Budapesti Corvinus Egyetem

Közgazdasági és Gazdaságinformatikai Doktori Iskola

Actuarial Contributions to the Study of Health Insurance

Ph.D. Thesis

Péter Németh

Supervisor:

Dr. Péter Vékás

Budapest, 2025

Table of contents

List of figures

List of tables

Acknowledgments

I owe my deepest gratitude to my supervisor, Dr. Péter Vékás, whose advice and guidance have accompanied my studies throughout. He was always available, encouraged me at every stage, and reviewed the manuscript more times than I can count to ensure that no errors remained.

I am also grateful to my co-authors. Within a large collaborative project we succeeded in producing a high-impact publication, and through our joint work I learned a great deal and grew substantially as a researcher.

I would like to thank my colleagues and my manager at Aegon for their valuable feedback and careful reviews, for giving me the opportunity to practice presenting my results, and for their patience during the challenging periods surrounding the submission of both the draft and the final dissertation.

Finally, I wish to express my special thanks to my family—my parents; my wife, Orsi; and my son, Dénes—for helping me through the difficult periods and for their unwavering support and motivation, which made it possible for me to complete this dissertation.

Introduction

This dissertation has been written with deliberate intent, not only to contribute to academic knowledge but also to reflect personal values and professional aspirations. Throughout the work, several choices were made with purpose—choices that go beyond methodology or topic selection. These choices were guided by a desire to represent the quality of work that can be expected of me in the future and to demonstrate the standards I hold myself to. One of the guiding principles in developing this dissertation has been the prioritization of quality over quantity.

Perhaps most importantly, the choice of topic speaks to both personal motivation and societal relevance. While I had multiple options for model development, I chose to focus on health insurance in order to highlight broader issues within the Hungarian healthcare system and to underscore the significance of this sector. The actuarial literature in Hungary remains limited in this field, which further motivated me to contribute in a meaningful way. By addressing a topic that is both close to my heart and important to the wider community, my aim is to demonstrate that academic work can—and should—serve a broader social purpose.

Health insurance pricing has become an increasingly relevant and pressing topic in Hungary in recent years. Although the publicly funded healthcare system continues to provide nearly comprehensive coverage for the population, the quality and reliability of care are steadily declining. A growing shortage of medical professionals, outdated technologies, and deteriorating infrastructure have all contributed to a significant drop in service standards. As a result, the private healthcare sector is expanding rapidly, responding to the rising demand for faster, more personalized, and higher-quality care.

Despite this growth, the range of available private health insurance products remains limited. Few companies offer such products in Hungary, and those that do often fail to incorporate key features commonly found in international markets—such as risk-based pricing, incentives for preventive care, and the use of validated health measures. This gap presents both a challenge and an opportunity: the challenge of developing comprehensive, reliable products, and the opportunity to build models and frameworks that better reflect health risks and behaviors in pricing.

In conducting this research, I chose to rely exclusively on pre-COVID data. The primary reason for this decision was to avoid the significant and often unpredictable fluctuations in healthcare utilization and costs that occurred during the pandemic. Additionally, the post-COVID period in Hungary has not yet generated sufficient stable data for robust analysis. By focusing on the pre-pandemic period, this dissertation aims to ensure consistency and reliability in the models presented.

The core objective of this dissertation is to demonstrate the importance of sound health insurance pricing mechanisms while introducing relevant models and practical applications. The work is structured around three studies. The first, published in an MTA A-category journal, defines the most important concepts and product types, effectively setting the framework for the entire dissertation. The second, which appeared in a Scimago Q1 D1 journal, focuses on the validation of the Patient Activation Measure (PAM-13)—a widely used instrument for assessing an individual's knowledge, skill, and confidence in managing their own health. Validating this measure within the Hungarian context was an essential step before applying it in further modeling. The third study, intended for publication following the defense of this dissertation, builds directly on this validation. Using the PAM-13 data, it demonstrates how such a measure can be integrated into pricing models through advanced statistical methods, providing an example of how validated tools can contribute to more accurate, fair, and dynamic insurance pricing.

By combining theory, validation, and application, this dissertation aims to contribute to both academic literature and practical development in the Hungarian private health insurance sector. It highlights the urgent need for innovation and structured thinking in a market that is poised for transformation.

Having outlined the context, motivation, and structure of this dissertation, the following section presents the specific research questions that guided each of the three studies. The main results of the three papers are briefly summarized below, providing an overview of the contribution each study makes to the overall aim of this dissertation.

The first study was published in the journal Statisztikai Szemle *(Németh, 2024).* Originally written in Hungarian, I only translated this paper for the thesis. This study aims to address the following research questions:

a) What role does data quality play in determining accurate health insurance pricing models in Hungary's private health sector?
b) How do demographic factors, such as age and income, influence the affordability and pricing of private health insurance in Hungary?
c) How can private health insurance products be designed to balance affordability with comprehensive coverage in countries facing significant public sector underfunding?

Results:

This study highlights the growing role of private health insurance in Hungary's healthcare sector as public healthcare services decline. As demand for private healthcare increases due to problems in the public sector such as long waiting times, outdated infrastructure and lack of resources, private insurance has emerged as a viable solution to cover health-related risks. The study also presents a comprehensive pricing model for various health insurance products using statistical data for different demographic groups, focusing primarily on a 40-year-old individual as a baseline.

With public health systems under pressure, private insurance products with targeted pricing could address unmet needs, especially for preventive and high-cost health events. In my study, I also suggest that the integration of awareness campaigns and subsidies could help the younger population to take up private health insurance and encourage a shift towards preventive health practices.

My study highlights the limitations of access to accurate, high-quality statistical data, which can affect the reliability of pricing models. Data revisions and differences in reporting practices complicate long-term projections and the accuracy of product pricing.

The calculated premiums are close to those offered by insurers, suggesting that the pricing models are representative of current market conditions. However, the relatively low penetration of private health insurance, especially for the middle class, reflects affordability concerns in the context of average Hungarian incomes, and is further compounded by distrust towards these products and the lack of adequate incentives for insurance agents.

The second study was published in the European Journal of Health Economics in 2022 *(Zrubka et al., 2022).* In this paper, my contribution included performing the calculations, particularly for the known-groups hypothesis, creating the result tables, reviewing the final text, and assisting in the review process by addressing feedback from the journal. This study aims to answer the following research questions:

d) How appropriate is the Patient Activation Measure (PAM), based on WHO and COSMIN guidelines, for assessing health activation levels in the Hungarian general population aged 40 and above?
e) How does patient activation, as measured by PAM-13, influence lifestyle-related health risks in the Hungarian population?
f) What is the impact of digital health literacy on patient activation and health-seeking behaviors in Hungary?

Results:

The sample comprised 779 respondents. Mean (± SD) age was 60.4 ± 10.6 years, 54% were female and 67% had chronic illness. Mean (± SD) PAM-13 score was 60.6 ± 10.0. We found good internal consistency (Cronbach alpha: 0.77), moderate test–retest reliability (ICC: 0.62; n = 75), a single-factor structure and good content validity: PAM-13 showed moderate correlation with the eHealth Literacy Scale (r = 0.40), and no correlation with age (r = 0.02), education (r = 0.04) or income (ρ = 0.04). Higher PAM-13 scores were associated with fewer lifestyle risks ($p < 0.001$), more frequent health information seeking ($p < 0.001$), participation in patient education ($p = 0.018$) and various online health-related behaviours. When controlling for health literacy, sociodemographic factors and health status, the association of higher PAM-13 scores with overall fewer lifestyle risks, normal body mass index, physical activity and adequate diet remained significant. Similar properties were observed in the subgroup of participants with chronic morbidity, but not in the age group 65+.

The third study has not yet been published. The aim is to publish it, like the previous articles, in an high-impact journal after defending the dissertation.

This study aims to address the following research questions:

g) How can the Patient Activation Measure (PAM) be integrated into health insurance pricing models to predict healthcare costs?
h) Does a higher Patient Activation Measure (PAM) score correlate with lower

health-related costs in the Hungarian population, or does it contribute to increased expenses?

i) How do demographic factors (age, income, education) interact with patient activation (PAM) in determining healthcare costs and insurance risk?

Results:

Evidence from several studies links higher PAM scores with reduced healthcare costs, confirming the value of PAM in predicting health engagement and the potential for cost savings. Activated patients have lower rates of hospitalisation and emergency department visits, which are major cost drivers.

Using five generalised linear models (GLMs), this study confirms that PAM is a significant predictor of healthcare costs in four models. The group with the highest PAM consistently showed lower cost multipliers, underlining the predictive validity of PAM for cost management in health insurance pricing.

Demographic factors such as age, income and education had a significant impact on costs. For example, more educated groups and those with higher incomes had lower predicted costs, reinforcing the link between socio-economic status and health management.

While all models passed validation tests, the correlations between predicted and actual costs were moderate, highlighting the complexity of cost prediction. The model that included the PAM, the Hungarian regions and was built on a dataset including outliers showed the strongest correlation, suggesting that including the PAM alongside socio-demographic factors provides more reliable estimates.

Overall, the study establishes PAM as a viable tool in health insurance modelling and suggests that improving patient activation could reduce health expenditure and support more sustainable insurance models in the Hungarian healthcare system. Overall, the study shows that insurance companies can consider the use of PAM in their product pricing.

Part I. A fundamental model of health insurance pricing

The health crisis caused by the coronavirus pandemic has highlighted how vulnerable we are. Meanwhile, health care is increasingly shifting towards the private sector, which is significantly increasing extra household expenditure. More and more private health care institutions are constantly emerging, while public institutions are declining. State-run clinics, departments and wards are being closed down or transformed due to lack of doctors or resources. Public care is characterised by long waiting lists, outdated and poor-quality equipment, and in many cases the buildings are in a particularly poor state of repair. All this creates opportunities for private health care to take hold. In this situation, insurance can be a particularly suitable means of covering risks.

I.1. Theoretical overview

I will present the pricing of a complete health insurance product through an example. The aim is to illustrate international practice, using premiums calculated for a 40-year-old person as an illustration.

For insurance, statistics are of paramount importance. For part of the calculations, I have relied mainly on *the Demographic Yearbook of the* Central Statistical Office (KSH) and the *National Cancer Register.* For some models it is particularly difficult to obtain statistical data of sufficient quality, a problem which I will always address. However, I consider it important to draw attention to the lack of adequate data, which poses a challenge for the pricing of the products concerned and thus for their creation. In these cases, I will give an approximate price based on an expert estimate, whereas where data is available - for example, for critical illness insurance - I will calculate the premium using the formulas set out in the relevant section, based on data for the Hungarian population. I will start by clarifying the basic assumptions related to health insurance, followed by a brief introduction to health insurance products, and finally I will share some thoughts on health schemes. The chapters are based on my previous work, using the latest available data *(Németh, 2016)*.

I.1.1. Basic assumptions

In the health insurance models presented in this article, I make the following assumptions. In the models under consideration, I ignore issues related to population change, such as migration. I have used the *2019 Demographic Yearbook of the KSH, as the* latest data from the *National Cancer Register* also refer to 2019. I have not taken into account the impact of the coronavirus pandemic on mortality in the pricing, calculating that in the long run, vaccines will reduce this impact and mortality trends will return to those seen before the coronavirus epidemic, with the pre-viral trend continuing.

For cash flows, I assume the following. For the income side, I assume earned income and pension income. By earned income I mean net income, excluding various public benefits and transfers. This item comes from the employer. By pension income I also mean the amount received, which is provided by the state. On the expenditure side, mention should be made of health care costs and costs related to long-term care. Health care costs include the cost of medicines and various medical treatments, hospital, outpatient and general practitioner care, and diagnostic costs, which are partly financed by social security and partly by patients. For the costs of long-term care, I include the costs of using specialist nurses and the costs of changes in living conditions.

By default, I ignore inflation, as it was not as critical in 2019 as it is today, and I assume that wages will increase as people progress in their careers, as work experience makes them more valuable employees.

Health-related costs, if not financed by public funds, can be covered in several ways. I would like to highlight three of them. The first is the so-called *out-of-pocket* expenditure, i.e. the amount paid directly to the provider when the service is used, which is financed by the patient himself. In the Hungarian system, this is often the case for dental care, and the co-payment for medicines, which is also a significant expense in Hungary.

The second form of expenditure is health savings, which is a form of savings that can be used directly to cover health expenditure. Savings are best suited to finance medium probability and medium cost health events, as shown in Table 1. In Hungary, these are health savings.

