www.vems.ch

The santésuisse cost-index is likely to be an ecological fallacy – a mathematical experiment

By Michel Romanens, MD
Citation: only with mentioned source: www.docfind.ch/WZWEcologicalFallacy.pdf

Olten, January 2019

Background
Costs generated by GPs are created by chance, partially, because GPs do not select their patients and partially because the disease presented varies unexpectedly in a given patient due to new diagnoses, complications and variations in the speed of healing processes.

Aim
We aim to assess the performance of the average cost at the physician level (santésuisse cost-index) to detect experimentally created uneconomicity.

Method
In order to test the average cost tool of santésuisse to detect uneconomical costs, we developed a random model with 1'00'000 patients having been treated by 1000 doctors over 1 year. For each patient, appropriate usual care costs regarding usual care were generated using normally distributed random samples within a prespecified range of costs defined by a mean of 300 Sfr. and a standard deviation of 50 Sfr. by MedCalc version 18.11. For each patient the costs having been spent by his GP was generated the same way using slightly higher costs with a mean of 400 Sfr. and a standard deviation of 60 Sfr, which then produced clearly defined amounts of overmedicalization per GP. Because of the heavy tail distribution of costs encountered in general medicine, each 10th of the 1 Mio patient received a proxy for a PCG defined by mean costs of 12'000 (SD 2000 Sfr) and some higher costs being spent by GP of 18'000 Sfr (SD 2000 Sfr). For each patient usual care costs and PCG costs were added and compared to the costs generated by the GP at the patient level. Next, for each GP, the average of appropriate costs and generated costs were compared and a true cost-index was generated by dividing generated by appropriate costs.

Uneconomicity cases were defined by an index above 110%. Further, the santésuisse cost-index was created by the average GP costs of all patients divided by the number of patients and each GPs cost were compared to this average cost-index by dividing the costs generated at the GP level by the average costs of all patients. The performance of the santésuisse cost-index was tested to detect the cases using area under the curve (AUC) receiver operating statistics (ROC-analysis with the DeLong method[1])

Results

The number of patients generated by the system was 1'000'000, which were distributed by chance among 1000 GP. The average appropriate costs per GP ranged between 1'156 Sfr to 1'831 Sfr and the average spent costs per GP ranged between 1'258 Sfr to 1'937 Sr. For the total of patients, appropriate costs were Sfr. 1'499'353'377 and costs spent by GPs was Sfr. 1'599'077'807 with a sum of inappropriate costs of 99'724'430 (6.7%).

For illustration, Figure 1 shows GP nr. 1 with the appropriate cost and the GP costs per patient (r^2 = 0.93) for his 1'005 patients. He had spent 1.4 Mio Sfr, appropriate costs were 1.3 Mio Sfr. resulting in an overmedicalization of 7.7% or overmedicalization costs of 100'469 Sfr. His santésuisse cost-index was 87% within the group of the 1'000 GPs.

The appropriate index per GP ranged between 100% and 114% (average 107%). The number of GPs that treated with an appropriate index of 110% or more was 53 (53 cases). The average cost of all patients spent by GPs was 1'599 Sfr, which corresponds to the santésuisse cost-index of 100%. The santésuisse index among the 1000 GPs ranged between 79% and 121%. The regression line was negative for an increase in inappropriate costs compared to the santésuisse index (y = -0.4381x + 146.74) and the association was minimal (r^2=0.016, Figure 2).

Results from ROC Analysis are shown in Fig. 3. Overall, the performance of the santésuisse index was weak (AUC 61%, p=0.003). According to the Youden's index, a santésuisse cost-index of 97% showed the best performance (sensitivity 55%, specificity 62% to detect ≥10% of inappropriate costs). At the cost-index of 130%, which was never reached in the model, therefore, sensitivity is 100% and specificity is 0%. For further details refer to the appendix.

