Institute and Faculty of Actuaries

F1: Predictive Modelling and Big Data for PMI Pricing and Analytics

Stephen Bishop
Daniel Cantzler

08 Mai 2015

Agenda

1. From Big Data to Analytics
2. Methods and Techniques
3. Customer Relationship Managements (CRM)
4. Pricing
5. Product Development
6. Fraud and Abuse
7. Disease Management Programs (DMP)
8. Q & A

Today the exponentially increasing amount of data is staggering…

MUNICH HEALTH

Munich RE

10 Terabytes
Sensor data produced by a jet every 30min of flight time

5 to 250 Gigabyte / hour
Produced by up to 100 sensors in modern cars now

3.200 millions
...daily of likes and comments in Facebook.

250 millions
.. of tweets per day

42% increase
of machine-generated data by 2020

81% growth
Global mobile traffic 2013

1 Terabyte
Structured trading data collected by the NYSE each day the market is open

...and opens up myriads of new applications, products and services

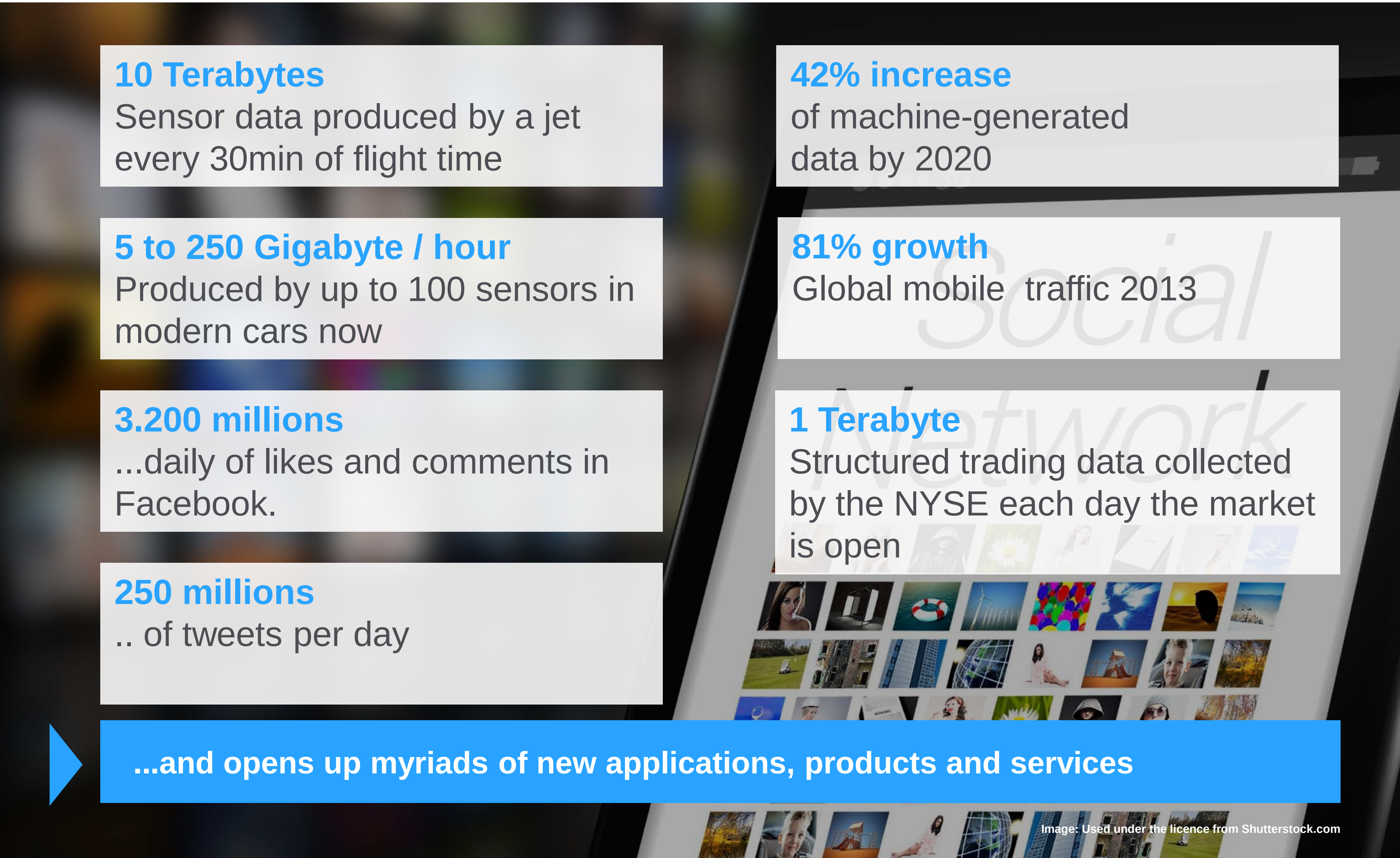
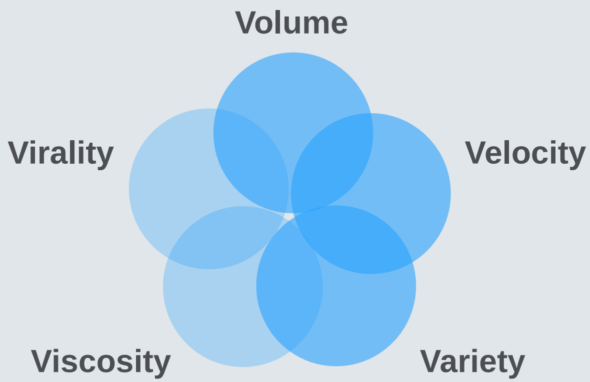


Image: Used under the licence from Shutterstock.com

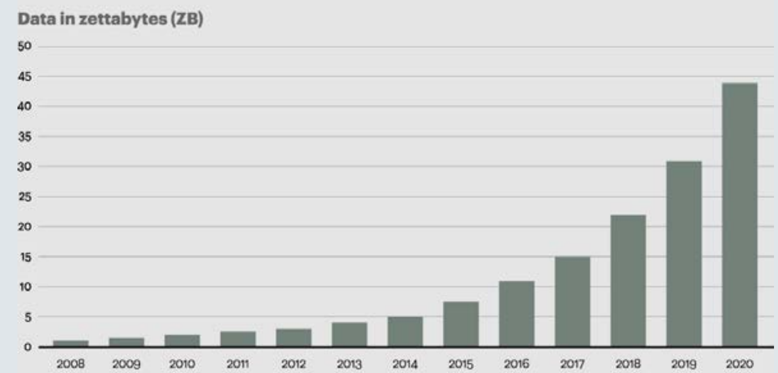
Big Data is a trend that will strongly influence the future

... Is defined by 5 V's



Source: Munich Re

... is growing strongly



Source: Oracle

Source: Munich Re

... is driven by modern technologies & trends

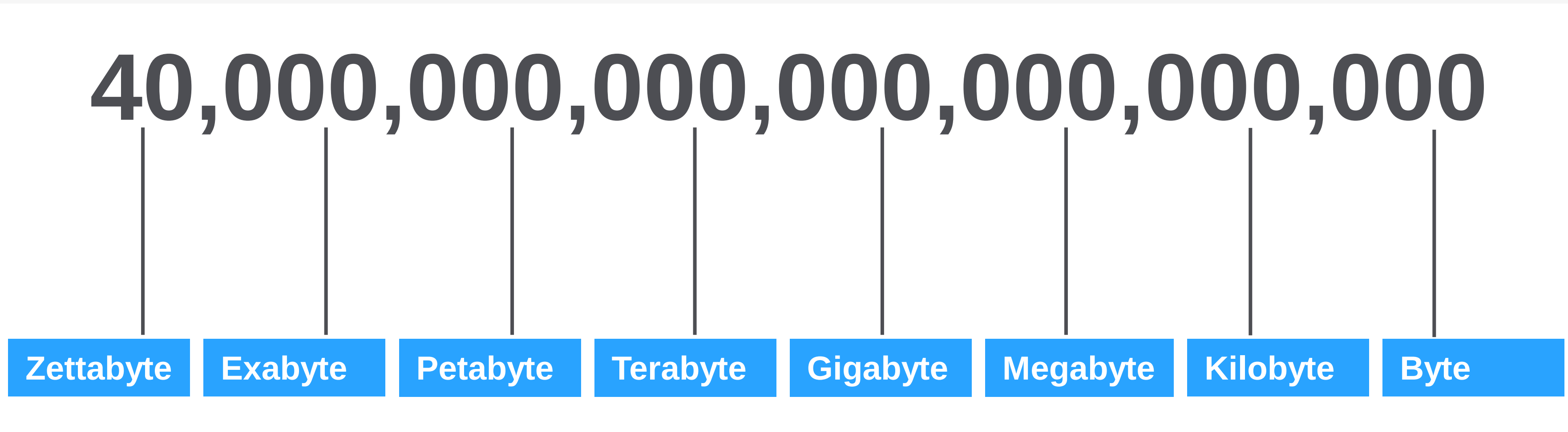
... creates a new market environment

Source: buttom right: Social Media Prism by ethority / http://facebook.com/SocialMediaPrism/https//:twitter.com/SoMePrism/ http://pinterest.com/someprism,/ Contactus for updates:prism@ethority.net

When does it become BIG data?