However, the third form of expenditure, insurance, is the most important for this study. This can be annual, multi-year or even life-long insurance. Table 1 shows the magnitude of costs for different financial strategies according to the probability of a health event occurring.(Pitacco, 2014)

Table 1 Strategies for financing health-related costs

Health-related event		Appropriate financing strategy
Cost	**Probability**	
Low	High	Out-of-pocket
Medium	Medium	Savings
High	Low	Insurance

Source: *Pitacco (2014).*

Notes: This table indicates the optimal strategy for different probabilities and costs associated with health-related events.

I.1.2. Financing options for health systems

Public health systems in European countries are predominantly financed by taxes and/or contributions. Private health insurance is financed by premiums, which can be individual or group, depending on the service provided, and with some exceptions, are contributory. I classify private health insurance according to their role as follows, based on *OECD (2004)* and *Boncz (2015):*

1) Private primary health insurance:
   a) Main private health insurance: when an individual is unable to obtain public health insurance.
   b) Substitutive private health insurance: health insurance that replaces or is equivalent to public health insurance.

2) Duplicated private health insurance: when private health insurance includes elements that are also included in public health insurance, the difference is the quality of the service or the choice of provider.

3) In the case of extra private health insurance, the individual has compulsory insurance and pays the specified contribution or tax, but chooses to take out extra insurance on top of the basic package provided by the state. There are two ways of doing this:
   a) It also provides supplementary insurance for benefits that would otherwise be paid in part or in full by the individual, a form of alternative insurance.
   b) Complementary insurance: covers extra health services that are not included in the basic package. Voluntary insurance may be based on the individual's share of the premium or on the replacement of a co-payment.

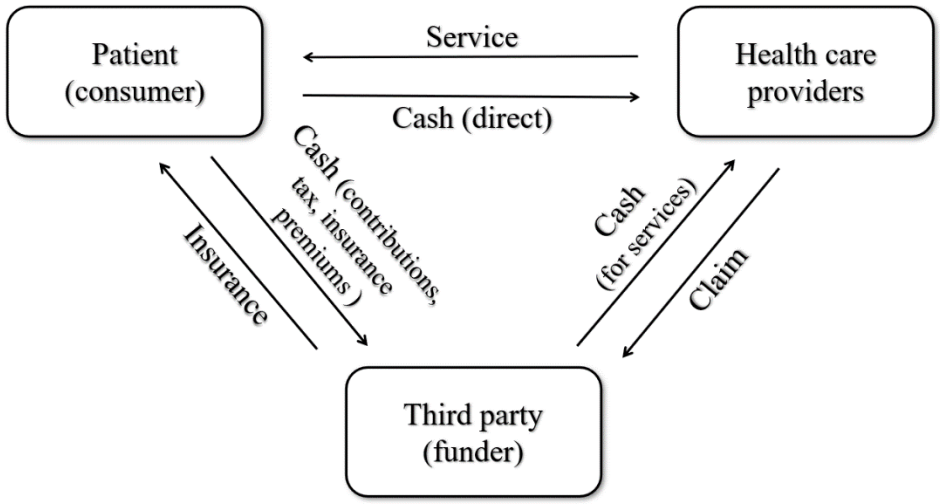
The health care market is different from other product markets. The main focus here is to develop a health system that aims to improve health and equality of opportunity, and to maximise the improvement with the resources available. It provides choice for the individual and is also concerned with affordability and transparency in the use of public funds *(Stiglitz, 2000)*. Market failures are common because competition is limited and the laws of the market are distorted. The majority of health care providers are not for profit, and there are soft barriers to entry and exit, and the consumer is disconnected from the funder. Market failures are mainly due to the following, which are discussed in more detail below:

- derived demand,
- ” free-rider effect “ and moral hazard,
- the principle of equal access,
- monopoly or oligopoly,
- information asymmetry,
- risk selection,
- adverse selection,
- the vulnerability of the patient,
- the demand generated by the service provider.

Derived demand arises from the fact that health as a product is intangible and cannot be priced, i.e. it has no market value, so we consider health services. Part of health services are public goods, so their supply cannot be organised on a purely market basis. Public goods cannot be charged for and we cannot prevent people from consuming them, so there is a free-rider effect or moral hazard, just think of the anti-plague measures. Some

health services are locally based, the provider side is an oligopoly market, there is little scope for enforcing free choice of doctors. Informational asymmetry means that in the case of business, one party has excess information over the other. The party with more information can very easily use its advantage to disadvantage the other party. In a simple case, the two players are the doctor and the patient. In this case, the patient may not be aware of the treatment methods available and may be easily persuaded to undergo a more expensive procedure, which is not necessarily a better treatment than the one available at a lower price. The health care market is a three-tier market: health care providers, patients and funders. In this case, the funder may be the patient, but it may also be private insurers or, where appropriate, the state. Figure 1 illustrates the actors in the health care market and their relationship to each other, showing why information asymmetry is a problem *(Vallyon, 2011).*

Figure 1 Health service and financing schemes



Source: *Stiglitz (2000).*

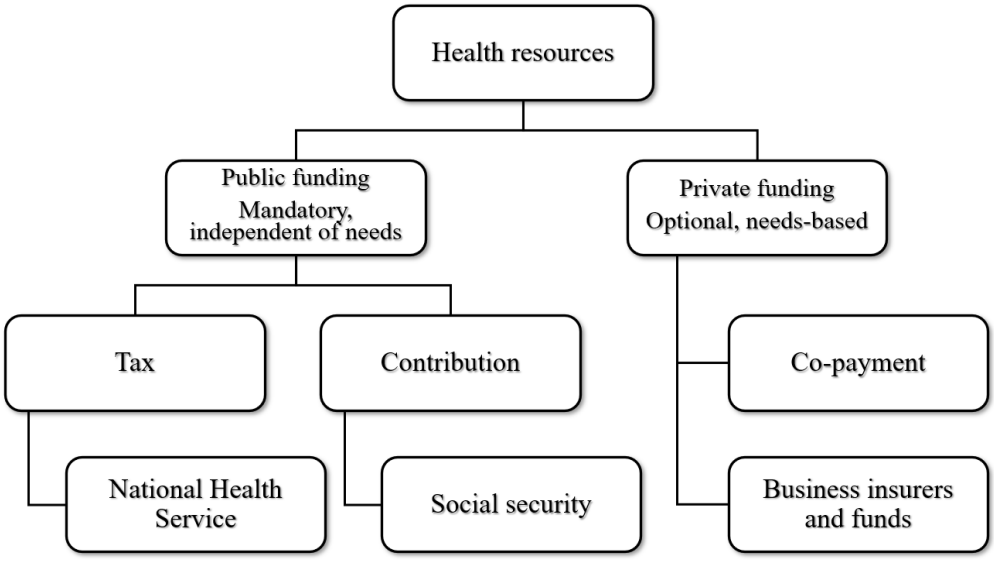
Notes: This figure illustrates the three parties and their relationships to demonstrate the information asymmetry.

The insurance premium is set by the insurer to cover the expected value of the loss due to the risk and the administrative costs and return, but this can lead to adverse selection. If the mechanism of risk equalisation is removed from the market, risk selection will emerge, with the result that poorer, sicker individuals are likely to be left uninsured

(Szabó-Gilly, 2003). Social insurance can operate not only in a single-insurer model but also in a multi-insurer model. This does not completely eliminate private insurance, which can operate as both basic and supplementary insurance. The advantages of social insurance include the application of the principle of solidarity, lower administrative costs, the elimination of harmful selection, and the management of externalities in the health market. Private insurance also has the advantages of choice, of encouraging further health savings, and of reinforcing individual responsibility and empowerment. Health insurance can be divided into two types of financing: public and private. Public health insurance is characterised by the principle of solidarity, while private health insurance is characterised by the principle of equivalence *(Vallyon, 2011).*

Health resources can be divided into two parts, public and private, as shown in Figure 2. It can be seen that public funding is made up of taxes and contributions. *Co-payment* refers to the fee paid by the patient for a service or product when using publicly funded care *(Kincses, 2006)*.

Figure 2 Classification of health care resources



Source: *Kincses (1999).*

Notes: This figure illustrates the healthcare system funding schemes, separating public and private funding and including their respective subcategories.

I.1.2.1. Financing characteristics of the Hungarian health care system

In the European Union, Member States are free to decide on taxes, contributions and services provided to their citizens, which is not a harmonisation issue. There are only coordination rules to ensure that citizens are not disadvantaged when crossing borders.

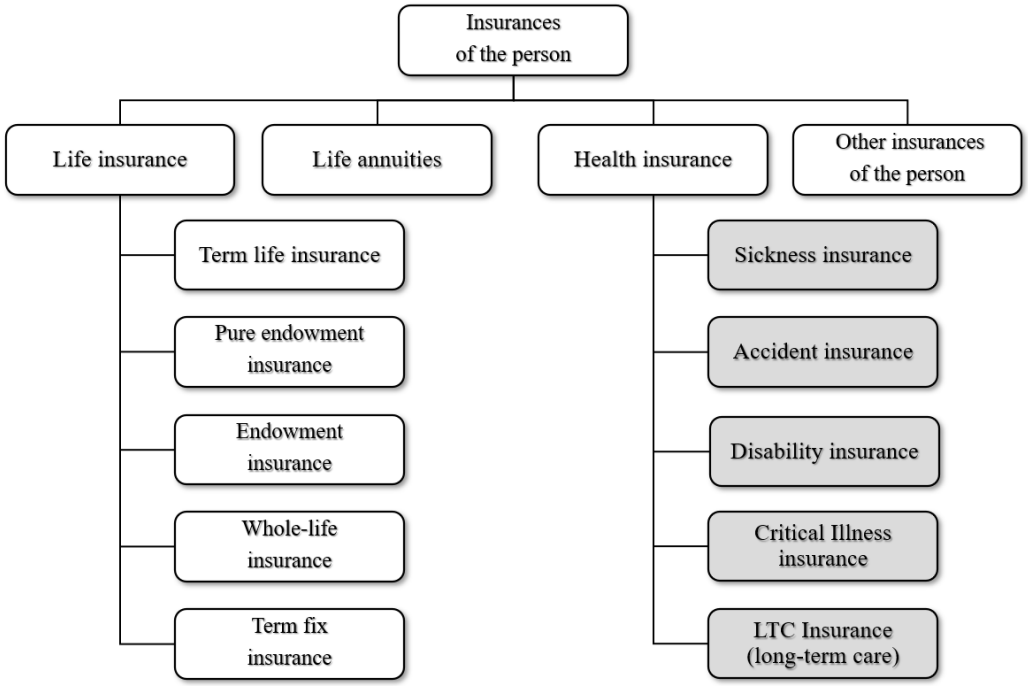
The Hungarian system puts solidarity first, meaning that access is open to all. However, this means that a pure market cannot operate, and it is up to the state to ensure the right of access. According to OECD data (OECD, 2019a), health expenditure in Hungary is around 6-7% of GDP. Private funding is between 34% and 36%. A significant proportion of this is neither insurance nor health insurance, but expenditure financed by households from their own resources.

Private health insurance is available through two different institutional schemes. One is through insurance with commercial insurers, and the other is through self-organised, not-for-profit health funds. In the first case, one should think in terms of cash benefits, while in the second case one should think in terms of benefits in kind, health-related services. Both play a complementary role in our country. In addition to social insurance, health insurance funds and insurers, other important aspects in Hungary include social assistance provided by the government and local authorities, donations offered by non-profit NGOs and private charities, sick leave provided by employers and expenses covered by households *(Vallyon, 2011)*. Health insurance currently has a low share of the Hungarian insurance market, while market players are constantly trying to develop new products and reach a wider audience. This shows their confidence in more positive results and their belief that there are still untapped areas in the health insurance market. In this context, insurers are trying to use all means to raise awareness of this product group among companies and consumers. According to market players, an improving economic environment would allow employers to increasingly provide for their employees through company cafeteria schemes and to encourage their employees to buy health insurance cover *(MABISZ, 2023)*.

I.1.3. Types of health insurance products

In this section, I will briefly describe the main types of health insurance, which will be discussed in the section on private health insurance and will also be covered in the product analysis. The reason for choosing private health insurance rather than social insurance is that the conditions of private health insurance are more freely customisable and easier to change. Figure 3 shows the grouping of personal insurance according to *Pitacco (2014)*, with each type of health insurance product as a subgroup.

Figure 3 Insurances of the person: basic products



Source: *Pitacco (2014).*

Notes: This figure depicts the different types of insurance that cover an individual. The most notable aspect here is the Health Insurance section.

In the following subsections, I turn to the characteristics of the different health insurance products.

