in interview rounds 1 and 2**

This section presents the key findings from the interviews conducted with the insurance fraud practitioners. The primary data source is the second interview round (round 2), as this focused specifically on fraudulent smartphone claims. The first interview round (round 1) had a broader scope, but in some cases relevant findings surfaced there as well. For example, round 1 touched upon the procedures of claims handling, which will be disclosed shortly. The participants will be referred to by their alias according to table 4.7.

Many cases of consensus were found with regards to how practitioners deemed fraud investigation should be conducted, and only a few areas of disagreement. The standard procedure for dealing with smartphone claims is to assess the report submitted by the claimant, screen the claimant's history of previous claims and e.g. ask for receipts or other proofs of ownership of the device. Using this information, the adjuster will make a first evaluation regarding of the likelihood for the claim to be fraudulent. Practitioners state that this judgement is heavily determined by what is found in the claimant's history of past claims. This is in line with the conclusions of the report from (Korsell, et al., 2015). Should suspicions arise at this stage, the adjuster will ask the claimant for more and more information until the adjuster can conclude that the claim is either legitimate or fraudulent. The claimant is obliged by law to provide evidence that beyond reasonable doubt support that the claim is legitimate. The principal strategy used by practitioners to reach a conclusion is to collect as much objective evidence as is attainable in order to validate the claimant's story. The amount of evidence adequate to deny payment is somewhat arbitrary, but what would stick in a court of law and the risk of bad publicity are two straining factors. Extending the adjustment process is a common method employed when the adjuster is suspicious but lacks evidence, to signal to the claimant that the

claim is thoroughly examined and to see if the claimant would do a “fraud walkaway”.

“The assessment can be thought of like a funnel, where you get more and more certain whether or not fraud can be dismissed or concluded” – S1

Smartphone claims are of relatively low value, meaning insurance companies find it difficult to economically motivate sending the claim to further assessment performed by fraud investigator. In one of the interviews, the interviewee disclosed that adjusters working at the company were not provided with the complete list of viable fraud indicators. The reason for this was apparently that the company has decided to keep the complete fraud detection system hidden from adjusters and accessible only to the fraud investigators.

#### 4.2.1.1 Claims report

*“When reporting a claim via our app on the phone, you first encounter a voice recorder and you record a message where you detail: what has happened; who is affected; and when it happened. Here, claimants will say just about anything. Someone might say ‘my phone is broken’, period, others are immensely detailed. Standard is “Hi, my phone broken. I dropped it on the ground. The screen is broken on the front and back. It still works. It happened last Sunday, and it affected me.”” – S2*

Above quote is an example of how a claim is reported, as explained by an adjuster. The claim report is normally filed either via a telephone call, via an online form, via telephone call, or via the recording of a voice message. If the adjuster can establish that the claim is covered for by the terms of the policy, the first fraud assessment of the claim will be based on the information provided here. In the claim report, there are many aspects to consider as it is an unstructured data source. Indicators identified here are e.g. *plausibility of story*<sup>VD02</sup> and whether the claimant is *claims-wise*<sup>CC04</sup>. Being claims-wise simply means that the claimant seems to be well-versed within insurance lingo, is unusually familiar with the routines of claims processing and the documents and additional information often requested by the adjuster. The assessment of VD02 is expressed by S1:

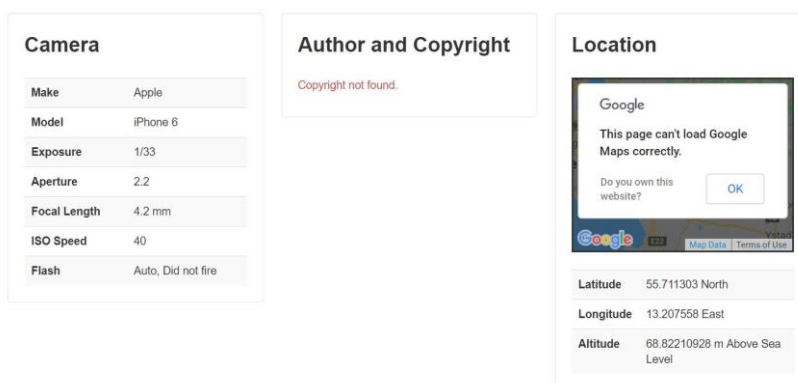
You will to some degree relate to your own experiences when assessing what to deem plausible. Fraudster stories often have a comical ending in comparison to a normal explanation. When making stuff up, you want to tell an unusual story, e.g. leaving the phone on the roof of the car and driving over it when hitting the breaks. Compare this to [the phone] falling out of the jacket when cycling or dropping it in the sea on a fishing trip, then the prior story is very odd and there are several strange aspects that require explanation.”



To verify a story, there are several data points to consider, e.g. witnesses <sup>VD01</sup>; documents such as receipts, tickets, bank or swish<sup>3</sup> statements <sup>VD05</sup>; or pictures on social media <sup>VD06</sup>. VD05 could be manipulated or tampered with in some sense <sup>VD07</sup>. Examination of such tampering can be performed using several tools; the exif-data extracted from pictures sent to the insurer show time and day when taken, if the image has been modified, GPS location, device maker and model etc. An example of the exif-information summary is seen in figure 4.8.

*"[...] you try finding markers in your customer database, adding rules to detect possible frauds which would otherwise pass undetected. Perhaps the same e-mail address was used in a former investigation, or the same car have been reported damaged." – S5*

Links between e-mails or home addresses among the insured collective is put forward by a representative from an industry organisation as examples of indicative data to examine, since there might be someone else that has committed fraud living at the same address, or the same phone has been used in separate claims by different claimants. If pictures of receipts and phone damages are stored by the insurer, it can be detected when the same pictures are sent in by another claimant on the claimed object. Checking the specific device being claimed for with other insurers using e.g. IMEI number can be done if there has been a similar claim in GSR in recent time, which provides a data source on the same topic. However, to perform network analysis regarding other aspects than the claim itself is not possible due to current data privacy regulation according to S7 <sup>CC19</sup>.



**Figure 4.8 Exif-data from picture taken with authors phone. (Screen shot taken by author of <http://metapicz.com/#landing>, 2019)**

<sup>3</sup> Swish is a Swedish payment service that enables quick transactions between phones

#### 4.2.1.2 Claims history

“An accelerating number of claims per policy is alarming. [Fraudsters] report claims a year after signing the policy, then six months go by, then three months and one month and so on. They get a feel for the system and try new approaches. When they are caught, they are terminated as customers. Then they move on to the next insurer and start over.” – S3

Along with the claims report and proof of ownership, historic claims behaviour is the most common data source to examine in IFD according to several interviewed practitioners. Both the insurer’s own database and GSR are used for this purpose. Often, it can be the claimed amount <sup>DD06</sup> that invokes investigation and lower value claims might not be investigated at all. However, if the claimant has a suspicious claims history, the claim is said to be investigated regardless of the amount, which is a good indicator of how much weight is put on this as a predictor. *Insurance company, policy type and date of claim* are reported to GSR at the point of claim and is available to view for all member companies in the format shown in figure 4.9.

“ It is a huge issue with people who claim in the first couple of months of the policy, it is an enormous number who do that.” - F1

Registrerade skador			
Personnr/Orgnr [REDACTED]			
Skadedatum	Försäkringsbolag	Skadenr	Handläggare
2017-09-25	22062 IF SKADEFÖRSÄKRING	177696447	AUTPRO
<b>Skadekoder</b> 0103 KOMBINERAT HEM, CYKEL			
Skadedatum	Försäkringsbolag	Skadenr	Handläggare
2016-03-01	22001 FOLKSAM SAK	[REDACTED] 160301	RN
<b>Skadekoder</b> 3100 OLYCKSFALL, GRUPP			
Skadedatum	Försäkringsbolag	Skadenr	Handläggare
2015-10-25	22062 IF SKADEFÖRSÄKRING	154468619	OLAPET
<b>Skadekoder</b> 0105 KOMBINERAT HEM, ÖVRIG STÖLD			
Skadedatum	Försäkringsbolag	Skadenr	Handläggare
2015-07-08	22001 FOLKSAM SAK	[REDACTED] 150708	3G
<b>Skadekoder</b> 3100 OLYCKSFALL, GRUPP			
Totalt 4 registrerade skador			

Figure 4.9 An instance of the information in GSR (photo taken by author of own statement, 2019)

The investigators describe that some individuals commit fraud over and over again, and therefore argues that the claims history is a very important indicator. The adjuster will also sometimes turn to other insurance companies and ask whether the

claimant in question has made any similar claims recently with that company<sup>CC17</sup>. Other actions the adjuster may call upon are; the time between underwriting and claim<sup>CD03</sup>; the total number of previous claims<sup>CC02</sup> and the number of times the claimant has changed insurer within the same kind of policy<sup>CC16</sup>. CC02 and CC17 can indicate that the claimant is a serial fraudster, especially if there has been a sudden acceleration of claims recently. CD03 can indicate that the damage had occurred before the policy was signed, in combination with CC17, that the claimant has attempted to defraud several insurers at once. This can be detected by asking around among other insurers to whom the claimant has reported similar claims, to see whether it is the same device and whether they were suspicious of the legitimacy of the claims. To be able to establish this, the IMEI number<sup>4 DD02</sup> is collected and compared.

