



Institute  
and Faculty  
of Actuaries

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

Stephen Bishop

Daniel Cantzler



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...

### 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

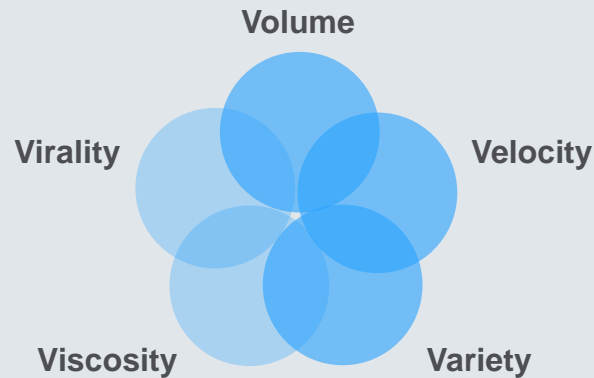
### 1 Terabyte

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

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

# Big Data is a trend that will strongly influence the future

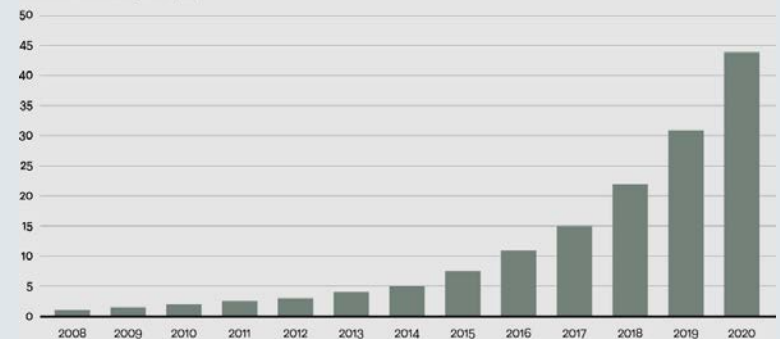
## ... Is defined by 5 V's



Source: Munich Re

## ... is growing strongly

Data in zettabytes (ZB)

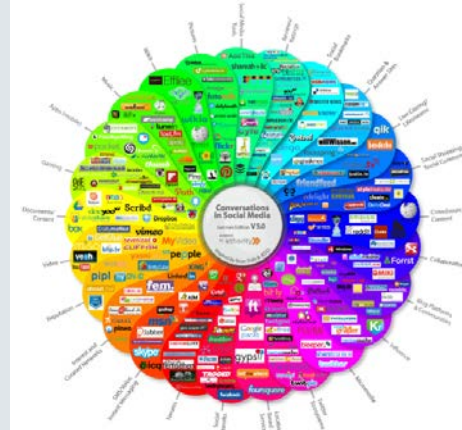


Source: Oracle



Source: Munich Re

## ... is driven by modern technologies & trends



## ... creates a new market environment

# When does it become BIG data?

40,000,000,000,000,000,000,000

Zettabyte

Exabyte

Petabyte

Terabyte

Gigabyte

Megabyte

Kilobyte

Byte

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

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

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

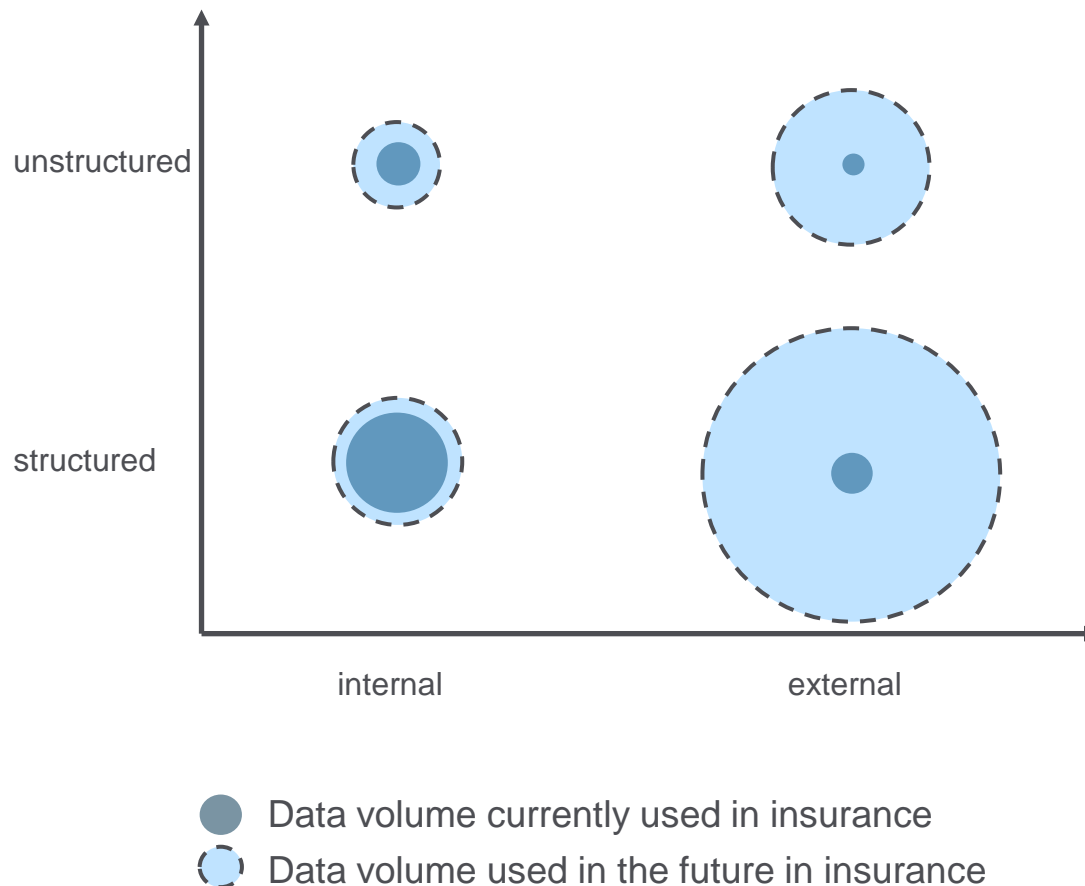
Decreasing prices for data storage

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

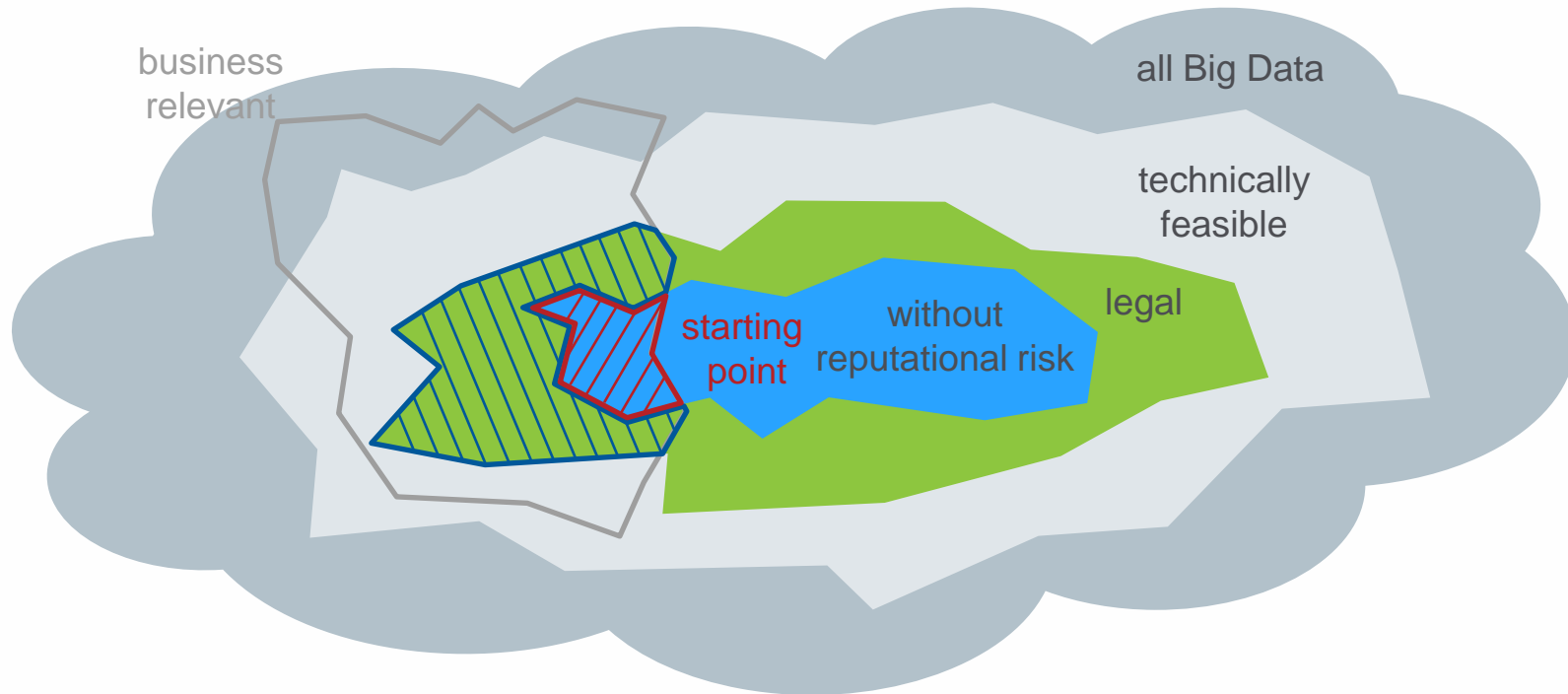
### Data relevant for insurance companies



- 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.