I.1.3.1. Sickness insurance

The insurer provides coverage in the event of the insured person's illness, typically in the form of reimbursement for medical expenses. This may include inpatient care, covering

all hospital-related costs such as treatments, emergency services, and medications. Outpatient services, diagnostic tests, and doctor-prescribed medications may also be reimbursed. Additionally, some policies offer a daily hospital allowance, which is a fixed benefit and not tied to actual expenses. When reimbursing medical expenses, risk-sharing is regulated through deductibles, i.e. the following methods (and any combination of them) are usually used:

- Deductible (fixed amount): the insurer sets a deductible in the contract, which is a fixed amount deducted from the amount of compensation. If the amount of the loss is less than the deductible, the insurer will not pay any compensation.
- Proportional deductible (expressed as a percentage): the insurer sets a deductible in the contract, expressed as a percentage, and deducts it from the amount of compensation.
- Stop-loss: the maximum amount an insured person has to pay out-of-pocket for his or her health care expenses. This can be set simply per claim or even per insurance period.

The signing of a contract is usually preceded by a health assessment. There are some types of health insurance products that can be taken out after a partial health assessment or even without a health assessment, for speed and flexibility. Of course, this entails a relatively high premium, as it is assumed that this option is sought by riskier customers. Another common way of dealing with adverse selection is to introduce a waiting period, during which the insurer is not obliged to pay the sum insured for an insured event. A well-designed insurance system also encourages health preservation through the assumption of risks of illness by insurers *(Asztalos, 1997).* Taking the above terms into account, cost sharing can be described as follows, based on *Pitacco (2014):*

$$x = u + y \tag{1}$$

$$u = \begin{cases} x & \text{if } x < D \\ \alpha(x - D) + D & \text{if } D \leq x < M \\ SL & \text{if } x \geq M \end{cases} \tag{2}$$

$$y = \begin{cases} 0 & \text{if } x < D \\ (1-\alpha)(x-D) & \text{if } D \le x < M \\ x - SL & \text{if } x \ge M \end{cases} \quad (3)$$

where

$$M = \frac{1}{\alpha}(SL - (1-\alpha)D).$$

$x =$ amount of costs

$u =$ out-of-pocket

$y =$ refund

$D =$ amount of co-payment

$\alpha =$ co-payment rate, $0 < \alpha \le 1$

$SL =$ stop-loss

If $\alpha = 1$, *then* $M = SL$, *or if* $D = 0$, *then* $M = \frac{SL}{\alpha}$.

I.1.3.2. Accident insurance

Accident insurance is an insurance contract that provides benefit to the insured person in the event of an accident or to the beneficiary in the event of death. The benefit can be a lump sum payment or an annuity.

An accident is an unusual, unexpected and unintentional event that occurs at a place and time for no apparent or intended reason. It is usually an event with a negative outcome that could have been avoided by detection before it occurred.

In the case of accident insurance, the following payments are possible:

- In the event of death, if the insured person dies as a result of an accident, a fixed amount of benefit is paid to the beneficiary.
- In the case of permanent disability, the amount of the payment depends on the percentage of health impairment.
- Reimbursement of accident-related medical expenses.
- A daily allowance in cases where the disability is temporary. It means a fixed amount of benefit for a limited period of time, usually from a few months to a year.

The duration of the insurance is usually annual, but it can sometimes be multiannual. In general, different amounts of insurance cover are available for accidents of different origin and nature. For this type of insurance, the insurance contract often contains exclusions, including accidents related to armed conflicts. There are also special accident insurance policies, such as travel or student accident insurance. Accident and sickness insurance is often used as a supplementary insurance to life insurance *(Asztalos, 1997; Jean, 2004).* They are quite common and there are many versions of these products on the market, but they are not at the top of the range in terms of premium income. A detailed picture for Hungary can be obtained from the latest quarterly MABISZ (Association of Hungarian Insurance Companies) report *(MABISZ, 2023)* (Table 2). It can be seen that accident insurance is in a prominent position in terms of the stock, health insurance is particularly rare.

Table 2 Insurance lines of business, Q4 2022

thousand HUF, piece

	Gross written premium	Single premium	Claims (cy)	Number of policies	Earned premium
Life insurance	60 592 593	22 058 566	47 523 180	554 938	42 373 689
Pure endowment insurance	5 226 338	391 320	6 351 268	26 633	4 720 567
Endowment insurance	95 093 052	64 833 072	136 026 159	186 743	28 741 641
Unit-linked life insurance	291 631 248	131 082 109	244 187 752	669 508	187 130 629
Life insurance	8 053 806	109 555	3 803 028	8 131	8 758 494
Pension insurance	135 088 690	18 360 682	20 304 854	451 861	116 967 097
UL Pension insurance	*105 192 011*	*14 779 344*	*14 834 068*	*344 787*	*90 429 296*
Other pension insurance	349 003	37 393	-1 752 752	1 325	336 290
Life annuity	1 899 917	1 783 902	1 259 336	718	111 221
Life insurance with a premium protection clause	9 869 426	900 807	15 234 937	36 656	4 301 586
Group premium protection life insurance	10 686 242	125 225	2 152 324	319	9 576 100
Group life insurance	27 491 356	101 544	8 085 985	6 094	27 928 200
Other life insurance	83 366	100	96 665	427	76 621
Accident and sickness insurance for life insurance*	38 055 707	111 775	9 996 652	2 397 340	40 072 926
Total life insurance	***646 065 037***	***239 784 275***	***483 272 736***	***1 943 353***	***431 022 135***
Private property insurance	165 708 596	168 728	54 524 857	3 296 398	169 807 426
Institutional property insurance	3 600 278	215 924	1 617 763	3 663	3 981 920
Corporate property insurance (except SME)	90 421 980	24 180 745	50 073 163	57 093	59 218 168
SME property insurance	22 103 814	1 751 127	3 536 292	131 059	21 155 955
Other property insurance	5 819 448	2 836 481	4 870 960	8 416	3 024 164
General liability insurance	18 717 243	4 191 260	3 324 515	57 895	15 193 196
Professional liability insurance	11 364 193	925 121	61 662	68 804	10 338 350
Premium protection insurance from non-life insurance contracts (except residential property insurance with credit protection clause)	29 421 749	1 292 247	5 874 758	24 021	24 605 499
Cargo insurance	7 243 978	2 953 518	917 671	26 908	4 627 093
Travel insurance	15 214 642	12 501 140	3 429 400	648	2 301 241
Accident insurance	12 865 586	584 085	3 527 027	618 345	12 779 424
Sickness insurance	10 450 850	17 827	5 320 316	32 394	9 920 617
Car insurance	275 746 138	1 904 917	119 485 617	5 959 784	281 966 272
Motor third-party liability insurance	*275 266 519*	*1 696 891*	*119 247 166*	*5 968 676*	*281 707 704*
Casco	137 091 282	1 140 731	80 431 155	1 025 416	140 959 802
Guarantee and bonding insurance	5 168 763	5 088 970	705 453	955	97 632
Extended guarantee	5 537 526	5 452 409	1 352 257	6 020	106 775
Legal expenses insurance	496 018	6 290	192 737	41 633	505 269
Funeral expenses insurance	238 733	0	181 374	6 381	231 419
Assistance insurance	4 006 376	1 044	1 153 042	1 740	3 820 372
Insurance against various financial losses	6 045 242	1 164 908	785 288	51 885	4 029 430
Total non-life insurance	***827 262 435***	***66 377 472***	***341 365 306***	***11 419 458***	***768 670 025***
Total companies	***1 473 327 472***	***306 161 747***	***824 638 042***	***13 362 811***	***1 199 692 160***

Source: *MABISZ (2023).*

*Notes: * The line "Accident and sickness insurance linked to life insurance" is not included in "Total life insurance". This table presents the aggregated results of Hungarian insurance companies for 2022, including gross written premiums (GWP) and claims. "CY" refers to the calendar year. Not all insurers provided detailed data.*

I.1.3.3. Disability insurance

With this type of insurance, the insurer covers both permanent and temporary disability. It is usually paid in the form of an annuity, but lump-sum payments are also common. It is available on the market in both individual and group insurance. In the case of individual insurance, it is also known as income protection insurance. There are three definitions of total disability *(Asztalos, 1997):*

- The insured is unable to continue his/her occupation.
- The insured person cannot continue in his/her profession or in any occupation related to his/her previous training.
- The insured cannot engage in any gainful activity.

Group insurance is usually taken out by companies for their employees. The two main types are short-term and long-term insurance. The amount of the benefit is often linked to a percentage of the pre-disability salary. Disability is periodically reviewed or strict rules are put in place to reduce moral hazard. Here too, of course, insurers index the premium and the sum assured (if specified), usually by a factor greater than one, in order to protect the policyholder against inflation. Of course, insurers do not necessarily multiply the premium and the sum assured by the same factor.

I.1.3.4. Critical illness insurance

Critical illness insurance pays a benefit in the event of diagnosis of a serious illness from a pre-defined list of such conditions. This list may of course change as medical science develops. Typically, this includes various cancers, heart attacks and strokes. Here, too, insurers often make use of waiting periods to reduce the adverse selection mechanisms. Critical illness cover is often only provided as a supplementary cover to life insurance, and not as a stand-alone insurance. It is important to stress that in this type of product, the insurer pays on the basis of the diagnosis and not on the severity of the illness. This means

that the service does not take into account medical costs or the financial loss caused by the illness. As medical science advances, life expectancy after a critical illness is becoming longer and the likelihood of survival and recovery is increasing. Sometimes the insured may wish to extend his/her policy after the cure, as it will cease after the first claim, but there may be other critical illnesses and the insured may wish to extend his/her policy at the same tariff without a new medical assessment. There are two ways to do this. The first is renewing the insurance, where the insurer will again impose a waiting period or include the illnesses that caused the previous insured event and related illnesses in the exclusions. The other option is when the insurance covers several illnesses in advance and the insurance only expires at the end of the term. Here again, the condition is that the insurer does not pay more than once for illnesses that have already occurred during the term or for illnesses related to it *(Dash-Grimshaw, 1993).* There are versions of this product where the payout depends on the severity of the illness or includes a death benefit. In the event of death, a distinction can be made between the case where the death of the insured occurs in connection with a critical illness and the case where it does not. If the insurer pays an annuity based on the severity of the illness or if the insurance includes these death payments, then it is worth using multi-state Markov chains for pricing purposes. Similar models are presented later for disability and LTC insurance *(Pasaribu et al., 2019).*

I.1.3.5. Long-term care insurance

This insurance is designed to cover long-term care costs. It can be taken out as a comprehensive or indemnity insurance. In the former case, the insurer pays a predetermined amount, while in the latter it reimburses care costs. The assessment of the severity of the disability also plays an important role. Among the various indices, one of the most widely used is the Activities of Daily Living (ADL) index *(Pitacco, 2014),* which takes into account the following activities of daily living: eating, cleaning, dressing, mobility, personal hygiene, toileting. Thus, based on how many of the above list of activities the insured is able to perform independently, they are classified into categories (0, 1, 2 or 3). 0 means no need for care according to the classification, 1 is category 1 if the insured cannot perform 3 of the above activities on their own, 2 when they cannot perform 4 or 5 activities on their own, and 3 is total disability when the insured needs assistance to perform all 6 activities.

This is similar to the Barthel index *(Pitacco, 2014)*, which is also widely used. Here, 13 activities are assessed, including mobility, eating, drinking, etc. These are used to determine the percentage level. A level of 40% corresponds to ADL level 1, 70% to ADL level 2 and 100% to ADL level 3. In addition to functional activity, instrumental activity is also measured, for which the Instrumental Activities of Daily Living (IADL) test *(Pitacco, 2014) is* used. This includes activities such as shopping, making a phone call, preparing a meal or housework. Long-term care (LTC) insurance can be classified in three ways in terms of how the service is financed. The first group includes insurance with a predetermined amount. There are fixed-amount services and services depending on the level of disability. The second class includes reimbursement of medical expenses. Lastly, the third category covers various care services, including reimbursement of nursing home fees. This is usually automatically indexed to inflation in most countries. All in all, such insurance is considered to be particularly expensive, and only the upper and upper middle classes of society can afford to buy such insurance *(Asztalos, 1997)*. As life expectancy increases, the insurer is expected to provide benefits for a longer period of time. This in turn increases premiums *(Kovács, 2017).*

I.2. Pricing of health insurance products

In this chapter, I present premium calculations for private health insurance products. The example used to illustrate these is middle-class young adults in Hungary, a segment that could benefit from this insurance in the long term and are able to afford it. The purpose of the article is to illustrate, to calculate the price of products in order to draw conclusions and to examine the feasibility of further developing the models in terms of pricing, given the right data. As this is a rather complex topic, simpler models are presented. This is mainly for reasons of scope, but also because of the limited access to data. Modelling is also complicated by the fact that different diseases are not completely independent of each other, i.e. there are cases where one disease increases the risk of another, but there are also examples where one disease is a direct consequence of another. These cases can be dealt with in two ways. One approach is to use simplified models that include only independent diseases. The other option is to produce a complex and rather extensive

model by examining each case and studying relevant medical and statistical research.[1] As the subject of my article is health insurance, the main interest of the various products is the size of the premiums for the services provided, and I will therefore use simpler models.