MUNICH HEALTH

Munich RE



- 43 zettabytes of data will probably be generated by 2020
- 300 times the volume in 2005

Source: IBM

Several developments and trends will drive the need of Big Data analytics in insurance

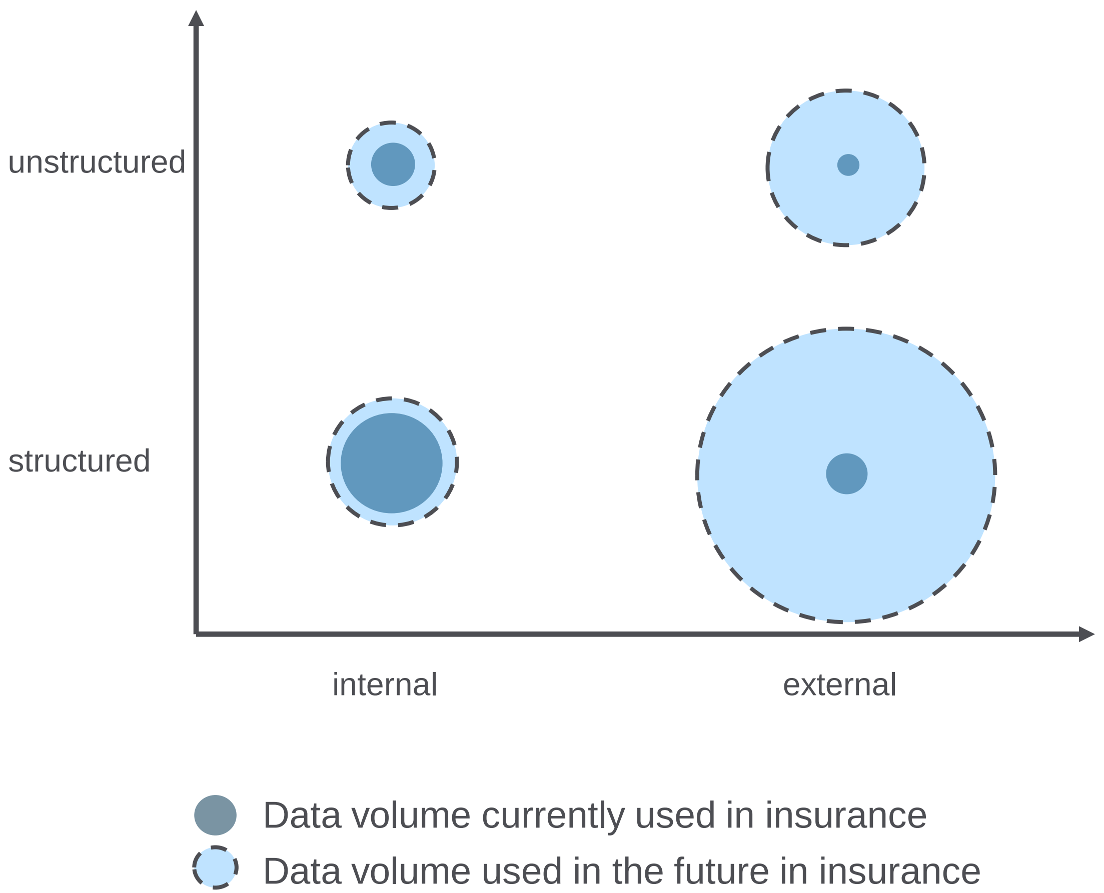
MUNICH HEALTH

Munich RE

Increasing legal & regulatory requirements	Increase in Big Data	Increasing quality of data capturing
High pressure on costs in developed markets	**Increasing importance of data driven decisions in the future**	Decreasing prices for data storage
High top-line competition in emerging markets		Increasing computing power

We expect to observe a strong increase in the availability of external and structured data

Data relevant for insurance companies

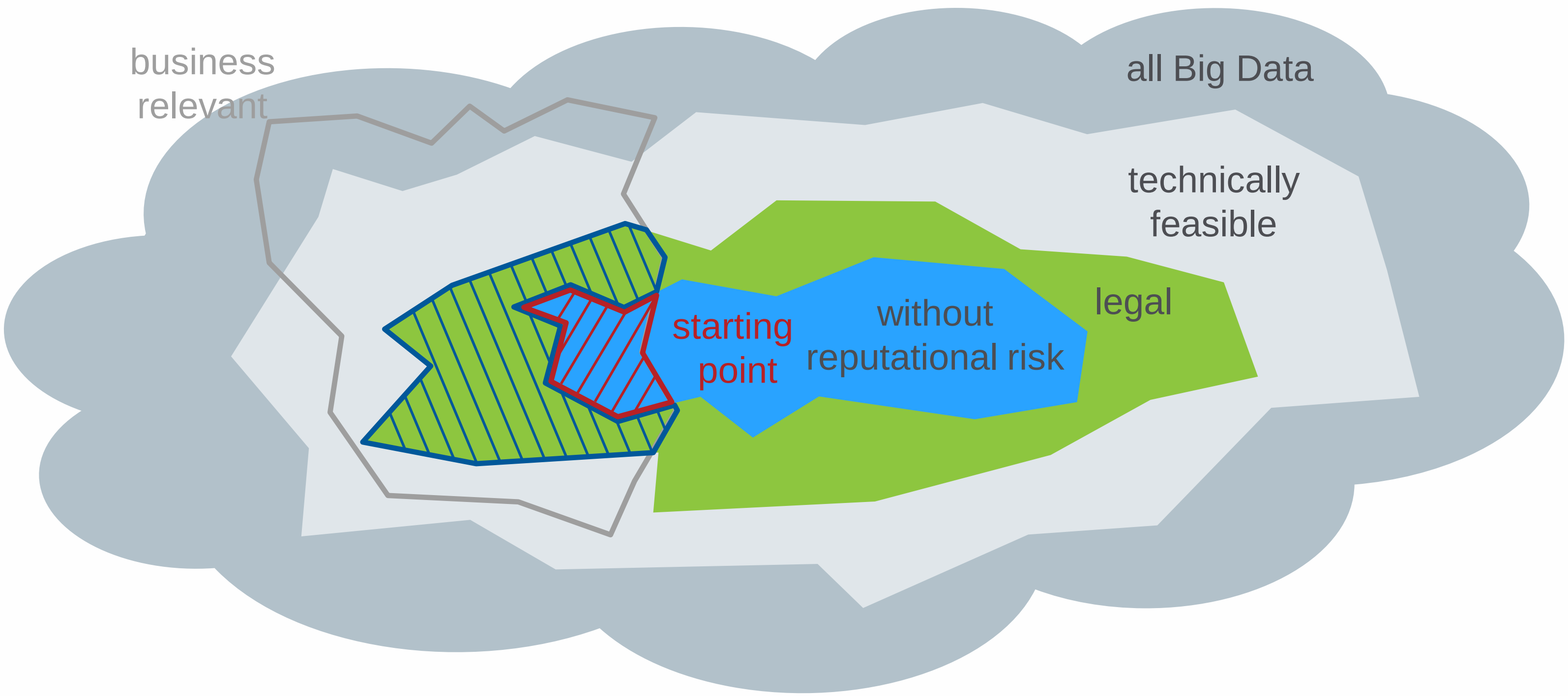


- Current focus are structured internal data.
- The volume of external data will increase strongly.
- The combination of int. and ext. data will gain more and more importance.

Source: Munich Re

Starting point for Big Data is business relevant and usable information

Subsections of Big Data

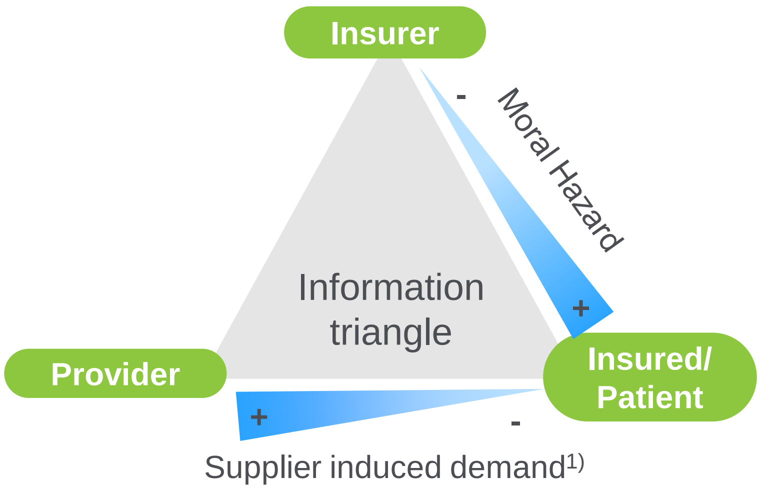


First step on the way to make use of Big Data is the identification of business relevant information which does not present the risk of reputational damage

Source: Munich Re

Beside Big Data, known and new developments require data Analytics in health insurance

Big Data



Information asymmetry between players in the market

Technical development (esp. hardware)

Legal / regulatory environment

1) Supplier induced demand: physicians choose methods and intensity of treatments which patients would not choose if they had the same information as the physician

Superior data Analytics is key to generate effective measures to successfully steer health care business

Source: Munich Re

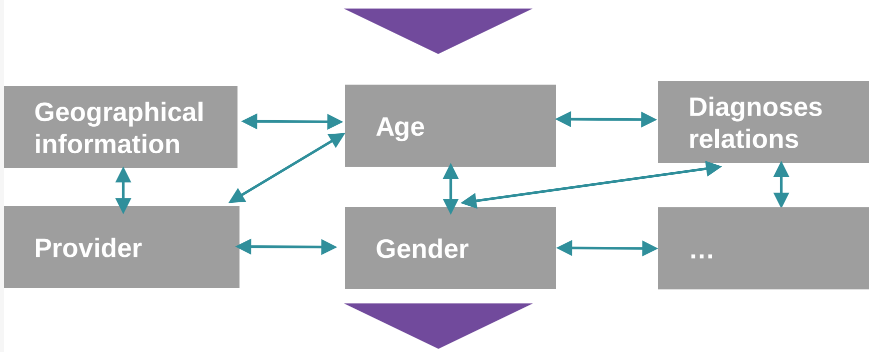
Business Analytics is based on different statistical approaches predicting individual outcomes

MUNICH HEALTH

Munich RE

Within Business Analytics different statistical approaches and parameters are used …

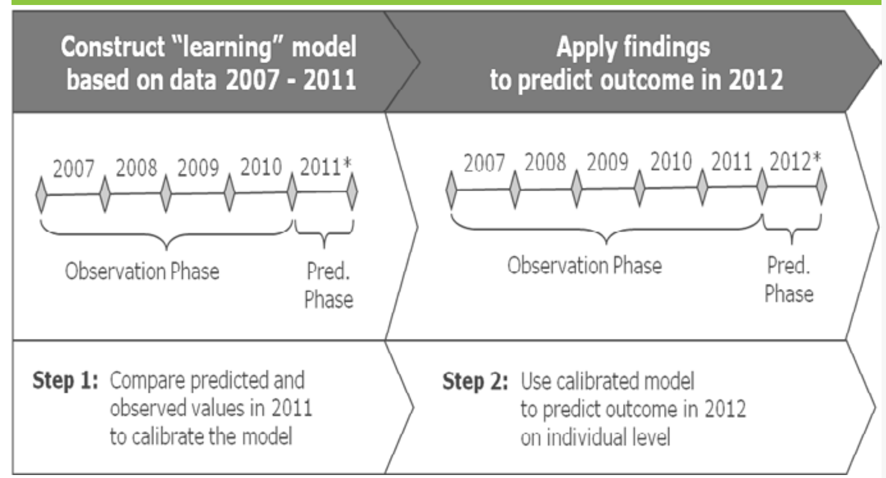
Medical knowledge



Techniques: regression, decision trees, etc.