*“In those cases, you will inspect the computer and then we check in GSR, the industry collective register: sometimes the customer has recently reported a claim including a computer at another insurer.[...] The investigator will then contact the investigator at the other insurer and ask for additional information ‘what kind of computer has been claimed for?’, if it turns out to be the same computer, this is enough to deny payment.” – S4*

#### 4.2.1.3 Proof of ownership

The final primary data which is acquired in the first assessment of the claim is the proof of ownership. This can be done in various ways, but the most common is to ask for receipts or pictures thereof, or in case the phone is bought second-hand, a bank statement or swish statement confirming the purchase. If none of the above can be provided, a picture of the phone box in which it was delivered, photos showing that the claimant is using the phone or a statement of phone usage<sup>DD11</sup> provided by the mobile network operator can sometimes suffice. There is always a risk that pictures and documents have been tampered with in some way<sup>VD07</sup>, for example by altering the purchase date of purchase to a more recent date, or by simply increasing price. The provided pictures or invoices can also have been downloaded from a search engine like Google.

“Googling ‘Fatura’ you will find thousands of Turkish invoices. Most often it is pure nonsense when you receive such an invoice as means for verifying damage. Typically, you download a fake invoice, insert your own name and file a claim saying that you rented a car in Turkey somewhere and have experienced a traffic accident. [Image software such as] Photoshop is also widely used to manipulate pictures of different kinds. Criminals are often lazy in my experience, they use the same invoice

---

<sup>4</sup> IMEI, an acronym for *International Mobile Equipment Identity*, is a unique number often printed on the battery or the back of the phone, used to identify valid mobile phone devices. It can be used for e.g. stopping a stolen phone from accessing the network in that country (IMEI info, 2019)

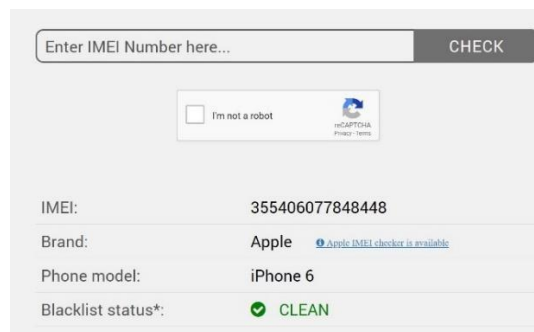
many times, trying at different insurers. It is then helpful to be in contact with the other insurers so that we quickly can be notified when new tricks appear.” – S3

#### 4.2.1.4 Theft, loss or damage

The three most commonly used reasons for submitting a smartphone insurance claim are damage, loss and theft, giving rise to three types of smartphone claims. The type determines the next data to look for after the first assessment has been made. If the claim is a theft, many respondents stated that they ask the claimant to report this to the police and provide the insurer with a copy of the police report<sup>VD04</sup>. The motive for this, however, is not obvious. One respondent, S2, said:

*“Police reports have lost much of their purpose. Previously, you had to go to the nearest police office, and if you have decided to commit fraud, you have to look a policeman in the eyes and flat out lie. Today you can file a police report online, which detracts the value of this.”*

According to this statement, demanding a police report is merely a way of testing how serious the claimant is with the claim. Another document which is often required by claimants to provide when a phone is out of possession is a block certificate<sup>DD08</sup> from the mobile network operator. This document ensures that neither the claimant, nor anyone else can use the phone for texting and making phone calls. The IMEI number can be used for a similar purpose, as the phone can be reported “blacklisted” if stolen. Figure 4.10 depicts how this can be examined. Another use of the IMEI number is to verify that the phone exists at all.



The screenshot shows a web interface for checking an IMEI number. At the top, there is a text input field labeled "Enter IMEI Number here..." and a "CHECK" button. Below the input field is a CAPTCHA area with the text "I'm not a robot" and a reCAPTCHA logo. The results are displayed in a table-like format:

IMEI:	355406077848448
Brand:	Apple <a href="#">Apple IMEI checker is available</a>
Phone model:	iPhone 6
Blacklist status*:	<span style="color: green;">✔ CLEAN</span>

**Figure 4.10 IMEI information (screen shot taken by author from <http://imeipro.info/>)**

If a smartphone is reported stolen or lost, one instance of fraud involves fabricating a series of events, sending in all proofs of ownership and usage but instead selling the device. When this is suspected, looking at online marketplaces, such as *Blocket*, *Tradera* or *E-bay*<sup>DD05</sup>, is a suggested approach.

*“We look to see if claimants posts items for sale at different forums. Facebook can be tough, since the claimant might be using aliases, and many have the same name etc. We look for ads put out from the same neighbourhood as the claimant lives.” – S1*

In many cases, smartphones are regarded as “*attractive to thieves*” by the terms and conditions of home insurance policies. This means that the claimant has a more extensive responsibility of supervising the device so that it is not stolen or lost due to careless behaviour. One example of this would be leaving the phone unattended at a café table while visiting the restroom. This would most likely not yield a compensation from the insurance company, as stated in the policy terms and conditions. Hence, many payments are denied because of this.

Smartphone that are reported damaged often involve a broken screen. In those cases, it was stated by respondents that an identified fraud modus is to sign an *allriskförsäkring*<sup>CC20</sup> (all risk insurance policy), after the damage has occurred to get the insurer to share their repair costs. Thus, the claim will look legitimate as all documents can be provided. However, in these cases the time between underwriting and first claim is reportedly often quite short<sup>CD03</sup>. An approach for countering this is taken by the company where respondent U1 I employed; they ask for pictures of the phone itself in the onboarding to conclude that the claimant is not trying to claim for a damage that occurred before he or she had bought the policy.

#### 4.2.1.5 Surge in claims in proximity to new iPhone releases<sup>CD04</sup>

As stated in the section 2.3, a surge of claims correlating to the release of new iPhone models has been reported and is known in the industry. However, in the interviews we find diverging views if this is still an issue that needs attention. S8 stated that their company informs the claimants in these periods that they are conducting extra careful assessments of incoming smartphone claims, which has proven to be a successful strategy for them:

*“Most fraudsters give up the claims when we inform them that we are conducting particularly thorough adjustments [in connection to an iPhone release]. We saw that 36 % of all smartphone claimants did not follow up on their claims when we started with that. It is a high enough share for us to discard the possibility that the main reason was an annoying adjustment process.”*

A few other respondents had an opposite understanding, here represented by S1:

*“These cycles do not appear anymore since the tech companies fail to create the same interest as they did before. Around Christmas time, we still see an influx of claims as people probably are low on money that time of the year. [...] When iPhone 4 and 5 was released, it was an enormous surge, but now people seem to keep their phones longer.” – S1*

Due to that, their company did not make any extra efforts around these presumed cycle peaks. One explanation for the decreased surge could be the one stated above, however S4 stated that they would not replace a lost or stolen phone with the latest release, but instead with one of same model, which would decrease the motive of “upgrading frauds”.

#### 4.2.1.6 *Fraudster characteristics*

The view of the fraudster profile is somewhat mixed when it comes to smartphone insurance frauds among the respondents. Almost all respondents believe that most of these frauds are committed by ordinary people and that it is probably equally many men and women. There is some disparity when it comes to age; some are confident that the average fraudster is between 20 and 30 years old, some believe it is closer to between 40 and 50 <sup>CC14</sup>. One proposal from S1 was that the fraudsters could be split between three groups:

- Young people between 20 and 30 who just want their phone repaired and deceitfully trying to get the insurer to pay for part of it.
- Young people between 20 and 30 who are trying to find ways to fund a lifestyle they cannot afford by making up claims with thefts or losses of smartphones.
- “Regular” insurance fraudsters, i.e. serial fraudsters who have a troublesome claims history and a dire financial situation with large debts they are unable to pay off with other means than with defrauded money. They try with the most expensive phone models.

Whether or not this description stacks up or not, no respondent was certain in their assumptions and almost in all cases left the caveat that more data was needed to properly assess this. This uncertainty of the perpetrator profile is interesting since one of the key indicators for selecting claims to further investigate is the number of previously submitted claims, an indicator that all respondents are very sure of. Other indicators revealing the claimant’s motives was found to be financial status, e.g. debts to Kronofogdemyndigheten (appr. enforcement authority) <sup>FS04</sup> and the payment history with the insurer <sup>FS06</sup>. Other data to reveal the financial status of the claimant was suggested by N1 as socio-economic status of the neighbourhood <sup>CC06</sup> and the mortgage status in terms of amount and time to maturity <sup>FS05</sup>. A suggestion from S8 regarding plausibility of the third group is that a person with a highly strained economy would perhaps not be able to afford a smartphone device worth upwards of 20 000 SEK, which would make it worthwhile to assess the personal finances, if those expensive models occur in a claim <sup>FS07</sup>.