In the following subsections, I will present the specific pricing methods for each product and their results.

I.2.1. Sickness insurance

In the case of sickness insurance, the aim is to set up an insurance policy to cover any medical costs that may arise as a result of illness. It would basically cover three services. A hospital daily allowance for intensive care unit treatment for up to 180 days a year, amounting to HUF 10 000 per day. A hospital daily allowance for sickness for a maximum of 180 days, amounting to HUF 5 000 per day. Also convalescence allowance for a maximum of 28 days, amounting to HUF 2 500 per day. As with accident insurance, data in this area is very limited. Again, without adequate data, I have only used an expert estimate to determine the premium. On this basis, the net annual premium for a 40-year-old person **is HUF 40 000.** In Part III, once sufficient data has been obtained, I will develop and calculate five sickness insurance models that are significantly more sophisticated and advanced than the basic model outlined here. The primary purpose of presenting this model is to demonstrate its structure and approach. Due to the lack of publicly available data, the estimates provided are based on my nearly decade-long personal experience in this market.

I.2.2. Accident insurance

In our country, as can be seen from the *MABISZ* data (Table 2), accident insurance stands out in terms of the number of policies included in the article. This type of insurance is widely available on the market. This is because insurers develop different products for different types of accidents. In Hungary, according to *Eurostat* data, domestic injuries top the accident ranking, followed by accidents at work, traffic accidents and sports accidents.

[1] One such possible approach, favoured by actuaries, is the classification of risks into homogeneous clusters *(Ágoston et al., 2019).*

I.2.2.1. Preparing the calculation

In the case of accident insurance, the aim is to develop a product that provides a service to the customer in the event of accidents in the home. As compulsory motor liability insurance may cover road accidents and employers may insure their employees in the event of accidents at work, these cases are excluded from the proposed product package. For the calculation, I will determine the probabilities of occurrence for the various household accidents, as in the previous insurance calculation, and assign the appropriate sum insured to cases of different severity. However, researching household accident statistics did not yield data of sufficient quality and quantity, as these data are not publicly available, so I had to look for another way.

I.2.2.2. Results

As the calculation method would not deviate from the previous method, the only information needed for the subsequent calculations is the annual net premium rate. I have therefore determined a realistic net premium by means of an expert estimate based on the gross premiums available online from insurers. Without available data, I was unable to perform the calculations myself, so I decided to explore the market. In the Hungarian insurance market, many companies offer accidental coverage through online products. By reviewing the terms and conditions, I gained insights into the pricing of these products. These prices represent gross values, and based on my experience, I estimated a fair loading of around 30-40%. Using this information, I approximated a fair net price. However, it is important to note that the primary focus here is the description of the methodology, with the prices serving only as an illustrative example to provide a clearer understanding of the product. The insurance pays HUF 5 million to the beneficiary in the event of accidental death, HUF 250 000 for surgery and HUF 50 000 for broken and fractured bones. Gross premiums range from HUF 25 000 to HUF 30 000 for a 40-year-old person. On this basis, the net annual premium is estimated at **HUF 20 000.**

I.2.3. Disability insurance

People take out disability insurance to avoid loss of income due to illness and accidents.

As with life insurance, disability insurance also uses indexation to protect the policyholder against inflation. An interesting feature of disability annuities is that the annuity rate can be reduced over time if the period of previously established disability entitlement is extended. This mechanism is intended to encourage return to work and the recovery of earnings (at least to some extent) *(Pitacco, 2014).*

In the health insurance package presented later as an example, disability insurance lasts until the insured retires, i.e. until the age of 65 in the case of this calculation. The nursing care insurance is valid after retirement age, i.e. after the age of 65. This breakdown is necessary as the products in the package are divided into disjoint parts, well separated from each other. And disability is linked to care, so these products are linked to separate life stages. For similar reasons, disability payments are not included in accident insurance.

I.2.4.1. Theoretical background

To prepare the calculation, we need to clarify a few concepts. These concepts can be found in the textbook by *Pitacco (2014).*

The insured period is the period during which the insurance cover is active in the sense that the benefit is only payable if the claim is made during this period. Generally, the period from the conclusion of the contract to the expiry of the contract (m). The waiting period is the period from the conclusion of the contract (c) during which the insurance cover is not yet active. It is intended to limit the effects of adverse selection. The deferred period is the period (f) for which the declared disability must be at least present before the insurer starts to pay the annuity. Thus, if a contract includes both a waiting period and a deferred period, the total length of the period of risk $m - c - f$.

The qualification period is usually stipulated for insurance policies that provide a lump sum payment in the event of permanent disability. The purpose is to enable the insurer to ascertain the permanent nature of the disability. The maximum benefit period is the upper limit of the period (s) for which benefits are reimbursable. It may exceed the period of insurance and may even last until the death of the insured. In the case of accidental disability, it is generally shorter than for those due to illness. The stopping time is the time from the start of the contract (τ) when the payment of the annuity stops. This often

coincides with the retirement age, so if an individual y at the start of the contract and the retirement age is ξ, then $\tau = \xi - y$.

After the definitions, I turn to the actuarial models of disability insurance. These formulas are based on the textbook by *Pitacco (2014)*. First, consider an insurance policy that pays an annuity b per year if the insured is disabled. Let i denote the condition of being inactive and let a denote the condition of being active. Let B_h be the variable which the insurer must assigned to the amount h paid by the insurer at that time:

$$B_h = \begin{cases} b & \text{if, at time } h, \text{state} = i \\ 0 & \text{if, at time } h, \text{state} \neq i \end{cases} \tag{9}$$

From here we get the present value of annuities:

$$Y = \sum_{h=1}^{m} B_h v^h, \tag{10}$$

where v is the discount factor.

Its expected value, i.e. the expected present value of future payments:

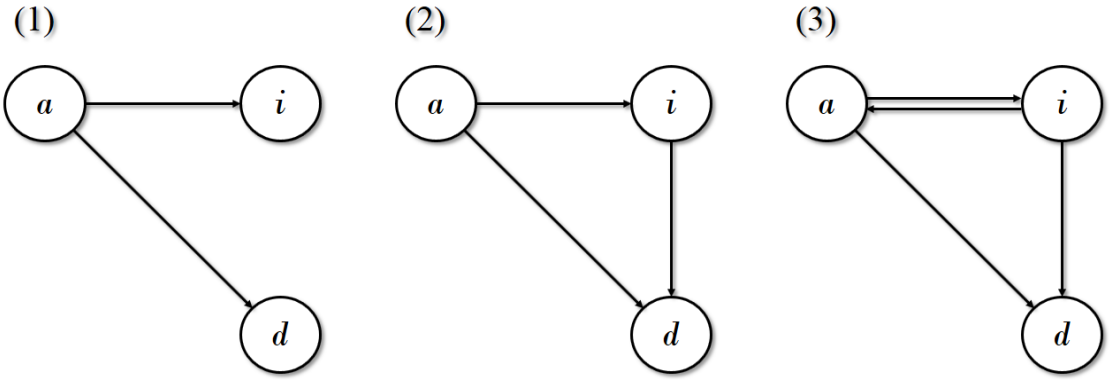
$$\mathbb{E}(Y) = \sum_{h=1}^{m} \mathbb{E}(B_h) v^h = \sum_{h=1}^{m} b \cdot {}_h p_y^{ai} \cdot v^h, \tag{11}$$

where ${}_h p_y^{ai}$ is the probability an individual aged y will become disabled at time $y + h$, and b is the amount of benefit due.

Multi-state models are often used to calculate disability insurance, so we now present a three-state model. The three states are a, i.e. active (healthy), the i, i.e. disabled, and d, i.e. dead. Multi-state models can be illustrated by graphs, where the directed edges are the transitions and the vertices are the states. The graph representation is illustrated using three examples, which are shown in Figure 4. In the first case, the insured receives a lump sum benefit if $a \longrightarrow i$ a transition occurs, i.e. permanent disability. This is the simplest structure. The second is more complicated, where an annuity is paid, and where the number of years of disability is taken into account, so that the moment of death is also relevant. While in the former case only the probability of disability is relevant, in the

latter case the death rate of the disabled is also an important factor. In the third case, the possibility of recovery is also taken into account, i.e. $i \longrightarrow a$ transition, when annuity payments also cease *(Pitacco, 2014).*

Figure 4 Three-state models



Source: *Pitacco (2014).*

Notes: This figure illustrates the three-state Markov model, where "a" represents active, "i" represents inactive, and "d" represents dead.

Let's first take a one-year period to determine the transition probabilities. Assume that in one year there is at most one change of state other than death. As an illustration, I have collected some cases in Table 3.

Table 3 Examples of transitions between states

State at age y	Transition(s)	State at age y + 1	Allowed by the model?
a	→	*i*	Yes
i	→	*a*	Yes
a	→ *i* →	*a*	No
a	→ *i* → *a* →	*i*	No
a	→	*d*	Yes
i	→	*d*	Yes
a	→ *i* →	*d*	Yes
i	→ *a* →	*d*	Yes
a	→ *i* → *a* →	*d*	No

Source: *Pitacco (2014).*

Notes: This table illustrates the transitions in the three-state Markov model, where "a" represents active, "i" represents inactive, and "d" represents dead.

Then the transition probabilities of a y-year-old person in state a: p_y^{aa}, that is, they remain active; q_y^{aa}, when he dies as active; p_y^{ai} when he or she becomes disabled; q_y^{ai} and, in the case of a person who dies as an invalid. These are all conditional probabilities, where the condition is the active status of the individual at the beginning of the duration. The following equations are then satisfied:

$$p_y^{aa} + p_y^{ai} = p_y^a \qquad q_y^{aa} + q_y^{ai} = q_y^a \qquad p_y^a + q_y^a = 1 \qquad p_y^{ai} + q_y^{ai} = w_y, \tag{12}$$

where p_y^a survival, q_y^a death and w_y and the probability of disability *(Pitacco, 2014).*

A similar definition of the y-year-old person's disability insurance probabilities: p_y^{ii}, q_y^{ii}, p_y^{ia}, q_y^{ia}, p_y^i, q_y^i, r_y, where r_y expresses the probability of recovery. We have assumed that only one transition other than death can occur, i.e. p_y^{aa} and p_y^{ii} there is no intermediate event. If we consider only permanent, definitive disability, then obviously $p_y^{ia} = q_y^{ia} = 0$. By arranging the probabilities defined so far into a stochastic matrix, we obtain the conditional probabilities of transitions, which are given in Table 4. The individual's status at age $y + 1$ only depends on his status at age y, so our model is a Markov process *(Pitacco, 2014).*

Table 4 Conditional probabilities of being in states

State at age y	State at age y + 1		
	a	*i*	*d*
a	p_y^{aa}	p_y^{ai}	q_y^a
i	p_y^{ia}	p_y^{ii}	q_y^i
d	0	0	1

Source: *Pitacco (2014).*

Notes: This table illustrates the conditional probabilities in the three-state Markov model, where "a" represents active, "i" represents inactive, and "d" represents dead. "y" denotes age, "p" represents the probability of transitioning from state y to state y+1, and "q" indicates the probability of dying in that state.

The previous model is extended to several years, where the y annual probability of health or disability of an individual aged $y+h$: ${}_hp_y^{aa}$ and ${}_hp_y^{ai}$, which can be written in the following recursive form using the Markov property:

$$ {}_hp_y^{aa} = {}_{h-1}p_y^{aa} \cdot p_{y+h-1}^{aa} + {}_{h-1}p_y^{ai} \cdot p_{y+h-1}^{ia} \tag{13} $$

$$ {}_hp_y^{ai} = {}_{h-1}p_y^{aa} \cdot p_{y+h-1}^{ai} + {}_{h-1}p_y^{ai} \cdot p_{y+h-1}^{ii} \tag{14} $$

where ${}_0p_y^{aa} = 1$ and ${}_0p_y^{ai} = 0$. Let ${}_hp_y^{\underline{aa}}$ and ${}_hp_y^{\underline{ii}}$ denote the probabilities that an individual remains in the states a or i throughout h periods:

$$ {}_hp_y^{\underline{aa}} = \prod_{k=0}^{h-1} p_{y+k}^{aa} \qquad {}_hp_y^{\underline{ii}} = \prod_{k=0}^{h-1} p_{y+k}^{ii} \tag{15} $$

Substituting this into equation (14):

$$ {}_hp_y^{ai} = \sum_{r=1}^{h} \left[{}_{h-r}p_y^{aa} \cdot p_{y+h-r}^{ai} \cdot {}_{r-1}p_{y+h-r+1}^{\underline{ii}} \right]. \tag{16} $$

So the probability that the y annual, a condition of an individual h in one year i is the probability (r) of the probability that $y+h-r$ to become disabled at age one ($y+h-r+1$ at age i-and remain disabled for the rest of the time, $r-1$ years *(Pitacco, 2014)*.