… to predict individual outcomes based on extrapolation of historical data

We construct models in two steps:



Predictive models for adverse outcomes on an individual basis

Source: Munich Re

Analytics enables data-driven business steering in various areas

MUNICH HEALTH

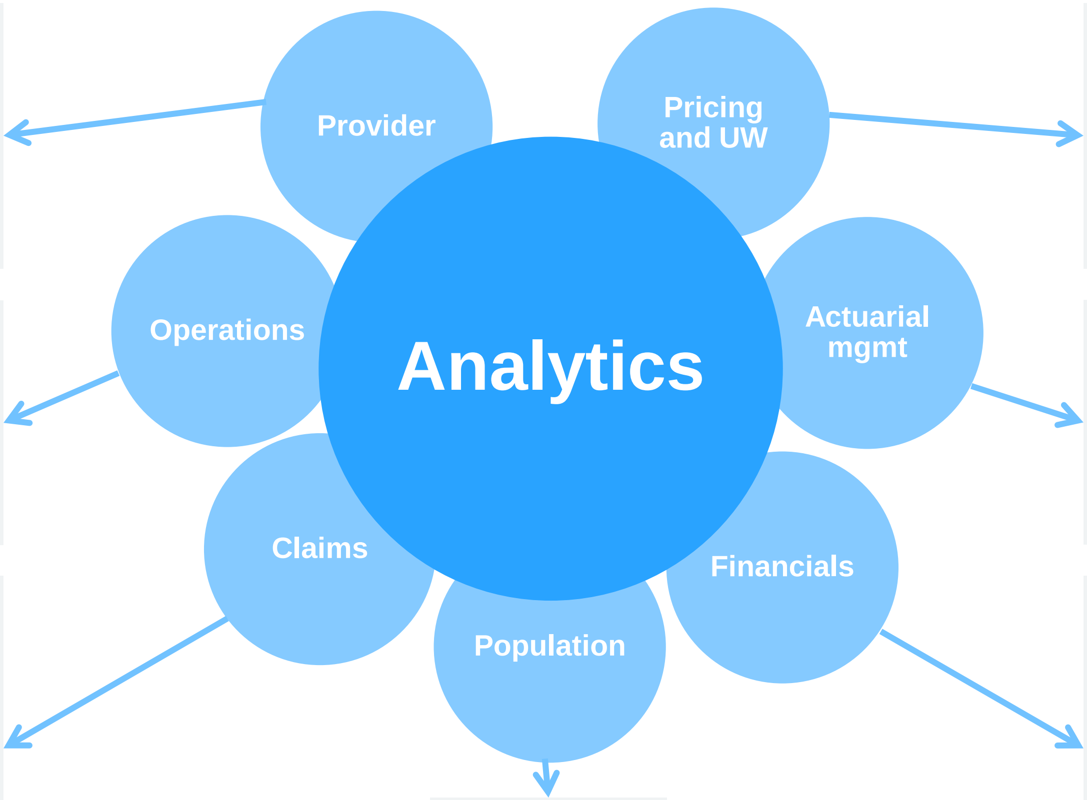
Munich RE

Impact of Analytics: High, Mid, Low

- NW mgmt
- Medical quality
- Discounts
- Access

- Claims / UW and policy processes
- Backlog mgmt

- Fraud
- Waste and abuse
- DMPs
- Medical trends / cost drivers



- Sales
- Marketing
- Lapse

- Product development
- Medical UW
- Price
- Renewal
- Standard UW

- Reserving
- Trends
- Incurred LR
- BC analysis

- Capital
- Forecast
- ORSA
- Financial LR
- Reporting
- Econ Ledger
- ALM
- Reinsurance

Methods and techniques

Image: used under license from Shutterstock.com

Advanced Analytics is based on other methods and data than traditional reporting

MUNICH HEALTH

Munich RE

COMPLEXITY & INFORMATION

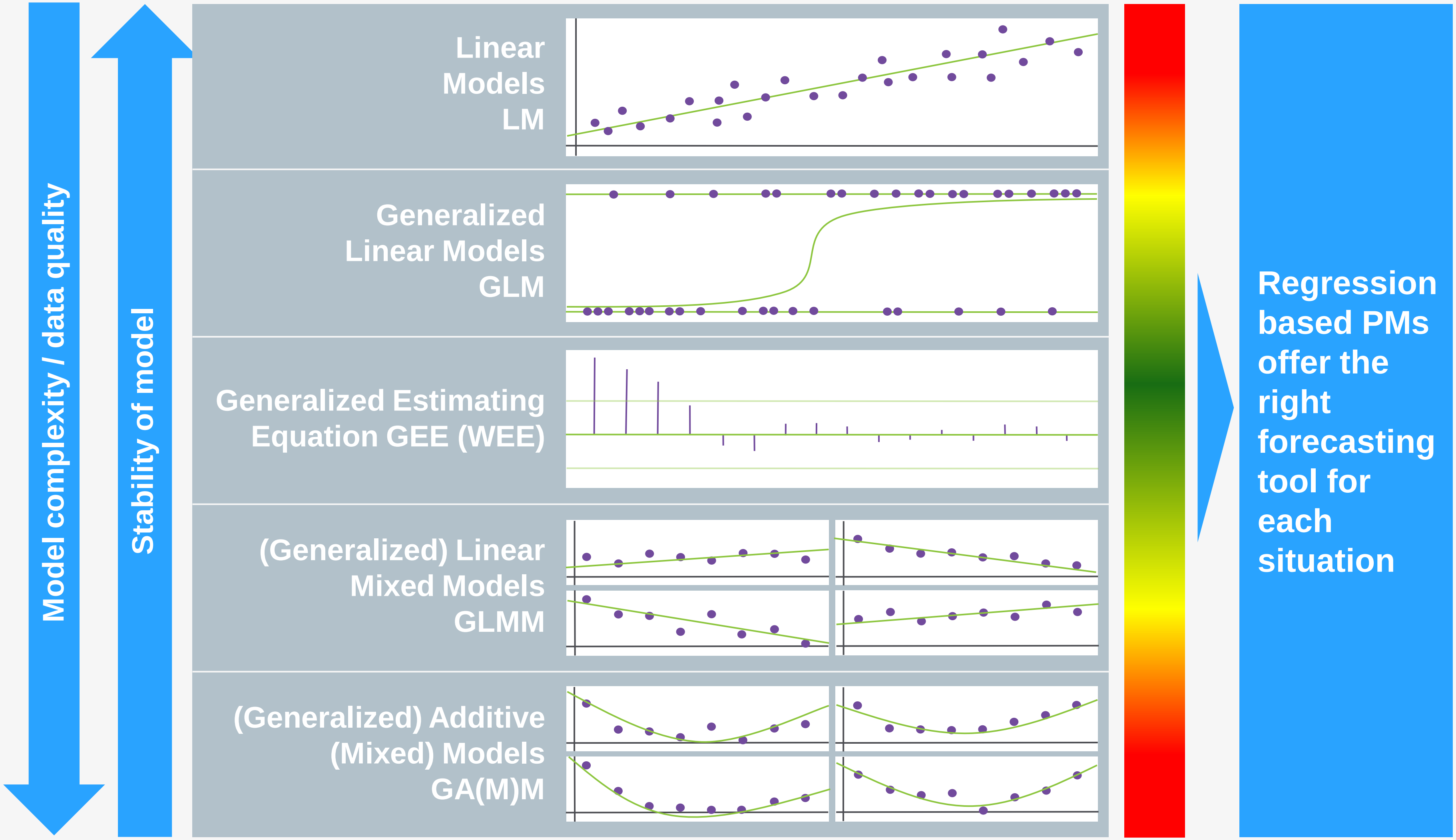
	Explain the past			Predict the future		
	Standard reporting	**Ad Hoc OLAP**	**Statistics**	**Fore-casting**	**Classical predictive models**	**Machine learning**
Question	What happened?		Why did it happen?	What will happen?	To whom will it happen?	To whom will it happen?
Example	BC development, member counts, …	Members per plan and region, ...	Large loss analysis, member groups, ...	LR- projection, seasonality analysis	Churn scores, fraud scores, ...	Medical paths, complex pattern, ...
Methods	Sum, max, min, average, count, …		Distribution theory, correlation analysis, ...	Time series theory, stochastic processes, ...	Regression models, decision trees, …	NeuralNetwork, random forests, SVM,..
Data	Already powerful with small sample size and little information		Medium sample size & little information		Larger sample size & more information	Very large sample size and a lot of information
	Business Intelligence		Business Analytics			
	Reporting		Statistical Analysis		Advanced Analytics	

Pros/Cons of different regression techniques for (health) insurance data

MUNICH HEALTH Munich RE

	Advantage	Disadvantage
Advanced Regression Analysis	+ **stable results even for small datasets and bad data quality** + **overfitting can be controlled** + **analyzes the significance of effects, quantifies the influence of all factors and excludes non-relevant factors by variable selection procedures** + **non-linear effects** can be respected (e.g. GAM) and controlled (e.g. degree of non- linearity by penalization techniques) + works also for not independent response (e.g. GEE)	**especially these 3 advantages make regression models a valuable tool for (health) insurance data** - requires pre-selection of reasonable inter-action terms before variable selection can be used - model convergence is not guaranteed, if complex structures are specified
Decision Trees / Random Forest	+ **detects complex relationships** in large datasets + returns the „variable importance“ of all factors + very stable algorithm produces always results	- tends to overfit training data in case of small datasets and bad data quality - remains a black box to the user and results can hardly be interpreted - does not check the significance of effects or spurious relationships
Neural Networks	+ **detect complex non-linear relationships** and **correlation structures** in large datasets + very stable algorithm + Very powerful with extremely large datasets and technical environment	- tend to overfit training data in case of small datasets and bad data quality - remain a black box to the user and results can hardly be interpreted - quality can only be evaluated by predictive results

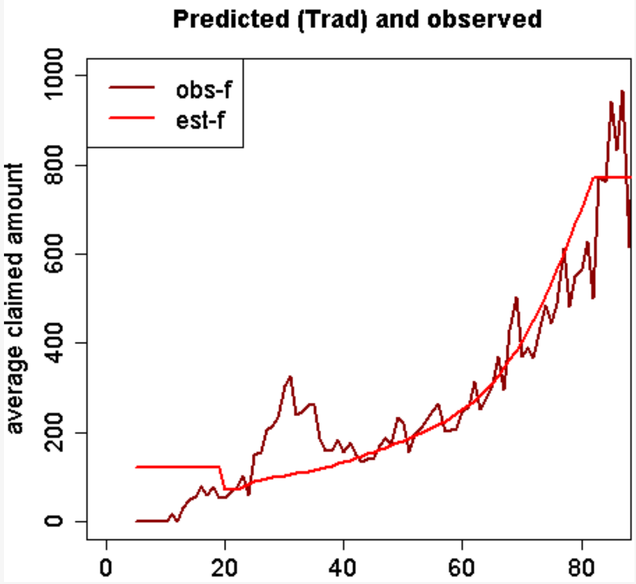
It is crucial to find the right trade-off between stability and complexity to optimize prediction power