”I might ask ‘does the claimant have a lot of debts?’ and if so, there is an incentive [for the claimant] to claim more than [being] entitled to. Often, they have a ton of debts, a gambling addiction, can’t take care of [their] personal finances and have no money left and then try to get money any way possible, and if that means making a dubious claim and getting paid then that is as good as it gets. [...] It correlates enormously with confirmed frauds in our data to have debts to Kronofogdemyndigheten, it is a strong correlation.” – S2

One aspect of the fraudster profiling regards doing risk assessment in the onboarding process. Not doing this might be a trade-off to worse assessments at point of claim, as argued by respondent N1. There are some predictors of an individual’s proneness

to fraud, or of the motives of defrauding the company, already at the point of onboarding, he argues. Knowing the risk in each of the policy holders can prepare the insurer for the level of thoroughness required when a claim is made. There is also the aspect of being able to generate new indicators, as there might be more fine-grained parameters, which can be examined when you have information about a person before they make a claim.

*“So it is a double edged sword in a way, where if you don’t have a good risk assessment solution in place you won’t be able to have a good fraud detection solution either because there’s a lot of unknowns or asymmetric information that enters your portfolio if you don’t have a good risk assessment solution in place. So just something to take into account as well. Detecting fraud is wonderful but you should also think of how to avoid it.” – N1*

Not doing a risk assessment in onboarding puts more pressure than needed on the adjuster at point of claim to evaluate the customer. Conducting a risk assessment in the onboarding and updating the information every interaction would decrease the amount of time that would need to be spent at the point of claim.

#### 4.2.1.7 Claimant behaviour in adjustment process

The claimant’s behaviour throughout the claims process is a recurring indicator that an adjuster or investigator will react to, which aligns well with the findings in (Korsell, et al., 2015) as stated in the background of this report. This can be summarized as eagerness of quick adjustment<sup>CC13</sup>, being emotional and referring to family situation or children to invoke empathy from adjuster<sup>CC18</sup> or providing very little information in their statements<sup>CC05</sup>.

*“One can wait out the claimant deliberately delaying the payment, ask for more documents, ask more questions, discuss with the person to see how they behave and often the fraudsters give up.” - F1*

If objective evidence of fraud is insufficient, but the gut feeling of fraud is very strong, prolonging the adjustment process is a common strategy among adjusters and investigators. In the compilation of SIFIs (4.1), this indicator is referred to as claimant endurance<sup>CC15</sup>. Reportedly, a deceitful claimant will often drop the claim after a while since they are too eager to wait for the payment, or because they believe they will expose themselves if they don’t. In the larger insurance companies, the adjuster will forward the case to an investigator, who normally contacts the claimant in order to assess how he or she behaves.

*“‘Fraud walk-away’ are common, we will ask for more and more evidence or proof of their story. If they are trying to deceive us, they might think ‘they’re on to me’ and they’ll often back away from the claim and cancel their policy.” – U1*

## 4.2.2 Findings from conference seminars

The conference *Insurance Innovators: Counter Fraud 2019* was divided into four main sessions, each containing a set of seminars with experts within insurance fraud lectures, talks, and panel discussions. Taking field notes from these seminars yielded data that can be found below.

During one seminar it was stated that unstructured data “...will eat structured data for breakfast” when it comes to fraud detection. What the speaker meant was that unstructured data contains far more information about both the claimant’s sentiment and the actual incident behind the claim. Unstructured data reportedly make up 80 percent of processed data, while structured data merely make up 20 percent.

### Examples of unstructured data

- Accident reports
- Police reports
- Witness statements
- Pictures & Videos
- E-mails & Letters
- Phone calls

**Figure 4.11 Examples of unstructured data gathered from IFD seminars (Hallqvist, 2019)**

A good example of unstructured data is the claim report itself that is provided by the claimant to the insurer. The text length in the claim and the level of detail in the claimant’s choice of wording has reportedly proven to be a strong fraud indicator.

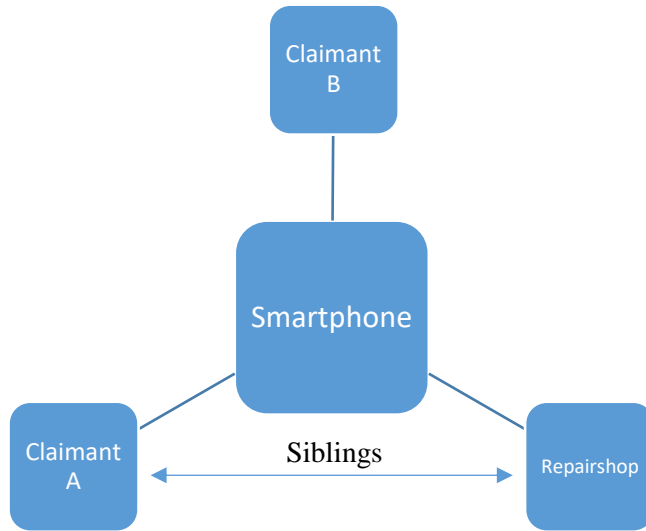
The lecture “A tech revolution in counter fraud – grasping the opportunities”, the speaker stated that if a customer meets the following criteria <sup>CC07</sup>, he or she is “...very unlikely to commit fraud”:

- Has 200+ LinkedIn contacts;
- Has an e-mail address that has been active 5+ years;
- Has a private Facebook and/or Twitter account

In addition to automated screenings of social media feeds to find evidence for injury fraud, the company also searches databases of e.g. marathon attendees to verify the story told by the claimant. If a suspicious event is detected, a screen dump is automatically taken, ensuring that the evidence will not disappear from the web.

An approach that has become increasingly popular within IFD is network analysis; mapping how claimants, 3<sup>rd</sup> party suppliers and involved items are connected. This coupled with geolocation analysis, via e.g. Google Maps, has proven to be an effective way of assessing e.g. suspicious motor claims. An example put forward by a data scientist working within IFD included a car accident where the involved parties were living on the same street, although the accident happened many miles away. Network analysis could also be used to see if involved parties are connected to a third party, e.g. a mechanic, via business relations. <sup>CC19</sup>

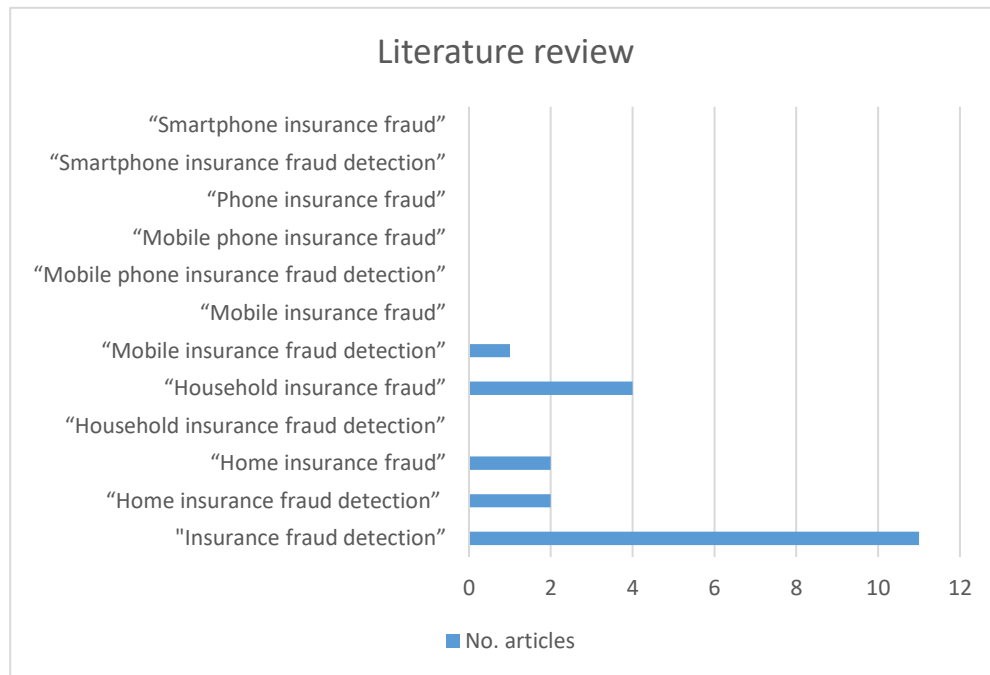




**Figure 4.12 Basic example of a network graph**

### 4.2.3 Findings from literature review

This section presents the key findings from the articles screened during the literature review, as well as examples of SIFI that were identified.



**Figure 4.13 Diagram of articles that surfaced from literature review**

Automobile insurance fraud detection seems to be the most well-understood subgroup of research, and has seen plenty of published articles since the 1990's. The majority of articles that surfaced from the literature review discuss automobile insurance fraud specifically. Although smartphones are categorized under household insurance policies in most cases, relevant fraud indicators were still identified from automobile IFD research. Household insurance fraud detection turned out to be not nearly as well-researched as automobile insurance, and not a single article was found concerning smartphone insurance fraud specifically. Only one published article, (M. Button, 2016), disclosed research findings on household insurance fraud on the level of discrete possession categories, where mobile phones were one category.