The expected present value of future payments where the policyholder is initially active (a status) individual aged y who is in receipt of a unit payment ($b=1$) in the event of disability:

$$ a_{y:m}^{ai} = \mathbb{E}(Y) = \sum_{h=1}^{m} v^h {}_hp_y^{ai}. \tag{17} $$

Using equation (16):

$$ a_{y:m}^{ai} = \sum_{h=1}^{m} v^h \sum_{r=1}^{h} \left[{}_{h-r}p_y^{aa} \cdot p_{y+h-r}^{ai} \cdot {}_{r-1}p_{y+h-r+1}^{\underline{ii}} \right] \tag{18} $$

Using the previous formula $j = h-r+1$-to obtain the following formula:

$$a_{y:m}^{ai} = \sum_{j=1}^{m}{}_{j-1}p_y^{aa} \cdot p_{y+j-1}^{ai} \sum_{h=j}^{m} v^h \cdot_{h-j} p_{y+j}^{\overline{ii}}. \quad (19)$$

I should note here that in the actuarial nominations a denotes an annuity deferred by one year, while $\ddot{a}$ denotes an annuity starting immediately. Which terminology I use for which product is always determined by the characteristics of the product.

Similarly, the expected present value of an invalidity pension can be written up, where the insurer pays an annual amount until the i condition of the insured person, but up to the end of the contract (m). So if the insured $y + j$ becomes disabled at the age of $m - j + 1$ years:

$$\ddot{a}_{y+j:m-j+1}^{i} = \sum_{h=j}^{m} v^{h-j} \cdot_{h-j} p_{y+j}^{\overline{ii}} \quad (20)$$

Using equation (19):

$$\begin{aligned} a_{y:m}^{ai} &= \sum_{j=1}^{m}{}_{j-1}p_y^{aa} \cdot p_{y+j-1}^{ai} \cdot v^j \cdot \ddot{a}_{y+j-1:m-j+1}^{i} \\ &= \sum_{j=1}^{m}{}_{j-1}p_y^{aa} \cdot v^{j-1} \cdot a_{y+j-1:1:m-j+1}^{ai} \end{aligned} \quad (21)$$

where $a_{y+j-1:1;m-j+1}^{ai} = p_{y+j-1}^{ai} \cdot v \cdot \ddot{a}_{y+j:m-j+1}^{i}$ is the expected annual cost of an insurer per $y + j + 1$ per active insured per year, called the premium for the insurance cover.

Maximum benefit period (s):

$$a_{y:m;s}^{ai} = \sum_{j=1}^{m}{}_{j-1}p_y^{aa} \cdot p_{y+j-1}^{ai} \cdot v^j \cdot \ddot{a}_{y+j:s}^{i} \quad (22)$$

Annuity benefit to be used for premium calculation:

$$\ddot{a}_{y:m'}^{aa} = \sum_{h=1}^{m'} v^{h-1} \cdot_{h-1} p_y^{aa} \quad (23)$$

We apply the principle of equivalence to the calculation of premiums and focus only on the net premium. In this way, the expected value of the present value of the premiums is

equal to the expected value of the present value of the payments.

The net single premium for a sum assured of b and a maximum benefit period of s years:

$$A_{y:m} = ba_{y:m}^{ai} \qquad A_{y:m;s} = ba_{y:m;s}^{ai} \tag{24}$$

In the case of regular premiums, the natural assumption is that the insured pays the premium when he or she is active. This premium payment period should be m' ($\leq m$). If no more than m-or s years, then the equivalence equation is satisfied:

$$P_{y,m(m')}\ddot{a}_{y:m'}^{aa} = ba_{y:m}^{ai} \text{ or } P_{y,m(m');s}\ddot{a}_{y:m'}^{aa} = ba_{y:m;s}^{ai} \tag{25}$$

Suppose that $m = m'$. Using equations (21), (23) and (25), we obtain:

$$P_{y,m(m)} = b\frac{\sum_{j=1}^{m} {}_{j-1}p_y^{aa} \cdot v^{j-1} \cdot a_{y+j-1:1;m-j+1}^{ai}}{\sum_{j=1}^{m} {}_{j-1}p_y^{aa} \cdot v^{j-1}} \tag{26}$$

Let $\varphi = \{a, i, d\}$ denote the simplest finite state space that can occur in a disability insurance policy, or $\tau = \{(a,i),(i,a),(a,d),(i,d)\}$ the set of possible direct transitions. Then the (φ, τ) pair is called a multi-state model. Denote $S(t)$ a t the random state at time. Then $\{S(t), t \geq 0\}$ or, in the discrete case $\{S(t), t = 0,1,\dots\}$ stochastic process, whose values are given by the set φ.

Consider first the discrete case, and suppose that any $u > t \geq 0$ integers, and any j, k it is satisfied that $\mathbb{P}\big(S(u) = k \big| S(t) = j \wedge H(t)\big) = \mathbb{P}(S(u) = k | S(t) = j)$, where $H(t)$ is an arbitrary hypothesis for the trajectory contained in the set $\{S(\tau), \tau < t\}$. This expresses precisely that the conditional probability depends only on the information about the "last" state, i.e. $\{S(t), t = 0,1,\dots\}$ is a Markov chain, which is what we assumed at the beginning. The p_y^{ai}, $y = x + t$ for example, with the notation now introduced, takes the following form: $p_{x+t}^{ai} = \mathbb{P}(S(t+1) = i | S(t) = a)$ *(Pitacco, 2014).*

Note that the process $\{S(t), t = 0,1,\dots\}$ can be endowed with a different probabilistic structure, so that it is not a Markov process but, for example, a counting process *(Löfdahl, 2013).*

A more sophisticated definition of state space allows a much more general treatment of dependence in the context of Markov chains, as the example in the previous section shows. The state space $\varphi = \{a, i^{(1)}, i^{(2)}, \dots, i^{(n)}, d\}$ would correspond to this.

In a continuous case, we work with transition intensities instead of one-year transition probabilities. Let $p^{jk}(t,u) = \mathbb{P}(S(u) = k | S(t) = j),\ j \neq k$. Then we interpret the transition intensity μ_t^{jk} as follows: $\mu_t^{jk} = \lim_{u \to t} \frac{p^{jk}(t,u)}{u-t}$. From the intensities, differential equations can be used (in practice, numerically) to determine the $p^{jk}(t,u)$ probabilities. In actuarial practice, intensities *are* estimated from statistical data on mortality, disability and recovery *(Pitacco, 2014).*

I.2.4.2. Results

Disability insurance products are data-intensive, since in reality the degree of disability is defined as a percentage (Haberman, 1999). Without sufficient quantity and quality of data, the calculation is based on an expert estimate. The chosen insurance covers a lump sum of HUF 5 million for a health impairment of more than 50%. The net annual premium for such insurance is estimated at **HUF 65 000** for a person aged 40**.**

I.2.4. Critical illness insurance

In Hungary, cancer is the leading cause of death alongside cardiovascular diseases, according to the Hungarian Statistical Office, so the aim is to create an insurance that provides services in the event of these diseases. The theoretical background and the formulas used are available in detail in *Banyár (2003)* and *Banyár-Vékás (2016).* For the insurance premium calculation of cancer, I have drawn on the work of *Kar (2016)*. Critical illness insurance, also known as dread disease insurance, can cover multiple severe illnesses, with the calculation method remaining unchanged while only the incidence rates increase. When compiling a list of these diseases, the key criterion is that they must be independent of each other. My selection was based on the availability and reliability of the *Rákregiszter* database, which provides a consistent and comprehensive dataset, making it ideal for demonstration purposes. The data used for the calculations can be found in *Appendix I.*

I.2.1.1. Theoretical background

Let K_x be the expected amount of damage per person. Then

$$K_x = \sum_{j=1}^{\tau} q_x^{(j)} SA_j,$$

where $q_x^{(j)}$ is the probability of occurrence of the j-th illness at age x over an one-year horizon, and SA_j is the amount of the benefit.

The previous formula SA_j depends on the type of product. In the case of a lump sum benefit, it is of course the same as the sum insured. For a fixed annuity for n years, let SA_{jar} indicate the annual benefit of the annuity, so that $SA_j = \frac{1-v^n}{1-v} SA_{jar}$, where $v = \frac{1}{1+i}$ and v is the discount factor, and i is the technical interest rate, in this case 2%. If it is a life annuity with a term of n years, then $SA_j = \sum_{y=0}^{n-1} p_y v^y SA_{jar}$, where p_y is the probability that a person diagnosed with a given type of disease lives for at least y years from the date of the diagnosis.

After K_x-is introduced, we determine the simple net premium based on the principle of equivalence. Let $l_{x+1} = l_x(1 - q_x)$, where l_x is the survival order in the life table and q_x is the one-year probability of death at age x . We ignore lapse here.

The net single premium is the following:

$$A_{x:n} = \frac{\sum_{j=0}^{n-1} l_{x+j} v^j K_{x+j}}{l_x} \quad (4)$$

For a whole-life product $n = \omega - x$.

The annual premium $P_{x:n}$ is determined using the principle of equivalence:

$$\ddot{a}_{x:n} P_{x:n} = A_{x:n} \quad (5)$$

where

$$\ddot{a}_{x:n} = \frac{\sum_{j=0}^{n-1} l_{x+j} v^j}{l_x}. \quad (6)$$

Expected claims payments usually increase with age, but premiums remain constant, so at the beginning of the policy the client pays more premiums than the claims need, but less at the end of the policy, so a reserve is needed, this is the ageing reserve, which is calculated prospectively: the present value of expected premium income is subtracted from the present value of expected benefit payouts. The principle of equivalence must be fulfilled not only at the beginning of the insurance but also at any other time. The equivalence equation at time m:

$$A_{x+m:n-m} = P_{x:n} \cdot \ddot{a}_{x+m:n-m} + {}_mV_x \tag{7}$$

Thus, the ageing reserve at time m is ${}_mV_x$:

$${}_mV_x = A_{x+m:n-m} - P_{x:n} \cdot \ddot{a}_{x+m:n-m} \tag{8}$$

I.2.1.2. Preparing the calculation

Data for the calculation are taken from *Eurostat, WHO,* the *KSH database* and the *National Cancer Registry.* Here are the most important data:

- *population per age group,* obtained *from* the *KSH databases*,
- the *number of* newly diagnosed *cases in* a given year obtained *from the National Cancer Registry* operated by the National Institute of Oncology for cancer and from the *KSH* and *WHO databases* for cardiovascular diseases,
- *the survival probabilities of* different diseases *(Tusnády et al., 2008),* based on the *National Cancer Registry* and the *KSH database*, and
- the *mortality table* based on the *KSH database*.

Population by year, age group and sex is available from the *Population projected* database (last updated on 24 October 2022). The long time series of data shows the population decline, which can also be identified from the interactive population pyramid of the KSH, which includes projections up to 2060. More complex models would fill an article on their own, but the aim is to present different health insurance products based on international trends, so a simpler model is sufficient. I will only deal with data from 2015 and beyond in the remainder of this chapter.

Due to the high number of errors in morbidity statistics, the data often require frequent corrections, which means that downloading the same dataset at different times can yield different results. KSH provides a detailed description of the data in the methodology of its tables. The planned revisions include routine (regular) revisions, and in the case of the *Morbidity data collection* linked to the KSH data collection (OSAP 1021), if a data error is detected, the data will be corrected at the next publication. For data revisions, if the data owner indicates a change from the previous year, the data are also corrected ex-post, or the monthly data are corrected at the end of the year based on the data owner's clarifications. The revised data are flagged. In the case of the OSAP 1549 data revision, in line with the practice of European cancer centres, the *National Cancer Registry* data are continuously revised, so fresh data should be treated as going back at least 3 years with the understanding that data are certain to change 1 year in the past, and highly likely to change 3-5 years in the past with the input of revisions. There have been no major revisions to these statistics in the last 10 years and none are expected. In the case of methodological changes, already published data are not revised retrospectively, but the limits of comparability over time are indicated to users (by separating the series with a line or a methodological note). Unannounced data revisions are among the unplanned revisions. They are only carried out in exceptional cases, namely when an unforeseen event (data error, technical problem, etc.) makes a data revision necessary. In the last 3 years, no unplanned revisions have taken place in this area of statistics. The data revision OSAP 2064 was due to organisational changes and legal reasons. The data content has been temporarily reduced and the comparability of the time series is limited. So if data are not corrected very often, errors can lead to bias. In databases of this nature and volume, administrative errors are quite common.