# Starting point for Big Data is business relevant and usable information

## Subsections of Big Data

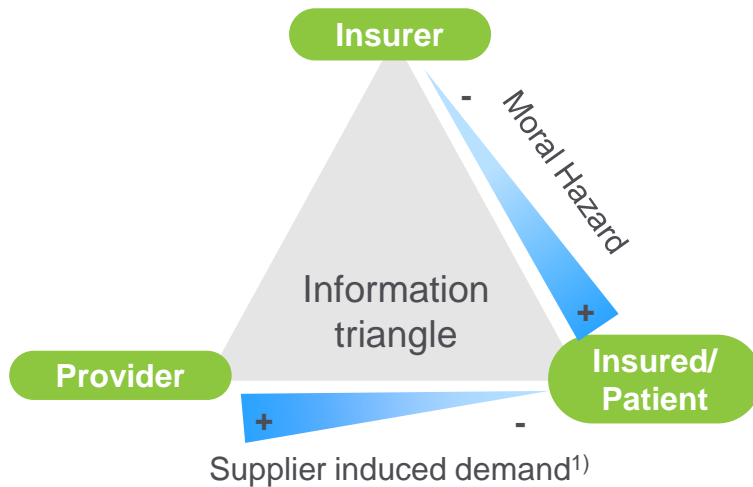


**First step on the way to make use of Big Data is the identification of business relevant information which does not present the risk of reputational damage**



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

## Big Data



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

Information asymmetry between players in the market

Technical development (esp. hardware)

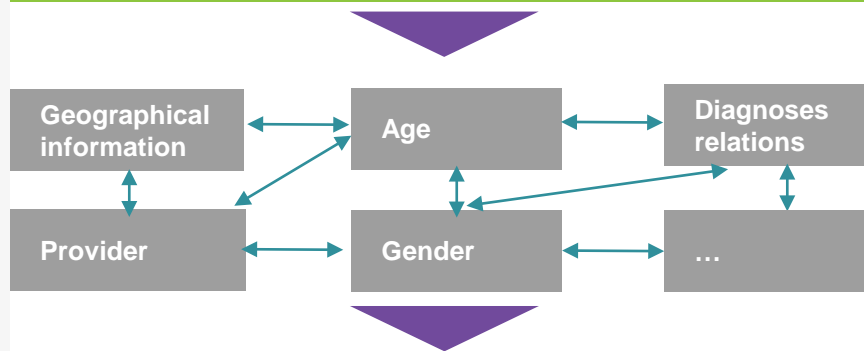
Legal / regulatory environment

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

# Business Analytics is based on different statistical approaches predicting individual outcomes

Within Business Analytics different statistical approaches and parameters are used ...

Medical knowledge

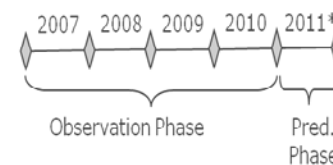


Techniques: regression, decision trees, etc.

... to predict individual outcomes based on extrapolation of historical data

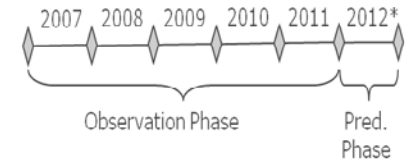
We construct models in two steps:

Construct "learning" model based on data 2007 - 2011



**Step 1:** Compare predicted and observed values in 2011 to calibrate the model

Apply findings to predict outcome in 2012

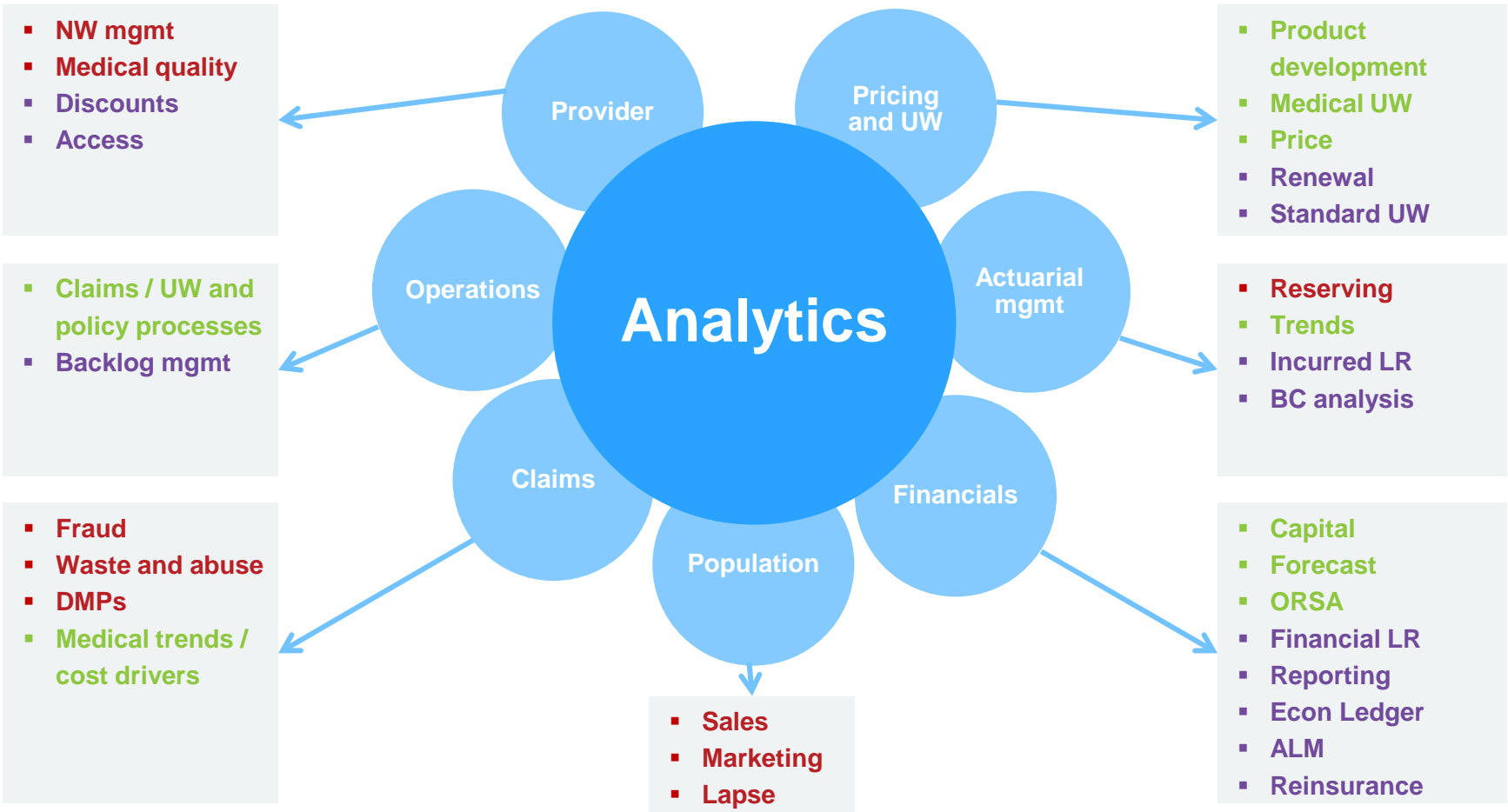


**Step 2:** Use calibrated model to predict outcome in 2012 on individual level

Predictive models for adverse outcomes on an individual basis

# Analytics enables data-driven business steering in various areas

## Impact of Analytics: High, Mid, Low



# Methods and techniques



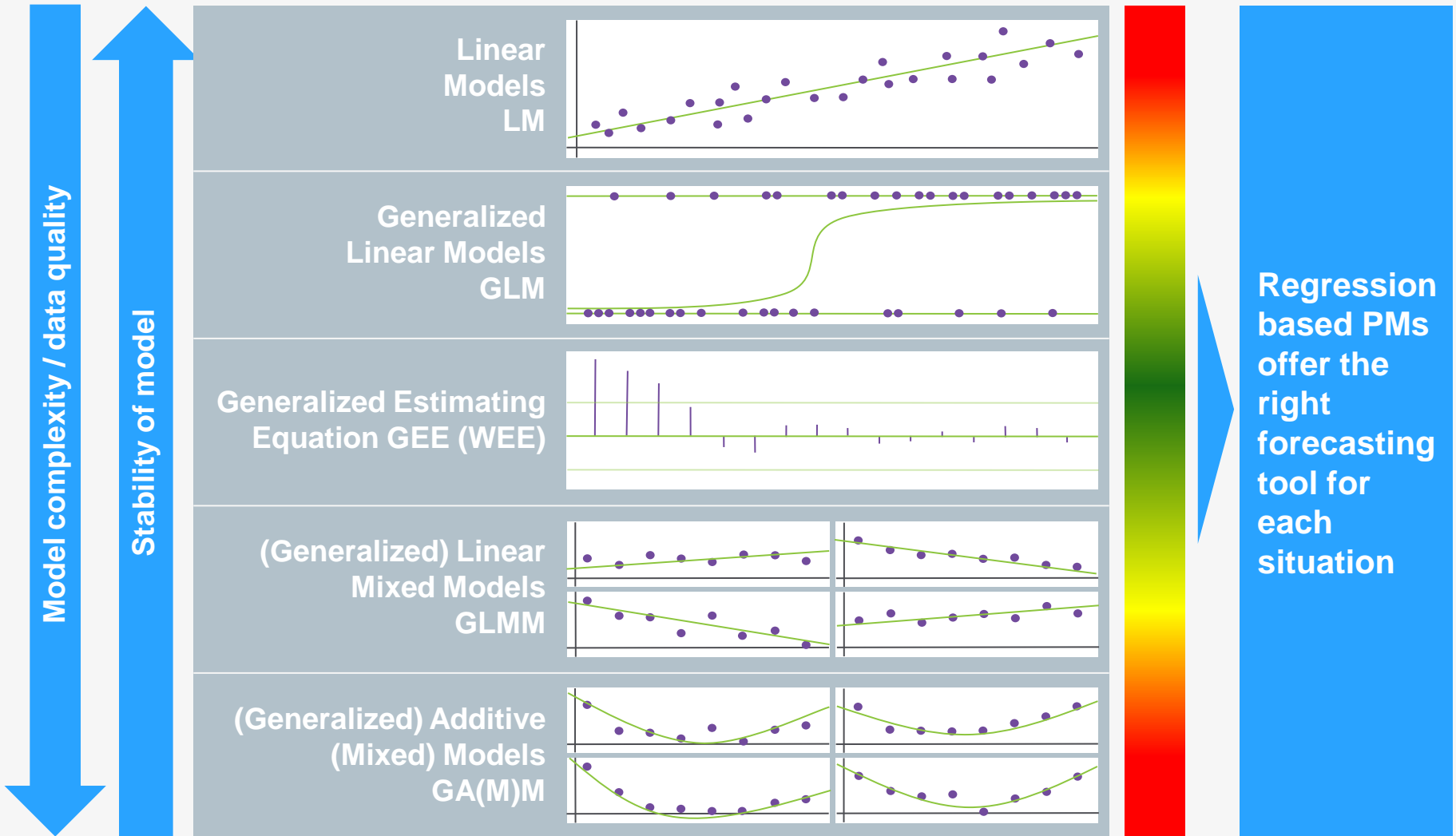
# Advanced Analytics is based on other methods and data than traditional reporting

	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

	Advantage	Disadvantage
Advanced Regression Analysis	<ul style="list-style-type: none"> <li>+ stable results even for small datasets and bad data quality</li> <li>+ overfitting can be controlled</li> <li>+ analyzes the significance of effects, quantifies the influence of all factors and excludes non-relevant factors by variable selection procedures</li> <li>+ <b>non-linear effects</b> can be respected (e.g. GAM) and controlled (e.g. degree of non-linearity by penalization techniques)</li> <li>+ works also for not independent response (e.g. GEE)</li> </ul>	<div> <p>especially these 3 advantages make regression models a valuable tool for (health) insurance data</p> <ul style="list-style-type: none"> <li>- requires pre-selection of reasonable interaction terms before variable selection can be used</li> <li>- model convergence is not guaranteed, if complex structures are specified</li> </ul> </div>
Decision Trees / Random Forest	<ul style="list-style-type: none"> <li>+ <b>detects complex relationships</b> in large datasets</li> <li>+ returns the „variable importance“ of all factors</li> <li>+ very stable algorithm produces always results</li> </ul>	<ul style="list-style-type: none"> <li>- tends to overfit training data in case of small datasets and bad data quality</li> <li>- remains a black box to the user and results can hardly be interpreted</li> <li>- does not check the significance of effects or spurious relationships</li> </ul>
Neural Networks	<ul style="list-style-type: none"> <li>+ <b>detect complex non-linear relationships and correlation structures</b> in large datasets</li> <li>+ very stable algorithm</li> <li>+ Very powerful with extremely large datasets and technical environment</li> </ul>	<ul style="list-style-type: none"> <li>- tend to overfit training data in case of small datasets and bad data quality</li> <li>- remain a black box to the user and results can hardly be interpreted</li> <li>- quality can only be evaluated by predictive results</li> </ul>

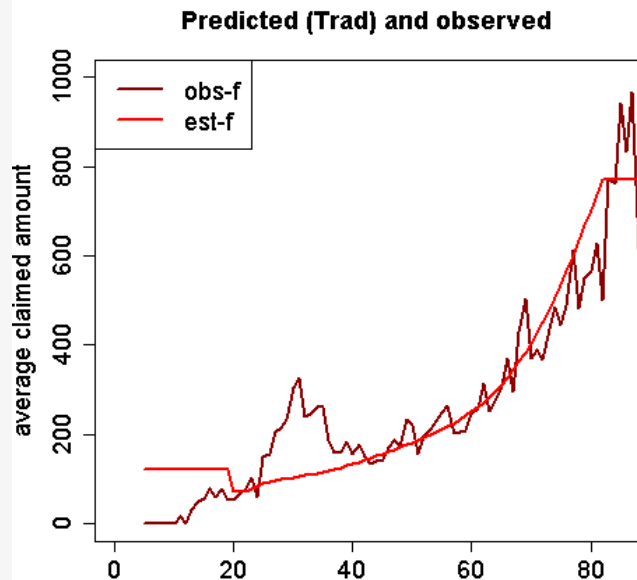
It is crucial to find the right trade-off between stability and complexity to optimize prediction power



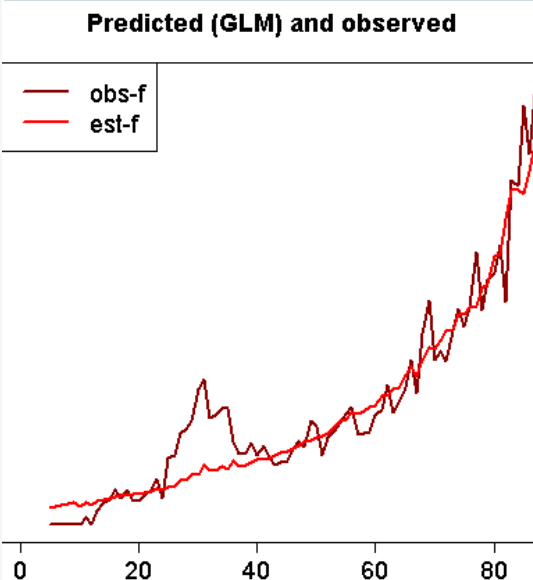
# Predicted vs. observed claimed amounts for particular subgroups allows optimal model choice

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

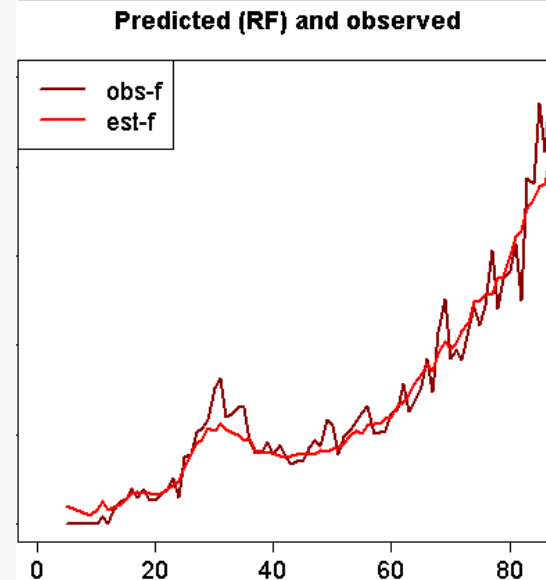
## Traditional approach



## 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)



# 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



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

	Acquisition	Up-selling	Cross-selling	Churn prevention	Win-back
Target	<p>Identification of relevant target groups (“good risks”)</p> <p>Optimized acquisition of new customers</p>	<p>Identification of customers with highest probability to extend their existing contract</p> <p>Higher penetration of health products</p>	<p>Identification of customers with highest probability to buy additional products</p> <p>Higher penetration of other products</p>	<p>Identification of customers with low product retention</p> <p>Avoidance of cancellation of “good risks”</p> <p>Increase of customer loyalty</p>	<p>Identification of former customers with high capital value and high win-back-probability</p> <p>Enhancement of the portfolio's profitability</p>
CRM solution	<ul style="list-style-type: none"> <li>Customer profiling and clustering based on external and internal data</li> <li>Sales channel preference analysis</li> <li>...</li> </ul>	<ul style="list-style-type: none"> <li>Scores to identify potential up-selling candidates</li> <li>Sales channel preference score</li> <li>Customer value analysis</li> <li>...</li> </ul>	<ul style="list-style-type: none"> <li>Scores to identify potential cross-selling candidates</li> <li>Sales channel preference score</li> <li>Customer value analysis</li> <li>...</li> </ul>	<ul style="list-style-type: none"> <li>Churn scores to identify “disloyal” customers</li> <li>Next best action analytics</li> <li>...</li> </ul>	<ul style="list-style-type: none"> <li>Win-back scores to identify customers most likely to return</li> <li>...</li> </ul>