(C. Phua, 2010) performed an extensive study where 51 published papers on different areas of fraud detection were analysed. They found that researchers are often in agreement when it comes to what indicators to use for many different types of fraud detection, one of which is household insurance fraud. Indicators for home insurance fraud is generally based on 1) behavioural variables such as: the claimed amount; the number of previous claims; how long the claimant had been a customer,

as well as 2) financial variables such as: income level; number of account overdrafts; average account balance. In contrast, motor insurance fraud indicators are usually binary variables grouped into categories such as accident, claimant, driver, vehicle etc. (C. Phua, 2010)

Acquiring datasets of real-world insurance claims is a very difficult task due to the strict data policies of insurance companies (C. Phua, 2010). This explains why the same dataset has been used in the development of so many fraud detection algorithms for motor insurance (Patrick L. Brockett, 1998) (Herberg I. Weisberg, 1991) (S. Viaene, 2002) (Derrig, 2002). It should be noted that the datasets used in published research is often quite small which can question the significance of some published results. For example, the model trained by (Bentley, 2000) was developed using only 49 cases of confirmed fraud and 1000 non-fraudulent cases.

There is no established word for what we in this report refer to as insurance fraud indicators. Predictors, red flags (S. Viaene, 2002), attributes (C. Phua, 2010), features (Derrig, 2002), and indicators (T. Ormerod, 2003) are all synonyms used among researchers. This sometimes make it difficult to compare results before one has concluded that the words being used are synonyms.

By including non-binary indicators (S. Viaene, 2002) increased the predictive performance of their automobile IFD algorithms significantly, providing a strong argument for an approach which combines conventional binary indicators with non-binary indicators for IFD. However, it is not known how much each individual non-binary indicator, such as e.g. age of claimant, contributed to the increase in predictive performance.

Papers that propose new data mining algorithms for IFD use numerical and categorical fraud indicators, which can be derived from the standard structured insurance documents and external sources such as credit institutes. Naturally, the algorithms themselves lack the basic instincts of experienced adjusters and investigators, which can make them vulnerable against deceivers that adjust the standard structured data they provide in order to not get caught by the algorithms. (Y. Wang, 2018) argues that this means expert experience should always be employed in combination with quantitative methods to achieve the most reliable results in fraud detection.

(M. Button, 2016) analysed a data set containing over 30 000 cases of confirmed household insurance fraud. To our knowledge, this is the largest data set ever used in an analysis of confirmed fraudulent household insurance claims. Their research offers important insights regarding e.g. the profile of the perpetrator. For example, the findings showed that household insurance fraud is primarily committed by ordinary people in everyday situations, rather than organized criminals.

A study conducted by (N. Morley, 2006) found that some claims adjusters tend to view fraud as an outlier phenomenon and a crime committed by organized criminals that appropriate large sums of money, rather than everyday people exaggerating

their claims. The study also found that adjusters' do not necessarily deem fraud detection as a part of their responsibility. In this context, it is important to note that it is not always in the interest of a claims adjuster to prioritize fraud detection. As (T. Ormerod, 2003) points out, the detection process can slow down adjusters, affect their productivity, and by extension also their salary if a commission salary model is used. Given their incentives, the adjusters are in some senses primed to not identify fraud.

#### 4.2.3.1 *Indicator data characteristics*

(T. Ormerod, 2003) argues that the fact that fraud indicators by default are static is problematic from a fraud detection point of view. According to (Viaene & Dedene, 2004), this static nature of fraud control can create a false sense of security, and that elements of unpredictability are required in order to keep the potential fraudsters in check. In an area such as medical diagnostics, the data anomalies one tries to identify (diseases, viruses or cancer cells etc) are relatively static and do not change over time. In contrast, insurance fraud is very much a dynamic phenomenon. (V. Cherkassky, 2002) argue that due to this static nature of fraud indicators, one should develop flexible data-driven strategies for identifying abnormal claims based on historical claims data. They mention fraud indicators based on property value, length of past insurance coverage, geographical location of the property as examples. The idea is that once a fraud type can be identified and detected in an effective way by the insurance companies, the characteristics of fraud will inevitably change. For this reason, some companies choose not to supply their adjusters with fraud indicators, since it can lead to even shorter longevity of viable indicators. The insurers fear that supplying their frontline adjusters with this information will inadvertently make it seep into the knowledge of the general public over time (T. Ormerod, 2003). (Derrig, 2002) calls this a perpetrator learning-curve, and argues that it makes effective methods powerless over time. Also, one cannot dismiss the possibility that someone working as an adjuster today might succumb to committing insurance fraud in the future, towards their own employer. In fact, some companies have translated this concern into an actual fraud indicator. "Is the claimant a previous employee?"<sup>CC03</sup> is one of the fraud indicators used by Zurich Municipal Insurance (Zurich Municipal Insurance, 2012).

#### 4.2.3.2 *Gender*

(M. Button, 2016) found a striking gender balance in the data of confirmed cases of household insurance fraud. 54 % were male and 46 % were female in a dataset of 31 010 confirmed fraudulent cases. This paints a much different picture than that of general insurance fraud statistics. In the general insurance fraud statistics 68 % of insurance fraudsters in Sweden in 2018 were male and only 32 % were female, even after the numbers were normalized to account overrepresentation of men in certain insurance categories such as motor insurance (Larmtjänst, 2019) This discrepancy alone provides an argument for why one should not rely on general insurance fraud

statistics when trying to understand the perpetrator profile for a specific type of fraud.

#### 4.2.3.3 Age

One article found that from 24 034 claims, the group of claimants between 31–50 years of age were responsible for 57 per cent of all fraudulent claims. It should be mentioned that this overrepresentation could be explained by the fact that this age group is more likely to be insured, and therefore comprise a large share of the fraud statistics (M. Button, 2016). A study of motor insurance fraud found a significant correlation between claimant's age<sup>CC14</sup> and the likelihood of fraud: The probability of committing fraud decreased as the claimant age increased (M. Artís, 2002).

As previously mentioned, (S. Viaene, 2002) increased the predictive performance of their automobile IFD algorithms significantly by including non-binary indicators like age and the time from when the claimant signed the policy to the time the claim was made<sup>CD03</sup>. This provides an argument for an approach which combines binary indicators with non-binary indicators like age for IFD.

#### 4.2.3.4 Claimed amount

(C. Phua, 2004) researched motor insurance fraud and suggested the use of a fraud indicator named *age\_price\_wsum*. The use of this indicator was based on the assumption that if a vehicle ages and its value stays relatively high over time, the probability of fraud when the vehicle gets claimed for is higher. The indicator is calculated as a weighted sum of the vehicle age and the vehicle price<sup>DD06</sup>. This indicator should be transferable to smartphone insurance claims as well, given that the same assumption is accepted. However, the variances in price and age among smartphones being claimed for would naturally be lower than the corresponding variances for cars.

One indicator for automobile insurance fraud suggested by (EB Belhadji, 2000) involves the case where a claimant's occupation does not seem to correspond with the high value of his or her car. The price of smartphones does not vary as much among as the price of cars, but the indicator could still be transferable to a degree. (G. Dionne, 2009) also showed that this is a significant indicator to consider in motor insurance fraud detection<sup>FS07</sup>.

(M. Button, 2016) found that a confirmed fraudulent household insurance claim was valued at £716 (~8900 SEK in 2016) on average, with a median of £500 (~6200 SEK in 2016). Mobile phones were the 3<sup>rd</sup> most commonly claimed object, accounting for 14.1 % of all claims.

#### 4.2.3.5 Time-to-claim

The home insurance dataset analysed by (M. Button, 2016) showed that about 50 % of claims were made by individuals who had bought their home insurance policy

less than 12 months prior to making the claim. 30 % were made by claimants that had bought the policy less than 6 months prior to making the claim <sup>CD02</sup>.

#### 4.2.3.6 *Number of previous claims*

When it comes to home insurance, the number of previous claims seems to be a poor fraud indicator. (M. Button, 2016) analysed 31 901 confirmed fraudulent household claims and found that 90 % were committed by claimants with 0-1 previous claim in their history of claims<sup>CC02</sup>. This is important to highlight, since it is an indicator that is commonly understood as important for IFD. This is incongruent with the practices of insurance companies, as well as multiple published papers. For example, (S. Viaene, 2005) found that a high number of claims to be one of most important indicators of fraud in three different fraud detection algorithms developed using an automobile insurance dataset.

#### 4.2.3.7 *Claimant behavior*

(M. Button, 2016) Button (2016) analysed 6 255 of confirmed fraudulent household insurance claims and found that 89,1 % of mobile phone claims were explained by claimants as *accidents*, while only 10,1 % as *theft* <sup>CD06</sup>.

Adjusters seem prone to be influenced by the claimant's attitude and general courtesy (N. Morley, 2006). If a claimant has an aggressive tone or is complaining to the adjuster, it seems as if the adjuster becomes increasingly alert for inconsistencies and tries harder to find evidence for the claim to be suspicious. (T. Ormerod, 2003) points out that this reactive behaviour by adjusters can very well be unfavourable for the company, since it can lead to false positives and thus a negative experience for genuine customers. (N. Morley, 2006) argues that this means technological approaches that seek to identify aggressive claimants automatically are ineffective, since adjusters are already so sensitive to it. Although true, it is important to note that this argument only holds as long as there are claims adjusters at all. In a future where automation plays an ever-increasing role in insurance, companies such as Hedvig might not even employ adjusters in the long term.