Although there are many databases on cancer, I chose the *National Cancer Registry* database. Some foreign databases, such as the *WHO database,* would also be suitable, but the article focuses specifically on the Hungarian situation, so it is advisable to rely on the more detailed *National Cancer Registry* for Hungarian data.

I used the statistics on various cancers from *the National Cancer* Registry to calculate the premiums for the insurance.

To obtain the probabilities of morbidity by disease, the number of cases per age group in

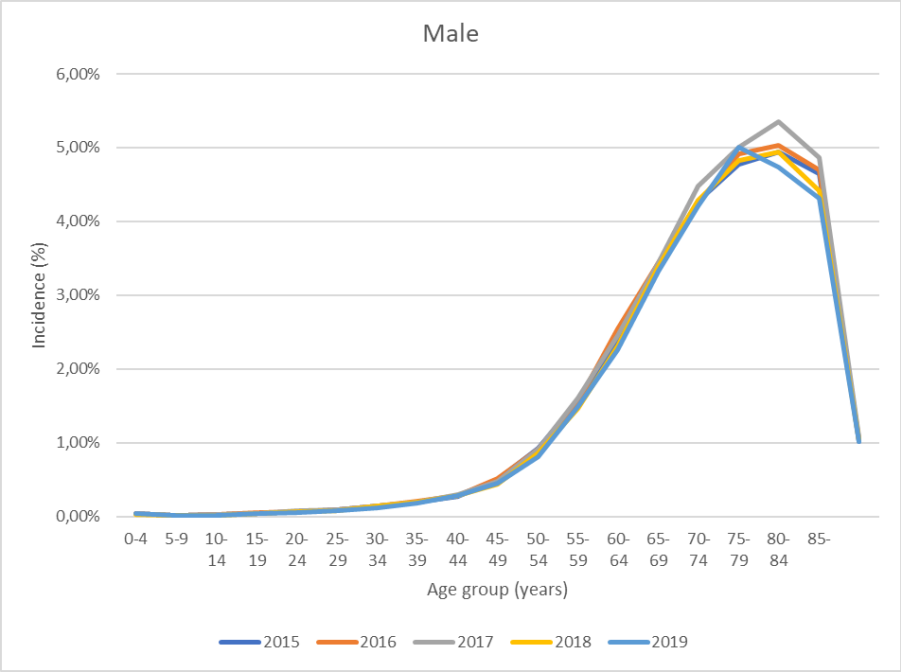
each year is needed, as well as the number of cases per age group by sex. For cancer, the data are from the period 2015-2019. Cancers and cardiovascular diseases are the most common causes of death, so I have included the main causes of death in the calculation of the product. It should also be noted here that the inclusion of any dreaded disease in the model does not materially change the calculation methodology, it only affects the premium level, only if the disease chosen is independent of cancer. In the model I take the disease into account regardless of its severity. For cancer, no significant trend can be observed when looking at incidence and mortality statistics over time. This is not surprising, as there have been no scientific advances, measures or events in this field in Hungary in the period under review that could justify a trend effect. If a trend effect were to be detected, it would be worthwhile to examine the probabilities of morbidity over time and build a predictive Lee-Carter model[2] *(Lee-Carter, 1992)* on the data, but instead I calculate the probabilities of morbidity for each disease group by taking the statistics for the 5 years mentioned above, taking into account the population for those 5 years (treating gender and age groups separately). An age group is always represented by the age in the middle, for example, the probability for 25- to 29-year-olds is associated with 27-year-olds. For sufficiently fine pricing, age-specific probabilities of morbidity are needed. This is done by fitting spline functions, i.e., piecewise polynomials, and the values of this spline function give the intermediate probabilities *(Móri, 2011).* To do this, an Excel extension called SRS1 Splines needs to be installed, and the data can be interpolated using the Cubic_Spline function. Finally, I generate unisex probabilities of morbidity according to the gender directive. The technical interest rate is 2% in all cases.

I.2.1.3. Results

I did the calculation as mentioned above. The resulting q_x-have been plotted as a function of age by disease type. For clarity, I have shown cancer separately. However, it should be noted that unisex probabilities are not always unisex, since breast malignancies (C50) and female genital malignancies (C51-C58) have female probabilities, while male malignancies (C60-C63) have male probabilities. The results are shown in Figures 5, 6 and 7.

[2] The Lee-Carter model is the most popular mortality prediction method, although since its creation it has been criticised and many improvements are known. These and their applications in Hungary are presented in *Vékás (2019a, 2019b, 2020), Gogola-Vékás (2020)* and *Szentkereszti-Vékás (2022).*

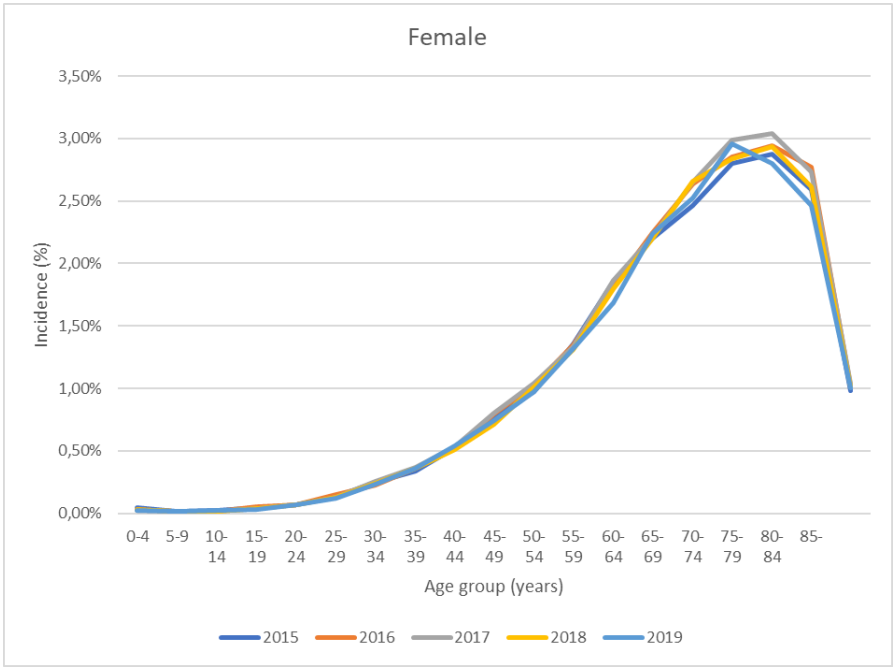
Figure 5 Male morbidity probabilities by age group



Source: own ed.

Notes: This figure illustrates the incidence rates by age group, with each group separated by 5-year intervals.

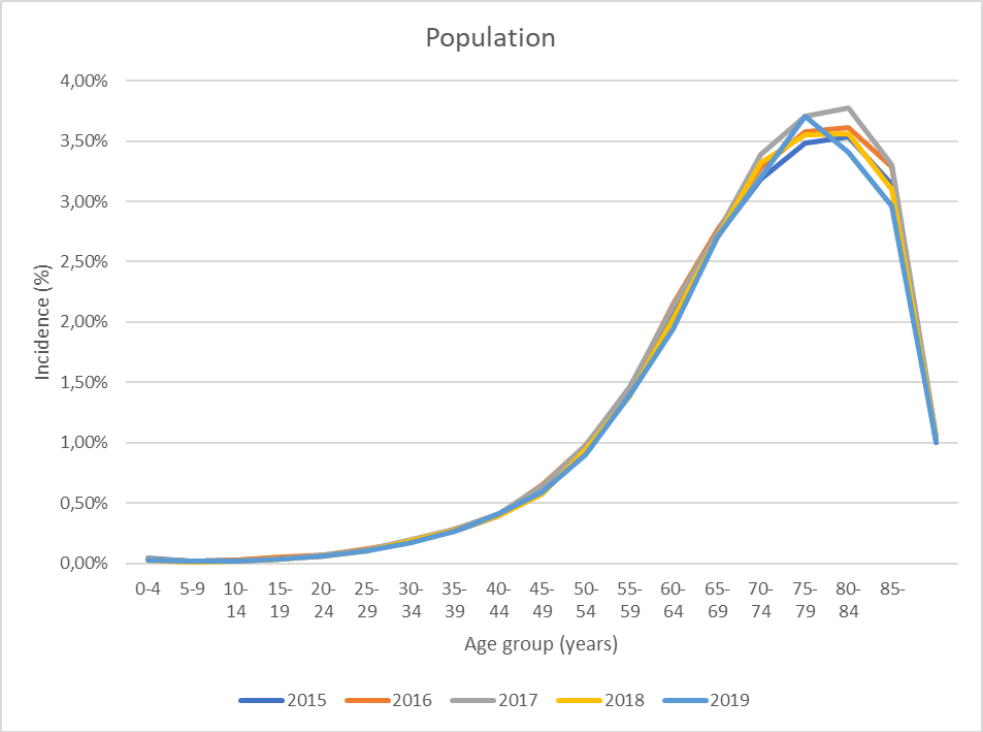
Figure 6 Female morbidity probabilities by age group



Source: own ed.

Notes: This figure illustrates the incidence rates by age group, with each group separated by 5-year intervals.

Figure 7 Population morbidity probabilities by age group



Source: own ed.

Notes: This figure illustrates the incidence rates by age group, with each group separated by 5-year intervals.

I used the calculator to determine the net single premium and the net annual premium for a *whole-life* insurance for a 40-year-old person. The product pays a lump sum of HUF 5 million in the event of cancer. In the case of cancer, the one-off net premium is **HUF 927 617,** while the annual net premium is **HUF 49 116.**

I.2.1.4. Further comments

It is worth starting with costs, such as initial and ongoing commission due to the agent, initial and ongoing cost, etc. I have not touched on these because there is no suitable general factual data available. Once these are included in the model, additional factors can be sought. It would also be worthwhile to include the results of the data collection by insurers in the model. For example, it would be good to introduce a regional multiplier. In the case of cancer, prevention and regular health screenings play a crucial role, so an index should be developed to reward customers who proactively maintain their health and can demonstrate, for example, that they have participated in such screenings. In the case

of cancer a death benefit can be introduced for death within one year. All we need are the survival probabilities. There have been many studies on this subject and on various procedures, but they all show that after 5 years from diagnosis, the number of deaths attributable to the disease falls, so we need survival probabilities of 1 to 5 years. Using data from the *National Cancer Registry* and *WHO* mortality data by disease, we can calculate average survival times (which is the average survival time until death) using the Gompertz survival model *(Tusnády et al., 2008) to* obtain the 1-5 year survival probabilities. There are several ways to extend the model, and the ones I have included so far are only an outline.

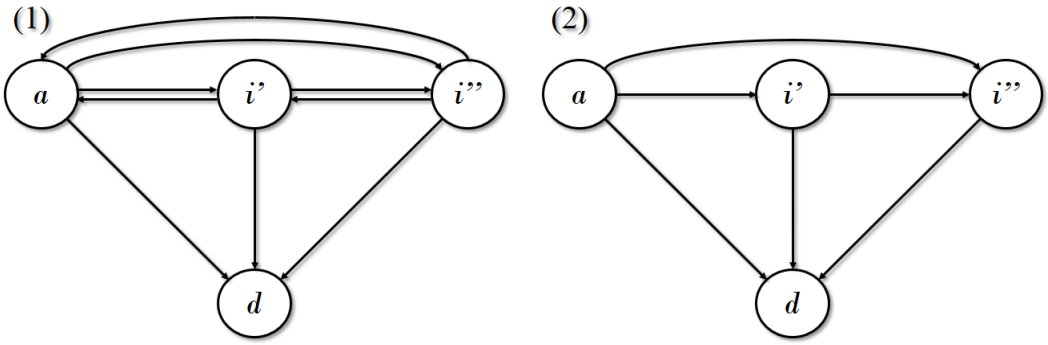
I.2.5. Long-term care insurance

Finally, the calculation of long-term care insurance is presented. As the previous chapter has shown, this is a complex model, which is accordingly very diverse on the market.