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

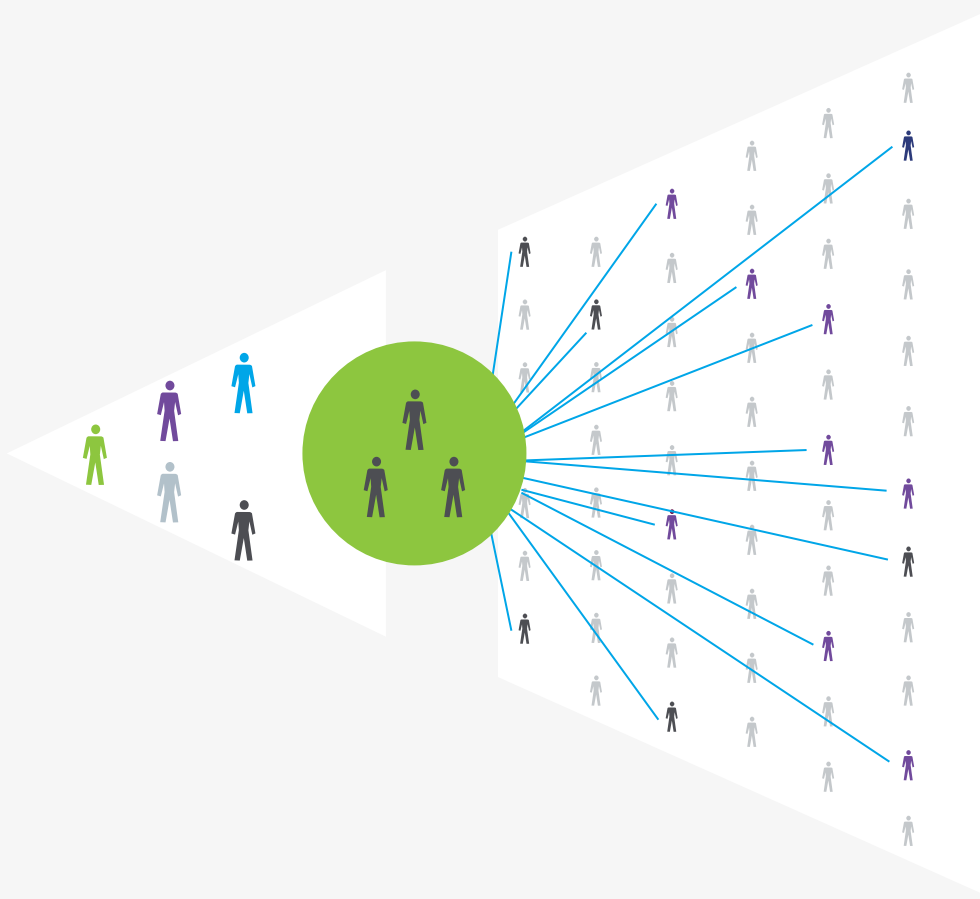
1. Existing customers

2. Identify and profile best customers

3. Find others like them

4. Model media preferences

5. Accelerate profitable growth



Direct Mail



E-Mail



Telemarketing



Mobile



Internet



Broadcast



Print Media

## 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**

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

## Churn-index is derived from score groups ...

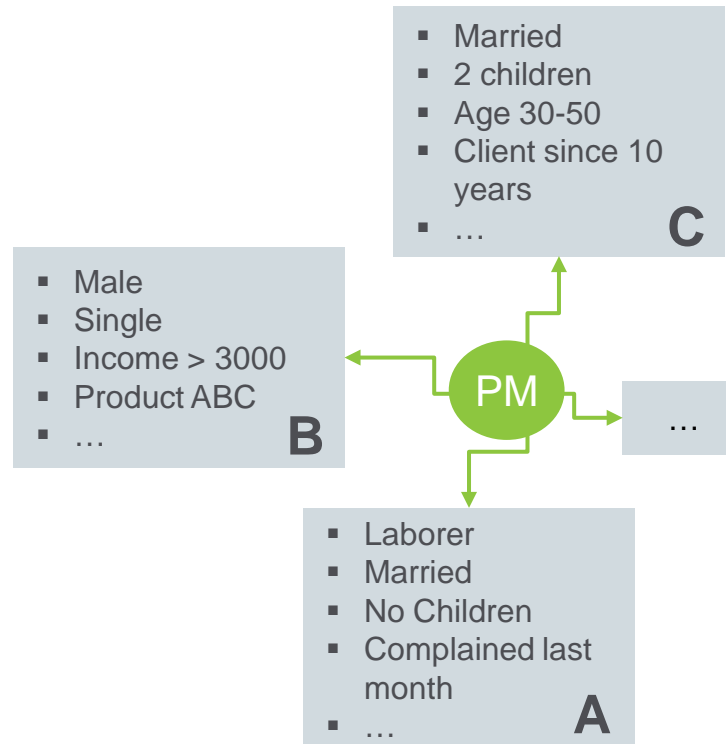
### Internal Data

- Insurance product
- Rate
- Gender
- ...

### External Data

- Socio-demographic
- Income status
- Family status
- ...