(EB Belhadji, 2000) proposes that If a claimant willingly accepts the blame for an accident, it should be viewed as an indicator of insurance fraud <sup>CC09</sup>.

(Vrij, 2004) proposes one should not focus on physical signals of the claimant, such as stuttering, sweating or fidgeting. Since this type of behaviour is conventionally seen as signs of deception, such behaviour often triggers suspicion among adjusters and investigators. This is reasonable considering the well-understood tendency people have to focus on evidence in support of their own hypothesis, instead of focusing on facts that could prove the hypothesis to be false. This phenomenon is generally known as confirmation bias (Nickerson, 1998).

(EB Belhadji, 2000) proposed the following indicator for automobile insurance fraud: *Shortly before the loss, the insured checked the extent of coverage with his or*

*her agent*. This is an example of an indicator that should be equally viable for smartphone IFD <sup>CC21</sup>.

#### 4.2.3.8 *Time of claim*

Several researchers have suggested that the timing of the claim should be of interest to insurance fraud detection efforts. (M. Button, 2016) analysed 32 924 fraudulent household insurance claims and found that there seem to be seasonality to consider within the data. In this dataset, September is the most common month in which fraudulent claims are made. It may be worth noting that September is also the month which has seen the most releases of new iPhone models. A subset of this data (5915 claims) contained information on object level. For this dataset, the peak month for making fraudulent mobile phone claims was August. The month that saw the least amount of fraudulent claims was February.

(C. Phua, 2004) speculate that the average insurance fraud perpetrator is more likely to commit fraud during holiday weeks because there is a general tendency of wanting to spend more money during such weeks and a notion that there is a lower chance of getting caught <sup>CD07</sup>.

## 4.3 Factors improving IFD capacity

### 4.3.1 **Industry collaboration**

*"Information of what objectively characterizes a fraudster"*

Above quote is what one of the speakers at Counter Fraud 2019 conference replied when asked what data he wanted but did not have, during a seminar. This information is hard to provide unless there is a close cooperation between the companies. If the combined data set of the members could be used to conduct research to answer the above stated question, detection of both organised and opportunistic fraudsters could increase. The subject was recurrent in the talks and lectures at the conference. However, market competition and data privacy concerns are claimed to hinder collaboration at this level. Insurance companies are, like other ordinary enterprises, competing over customers by offering low prices. This makes one of the key means of competition the data they have, as more accurate predictions enable better discrimination of risk. Hence, insurers are not keen on sharing this with their competitors. There is, to some extent, a cooperation in data sharing in Sweden through GSR as mentioned, but the content of that can be extracted by the members is very limited and only shows *Insurance company, policy type and date of claim*. According to S6, there is no plan on elaborating GSR currently, and refers to GDPR as a regulation that would hinder development of the register. In Massachusetts, a similar register is found called DCD (Detail Claim Database)

which includes reports from all automobile bodily injury claims closing January 1, 1994 and subsequent from the state (D'Arcy, 2005). Examples of content include injury information and expenses for related medical treatment. The most important feature of the DCD is that the results of research on this data set can be freely shared by researchers, leading to cooperative advances in the development of predictive models. This would be desirable also in Sweden for the IFD capacity to increase. To contrast, discussions at the Insurance Innovation Summit 2019 conference regarded how the use of data is communicated, which needs to be considered and it is stated that insurers should watch out for being suspected of analysing behaviours "behind the scenes". Also considering what data should be gathered is important; geolocation and genetics are two proposed examples of data that would be difficult to explain why they are collected and stored. There was an expressed belief that the insurers' need for data is less problematic than the tech giant's like Facebook and Google as the insights can be used to mitigate risk for consumers by warning of danger.

There are also problematic aspects of the fact that the insurers conduct their own fraud investigations, as they are vulnerable to serial fraudsters that jump between insurers. Among the practitioners, a version of "you will have to try to defraud someone else" was a recurrent phrase when describing what they tell someone who is caught trying to defraud them. This is suboptimal regarding how to deal with fraudsters who continually defrauds the companies. At the Counter Fraud 2019 conference there was discussions of a centralized investigation unit at the police of City of London, called Insurance Fraud Enforcement Department, IFED, which is funded by the insurance companies and all cases suspected for fraud are sent there. Due to this, there is a central collection of data on insurance fraud, which has the benefit that insurers can be alerted on an early stage that they are being targeted by organised criminal gangs who file claims en masse. This concept is suggested as a means within IFD also in Sweden, although some practitioners say this would not comply with Swedish law. Another aspect is that of organized criminals who operates across borders, which motivates this to be an issue of international collaboration. At the Counter Fraud 2019 conference a representative of Europol, the head of Analytical Project fraud in the seminar *insight from beyond insurance: broadening the conversation* brought up the issue of cross-national insurance fraud. As organised criminals often cross-national borders to avoid detection, this is essential to increase the detection rate of this kind of fraud. The IMEI number could be used in this purpose to create a database of phones reported stolen.

#### **4.3.2 Data acquisition strategy and improved analytical tools**

In the fight against fraud, many answers are hidden in the vast amounts of data available to insurers. Most companies have for long had routines and processes in place for the analysis of *structured* data such as payment data, claims data, credit scores and personal information. Many of the parameters that can be extracted from



this type of data have proven to be highly correlated with fraud. Moving forward, most companies at the Counter Fraud 2019 conference are more focused on understanding their *unstructured* data, an area where the current capabilities within the industry appear to vary considerably. Unstructured data includes e.g. documents, phone calls, social media patterns, images, videos etc. An example of unstructured data is the claim report itself. Several vendors are currently focused on exploring how voice data can be used to counter fraud by conducting tonality analysis of claims made via voice. Data acquisition was also a subject of discussion, and many participants expressed concern regarding personal integrity and data sharing consent with the methods proposed of looking at socio-economic background as an example of an indicator of fraud, which would be highly discriminatory to many people.

Although many companies at the Counter Fraud 2019 conference claim they use Artificial Intelligence, AI, and Machine Learning, ML, far from all have made it an integral part of their fraud detection systems. Several considerations must be made in order to get a realistic understanding of the possibilities of using ML to counter fraud. One aspect is that fraud data inherently has a highly unbalanced class as most claims are deemed legitimate. There are methods for handling this however, such as over- and under sampling. It is also important to consider that historic data of legitimate claims most likely contains undetected fraudulent claims, making it even more problematic to train algorithms using conventional ML methods. AI is allegedly superior to traditional statistical methods when it comes to analysing data with many features, which is the case both when regarding indicators of fraud and unstructured data itself. The speaker's response to a question of what type of claim that would be suitable for AI driven fraud analysis were

*“claims characterized by high volume, easily reported from the customer, where 2-3 pictures are provided of for example broken glass, pipes and the likes. It is also good to use AI as early as possible in the claim for nudging”*

These are typical characteristics of smartphone claims, as described in the background of this report. Another suggestion of applications for AI was smart image analysis used for assessing damage severity. If, for example, a car has been subject to an accident resulting in a broken window, the mechanic would be asked to snap pictures of the damage before and after the repair, which mitigates the risk for insurers to overpay for repairs. This could be implemented for smartphone screen damages as well.

One discussion At the Innovation Summit 2019 conference of the use of AI involved the traditional way of doing underwriting, pointing out that it stems from the intuition developed after underwriting the same risk for 30 years, which resembles the way traditional adjusting is done. When letting AI make suggestions to decisions it allegedly outperformed the underwriter since AI is great on analysis of historic data where it can take in massive amounts of data to find patterns, which is an attribute very applicable to fraud assessment. However, when it comes to taking in the present and predicting the future humans still outperform AI as they are better

at exceptions and new knowledge. Therefore, man and machine work best when they are on the same team.

The speaker of at seminar *AI and advanced analytics: new weapons in the fight against fraud* at the Counter Fraud 2019 conference, claimed that they use Logistic Regression in their data mining model and argued that the fact that it is both easy to understand and explain to others is a huge benefit of the model. However, a notable downside of Logistic Regression is what is known as “overfitting” on rare events. For example, a catastrophic event where a fire destroyed hundreds of cars in a parking lot lead to a temporary surge in motor claims which in turn created a disproportionate impact on the Logistic Regression model. The company is also experimenting with the use of the ML methods Random Forest (RF) and Neural Networks (NN). The speaker notes that the upside of using RF is that you can always trace why a certain claim was marked as fraudulent, although the downside is that it requires a lot of computational time. NN, on the other hand, has the benefit of a much higher level of analytical complexity, while the downside is that it is dependent on so called black-box decisions that cannot be explained. When it comes to the future development of AI solutions to counter fraud, many agree that collaboration will be extremely important. A Principal Data Scientist argued for cross-industry collaboration, and particularly point to healthcare as an important industry to collaborate with. Rare diseases bear close resemblance to insurance fraud in terms of data characteristics as both are examples of very uncommon instances of huge data sets, making it difficult to train an ML algorithm to identify them. She believed a lot could be learned if data scientists from the two industries would collaborate. The speaker is optimistic about the capabilities AI can enable in fraud detection and says ML can be effective in identifying both *opportunistic* and *planned* fraud.