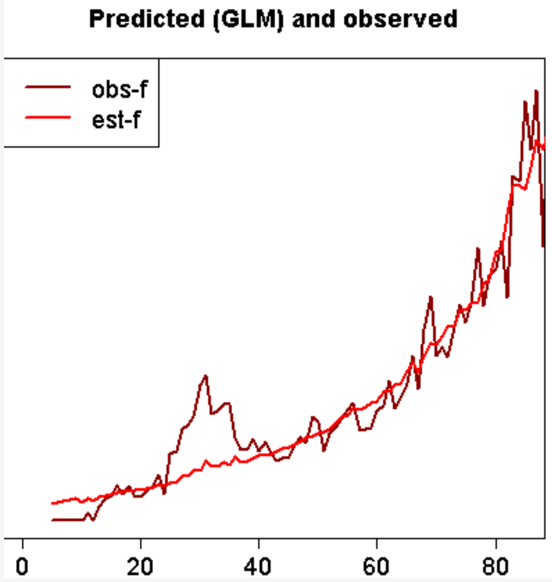
Predicted vs. observed claimed amounts for particular subgroups allows optimal model choice

Observed and predicted average claimed amounts in 2012 itemized by age and gender (here only women)

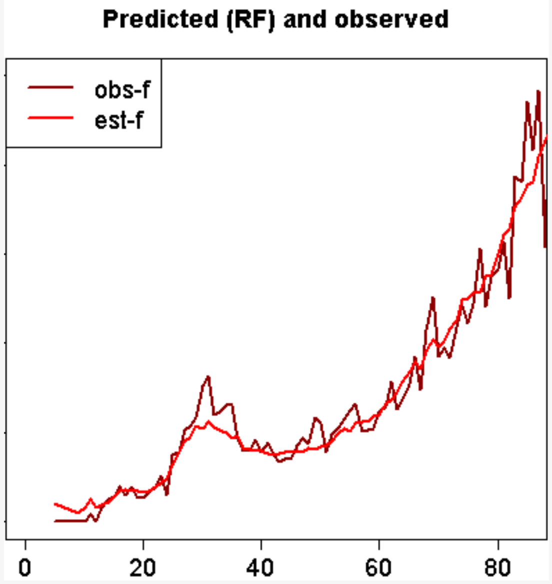
Traditional approach



GLM



Random Forest



The traditional approach takes age and gender into account and therefore mostly performs quite good on average. Only the random forest detects the peak for women in their thirties (pregnancy treatments)

Source: Munich Re

The most important variables to predict the claimed amount vary over the statistical methods

Zero-inflated gamma (GLM)

Logit model:

- number of hospital services
- number of diagnosis in ICD-chapter 13,

5 (mental behavioural disorders)
and 10 (respiratory system)

- unemployment rate

Gamma model:

- length of stay
- age
- unemployment rate
- education rate
- residence area

Random forest

- previous claimed amount
- average payment delay
- number of hospital services
- age
- number of diagnosis in ICD-chapter 13 (musculoskeletal system / connective tissue)
- number of chronic diseases

CRM

Image: plainpicture/Photo Alto/Sigrid Olsson

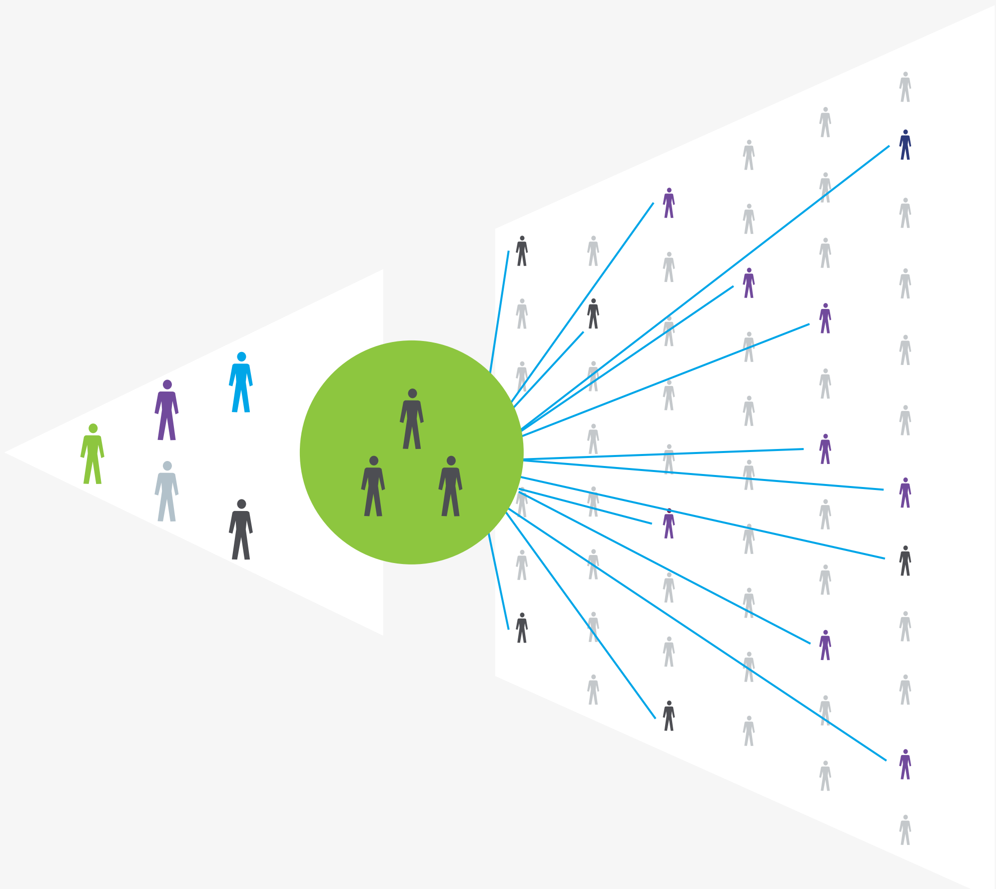
Business analytics improves the effectiveness of CRM throughout a customer's life cycle

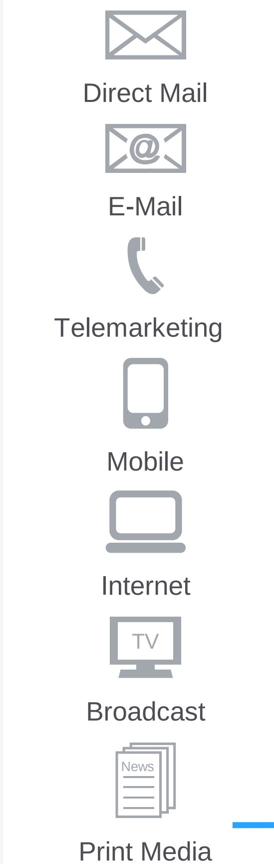
	Acquisition	Up-selling	Cross-selling	Churn prevention	Win-back
Target	Identification of relevant target groups ("good risks") Optimized acquisition of new customers	Identification of customers with highest probability to extend their existing contract Higher penetration of health products	Identification of customers with highest probability to buy additional products Higher penetration of other products	Identification of customers with low product retention Avoidance of cancellation of "good risks" Increase of customer loyalty	Identification of former customers with high capital value and high win-back-probability Enhancement of the portfolio's profitability
CRM solution	▪ Customer profiling and clustering based on external and internal data ▪ Sales channel preference analysis ▪ …	▪ Scores to identify potential up-selling candidates ▪ Sales channel preference score ▪ Costumer value analysis ▪ …	▪ Scores to identify potential cross-selling candidates ▪ Sales channel preference score ▪ Costumer value analysis ▪ …	▪ Churn scores to identify "disloyal" customers ▪ Next best action analytics ▪ …	▪ Win-back scores to identify customers most likely to return ▪ …

Source: acxiom

Client segmentation combined with Predictive Models for media preference accelerate profitable growth

1. Existing customers	2. Identify and profile best customers	3. Find others like them	4. Model media preferences	5. Accelerate profitable growth





Steps

1. Identify customers' media preferences
2. Predict customers' channel response
3. Maximize channel productivity
4. Accelerate return on marketing investment

Predictive Modeling for media allocation and analysis of channel interplay

Source: acxiom

Churn prevention is optimized by constructing score groups based on internal and external data information

MUNICH HEALTH

Munich RE

Churn-index is derived from score groups …

Internal Data

- Insurance product
- Rate
- Gender
- …

External Data

- Socio-demographic
- Income status
- Family status
- …

Score groups

C
- Married
- 2 children
- Age 30-50
- Client since 10 years
- …

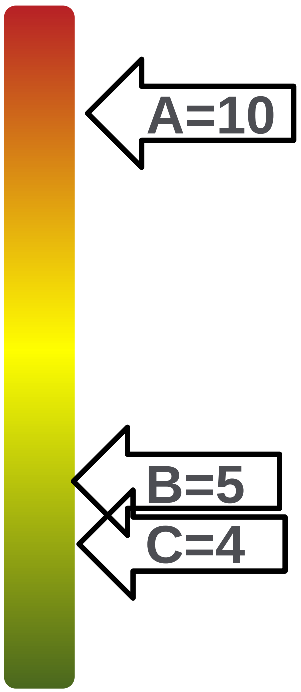
B
- Male
- Single
- Income > 3000
- Product ABC
- …