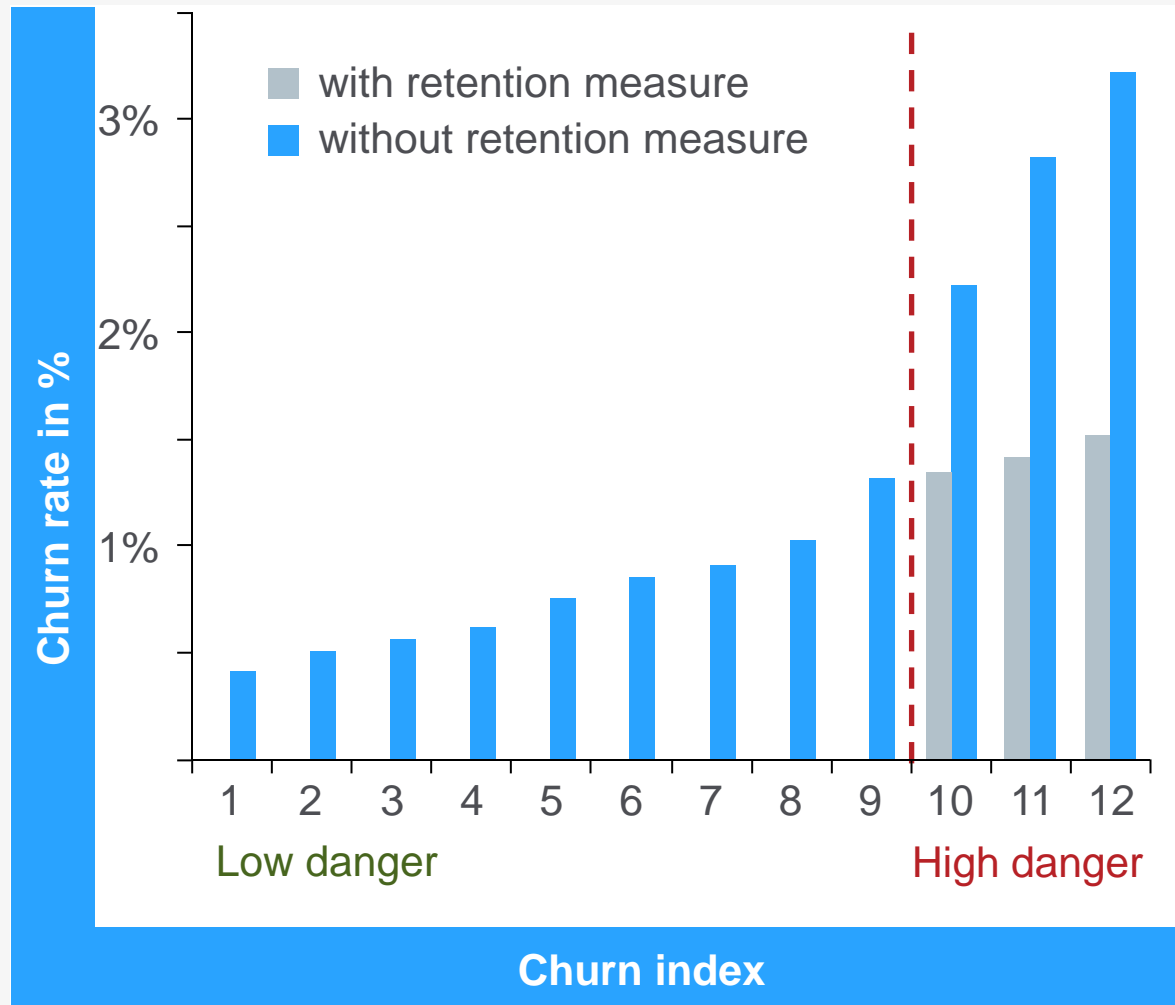
### Score groups



### Churn index



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



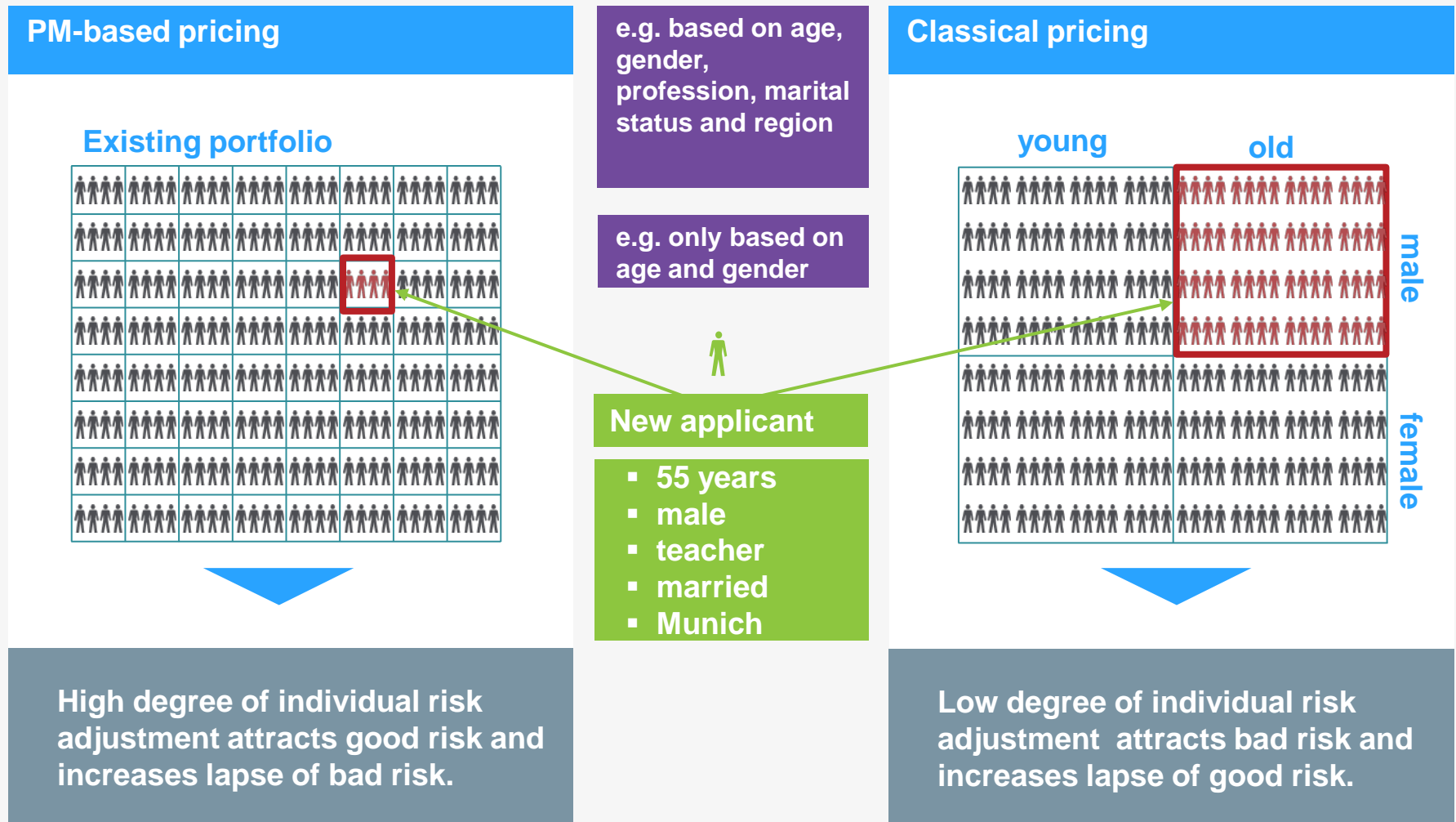
## Success

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

# Pricing

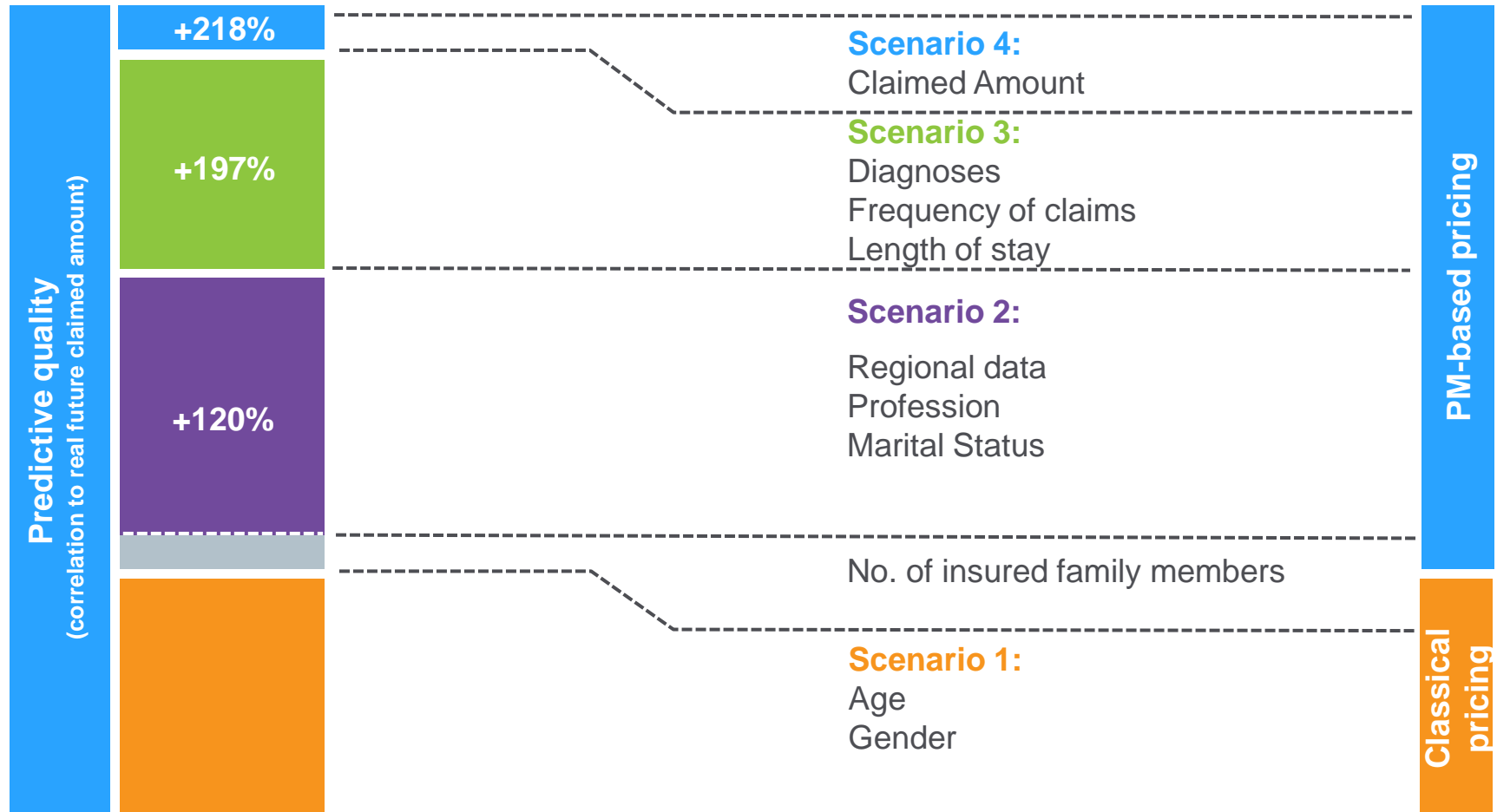


# PM-based pricing yields a stronger differentiation of the risk structure to optimize premium calculation





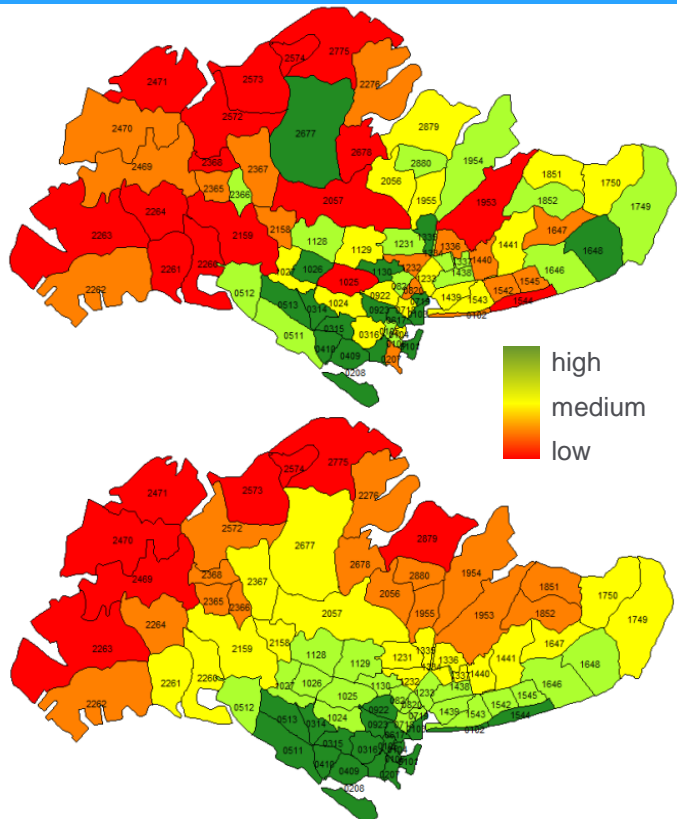
# Predictive quality depends on information used