### 4.3.3 Legacy systems

In seminar 2 of the Counter Fraud 2019 conference in London, *AI and advanced analytics: new weapons in the fight against fraud*, one speaker was critical to that the industry continues to move slowly in updating their legacy systems. Legacy systems refer to internal systems that have not been adequately entertained and updated to be used in an efficient manner. Issues with legacy systems include e.g. a bias stemming from being the starting point of chosen parameters, which is what is pointed to, in particular that some business rules have not been updated since the 1970’s. As these systems are often integrated in the whole organization, they can be very expensive to replace which is why they are often kept longer than they deserve. The discussions revealed that the many of the insurance companies did not have a standard way of reporting most data, and different divisions had their own way of doing it. This is a large issue for developing efficient fraud detection systems as new features, data or analytic tools may not be possible to integrate. Legacy systems can hold back the analytic capability and slow down the claims process, which is why it

is an important area to improve. When asked what kind of data, that the speakers were currently in possession of, which could be put to better use to increase fraud prevention capabilities, one answer was:

*"Data quality in general, as we don't have a standard way of reporting. The different internal divisions choose to report in a manner that suits them, making it impossible to use each other's data."*

## 4.4 Analysis of SIFI

In this section, a few notable examples of indicators are highlighted where divergencies and congruencies were identified among the different method areas.

### 4.4.1 Number of previous claims

In the case of the indicator CC02, (number of previous insurance claims), there is striking divergence between theory and practice prima facie. From the literature review we found that a clear majority of confirmed fraudulent household insurance claims (90 %) were made by claimants who had made 0-1 previous claims (M. Button, 2016). The practitioners, though well-aware that most frauds probably were committed by regular policy holders, spoke of a high number of claims as one of the most reliable fraud indicators. It is important to note that these two views are not necessarily in disagreement, as they might describe two different fraudster profiles. We conclude that the number of claims is probably not a bad indicator of smartphone insurance fraud per se, but rather that this indicator will probably not contribute to the detection of most fraudulent smartphone claims. Hence, it is important that CC02 is complimented with indicators that are more indicative of opportunistic insurance fraud.

While the aforementioned findings of (M. Button, 2016) are important to consider, it should be mentioned that there are many examples from the literature review that do mention the correlation between a plentiful claims history and the likelihood of fraud, e.g. (Patrick L. Brockett, 1998), (Derrig, 2002) and (S. Viaene, 2005). One article also mentions this indicator in the context of household insurance claims specifically, in the case of a household insurance policy holder making 3 or more claims over a 3-year period, as an example of behavior indicative of fraud (V. Cherkassky, 2002). However, it is unclear from where they drew this conception, as no empirical results were presented, or other source referenced. A possible explanation is that the conception was drawn from industry experts. All things considered; it seems likely that CC02 is a good example of a widespread fraud indicator, but that the actual utility of the indicator is not as high in smartphone IFD as in other areas of IFD.

#### 4.4.2 Opportunism in abundance

Another important insight drawn from the empirical findings from (M. Button, 2016) is that household insurance fraudsters are first and foremost opportunistic ordinary people, rather than organized criminals. This is reminiscent of several reports from interviewed the practitioners, and adds further support to (Persson & Bongehiell, 1998) when they declared that opportunistic fraudsters are indistinguishable from the rest of the insured collective. These findings should be considered in the light of (N. Morley, 2006), who found clear tendencies among adjusters to predominantly associate fraud with professional criminals rather than everyday people exaggerating their claims. From this we gather that adjusters seem to be ill-equipped to deal with cases of household insurance fraud in particular. Adjusters seem much better suited to identify fraud using indicators such as CC04 (appears to be claims wise and CC16 (number of policy changes).

#### 4.4.3 Economic incentives driving fraudulent claimants

Practitioners are looking for signals of economic motive for committing fraud, e.g. if the claimant is in financial strain, which can be established using the indicators FS01 - FS07. The economic motive, can also be present without dire personal finances, as pointed out by one of the practitioners, suggesting that a potential type of fraudsters could be 20-30-year-olds who are attempting to fund a lifestyle they cannot really afford. To capture this, other data sources revealing the economic motive could be pursued, for example using social media analysis, as suggested by e.g. S1. In general IFD, the economic motive is often claimed to be present, as pointed out by S2, and this could be true with the same reasoning as of the number of claims above. The claimed amount in most smartphone cases would not much help a person with heavy debts to Kronofogden, but if the smartphone costing upwards of 20 000 SEK, this would make it worthwhile to assess the personal finances as proposed by S8. If those expensive models occur in a claim, it would therefore probably be more important to examine the personal finances.

Practitioners are looking for signals of economic motive for committing fraud, e.g. if the claimant is in financial strain, which can be established using the indicators FS01 - FS07. The economic motive, however, can also be present without dire personal finances, as pointed out by one of the practitioners who suggested that a potential type of fraudsters could be 20-30-year-olds who are attempting to fund a lifestyle they cannot afford. To capture this, other data sources revealing the economic motive could be pursued, for example using social media analysis, as suggested by e.g. S1. One of the example of an economically motivated fraudster profile proposed by S8 was a person with a highly strained economy that would probably not be able to afford a smartphone costing upwards of 20 000 SEK, which would make it worthwhile to assess the personal finances, if those expensive models occur in a claim<sup>FS07</sup>. Similarly, (EB Belhadji, 2000) proposed a fraud indicator for

motor insurance that basically reads “claimant’s occupation does not seem to correspond with the high value of his or her car.”

#### 4.4.4 Poor controls on standard cases

When asked how large share of all incoming smartphone claims are estimated to be fraudulent, the replies range between 5 % and 30 %, with some caveats regarding type of claim and time of the year, indicating that there is very little knowledge of what to be thought of as a good result regarding solved cases; if being able to confirm fraud in 2 % of all claims, this appear very differently if the real share is 5 % compared to if it is 30 %. The statement that 36 % of claimants did a “fraud walk-away” when being extra thoroughly controlled in periods around claim cycle peaks, could also indicate that selection of cases is normally done poorly, especially if the statement that the claim cycle peaks now are less prevalent is true. There are, of course, other explanations for low clearance rate. Some concern the motive for controls: low economic incentive for the insurer to investigate due to the low value in the claims, low economic value for the adjuster who might get commission from sales and worries about productivity and customer satisfaction (T. Ormerod, 2003). Other explanation lies with concerns of intruding on data privacy regulations and the time required to gather, clean, visualize and analyse the data.

Some insurers are afraid that the knowledge of the used indicators will reach the general public in short time if they offer their adjusters information of used indicators (T. Ormerod, 2003). This is not as problematic when it comes to claims that are always examined by investigators, such as car accident claims and house fire claims. However, when it comes to claims like smartphone claims, this tactic causes the adjusters to have lesser tools for evaluating the possibility of fraud at point of claim. Rather, adjusters would need to be educated in fraud analytics in this area to be able to do better assessments. Considering that the vast majority of smartphone fraudsters are ordinary policy holders with one claim or less in their history and thus presumably have an opportunistic motive, the public would probably not be looking for this information anyway. On another note, the problem of unaware adjusters would be greater if the set of indicators were to be static, which should not be the case for smartphone insurance fraud as (T. Ormerod, 2003) points out. For this reason, to discover changes in fraud behaviour, the relevance of the suggested data in the SIFI tables should be re-evaluated continuously. (V. Cherkassky, 2002) suggests that as soon as a fraud type is identified, this will be learned by fraudsters who change the modus and new indicators would need to form. To keep the fraudsters in check, (Viaene & Dedene, 2004) suggests that elements of unpredictability should be incorporated. One way to do that while possibly also gaining new indicators is to start with selecting claims randomly for more thorough controls. Other suggestions for discovering new modus is to use more of unstructured data in the analysis, which can potentially generate more information of customer behavioural patterns. More customer interactions would also generate

more data points and could therefore also contribute to this. Lastly, the industry wide collaboration is particularly important in communicating new modus to the insurance companies. They do this currently in Larmtjänst in Sweden, but there is a need for a higher degree in some claim types such as smartphones.

#### **4.4.5 Man vs. machine**

Adjusters are allegedly reacting if the claimant is behaving aggressively, impatiently or emotionally and it is confirmed by many of the practitioners that this has a correlation with many frauds they have detected. However, this might be one of the areas where the approach of intuition is weak. According to Morley (2006), adjusters are prone to get more sensitive to other types of indicators if the adjuster also experiences an assertive claimant. Thus, claimants who act like this would be examined to a greater degree and has therefore a higher risk of getting caught. This is an example of confirmation bias, and risks reaffirming an inaccurate hypothesis so that the adjuster misses other more important indicators. As is brought up in the section of improvements of capacity regarding analytical tools, tonality analysis is suggested to be used for assessing the written or spoken claims report. In a future where automation plays an ever-increasing role in insurance, companies such as Hedvig might not even employ adjusters in the long term. Using an AI algorithm for the assessment could perhaps mitigate the problem of personal bias, but on the other hand, there are also issues with being too reliant on automated IFD systems using data mining algorithms. Algorithms lack the instinct to interpret previously unfamiliar behaviour compared humans. This is supported both by in the literature review (Y. Wang, 2018) and in the Insurance Innovation Summit conference, both concluding that expert experience in combination with quantitative methods and employment of ML is desirable to reach the best results in IFD.