PM

…

A
- Laborer
- Married
- No Children
- Complained last month
- …

Churn index

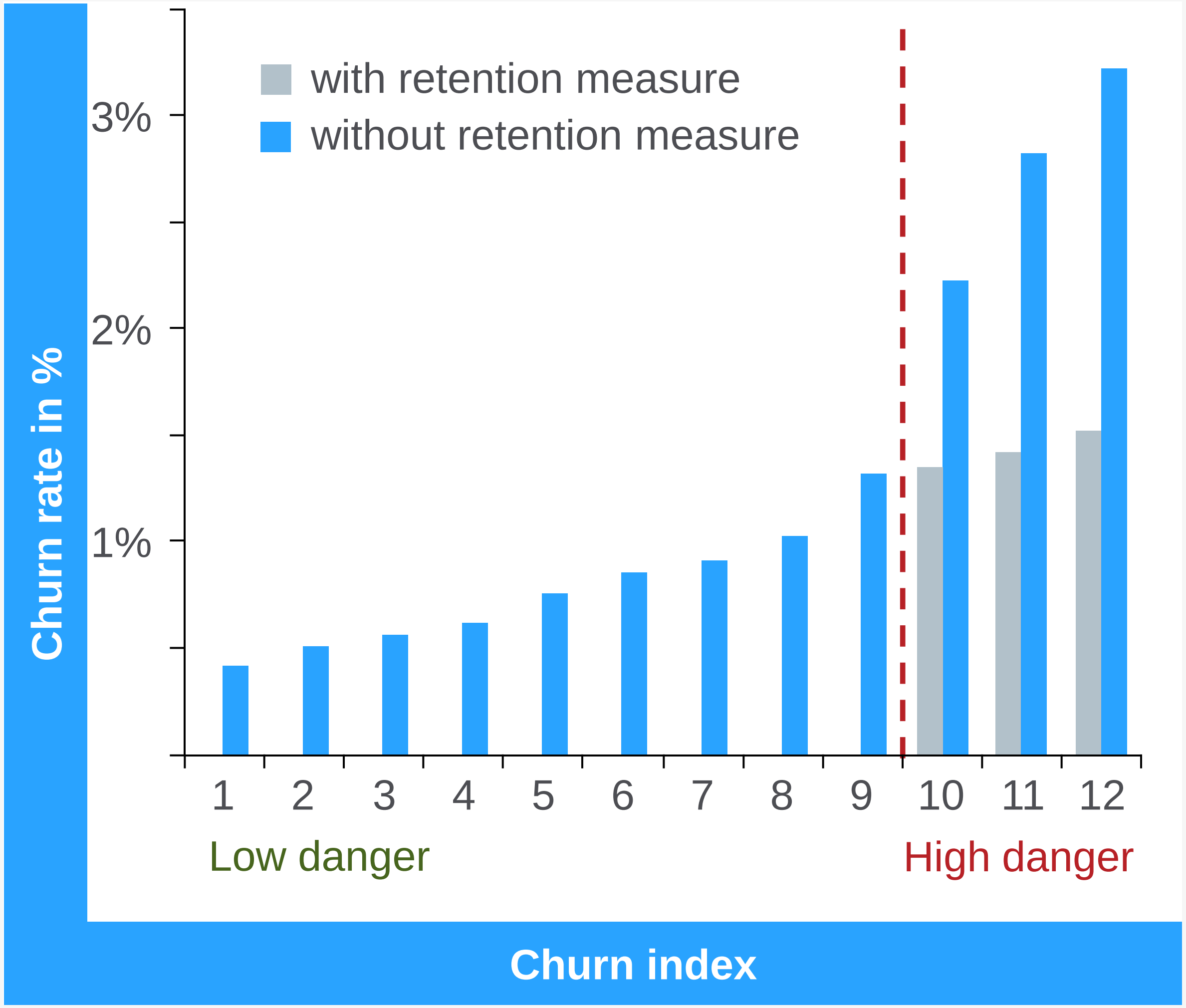


Source: acxiom

Customer retention measures based on external churn prevention models have proven their effectiveness

MUNICH HEALTH

Munich RE



Success

- Reducing **costs by 75%** through targeted measures
- Reducing **the churn rate by 47.5%** in the three highest churn index classes

Source: acxiom

Pricing

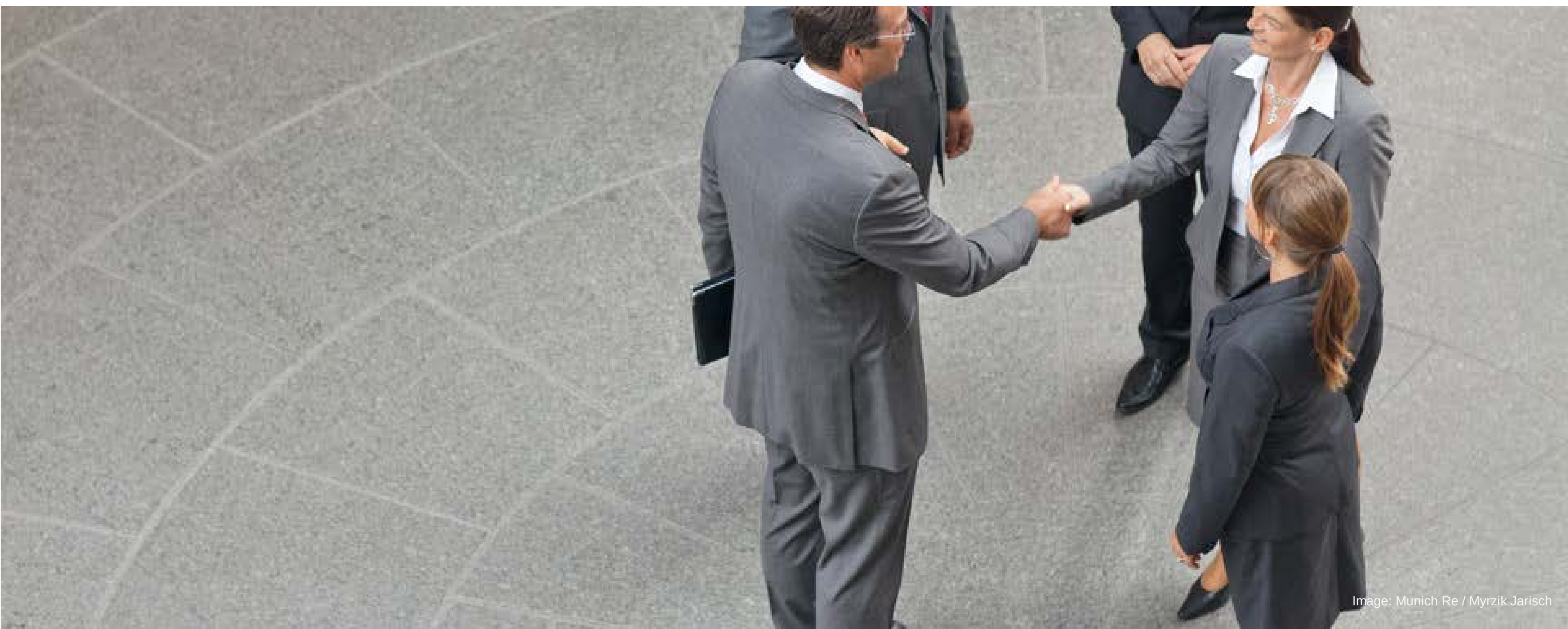


Image: Munich Re / Myrzik Jarisch

PM-based pricing yields a stronger differentiation of the risk structure to optimize premium calculation

MUNICH HEALTH

Munich RE

PM-based pricing

Existing portfolio

High degree of individual risk adjustment attracts good risk and increases lapse of bad risk.

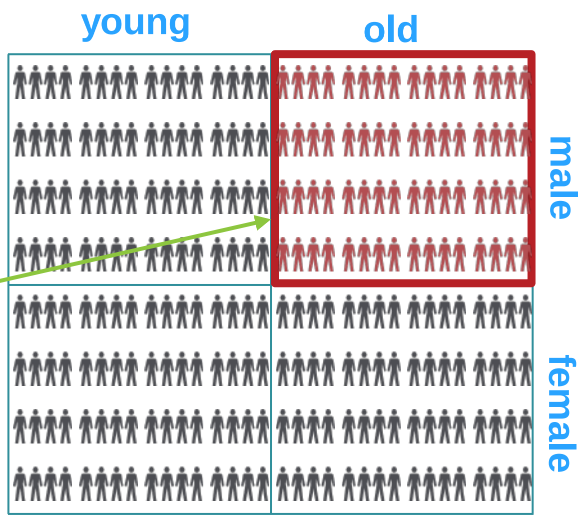
e.g. based on age, gender, profession, marital status and region

e.g. only based on age and gender

New applicant

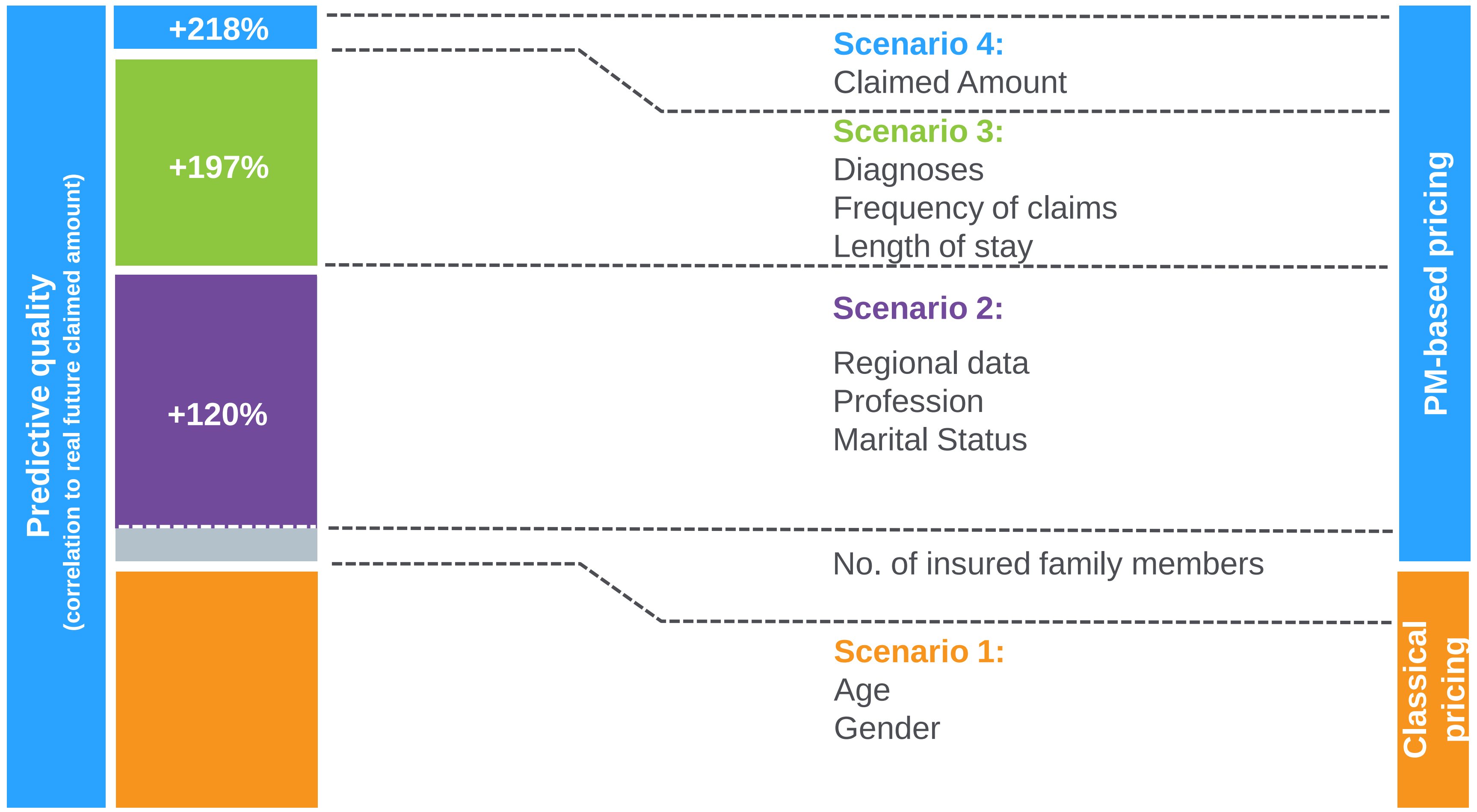
- 55 years
- male
- teacher
- married
- Munich

Classical pricing



Low degree of individual risk adjustment attracts bad risk and increases lapse of good risk.