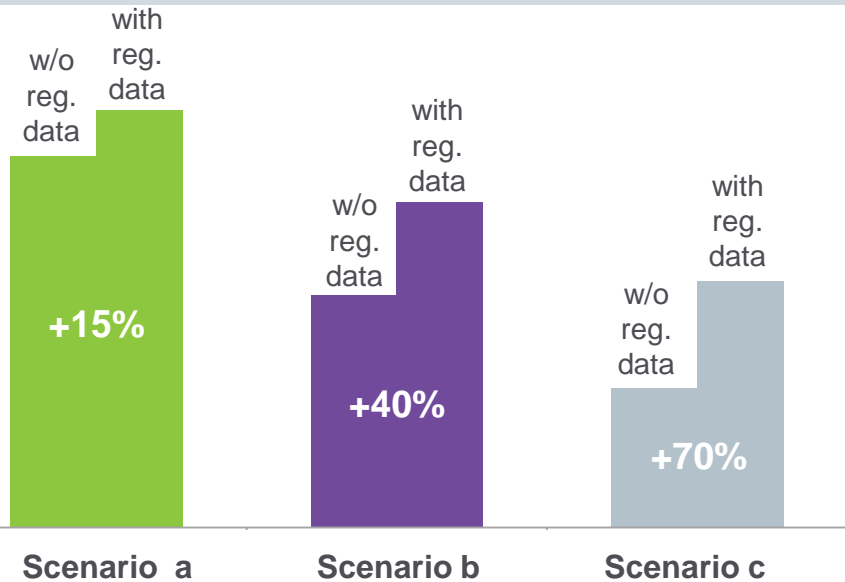
# Usage of clients' address data combined with ext. regional data further improves predictive quality

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

Example: Income per capita in Singapore



Predictive quality



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

- 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



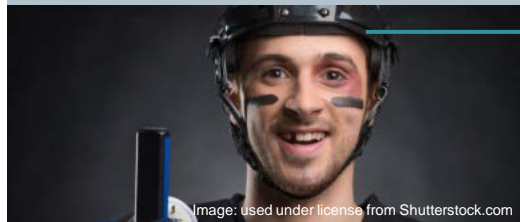
Smoker Association

Members

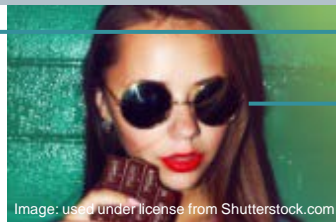
Join group

Create Group

### Profile pictures



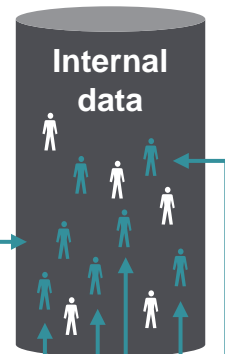
Usage of Image Mining!



### Likes and posts

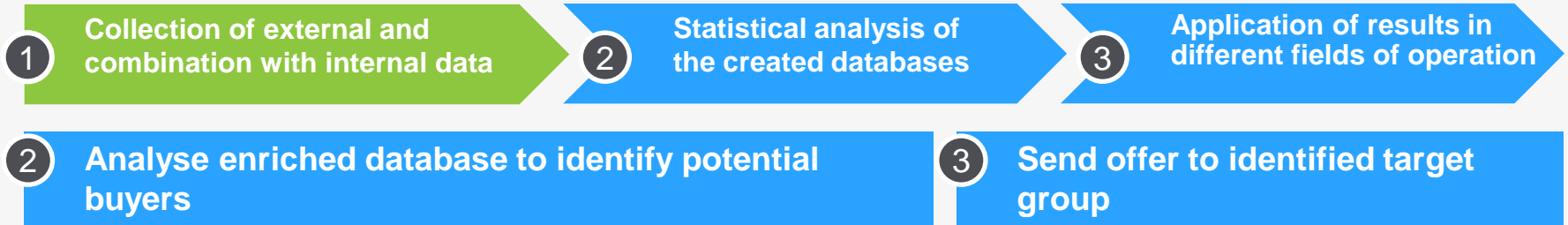


Exact matching or similarity matching



# Social Media Analysis for Product Development:

## Example 1 - Supplementary dental cover

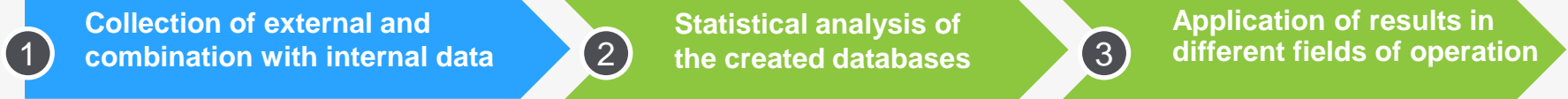


```
10  
17 • ##### Friends #####  
18 # Extract information about my profile  
19 me <- getUsers("me", token = fb_auth)  
20  
21  
22 # Extract list of friends with their information  
23 friends <- getFriends(token = fb_auth)  
24  
25 # Extract list of likes (e.g. first 100) of a facebook friend (e.g. friend with ID = 100000306460607)  
26 likes_07 <- getLikes(100000306460607, n=100, token = fb_auth)  
27  
28 # Export into CSV format  
29 write.csv(me, "me.csv", as.is=T, row.names=F, col.names=T)  
30 write.csv(friends, "friends.csv", as.is=T, row.names=F, col.names=T)  
31 write.csv(likes_07, "likes_07.csv", as.is=T, row.names=F, col.names=T)  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  