Figure 1: Illustrative example of GP nr. 1

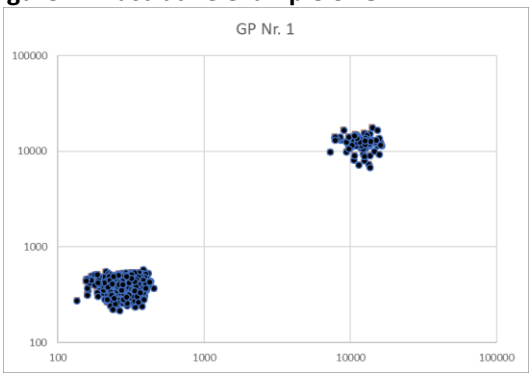


Figure 2: Correlation between Appropriate Index and santésuisse Index

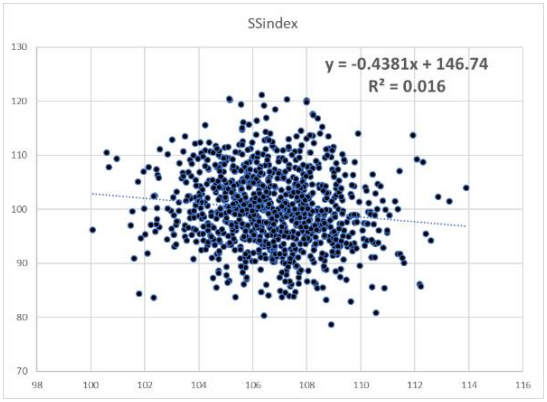
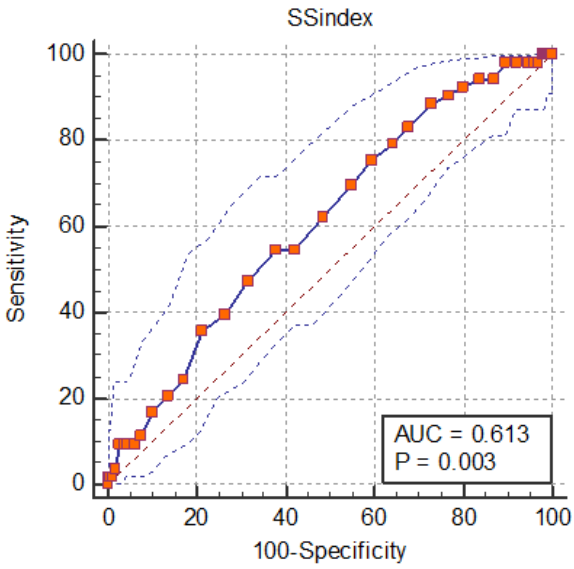


Figure 3: ROC analysis comparing overmedicalization (cases) with the santésuisse cost-index



Discussion

Using an experimental design, we generated a random for appropriate costs in 1 Mio patients and grouped these patients among 1'000 GPs, who overmedicalized these patients by a known amount of money. When we compared the true cost-index of overmedicalization with the santésuisse cost-index, we found a very poor correlation (r^2=0.016) and a counterintuitive tendency for lower santésuisse cost-indices with higher true cost-indices (Figure 2). From this, there is no scientific proof that the santésuisse cost-index is capable to filter out uneconomically generated costs.

At a practical level, where the santésuisse cost-index of 130% is used to detect uneconomicity as a filter for Swiss GPs, in a group of GPs defined as costs > 110% above medically justified costs, the specificity was 0% and ROC analysis revealed a poor performance to detect true cases of uneconomicity (AUC 61%, p=0.003).

Because in general, GPs cannot select neither their patients nor their diseases, our random model may be appropriate to reflect the cost-reality imposed on GPs and it can be expected, that uneconomical behavior takes place at lower cost-levels than suggested by the santésuisse cost-index.

Because we directly defined necessary costs at the patient level, we could eliminate the need for explaining variables such as hospitalizations in previous years, gender, age or PCGs. Whether the addition of such approximations may increase the performance of the new regression-index of santésuisse is unknown.

Our findings are intriguing and deserve further explanation. The negative correlation between true overmedicalization and the santésuisse cost-index (derived from the group level) appears to be a form of reverse causality due to omitted variables (true overmedicalization approximated by variance in inter-physician cost averages). Because santésuisse creates the cost-index at the physician level and uses aggregations of data, that are presumed to reflect individual data, a Robinson's paradox is likely to occur with all forms of santésuisse cost-indices.