I.2.5.1. Theoretical introduction

We are focusing on LTC insurance policies that both define benefits according to the level of disability and are presented as a stand-alone product. Similar to disability insurance, we assume a multi-state system. However, we now consider several levels of disability (i', i''). This gives a four-state model, i.e. the insured can be active (a), mildly disabled (i'), severely disabled (i'') or dead (d). Figure 8 gives an indication of the possible outcomes. Compared to the disability model, we simplify as follows: we exclude the possibility of recovery, i.e. we assume the second part of Figure 8. This is done because we assume that disability occurs as a result of old age, which is an irreversible process. Mild disability can be achieved as follows: $a \to i'$; and severe disability can be achieved in two ways: $a \to i''$, and $a \to i'$, then $i' \to i''$. In the latter case, two years are required, since only one change of condition can occur in a year *(Pitacco, 2014).*

Figure 8 Four-state models for LTC



Source: *Pitacco (2014).*

Notes: This table illustrates the transitions in the four-state Markov model, where "a" represents active, "i'" represents mild disability, "i"" represents severe disability, and "d" represents dead.

The one-year transition matrix is shown in Table 5.

Table 5 Conditional probabilities related to LTCI products

State at age y	State at age y + 1			
	a	*i'*	*i''*	*d*
a	p_y^{aa}	$p_y^{ai'}$	$p_y^{ai''}$	q_y^{a}
i'	0	$p_y^{i'i'}$	$p_y^{i'i''}$	$q_y^{i'}$
i''	0	0	$p_y^{i''i''}$	$q_y^{i''}$
d	0	0	0	1

Source: *Pitacco (2014).*

Notes: This table illustrates the conditional probabilities in the four-state Markov model, where "a" represents active, "i'" represents mild disability, "i"" represents severe disability, and "d" represents dead. "y" denotes age, "p" represents the probability of transitioning from state y to state y+1, and "q" represents the probability of dying in that state.

From these, and using formulae (13) and (14), we obtain the following for the multi-year transition probabilities:

$$ {}_hp_y^{aa} =_{h-1} p_y^{aa} \cdot p_{y+h-1}^{aa} \tag{27}$$

$$ {}_hp_y^{ai'} = {}_{h-1}p_y^{ai'} \cdot p_{y+h-1}^{i'i'} + {}_{h-1}p_y^{aa} \cdot p_{y+h-1}^{ai'} \quad (28) $$

$$ {}_hp_y^{ai''} = {}_{h-1}p_y^{ai''} \cdot p_{y+h-1}^{i''i''} + {}_{h-1}p_y^{ai'} \cdot p_{y+h-1}^{i'ii''} + {}_{h-1}p_y^{aa} \cdot p_{y+h-1}^{ai''} \quad (29) $$

They show the probabilities of a y-year-old individual being in various possible states at the age of $y + h$. The three equations are expressed in terms of the probabilities of one year earlier $({}_{h-1}p_{y}^{\cdot\cdot})$. We transform them further to see the probabilities of all the years that have passed. We can then also express when the individual becomes disabled *(Pitacco, 2014).*

$$ {}_hp_y^{ai'} = \sum_{r=1}^{h} \left[{}_{h-r}p_y^{aa} \cdot p_{y+h-r}^{ai'} \cdot \prod_{g=1}^{r-1} p_{y+h-r+g}^{i'i'} \right] \quad (30) $$

$$ {}_hp_y^{ai''} = \sum_{r=1}^{h} \left[{}_{h-r}p_y^{aa} \cdot p_{y+h-r}^{ai''} \cdot \prod_{g=1}^{r-1} p_{y+h-r+g}^{i''i''} + {}_{h-r}p_y^{ai'} \cdot p_{y+h-r}^{i'i''} \cdot \prod_{g=1}^{r-1} p_{y+h-r+g}^{i''i''} \right] \quad (31) $$

The probability that a y-year-old active individual is mildly disabled throughout the next h years is the probability that he becomes mildly disabled in one year and remains as such for the remainder of the period. Similarly, the probability that a y -year-old individual will become severely disabled throughout the next h years is the probability that he becomes severely disabled in the next period and remains as such for the remainder of the period. This can happen starting from both the active and the mildly disabled states *(Pitacco, 2014).*

Let's assume that the insurer makes the following annuity payments: $b' = 1$, if the insured's disability status is i', or $b'' = 1 + \beta$, if the insured's disability status is i'' $(\beta > 0)$. Let $a_y^{ai'i''}(\beta)$ indicate the expected present value of future payments, which, because of the principle of equivalence, is the single premium. Based on formula (11):

$$ \begin{aligned} a_y^{ai'i''}(\beta) &= \sum_{h=1}^{+\infty} \left({}_hp_y^{ai'} + (1+\beta) {}_hp_y^{ai''} \right) v^{h=} \\ &= \sum_{h=1}^{+\infty} \Big[\sum_{r=1}^{h} {}_{h-r}p_y^{aa} \cdot p_{y+h-r}^{ai'} \cdot \prod_{g=1}^{r-1} p_{y+h-r+g}^{i'i''} + \\ &+ (1+\beta) \sum_{r=1}^{h} {}_{h-r}p_y^{aa} \cdot p_{y+h-r}^{ai''} \cdot \prod_{g=1}^{r-1} p_{y+h-r+g}^{i''i''} + \\ &+ (1+\beta) \sum_{r=1}^{h} {}_{h-r}p_y^{ai'} \cdot p_{y+h-r}^{i'i''} \cdot \prod_{g=1}^{r-1} p_{y+h-r+g}^{i''i''} \Big] v^h \end{aligned} \quad (32) $$

Using the derivation analogy of equation (21), we obtain:

$$\begin{aligned} a_y^{ai'i''}(\beta) \quad &= \sum_{j=1}^{+\infty} \left({}_{j-1}p_y^{aa} \cdot p_{y+j-1}^{ai'} \cdot v^j \cdot \ddot{a}_{y+j}^{i'} \right) + \\ &+(1+\beta) \sum_{j=1}^{+\infty} \left({}_{j-1}p_y^{aa} \cdot p_{y+j-1}^{ai''} \cdot v^j \ddot{a}_{y+j}^{i''} \right) + \\ &+(1+\beta) \sum_{j=2}^{+\infty} \left({}_{j-1}p_y^{ai'} \cdot p_{y+j-1}^{i'i''} \cdot v^j \ddot{a}_{y+j}^{i''} \right) \end{aligned} \tag{33}$$

Assume that the regular premium is paid for at most m' years, as long as the insured is healthy (in condition a). Then the annual premium is (P), where benefit $\ddot{a}_{y:m'}^{aa}$ is defined by equation (23) (Shao, 2015), is:

$$P = \frac{a_y^{ai'i''}(\beta)}{\ddot{a}_{y:m'}^{aa}} \tag{34}$$

I.2.5.2. Results

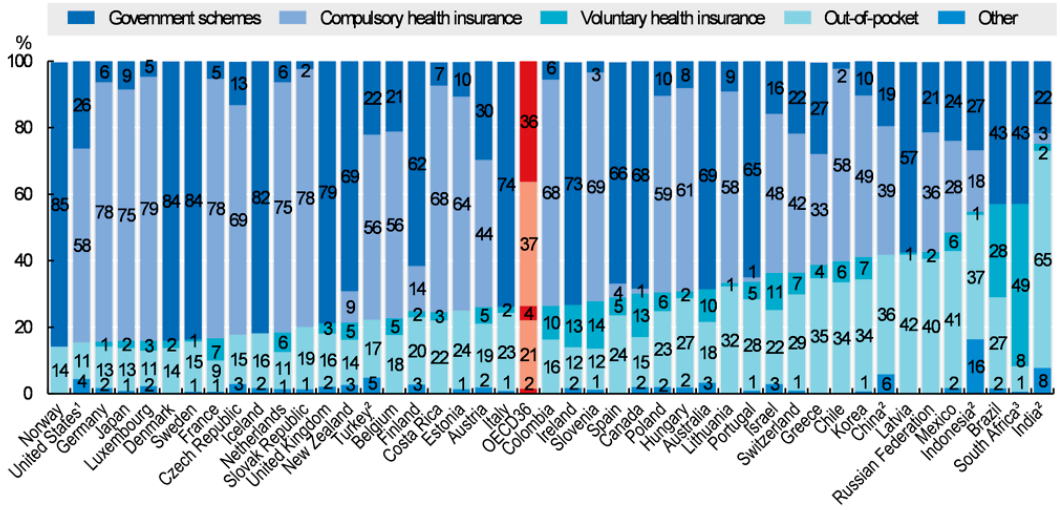
The long-term care insurance discussed here distinguishes between two conditions among people in need of care. The first condition is the milder one, where the patient is unable to take care of his/her daily activities and receives a monthly benefit of HUF 50 000, or HUF 150 000 if the patient is unable to take care of himself/herself. In general, LTC insurance is considered to be very expensive *(Faragó, 2012)*. The net annual premium calculated in this way is estimated at **HUF 31 000.** To my knowledge, the Hungarian market currently does not offer long-term care insurance. In an effort to find relevant data for estimating the premium, I searched for academic papers but found only a single source—a master's thesis from 2016 (Fónagy, 2016). This study calculates the probabilities of transitions between different states. Based on these probabilities, my expert judgment—formed through my experience and professional impressions—aligns closely with the findings. Therefore, I believe the provided example price offers a reasonable estimate of how much this coverage could cost annually.

I.3. The potential of complex health insurance products in a social security system

In the previous chapter, I presented the calculations and issues related to the different health insurance products. The calculations were based on a 40-year-old person. If we look at the financing schemes of Western countries such as Germany, France, the

Netherlands or the UK, we can see that the share of Government schemes and Compulsory health insurance is higher than in our country, and that we have a significant amount of out-of-pocket expenditure, as shown in Figure 9.

Figure 9 Health expenditure by type of financing, 2017 (or nearest year)



Source: *OECD Health Statistics (2019a).*

Notes: 1. All spending by private health insurance companies in the United States is reported under compulsory health insurance. 2. Health payment schemes unable to be disaggregated into voluntary health insurance, NPISH and enterprise financing are reported under other. 3. Voluntary payment schemes unable to be disaggregated are reported under voluntary health insurance.

Looking at the KSH data, there is an upward trend in per capita health expenditure in Hungary, which has been increasing year on year.[3] To remedy this situation, I propose that public intervention is needed. The premiums I have calculated are in line with the premiums charged by insurers, but demand has not increased in recent years. The analysis of the regulation of private health insurance is made rather difficult by the fact that individual health insurance policies can vary widely. The function of each insurance, its role in the health care system and its penetration vary widely. A private health insurance

[3] It is important to note that, in ageing societies, the sustainability of pension systems is a key issue alongside health care systems. For a more detailed overview of the sustainability challenges of the Hungarian public pension system and their modelling options, see for example *Németh-Németh-Vékás (2020a, 2020b)* and *Kovács-Rétallér-Vékás (2015).*

policy, as I have shown in my model, can be either a payment or indemnity insurance policy.

At the 6th International Insurance Conference of MABISZ, Csaba Lantos pointed out that, in his opinion, doctors actually control the health care market. Almost half of public sector spending, 422 billion forints, is made up of personnel costs, health workers, ambulance services and GPs. The net salary of public doctors in 2012 amounted to 86 billion forints, plus a part of the amount paid to private providers, estimated at 40%, or around 63 billion forints. Roughly 70% of this is estimated to be the net income of private practices, or around HUF 44 billion. If we add to this the parasolvency estimated by the KSH at around HUF 70 billion and the HUF 144 billion of income from "home prescriptions", the total income for doctors is HUF 344 billion. According to the KSH, the number of doctors is estimated at 26,000, giving an average net monthly income of HUF 1 083 519 per doctor. The amount spent on financing services paid by insurers is estimated to be no more than HUF 1 billion, with voluntary health funds spending over HUF 10 billion, so there is still room for improvement in the health insurance market, which would require as a first step changing the chaotic situation of the domestic health care system, which is no small task *(Lantos, 2015). The* most significant change since the establishment of the report is the creation of legislation to separate public and private practice. Private providers are becoming more numerous. The fastest growth has been in and around Budapest. Of course, the growth in the premium income of health and self-help funds is proportional to this. According to KSH Table *4.1.1.2 Health expenditure by subsystem, % of GDP, health investment expenditure,* government subsystems spent a total of HUF 2 952.8 billion in 2021, while expenditure by voluntary health financing subsystems was HUF 118.5 billion in the same year.