```

# Social Media Analysis for Product Development:

## Example 2 - Supplementary inpatient cover

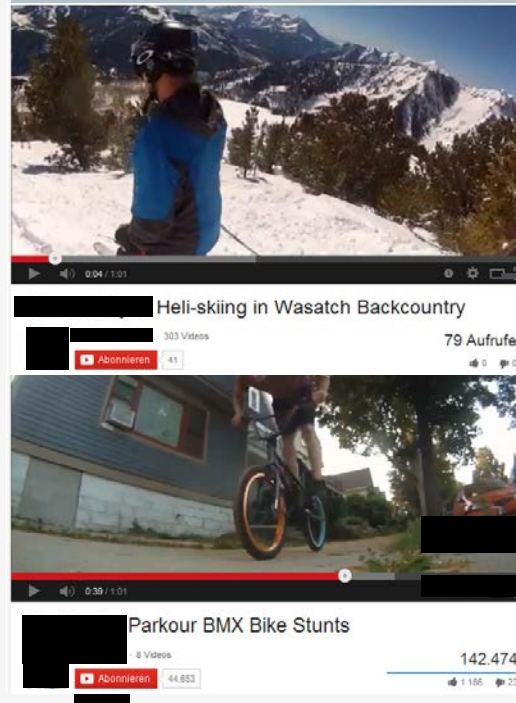


1 Identify risk groups through relevant activities & match with internal data, e.g. via YouTube

### 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 Lugging
8	Mountain Climbing
9	BMX →
10	White-Water Rafting

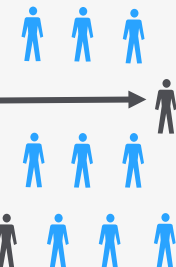
### YouTube uploads



### Usage of Video Mining!

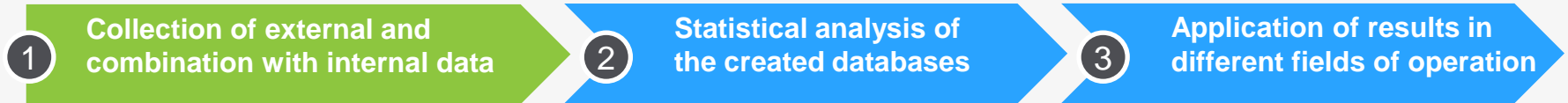
Exact matching or similarity matching

Internal Data





# Analytical approaches can be applied after combining internal and external information



## 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

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



**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**

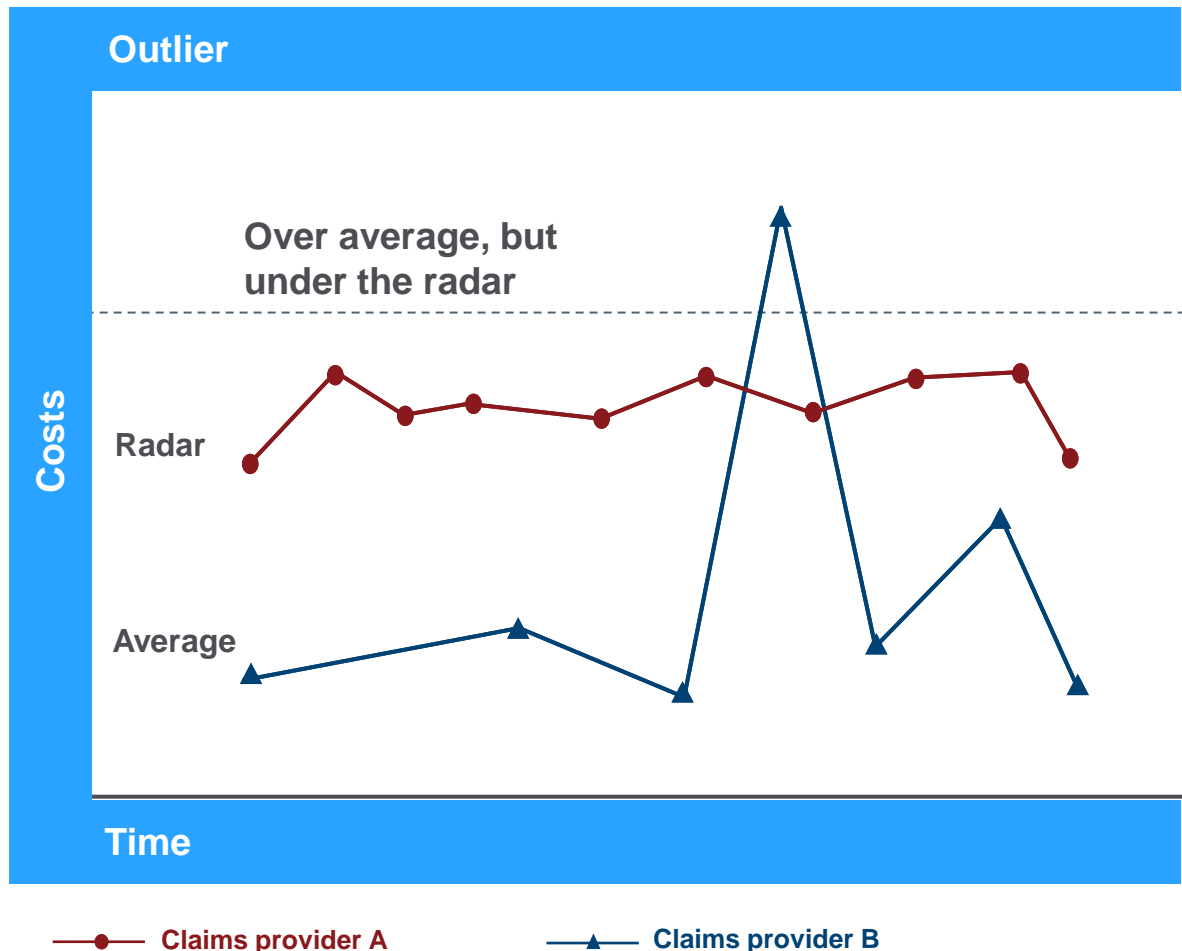
**Analysis of merged database requires analytical and other experts**

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

# Fraud and abuse



# 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

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

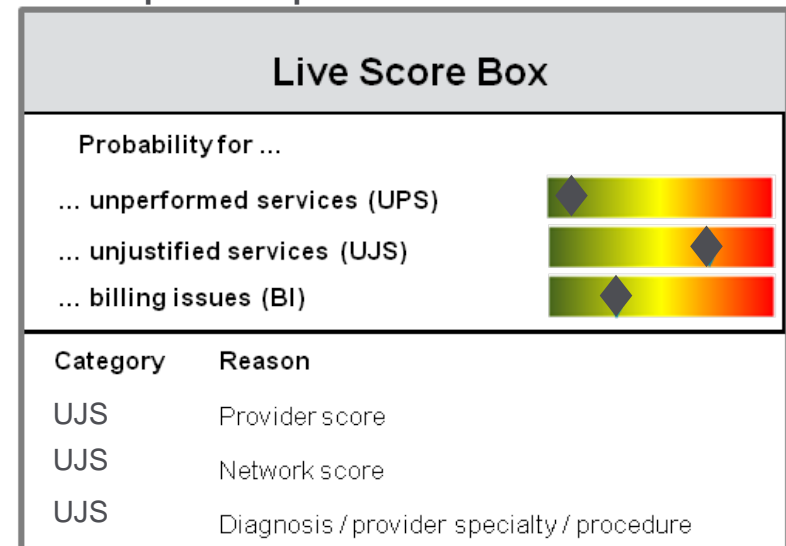
## Deterministic rules (n Rule Types)

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**



# Example from Indonesia: Infectious diseases

## Scoring

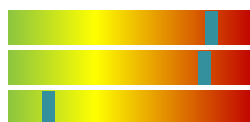
### Live Score Box

#### 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%



## 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.

- 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.

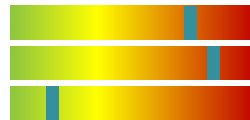
# Example from Germany: Referral system

## 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

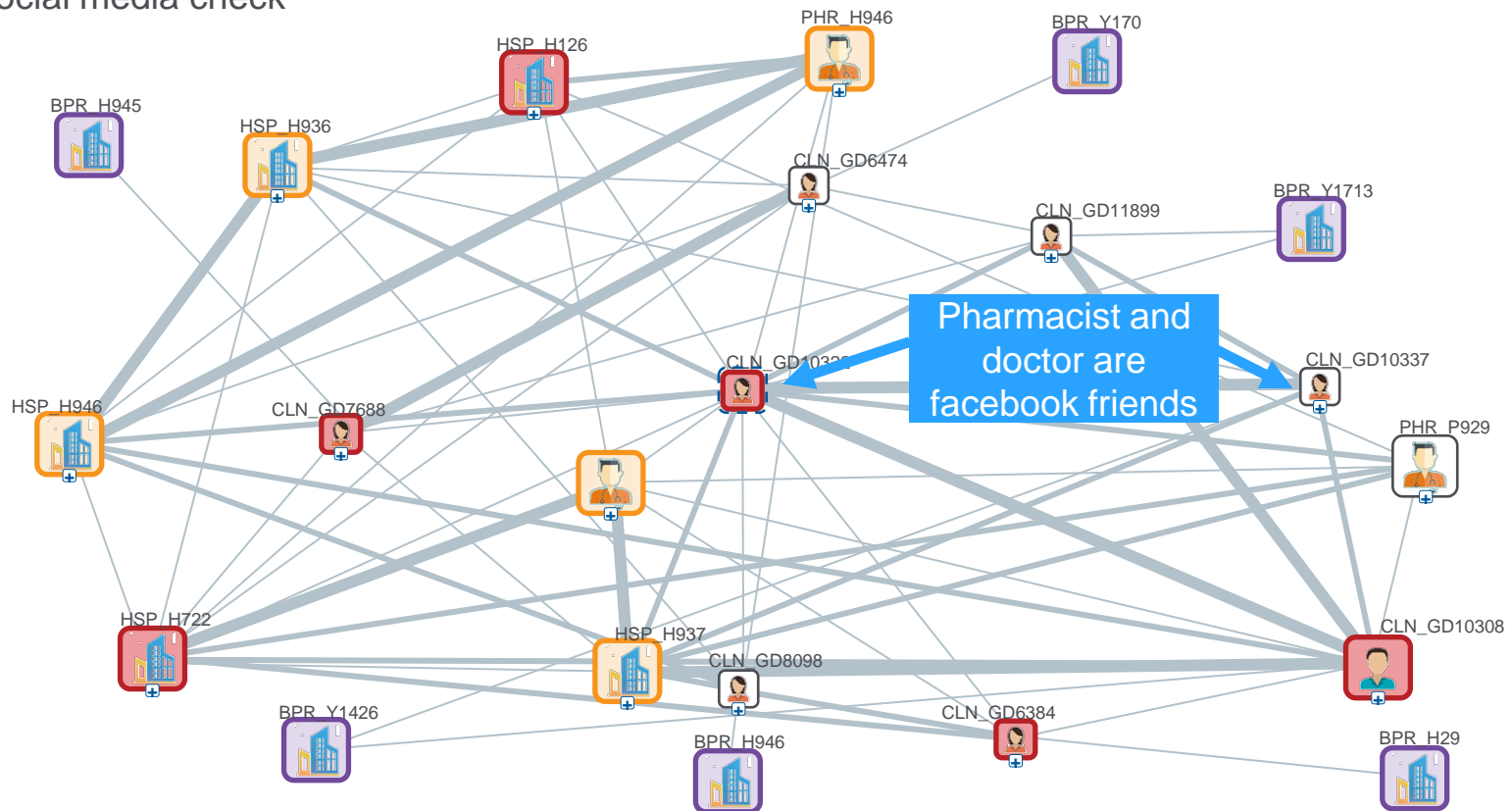


- After visiting general practitioner 610, many of his patients, drive 150km (from Stuttgart to the Black Forest) to visit alternative practitioner 433.
- The alternative practitioner prescribes a lot of cold remedies.
- The prescriptions are nearly always filled in a nearby pharmacy (753).
- Compared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