Predictive quality depends on information used

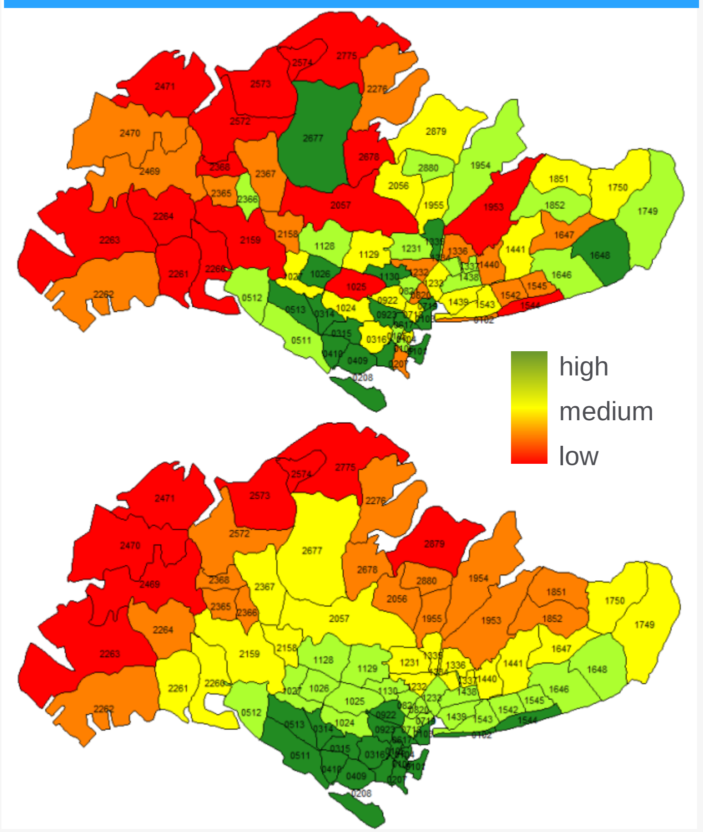


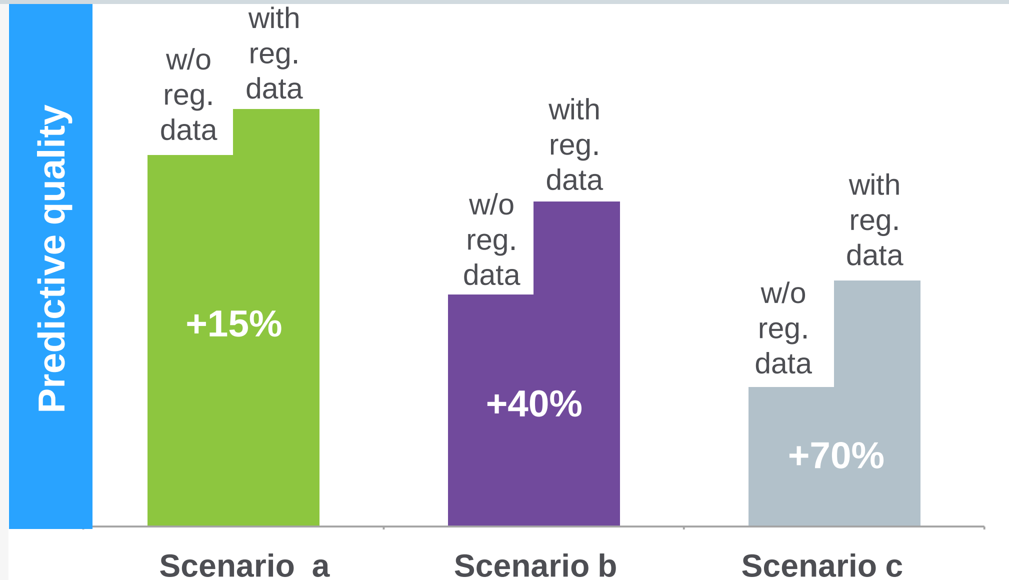
Source: Munich Re

Usage of clients' address data combined with ext. regional data further improves predictive quality

In order to balance random variations of external regional data, spatial smoothing techniques are applied:

Example:Income per capita in Singapore





External socio-demographic and -economic regional data strongly contribute to a more precise prediction of future costs, especially if preceding costs and the claims history are not taken into account.

BUT: External regional data can only be allocated if clients' address data are available!!

Product development

Social Media Analysis for Product Development:
Example 1 - Supplementary dental cover

MUNICH HEALTH

Munich RE

1 **Collection of external and combination with internal data**

2 **Statistical analysis of the created databases**

3 **Application of results in different fields of operation**

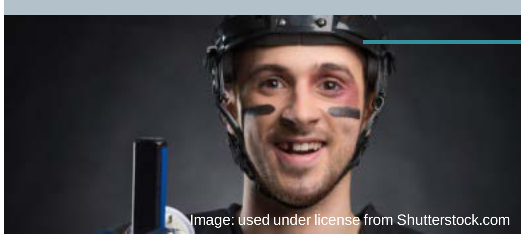
1 Identify risk groups through relevant activities & match with internal data, e.g. via facebook

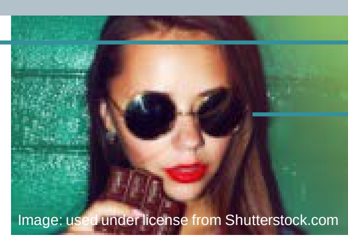
Membership in facebook groups

Smoker Association | Members | Photos | Join group | + Create Group

Profile pictures

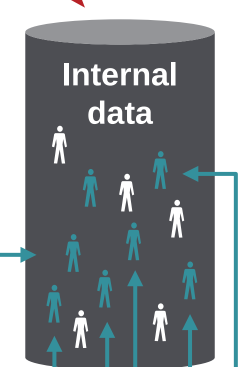
Usage of Image Mining!





Likes and posts

Exact matching or similarity matching



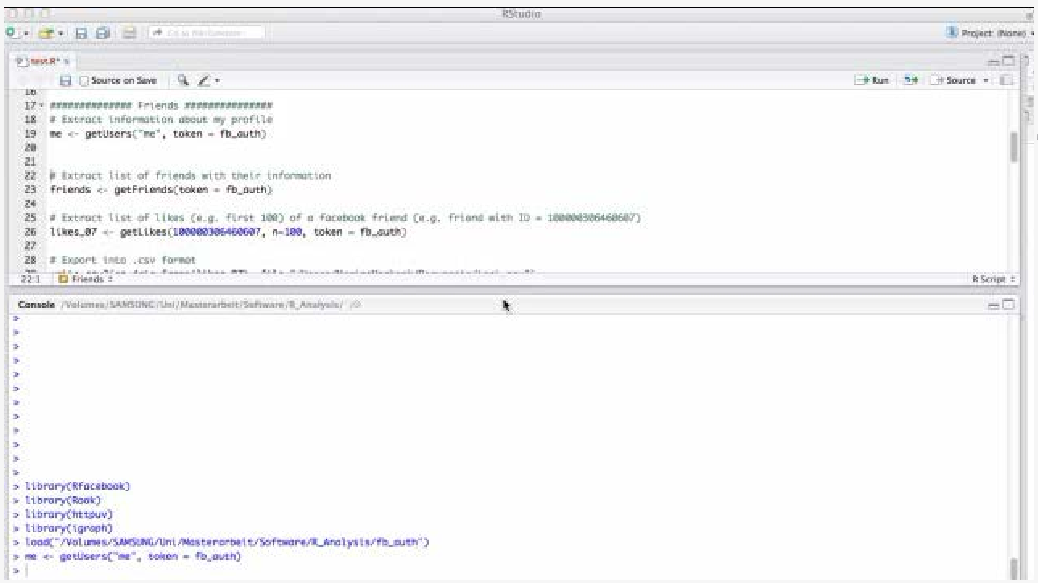
Social Media Analysis for Product Development:
Example 1 - Supplementary dental cover

1 Collection of external and combination with internal data

2 Statistical analysis of the created databases

3 Application of results in different fields of operation

2 Analyse enriched database to identify potential buyers



3 Send offer to identified target group

79%

76%

69%

1%

List of persons with highest probability to buy supplementary dental cover including individual risk profile

Existing clients:
Targeted up-selling campaign by sales department

Potential clients:
Design specific product to attract likely buyers with moderate risk

Source: Munich Re

Social Media Analysis for Product Development:
Example 2 - Supplementary inpatient cover

MUNICH HEALTH

Munich RE

1 **Collection of external and combination with internal data**

2 **Statistical analysis of the created databases**

3 **Application of results in different fields of operation**

1 **Identify risk groups through relevant activities & match with internal data, e.g. via YouTube**

List of 10 most dangerous sports*

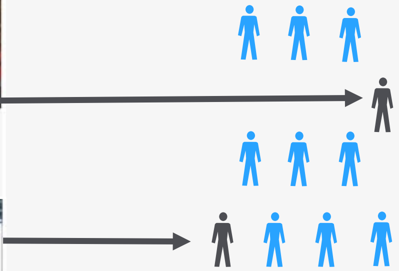
Rank	Sports
1	Base Jumping
2	**Heli-Skiing**
3	Diving
4	Cave Diving
5	Bull Riding
6	Big Wave Surfing
7	Street Luging
8	Mountain Climbing
9	**BMX**
10	White-Water Rafting

YouTube uploads

Usage of Video Mining!

Exact matching or similarity matching

Internal Data



Analytical approaches can be applied after combining internal and external information

MUNICH HEALTH   Munich RE

1 Collection of external and combination with internal data

2 Statistical analysis of the created databases

3 Application of results in different fields of operation

Exact matching

Existing client

John Baker
- male
- born 20.12.1968
- Toronto, M4B 1B3

John Baker
- male
- born 1968
- often located in M4B 1B3
- Member of smokers group in facebook, profile picture with cigarette

Similarity matching

Existing client

John Baker
25 years
likes NFL, skiing, iPhone
bought accident insurance

Simon Waters
26 years
likes MLB, iPhone
searched for life insurance

...