Much of the data used by practitioners are used with the purpose of verifying the story of the claimant, to ensure that the claim is legitimate, which is probably a good approach when finding a fraudulent case. The area where the practitioners could improve most lies within selecting fraudulent cases with good precision. Exploring unstructured data to a greater extent is probably a good first step. Investments in new technology seems to be needed to update the legacy systems regardless of this to increase the quality of the data and thus the possibility to analyse it.

#### **4.4.6 Endure or walkaway**

The indicator CC15, “endurance”, is several respondents proposed as an alternative when the adjuster or investigator cannot find objective evidence to support a particularly strong suspicion. The quote from F1 was very telling in this regard. The assumption is that honest claimants will be patient as they know they have a right to payment, while fraudsters will either become nervous and say something that

might expose them or tire of waiting for the payment and do a so-called “fraud walkaway”. This is a risky assumption and there is always a chance that the insurer loses honest customers with this tactic.

#### **4.4.7 Fraud due to the release of new smartphone models**

Some of the practitioners stated that there would be more frauds attempted in timewise proximity to the releases of new smartphone models<sup>CD04</sup>. S8 provided support for this when stating that 36 % of claimants did “fraud walkaways” when the company performed extra thorough investigations in such periods. The literature suggests that this is in fact a good data point to include (Button, 2016), but there are also countering statements from practitioners claiming that the presence of the peak has decreased and is not as strong as it was a few years ago. However, that does not mean the trend cannot turn again, come the release of a new smartphone that consumers desire. It is suggested that this indicator should be used, albeit with some extra caution, and that its indicative performance should be tracked.

# 5 Discussion and conclusions

## 5.1 Discussion of results

### 5.1.1 Generalizability

Although this report revolves around fraudulent smartphone claims specifically, the identified fraud indicators could potentially be applicable within other areas of insurance as well. It appears likely that the results could prove useful for the detection of fraudulent claims of other devices, such as laptops and tablets. However, it is important to note that seemingly small differences in device characteristics can probably have a notable impact on fraud detection capability. For instance, neither laptops nor tablets have as practical, ubiquitous and universally understood identifiers like smartphones do with IMEI numbers. For one thing, tablets are not as abundant as smartphones among consumers, and can often have shared ownership. Another consideration has to do with price, where smartphones are quite uniform relative to laptops. Also, in terms of quantity, the number of smartphones in circulation are much greater than the number of laptops and tablets.

One benefit of IMEI is the fact that it has its own supporting infrastructure, that, among other things allows the “blacklisting” of notorious devices, i.e. devices that have previously been used in e.g. insurance fraud or money laundering schemes. Other aspects that could hinder generalizability of the results can be drawn from e.g. the price difference between laptops and smartphones. Laptops can be significantly more expensive than smartphones, which could mean that other psychological and behavioural dynamics are at play that governs decision making. As for tablets, the benefit of developing effective IFD systems are not as clear as for smartphones and laptops, since tablets are not as abundant among consumers, and are often shared among multiple users, for instance in a family. The degree to which the results are generalizable is a good example of a suitable area to be investigated in future research.

To validate the results, the synthesized table could have been sent to all participants to get a second opinion, which would have increased the credibility of the results. A natural step from this report is to proceed with building an automated fraud detection system and a suggestion is to examine which Machine Learning methods would be most suitable to employ. Another suggestion is to explore other dimensions of the suggested indicators in e.g. if it is *feasible* and *legal* to collect and



store or what the *customer attitudes* are towards collecting the data. Also, in regard to the customers, what are their attitudes to what information about them they think is ok for the insurer to look at, and whether that would change if there was a compensation for it in terms of cheaper premiums are other aspects of interest.

### 5.1.2 Interviews

Investigators from three of the four largest insurance companies in Sweden were included in the study. The four largest companies make up for 80% of the total market share of home insurance policy holders in Sweden, so their combined experience should be vast within the research area in question. They could also be considered to hold well-established opinions within fraud detection. Thus, for the purpose of describing general practices for IFD we can confidently say that the sample sufficed. However, it is primarily the adjusters who work with smartphone claims on a daily basis, and an investigator is normally only brought in to deal with complex and suspicious claims. By including more adjusters in the interview sample, we could therefore have increased the authenticity of the results. Getting access to the many attendees of the two conferences was a key reason for attending, as this gave us a uniquely diverse set of potential interview candidates. Our attempt of reaching out to them did, however, not have a high response rate. Only 5 out of 104 responded to our interview request, and only 3 agreed to be interviewed. A possible reason for the low response rate could be explained by the fact that we had to use the conference attendee's forum to reach out. Via this platform, e-mails were sent, and it is not unlikely that many e-mails got caught in the spam filters of those who were contacted. The report would have had a more grounded result if the response rate had been higher. The shaping of questions for round 2 was purposefully affected by the results from round 1, and this enabled us to ask intricate and precise questions about defrauding strategies. However, as the interviews were kept around 20 minutes, there was not much time for elaboration of the replies. Extending the interviewing time could have increased the conformability of the results.

### 5.1.3 Conferences

The seminars were very useful for gathering high-level insights of the trends and challenges within insurance in general and IFD in particular. However, they did not disclose many examples of viable fraud indicators and one possible explanation for this comes from purely competitive reasoning: although the conference was a collegial arrangement, the reality is that many of the participating companies compete for the same market share. Openly disclosing fraud indicators could pose a real threat to competitive capabilities.

#### 5.1.4 Literature

The data gathered from the literature review is by no means comprehensive since it does not fully encompass all published research in the research area. However, it still serves the overarching purpose of triangulation in our research method and it enabled validation and further analysis of the results from the other method areas. As stated in the introduction to section 4.4.3, automobile IFD seems to be the most well researched area of insurance fraud and stood for the majority of the papers found in the literature review conducted in this report. Much fewer papers were found on home insurance fraud (the policy in which smartphones would be included) and no papers concerned smartphone fraud specifically; only one published article included smartphones as a discrete possession category. More research on household insurance fraud in general would therefore be suggested for further research. The difficulty in getting hold of good real-world claims data set due to data privacy policies is a major restriction on the ability to do research on insurance fraud. The same data set has therefore been reused by many researchers, or the research has been done on very small datasets. This can certainly have had impact on our result, but the spread in publication date and country of origin should still offer good validity to it. More insights and potentially more predictors could possibly have been identified if the literature review would have been expanded. There is a chance that including *laptop*, *computer* and *tablet* would have generated more articles and including these would have increased the transferability of the results. The expansion could also have been done by increasing the size of TIER1, or by including more snowballing iterations, leading to even more articles sets like TIER3 or even TIER4. The fact that some articles of TIER1 had a few authors in common is not ideal either. Thus, a comprehensive literature review including more search terms and more iterations is left as a suggestion for further research.

## 5.2 Conclusions

This report has explored a variety of areas from which insights regarding smartphone IFD have been extracted. It has identified 51 distinct SIFIs spread across 5 categories. The compilation of identified SIFIs can add value by making the manual procedures of frontline claims adjusters more efficient and effective. Many of the identified indicators should also be useful for training automated IFD systems. The degree to which this is feasible depends on one's current automation capabilities, for example the ability to utilize unstructured data. Furthermore, the report has explored how the capacity of smartphone IFD can be improved, by considering the potential of industry collaboration, data acquisition strategies, analytical tools, and the hinders of legacy systems. The triangulation method proved useful and enabled the observation of several congruencies and divergencies between the theory and practice of smartphone IFD. The results would benefit from further research of e.g. suitable data mining IFD algorithms for which the indicators can be used as feature variables.