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

## 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

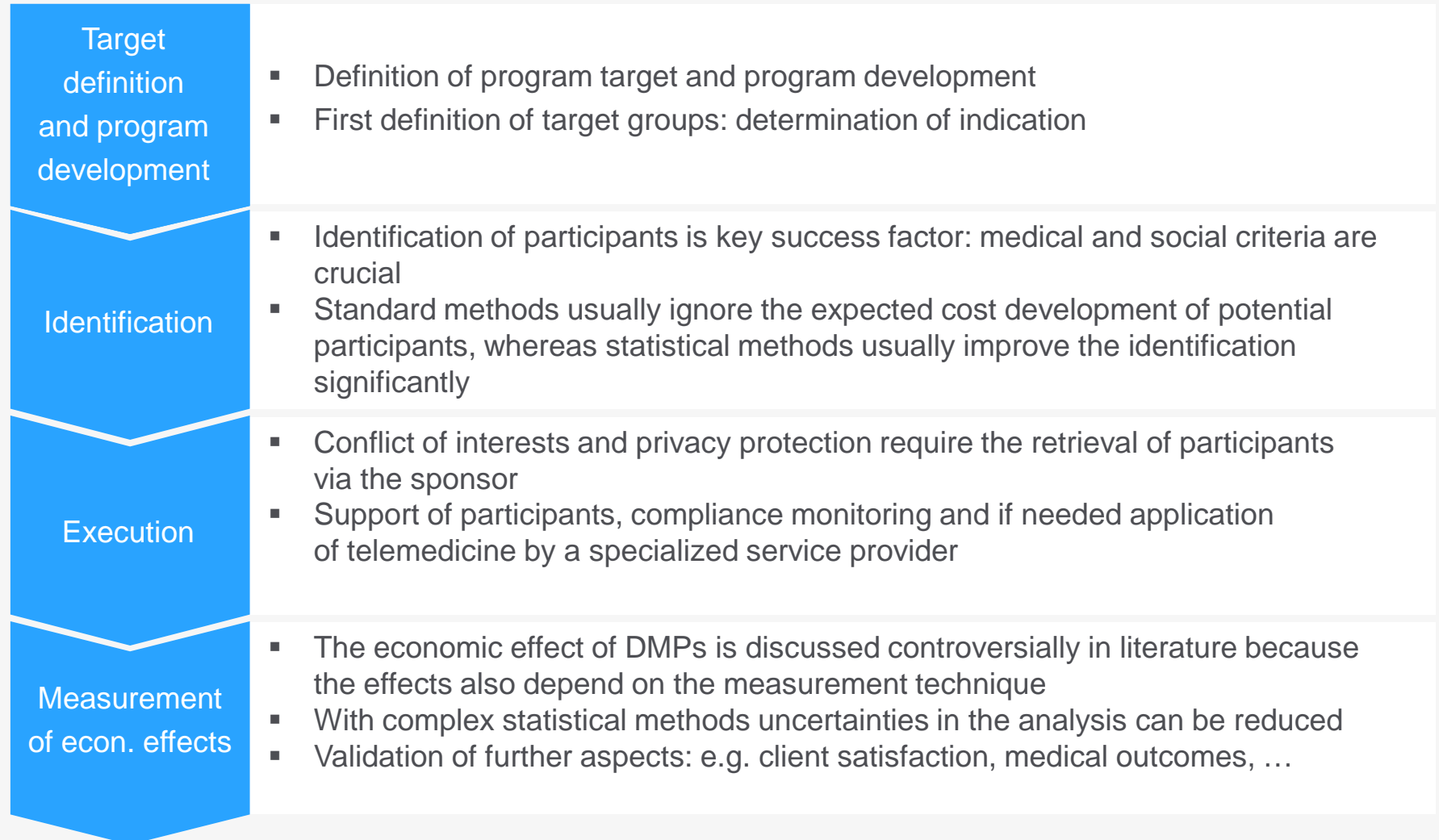
... however they can make the haystack considerably smaller!



## DMP (Disease management programs)

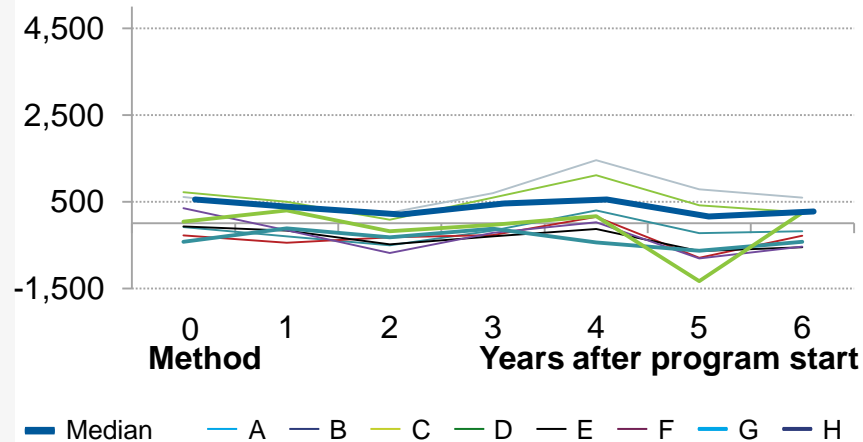


# Simplified process of every health program and application of statistical methods



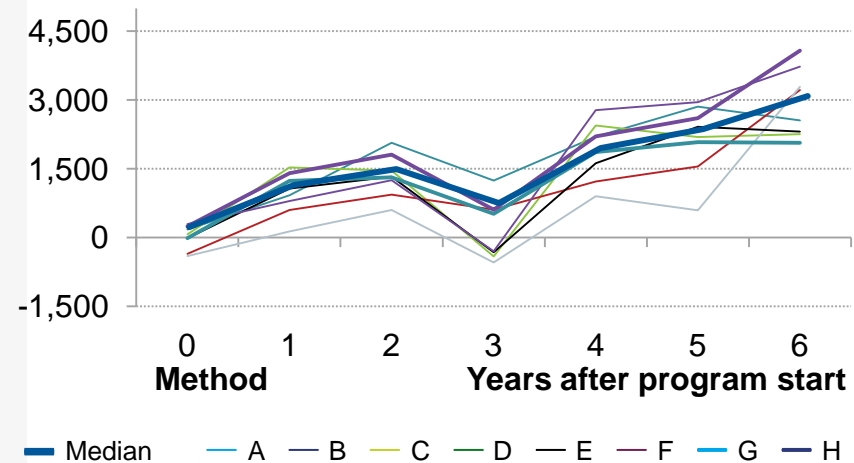
## Realizable savings based on different measurement methods

### Diabetes (savings in EUR)



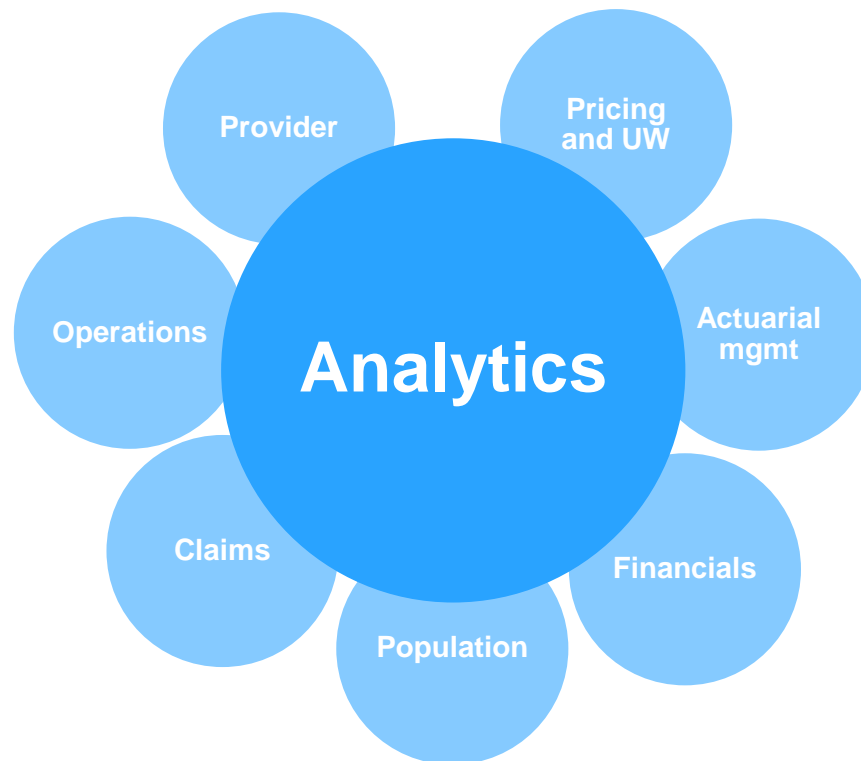
The results strongly depend on the measurement method. Most methods do not indicate a long-term saving potential.

### 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.**



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.

# Questions

# Comments

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.



Institute  
and Faculty  
of Actuaries