Peter Miller
26 years
likes NFL, free skiing
searched for accident insurance

- Clustering and profiling based on behavior and needs (and customer value/ claims experience)
- Life-Style segmentation
- Economic segmentation based on commercial data

Collection of external data – Compilation of filtered information into a structured tabular form

MUNICH HEALTH

Munich RE

Name	Date of birth	Place of residence	Hobbies/ Interests / Events	Groups / Likes in connection with		Posts / comments etc. containing information in connection with		Images / Videos	
				Dental	Hospita-lization	Dental	Hospita-lization	Dental	Hospita-lization
John Baker	20.12.1968	Toronto ON (Canada)	Watching movies	Marlboro	Bungee Jumping, Skydiving	-	+ Freefall	2 x cigarette	-
Sophie Williams	13.04.1972	Princeton NJ (USA)	Attended: survival tour	Snickers, Bounty	-	-	+ Survival Tour	-	3 x Chocolate Bar
Michael Smith	04.08.1966	Miami FL (USA)	Paintball, Attended: UFC	Street Fights		+ Boxing		-	-
Michael Smith	04.08.1966	San Diego CA (USA)	Attended: Wine tasting	-	-	+ Red wine	-	-	-

Extracted structured data

Extracted and processed unstructured data

Key take aways

Social media data is easily to extract, based on freeware

Data cleansing and structuring is a manual process

Merging internal and external data requires analytical capacities

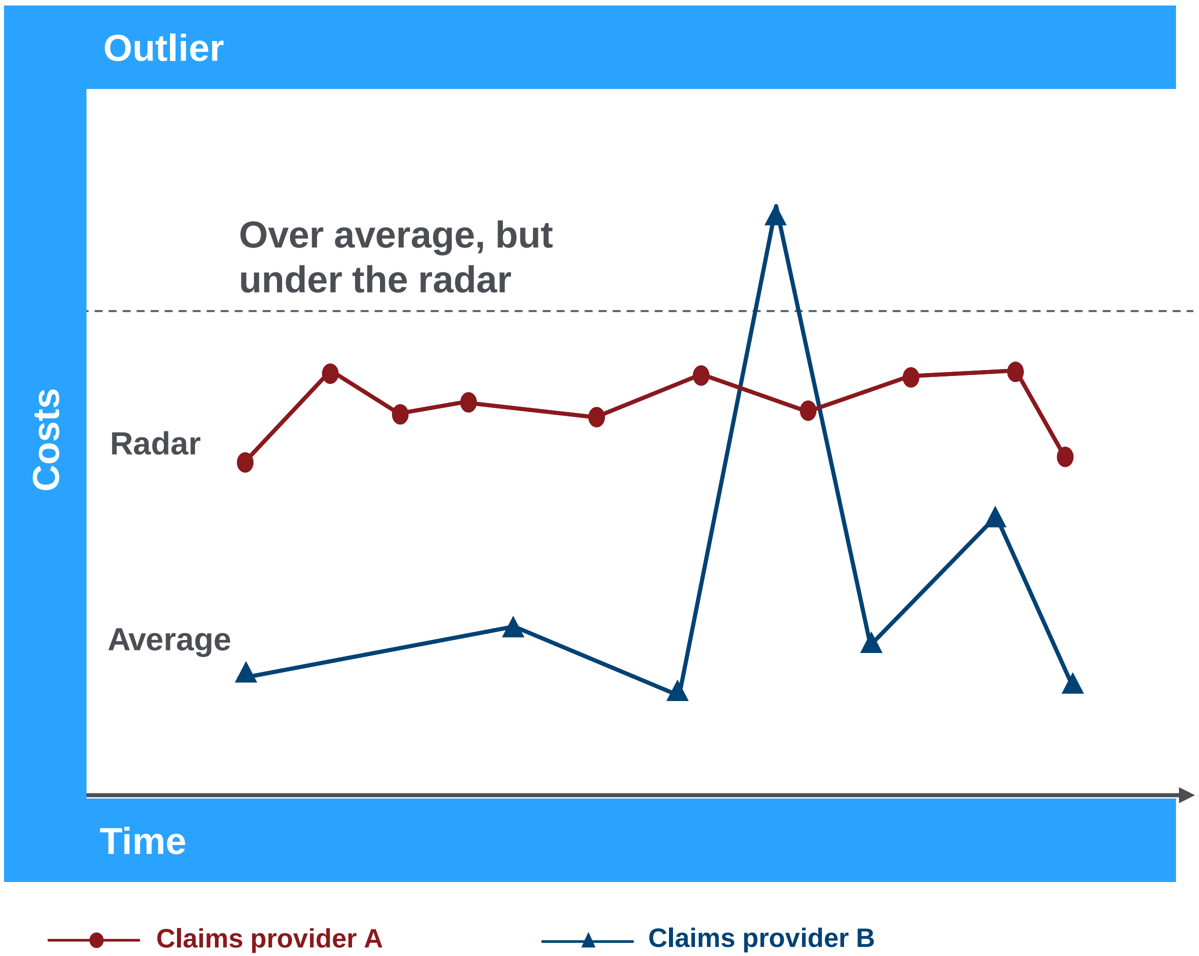
Analysis of merged database requires analytical and other experts

Building up internal knowledge about social media data analysis is a key issue!

Fraud and abuse

Image: plainpicture/fStop/Sven Hagolani

Challenge: identification of professional fraud and abuse



Outlier will be controlled

Using simple analysis leads to:

- outlier claim of **Provider B** will be investigated
- abusive behavior of **Provider A** will not be identified

Different methods and techniques are needed to control fraud and abuse

Source: Munich Re

Further development: deterministic and probabilistic rules make real-time decisions possible even today

MUNICH HEALTH
Munich RE

Deterministic rules (n Rule Types)

Rule Type	Diagnosis - Treatment	Diagnosis - Drug	Treatment - Gender
Rule Edits	Appendicitis & appendectomy	Respiratory infection & Augmentin tb.	Ultrasound prostate & male
	Appendicitis & tissue sample	Respiratory infection & Norvasc tb.	Ultrasound prostate & female
≈ 2,500,000 Rule Edits			

- Fully-automated check of medical services
- Recognition of medical invoices for false or unnecessary treatments / prescriptions (international experience value: 12% of the invoice amount)
- Integration of the insured person's claims history into the decision-making process

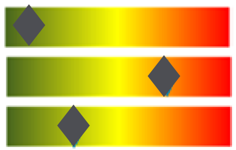
Probabilistic scores

- **Real-time decisions on submitted invoices**
- **Calculation of the scores in three categories:**
- Unperformed services
- Unjustified services
- Billing issues
- **Basis for decision: Invoice data and results of the retrospective report**

Live Score Box

Probability for ...

... unperformed services (UPS)
... unjustified services (UJS)
... billing issues (BI)



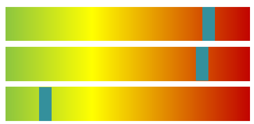
Category	Reason
UJS	Provider score
UJS	Network score
UJS	Diagnosis / provider specialty / procedure

Example from Indonesia: Infectious diseases

Scoring

Live Score Box

Probability for …
Unperformed services (UP)
Unjustified services (UJ)
Other billing issues (BI)



Category	**Pattern**	**Report**
Provider	UP, UJ	Risk adjusted LoS
Provider	UJ	# certain diagnoses
F&A stat.	UP	# reportable diseases

Specific evaluation

Rank	Hospital	No. of claims	Sum real LoS	Sum exp. LoS	Ratio
1	Hospital 246	920	6,433	3,078	209%
2	Hospital 811	571	3,740	1,833	204%
3	Hospital 611	1,814	11,744	6,711	175%
4	Hospital 309	650	3,569	2,124	168%
5	Hospital 081	1243	7,208	4,742	152%
6	Hospital 743	576	3,492	2,376	147%

- Under consideration of the patients' risk profiles, the average LoS is too long in hospital 246.
- The number of infectious diseases (esp. Cholera and Dengue fever) treated by the hospital is very high compared to other general hospitals.
- The number of Cholera cases billed by the hospital is not realistic compared to the number of reported Cholera cases in Indonesia (WHO figures) and the hospitals market share.

Finding of investigation

- **The hospital** often billed treatments of infectious diseases, because optimal treatment period is not clearly defined.
- In many cases, mild diarrhea was coded as Cholera in order to charge additional room costs.
- In some cases, patients left **the hospital** earlier than reported.

Source: Munich Re

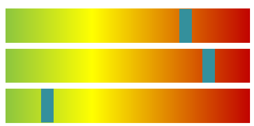
Example from Germany: Referral system

MUNICH HEALTH Munich RE

Scoring

Live Score Box

Probability for …
Unperformed services (UP)
Unjustified services (UJ)
Other billing issues (BI)

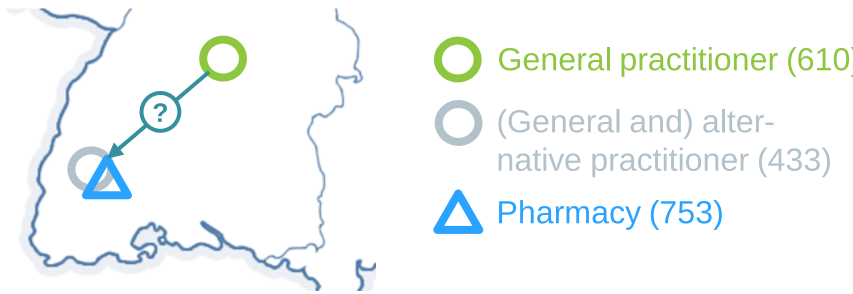


Category	Pattern	Report
Network	UJ, UP	Provider-member-distance
Network	UJ, UP	Network graph
F&A stat.	UJ	# specific drug group

Finding of investigation

- The providers built up a referral system.
- The **GP** recommended his patients to visit the **alternative practitioner** pro-mising cheap homeopathic treatment.
- Patients received non-covered homeopathic products instead of billed cold remedies in the **pharmacy**.

Specific evaluation



- After visiting general practitioner 610, many of his patients, drive 150km (from Stuttgart to the Black Forest) to visit alternative practitioner 433.
- The alternative practitioner prescribes a lot of cold remedies.
- The prescriptions are nearly always filled in a nearby pharmacy (753).
- Compared to the average pharmacy, pharmacy 753 bills clearly more cold remedies.
- The three providers are facebook friends.