Conclusion

Using an experimental model, where necessary treatment costs are known and were compared to randomly assigned overmedicalization costs per GP, an average cost-index of 130% never detected uneconomicity. Therefore, a high santésuisse cost-index of 130% is not suited for the detection of uneconomical medical treatment, and it is unclear, what exactly the santésuisse cost-index benchmark represents in reality. Because our model avoids the fallacies that may arise with inappropriate comparison groups, it may be argued that very high cost-indexes (> 130%) are artificial effects generated by inappropriate comparison groups and this point certainly deserves further scientific research. It appears that the only way to avoid a Robinson's ecological fallacy with santésuisse cost-indices is by random clinical audits [2], since modeling expected costs at the patient level, as we did in this experiment, has a questionable feasibility at the individual physician level [3,4].

From this experiment it becomes clear, that the santésuisse cost-index is not valid to detect overmedicalization. Asking GPs to explain their overmedicalization to santésuisse is likely to be impossible, because our experiment shows, that an index of 130% or more does not reflect overmedicalization and may even reflect the contrary of it.

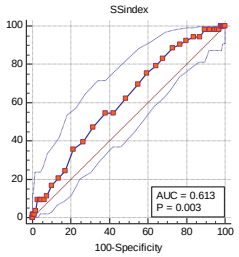
References

1. Delong E, Delong D, Clarke-Pearson D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845.
2. Romanens M. Inter-Institutional Profiling and Patient Safety : *J Ethics Med*. 2010;(1). http://www.docfind.ch/AuditsInMedicine.pdf.
3. Adams JLJJL, Mehrotra A, Thomas JW, McGlynn EAE. Physician cost profiling--reliability and risk of misclassification. *N Engl J Med*. 2010;362(11):1014-1021. doi:10.1056/NEJMsa0906323.
4. Mehrotra A, Adams JL, Thomas JW, Mcglynn EA. Cost Profiling in Health Care: Should the Focus be on Individual Physicians or Physician Groups? doi:10.1377/hlthaff.2009.1091.

Appendix:

ROC curve

Variable	SSindex
Classification variable	Cases



Sample size	1000
Positive group [a]	53 (5.30%)
Negative group [b]	947 (94.70%)

[a] Cases = 1
[b] Cases = 0

Disease prevalence (%)	5.3

Area under the ROC curve (AUC)

Area under the ROC curve (AUC)	0.613
Standard Error [a]	0.0378
95% Confidence interval [b]	0.582 to 0.644
z statistic	3.001
Significance level P (Area=0.5)	0.0027

[a] DeLong et al., 1988
[b] Binomial exact

Youden index

Youden index J	0.1681
Associated criterion	≤97
Sensitivity	54.72
Specificity	62.09

Optimal criterion

Optimal criterion [a]	<79
Sensitivity	0
Specificity	100

[a] Taking into account disease prevalence (5.30%) and estimated costs:
cost False Positive: 1; cost False Negative: 1
cost True Positive: 0; cost True Negative: 0

Criterion values and coordinates of the ROC curve [Show]

Criterion values and coordinates of the ROC curve [Hide]