# References

- AR. Bologa, R. B. A. F., 2013. Big Data and Specific Analysis Methods for Insurance Fraud Detection. *Database Systems Journal*.
- Assurant, 2012. *Smartphone insurance fraud on the rise*. [Online]  
Available at: <http://www.assurantsolutions.co.uk/connected-insight/smartphone-insurance-fraud-rise/>  
[Accessed 21 05 2019].
- Bentley, P., 2000. "Evolutionary, my dear Watson" - Investigating Committee-based Evolution of Fuzzy Rules for the Detection of Suspicious Insurance Claims. *Annual Conference on Genetic and Evolutionary Computation*, Issue 2, pp. 702-709.
- Bryman, A. & Bell, E., 2015. *Business Research Methods*. 4th ed. Oxford: Oxford University Press.
- C. Phua, D. A. V. L., 2004. Minority report in fraud detection: Classification of skewed data. *Explorations Newsletter - Special issue on learning from imbalanced datasets*, 1 June, 6(1), pp. 50-59.
- C. Phua, V. L. K. S. R. G., 2010. *A Comprehensive Survey of Data Mining-based Fraud Detection Research*, Melbourne: Baycorp Advantage.
- Cifas, 2019. *Cifas members report a 27% rise in false insurance claims across the UK in the past year, with spikes in household and motor insurance*, s.l.: Cifas.
- Derrig, R. A., 2002. Insurance fraud. *The Journal of Risk and Insurance*, 25 October, 69(3), pp. 271-287.
- EB Belhadji, G. D. F. T., 2000. A Model for the Detection of Insurance Fraud. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 25(4), p. 517-538.
- G. Dionne, F. G. P. P., 2009. Optimal Auditing with Scoring: Theory and Application to Insurance Fraud. *Management Science*, 55(1), pp. 58-70.
- Goodman, L. A., 1961. Snowball Sampling. *The Annals of Mathematical Statistics*, March, Volume 32, pp. 148-170.
- Gray, A., 2012. *Financial Times*. [Online]  
Available at: <https://www.ft.com/content/9ba2857c-9619-11e1-9d9d->

00144feab49a

[Accessed 25 May 2019].

- Gray, A., 2012. *Mobile phone insurance fraud soars*. [Online]  
Available at: <https://www.ft.com/content/9ba2857c-9619-11e1-9d9d-00144feab49a>  
[Accessed 27 05 2019].
- Hedvig, 2019. [Interview] (Mars 2019).
- Herberg I. Weisberg, R. A. D., 1991. Fraud and Automobile Insurance: A Report on Bodily Injury Liability Claims in Massachusetts. *Journal of Insurance Regulation*, 9(4), pp. 497-543.
- IMEI info, 2019. *IMEI.info*. [Online]  
Available at: <https://www.imei.info/faq-what-is-IMEI/>  
[Accessed 26 May 2019].
- Insurance Edge, 2018. *Opinion: Tackling smartphone insurance fraud*, s.l.: Insurance Edge.
- Internetstiftelsen, 2018. *Svenskarna och internet 2018*, s.l.: s.n.
- Jane Webster, R. W., 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), pp. xiii-xxiii.
- Justitiedepartementet L2, 2005. *Försäkringsavtalslagen*, Stockholm: Sveriges Riksdag.
- Korsell, L., Stenström, A. & Jonsson, A., 2015. *Försäkringsbedrägerier - En Selektionsstudie*, Stockholm: Brottsförebyggande rådet (BRÅ).
- Larmtjänst, 2019. *Försäkringsbedrägerier i Sverige 2018*, Stockholm: Larmtjänst och Svensk Försäkring.
- M. Artís, M. A. M. G., 2002. Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims. *Journal of Risk and Insurance*, 69(3), pp. 325-340.
- M. Button, F. P. B., 2016. 'All walks of life': A profile of household insurance fraudsters in the United Kingdom. *Security Journal*, 29(3), pp. 501-519.
- N. Morley, T. O. L. B., 2006. How the detection of insurance fraud succeeds and fails. *Psychology, Crime and Law*, 12(2), pp. 163-180.
- Nickerson, R., 1998. *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, s.l.: s.n.

- Patrick L. Brockett, X. X. a. R. A. D., 1998. Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud. *The Journal of Risk and Insurance*, 65(2), pp. 245-274.
- Persson, L. G. & Bongenhielm, B., 1998. *Försäkringsbedrägerier - En Kriminologisk Kartläggning*. Stockholm: Sveriges Försäkringsförbund.
- S. Viaene, G. D. R. D., 2005. Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications*, 29(3), pp. 653-666.
- S. Viaene, R. D. B., 2002. A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *The Journal of Risk and Insurance*, 69(3), pp. 373-421.
- Svensk Försäkring, 2019. *Försäkringar i Sverige 2019*, Stockholm: s.n.
- T. Ormerod, N. M. L. B. C. L. C. S., 2003. Using Ethnography To Design a Mass Detection Tool (MDT) For The Early Discovery of Insurance Fraud. *Extended Abstracts on Human Factors in Computing Systems*, pp. 650-651.
- V. Cherkassky, Y. M., 2002. *Multiple Model Estimation: A New Formulation for Predictive Learning*. Minneapolis: Department of Electrical and Computer Engineering University of Minnesota Minneapolis .
- Viaene, S. & Dedene, G., 2004. Insurance Fraud: Issues and Challenges. *The Geneva Papers on Risk and Insurance - Issues and Practices*, April, 29(2), pp. 313-333.
- Wohlin, C., 2014. *Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering*, s.l.: s.n.
- Wohlin, C., 2016. *Second-Generation Systematic Literature Studies*, Karlskrona: Blekinge Institute of Technology.
- Vrij, A., 2004. Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology*, 9(2), pp. 159-181.
- Y. Wang, W. X., 2018. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, Volume 105, pp. 87-95.
- Zurich Municipal Insurance, 2012. *Red Flag Indicators - Quick reference guide*. s.l., s.n.
- Zurich Municipal, n.d. *Red Flag Indicators*, s.l.: s.n.

# Appendix A Interviews

## Practitioner interview round 1

### *Participants*

<b>Alias</b>	<b>Nation of employment</b>	<b>Profession</b>
S1	Sweden	Head of Claims
S2	Sweden	Head of Claims Adjusting
S3	Sweden	Private Insurance Fraud Investigator
S4	Sweden	Head of Insurance Fraud Investigation
S5	Sweden	CEO of industry association
N1	The Netherlands	Business developer at insurance fraud detection solutions supplier
S6	Sweden	Administrator at industry association

### *Interview guide*

1. *Briefly, please describe your background*
2. *Please describe the fraud detection routines and investigations carried out by your company*
3. *Are there any challenges associated with these routines and systems?*
4. *Are there any attempts of automating these processes?*
5. *Does your company screen the claimant's online activity for patterns indicative of fraud?*
6. *How do you use external data registers to assess the likelihood of fraud?*
7. *Do you think there are any distinctions between the nature of insurance fraud and other types of fraud?*
8. *What are some characteristics of a typical fraudster? What do you look for when screening the different databases?*
9. *Why do you think fraud detection isn't a higher priority in the industry?*
10. *Soft fraud or hard fraud – which is the bigger issue and why?*
11. *Finally, please recommend someone you think we should get in touch with for our research.*



Practitioner interview round 2

*Participants*

<b>Alias</b>	<b>Nation of employment</b>	<b>Profession</b>
S1	Sweden	Head of Claims
S2	Sweden	Head of Claims Adjusting
S3	Sweden	Private Insurance Fraud Investigator
S4	Sweden	Head of Insurance Fraud Investigation
S7	Sweden	Head of Insurance Fraud Investigation
S8	Sweden	Head of Insurance Fraud Investigation
U1	UK	Head of fraud
W11	Switzerland	Data Scientist
F1	France	Data Scientist

## Interview guide

1. *What share of total smartphone claims would you estimate to be fraudulent?*
2. *Generally, common data which is evaluated in a claim is the claimant's behaviour when filing the claim (stress, assertiveness, too detailed description), personal financial situation and claims history.*
  - a. *What data would you currently use to establish smartphone fraud? Out of these – what data would make you suspicious and what would be sufficient to deny payment?*
  - b. *Out of these – what data would you access at first interaction with the claimant?*
  - c. *Would it be meaningful to divide the data on type of claim (theft, broken screen, lost) and, if so, how would that look like?*
3. *Is some type of smartphone claims distinctly more prevalent concerning attempts to hard fraud (i.e. arranged claims)?*
  - a. *If so, do you look for this in an investigation currently?*
4. *Is some type of smartphone claims distinctly more prevalent concerning attempts to soft fraud (i.e. opportunistic/embellished claims)?*
  - a. *If so, do you look for this in an investigation currently?*
5. *Adding items to a claim to mitigate the deductible is prevalent in insurance fraud in general. Have you experienced there to be a common such a tactic in smartphone claims?*
  - a. *If so, do you look for this in an investigation currently?*
6. *In a car accident involving collision with wild life, e.g. deer, just stating that you have "swerved" and driven into a ditch can invoke investigation. Not because the use of the word in itself is a predictor of fraud, but because it can be technically compared to the stated distance to the animal when taking action and velocity of the vehicle, i.e. it becomes possible to investigate. Is there anything similar in smartphone claims?*
  - a. *If so, would you look for this in an investigation currently?*
7. *What is your view of the profile of the typical smartphone fraudster?*
8. *Several reports show that there are clear periodicities in amounts of smartphone claims, e.g. in connection to iPhone releases, Christmas or Black Friday. Do you take this into consideration currently, and if so – how?*
9. *Asking for before and after pictures of a car from a mechanic who is hired to repair car damages in order to prevent third party fraud in motor insurance has become popular among insurers. An example of this type of fraud in smartphone claims could be that a repair shop does more than adequate repairs or charges more than market price when knowing it is an insurance company that pays. Is third party fraud in smartphone claims a problem in your opinion and do you take any measures for countering it?*