Source: Munich Re

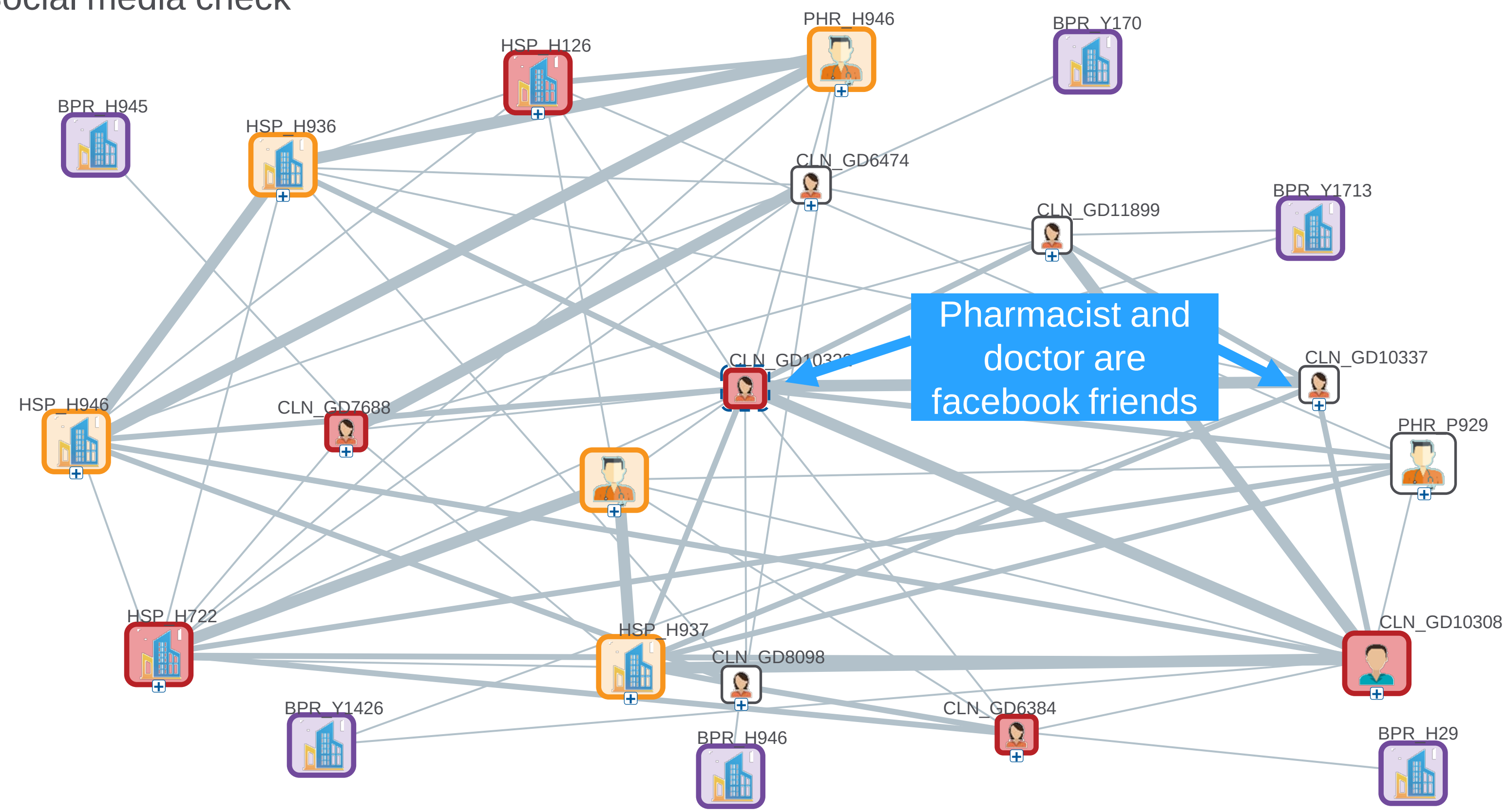
Network reports – Social media analysis checks links between players in the market

MUNICH HEALTH

Munich RE

Two Steps

1. Internal network analysis
2. Social media check



Analytical techniques facilitate an efficient identification of fraud and abuse

The identification of fraud and abuse can be compared with searching a needle in the haystack.

Image: used under license from Shutterstock.com

Analytical methods will not be able to find the needle …

Image: used under license from Shutterstock.com

… however they can make the haystack considerably smaller!

DMP (Disease management programs)

Simplified process of every health program and application of statistical methods

MUNICH HEALTH

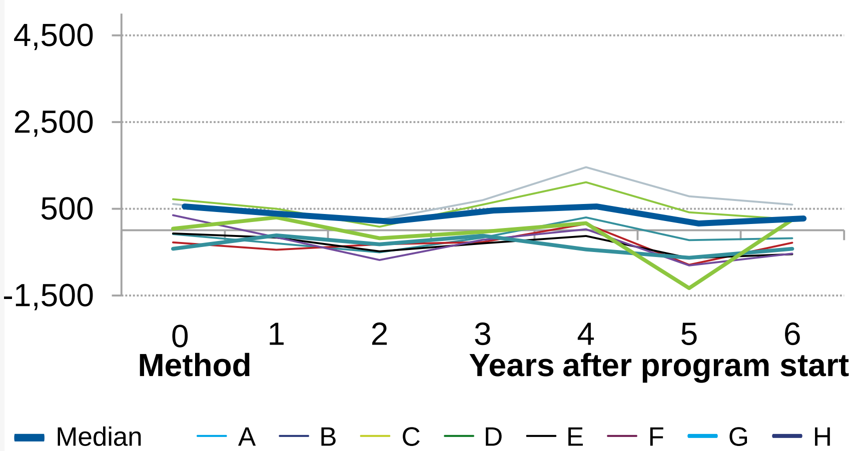
Munich RE

Target definition and program development	▪ Definition of program target and program development ▪ First definition of target groups: determination of indication
Identification	▪ Identification of participants is key success factor: medical and social criteria are crucial ▪ Standard methods usually ignore the expected cost development of potential participants, whereas statistical methods usually improve the identification significantly
Execution	▪ Conflict of interests and privacy protection require the retrieval of participants via the sponsor ▪ Support of participants, compliance monitoring and if needed application of telemedicine by a specialized service provider
Measurement of econ. effects	▪ The economic effect of DMPs is discussed controversially in literature because the effects also depend on the measurement technique ▪ With complex statistical methods uncertainties in the analysis can be reduced ▪ Validation of further aspects: e.g. client satisfaction, medical outcomes, …

DMP assessment

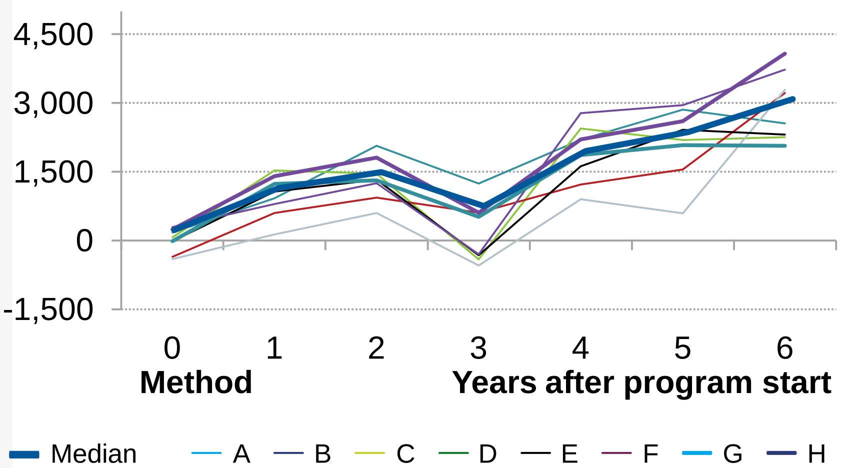
Realizable savings based on different measurement methods

Diabetes (savings in EUR)



The results strongly depend on the measurement method. Most methods do not indicate a long-term saving potential.

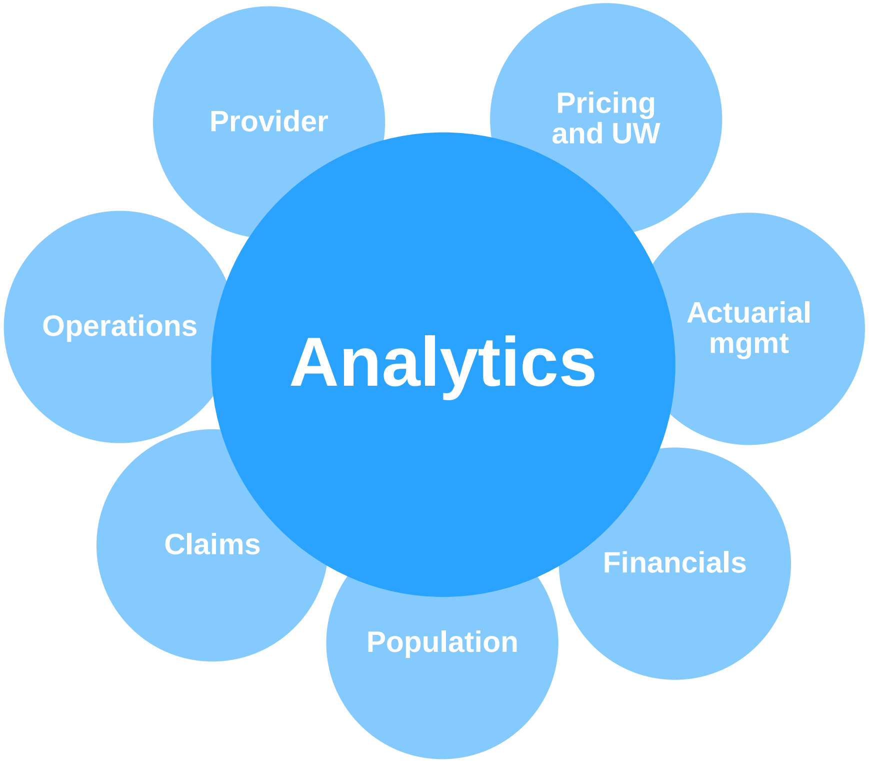
Chronic heart failure (savings in EUR)



Even though there is also a strong variation between different methods, a saving trend starting from the first program year can be detected.

For a reliable measurement, different methods need to be applied.

Closing


Provider
Pricing and UW
Operations
Analytics
Actuarial mgmt
Claims
Financials
Population

Disclaimer

Münchener Rückversicherungs-Gesellschaft (Munich Reinsurance Company) is a reinsurance company organised under the laws of Germany. In some countries, including in the United States, Munich Reinsurance Company holds the status of an unauthorised reinsurer. Policies are underwritten by Munich Reinsurance Company or its affiliated insurance and reinsurance subsidiaries. Certain coverages are not available in all jurisdictions.

Any description in this document is for general information purposes only and does not constitute an offer to sell or a solicitation of an offer to buy any product.

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.