Criterion	Sensitivity	95% CI	Specificity	95% CI	+LR	95% CI
<79	0	0.0 - 6.7	100	99.6 - 100.0		
≤79	0	0.0 - 6.7	99.89	99.4 - 100.0	0	
≤80	0	0.0 - 6.7	99.79	99.2 - 100.0	0	
≤81	1.89	0.05 - 10.1	99.79	99.2 - 100.0	8.93	0.8 - 97.0
≤83	1.89	0.05 - 10.1	99.58	98.9 - 99.9	4.47	0.5 - 39.3
≤84	1.89	0.05 - 10.1	98.84	97.9 - 99.4	1.62	0.2 - 12.3
≤85	3.77	0.5 - 13.0	98.31	97.3 - 99.0	2.23	0.5 - 9.5
≤86	9.43	3.1 - 20.7	97.25	96.0 - 98.2	3.44	1.4 - 8.6
≤87	9.43	3.1 - 20.7	96.09	94.7 - 97.2	2.41	1.0 - 5.9
≤88	9.43	3.1 - 20.7	95.35	93.8 - 96.6	2.03	0.8 - 4.9
≤89	9.43	3.1 - 20.7	93.77	92.0 - 95.2	1.51	0.6 - 3.6
≤90	11.32	4.3 - 23.0	92.4	90.5 - 94.0	1.49	0.7 - 3.3
≤91	16.98	8.1 - 29.8	89.76	87.6 - 91.6	1.66	0.9 - 3.1
≤92	20.75	10.8 - 34.1	86.27	83.9 - 88.4	1.51	0.9 - 2.6
≤93	24.53	13.8 - 38.3	83	80.5 - 85.3	1.44	0.9 - 2.4
≤94	35.85	23.1 - 50.2	78.78	76.0 - 81.3	1.69	1.2 - 2.5
≤95	39.62	26.5 - 54.0	73.5	70.6 - 76.3	1.49	1.1 - 2.1
≤96	47.17	33.3 - 61.4	68.43	65.4 - 71.4	1.49	1.1 - 2.0
≤97	54.72	40.4 - 68.4	62.09	58.9 - 65.2	1.44	1.1 - 1.9
≤98	54.72	40.4 - 68.4	57.97	54.8 - 61.1	1.3	1.0 - 1.7
≤99	62.26	47.9 - 75.2	51.53	48.3 - 54.8	1.28	1.0 - 1.6
≤100	69.81	55.7 - 81.7	45.09	41.9 - 48.3	1.27	1.1 - 1.5
≤101	75.47	61.7 - 86.2	40.65	37.5 - 43.9	1.27	1.1 - 1.5
≤102	79.25	65.9 - 89.2	35.69	32.6 - 38.8	1.23	1.1 - 1.4
≤103	83.02	70.2 - 91.9	32.1	29.1 - 35.2	1.22	1.1 - 1.4
≤104	88.68	77.0 - 95.7	26.93	24.1 - 29.9	1.21	1.1 - 1.3
≤105	90.57	79.3 - 96.9	23.34	20.7 - 26.2	1.18	1.1 - 1.3
≤106	92.45	81.8 - 97.9	20.06	17.6 - 22.8	1.16	1.1 - 1.3
≤107	94.34	84.3 - 98.8	16.26	14.0 - 18.8	1.13	1.0 - 1.2
≤108	94.34	84.3 - 98.8	12.99	10.9 - 15.3	1.08	1.0 - 1.2
≤109	98.11	89.9 - 100.0	10.56	8.7 - 12.7	1.1	1.1 - 1.1
≤110	98.11	89.9 - 100.0	7.81	6.2 - 9.7	1.06	1.0 - 1.1
≤111	98.11	89.9 - 100.0	5.39	4.0 - 7.0	1.04	1.0 - 1.1
≤112	98.11	89.9 - 100.0	4.12	2.9 - 5.6	1.02	1.0 - 1.1
≤113	98.11	89.9 - 100.0	3.06	2.1 - 4.4	1.01	1.0 - 1.1
≤114	100	93.3 - 100.0	2.22	1.4 - 3.4	1.02	1.0 - 1.0
≤115	100	93.3 - 100.0	1.8	1.0 - 2.9	1.02	1.0 - 1.0
≤116	100	93.3 - 100.0	1.37	0.7 - 2.3	1.01	1.0 - 1.0
≤117	100	93.3 - 100.0	1.06	0.5 - 1.9	1.01	1.0 - 1.0
≤118	100	93.3 - 100.0	0.84	0.4 - 1.7	1.01	1.0 - 1.0
≤119	100	93.3 - 100.0	0.63	0.2 - 1.4	1.01	1.0 - 1.0
≤120	100	93.3 - 100.0	0.11	0.003 - 0.6	1	1.0 - 1.0
≤121	100	93.3 - 100.0	0	0.0 - 0.4	1	1.0 - 1.0