Most studies on health care also address the issue of so called ’hálapénz’ (Baji, 2012a and 2012b). Hálapénz (literally "gratuity money") refers to informal payments made by patients to medical professionals in Hungary, often as a gesture of gratitude but sometimes with the expectation of better treatment or preferential access to healthcare services. The state is trying to reduce ’hálapénz’ through legislation, but the problem has not been fully addressed, as ’hálapénz’ has been replaced by private health care and the private insurance market is still very limited, with few providers. Private insurance could also be used to combat ’hálapénz’, which would help to increase transparency in the

health care system and create a more level playing field for citizens (Baji, 2010). According to the KSH database, the **average net** monthly **income** in Hungary is 483,900 HUF, which is **5,806,800 HUF per year** as of December 2024. **The calculated health insurance premium would be 3.53% of** this amount.

The living wage issue is not the subject of my study, but it is clear why there has been no growth in the health insurance market in recent years. A living wage is the minimum income necessary for a worker to meet their basic needs, including food, housing, healthcare, education, transportation, and other essentials, while also allowing for some discretionary spending and savings. Unlike the minimum wage, which is often set by law and may not account for actual living costs, a living wage is based on the real cost of living in a specific location. It ensures that workers and their families can maintain a decent standard of living without relying on government aid or excessive working hours. Research shows that the cheapest health care is the one that invests in prevention. Some examples of these studies are Holdroyd et al. (2024) and Bandeira et al. (2025). That is why my proposal would be for a form of subsidy that would support the younger generations to take care of their health and take out private health insurance. My assumption is that the high level of subsidy would not increase the moral hazard very much, but awareness-raising campaigns would also be needed. In my opinion, a regular premium scheme would be the most appropriate, as the continuous payment of premiums would remind the insured of the importance of looking after their health.

Part II. Validation of the PAM-13 instrument in the Hungarian general population

II.1. Background

In our days more and more people live with chronic conditions, which will account for nearly three quarters of global mortality in 2020 *(WHO, 2019a)*. The quest for sustainable financing of health care systems needs to embrace efficient approaches in the management of chronic conditions *(OECD, 2015)*. Over the last decades, patient-centredness has gained considerable attention in medicine, which, by putting patients' values and preferences in the forefront of medical decision-making, aims for their more efficient involvement as partners in the health production process *(Institute of Medicine, 2005)* and *(WHO, 2015)*. The active involvement of patients is particularly important in the reduction of modifiable lifestyle-related risks, which contribute to a considerable share of excess mortality from chronic conditions *(WHO, 2016)*. Since large-scale policy measures such as changes of income or education, that can influence important individual determinants of health and healthy behaviours potentially span in time over generations, personally acquired potentials, which can develop over later courses of life, such as knowledge, skills, positive emotions and engagement are of particular importance from both public health and health economic perspective *(WHO, 2019b)* and *(Bircher et al., 2014)*. A number of theories, such as internal locus of control *(Wallston et al.,1978)*, self-efficacy *(Bandura, 1977)* or self-management *(Lorig, 1993)* or the transtheoretical model of change *(Prochaska et al., 2008)* have addressed the drivers of change in health behaviours, and a number care delivery models, such as the Chronic Care Model promoted systematic improvements involving patient-centredness, support for self-management, evidence-based proactive interventions, integrated team-care and supportive information technology solutions *(Coleman et al., 2009)*. Digital health interventions have been shown to be effective in promoting healthy behaviours through patient education or supporting behaviour change, and upon the demonstration of adequate supportive evidence, authorities are now considering their adoption among publicly financed health technologies *(Morton et al.,2017)* and *(NICE, 2019)*.

The Patient-Activation Measure (PAM) has been developed to serve as a reliable tool that can measure the skills, knowledge and motivation of patients that are necessary for their

effective contribution to their own care, and eventually predict better outcomes *(Hibbard, 2004)*. Since its development, PAM has become an officially adopted patient reported outcome measure (PROM) by the National Institute of Health of the US and the National Health Service of the UK. Its validated versions have been available in over 20 countries and it has been applied in over 500 studies worldwide *(Insignia Health, 2019)*. It has been demonstrated that higher PAM values are associated with better health outcomes *(Mosen et al., 2007)*, fewer lifestyle-related risks *(Harvey et al., 2012)*, better adherence to therapy *(Graffigna et al., 2017)*, and lower use of health care resources *(Harvey et al., 2012)* and *(Remmers et al., 2009)*. Furthermore, it has been shown that patient activation can be improved via digital health interventions *(Aziz et al., 2019)* and *(Schnock et al., 2019)* as well as offline patient-support programs *(Carroll et al., 2019)*.

In accordance with national policies aiming to reduce lifestyle related excess mortality as well as the advancement of digital health *(Korm. határozat, 2017)*, our aim was to adapt and validate the 13-item PAM-13 tool in the Hungarian general population to serve as a widely tested and internationally recognised instrument in the development or monitoring of evidence-based health-promotion interventions.

II.2 Methods

II.2.1. Study design and participants

In April 2020 we conducted an online survey recruiting 900 respondents from a large online panel via quota-based sampling with strata set according to the 2011 population census *(KSH, 2011)*. Our sample was representative of the 40+ years old Hungarian population in terms of age groups (40-49, 50-59, 60-69 and 70+ years old), gender, education (primary, secondary and tertiary), geographic region (7 NUTS-2 regions) and type of residence (village, town or capital city). After 10 days, 100 respondents were randomly selected for repeated administration of the entire survey. We considered 10 days lag sufficient to prevent recall while capturing a stable PAM-13 status *(Skolasky et al., 2009), (Kosar et al., 2019)* and *(Terwee et al., 2007)*. Ethical approval was granted by the National Medical Research Council (TUKEB, ID: 49702-3/2019/EKU) and a research license were obtained for PAM13. Recruitment and data collection were performed by an online research company.

II.2.2. Measurement tools and survey items

PAM-13 and development of the Hungarian language version

PAM assesses one's knowledge, skill and confidence for self-management. The original version consisted of 22 items. The short version PAM-13 instrument consists of 13 items, scored on a 4-point Likert scale and a 5th 'not applicable' option. For a valid PAM-13 score, up to 3 'not applicable' responses were allowed. Using a proprietary scoring algorithm based on Rasch analysis, PAM-13 is scored on a scale of 0-100, where lower values suggest less likelihood that patients engage in effective self-management. As the PAM-13 scoring algorithm accounts for item difficulty and response patterns through Rasch analysis, no additional standardization is required, and even when different Likert scales are used in later adaptations, they similarly do not necessitate further standardization. Based on their PAM-13 scores, patients can be grouped into four PAM-13 levels. At level 1, patients may not understand the need to take active role in their health. At level 2, their confidence or skill is probably too low to take action. At level 3 patients are beginning to take action and at level 4, they may endure in self-management even in difficult times *(Hibbard et al., 2014), (Hibbard et al., 2005)* and *(Lindsay et al., 2018)*. In this analysis we use both the 0-100 score and the 4 levels of activation.

The Hungarian language version of the PAM-13 questionnaire was produced in accordance with the WHO guidelines for the translation and adaptation of instruments *(WHO, 2019c)*. Forward- translation was performed by two independent experts, the back-translation was carried out by two bilingual translators and elaborated by two researchers (*Ágota Dobos* and *Zsombor Zrubka*) against the original version of the instrument. The draft instrument was piloted along with cognitive debriefing on 10 respondents including both males and females from different age groups. The literal translations were overridden at several questions with natural phrases that were considered to be acceptable for the broadest audience, yet conceptually equivalent with the original questionnaire. The pre-final version was consulted with the developers of the PAM instrument. The 4-level Likert scale response options of the original instrument ("Disagree strongly", "Disagree", "Agree", "Agree strongly") were slightly modified to mark more precisely the scale degrees in a Hungarian context ("Completely disagree", "Rather disagree", "Rather agree", "Completely agree"). The Hungarian PAM-13 is attached in Appendix II.1.

II.2.2.1. eHEALS

The eHealth Literacy Scale (eHEALS) measures self-reported eHealth literacy using eight 5-point Likert scale items. The eHEALS score (range 8-40) is calculated by summing individual item scores, with higher values indicating greater skill levels. The Hungarian eHEALS instrument has been validated in the general population via an online survey *(Zrubka et al., 2019)* and *(Norman et al., 2006)*. Since eHEALS showed weak correlation with objective performance tests *(van der Vaart et al., 2011)* and *(Neter et al., 2017)*, we also measured performed health literacy in our survey.

II.2.2.2. Newest Vital Sign

The Newest Vital Sign (NVS) is a frequently used screening instrument for performed health literacy. Respondents need to answer six questions by interpreting the information from an ice-cream nutritional label and performing simple arithmetic tasks. For some questions correct answers can be formulated in several ways. The number of correctly answered items are counted. A score of 0-1 indicates limited- , 2-3 indicated probably limited- , and 4-6 indicates adequate health literacy *(Weiss et al., 2005)*. We adapted the Hungarian NVS for online administration *(Koltai et al., 2016)*. Instead of offering multiple-choice options in the online adaptation *(Mansfield et al.,2018)*, we specified the measurement unit for answers in the questions and evaluated the accuracy of free-text answers.

II.2.2.3. Minimum European Health Module

We also inquired respondents' health by the Minimum European Health Module (MEHM). The MEHM consists of three questions. Self-perceived health evaluates current health on a 5-point Likert scale (Very good; Good; Fair; Bad; Very Bad). The Global Activity Limitation Indicator (GALI) asks limitations in activities over the past 6 months due to a health problem (Not limited, Limited but not severely, Severely limited). A final item (Chronic Morbidity) inquires the presence of long-standing health problems *(Cox et al., 2009)*.

II.2.2.4. Health-related information seeking and online behaviours

We constructed seven items to assess the frequency of various health-related information seeking and online behaviours over the past 12 months. Item 1 inquired about health information seeking (HIS) in general and item 2 about participation at patient-education programs (PPE). Items 3-7 inquired about health-related use of the internet or mobile devices in different functional domains, motivated by the classification of the evidence standards framework of digital health interventions proposed by The National Institute for Health and Care Excellence (NICE) *(Nice, 2019)*. Item 3 inquired general online health-related administration (OHA), item 4 about online health-related information seeking (OHIS), item 5 about online health-related communication with HCPs, helpers or peer patients (OHC), item 6 about online health prevention activities (OHP) and item 7 about online disease management activities (ODA). All items were scored on a 6-point Likert scale (never, few times past year, bimonthly, monthly, several times per month, at least once a week).

II.2.2.5. Demographic variables

We recorded basic demographic variables and defined the following subgroups: age (40-49,50-59, 60-69, 70+ year olds), gender (male, female), family status (single, married, domestic partnership, divorced, widowed and other), education (primary, secondary, tertiary), health professional qualification (yes, no), type of settlement (capital, town, village), NUTS2 region and the place of residence based on postcode. Net monthly household income was queried in 11 range categories, and per-capita household income was calculated by dividing the category mid-range values by the number of household members, without adjustment for the number of children. The mid-range value of the upper open category was calculated by fitting the Pareto curve as proposed by Parker and Fenwick *(Parker et al., 1983)*. Local currency values were transformed to Euros using the average exchange rate for the period of Apr 1, 2019 Apr 1,-2020 (330.7 HUF/EUR) *(ECB, 2020)*.

II.2.2.6. Lifestyle risk factors

We recorded the most common modifiable risk factors for all-cause mortality and chronic conditions, such as BMI, smoking, alcohol intake, dietary habits, physical activity and

sedentary behaviour [42-46]*(Stein et al., 2004), (Lim et al., 2015a), (Ng et al., 2020), (WHO, 2014)* and *(McGorrian et al., 2011)*. Lifestyle-risks were inquired via single-question self-reported items. In order to represent similar "severity levels", the following cut-off values were chosen that represent approximately 1.4-fold or greater relative risk increase for overall mortality: BMI <18.5 or BMI ≥30 (BMI) *(Di Angelantonio et al., 2016)*, current smoking (Smoking) *(Taghizadeh et al., 2016)* and *(Pirie et al., 2013)*, sedentary behaviour ≥8 hours per day with <150 min exercise per week or no exercise at all (Physical activity) *(Stamatakis et al., 2019)*, no fruit and / or vegetable intake (Diet) *(Wang et al., 2014)* and binge drinking ≥1 day per week (Alcohol) *(Xi et al., 2017)*. Binge drinking was defined as >5 drinks / occasion for men and >4 drinks / occasion for women *(Xi et al., 2017)* and *(NIAAA, 2004)*. We also generated a lifestyle risk index (LRI) by adding the number of lifestyle risk factors for each patient. Based on their lifestyle risk index, respondents were assigned to risk groups using stringent- (no lifestyle risk vs any lifestyle risks) and relaxed (0-1 vs 2-4 lifestyle risk factors) criteria.

II.2.2.7. Preventive behaviours

We considered the participation at screenings and vaccination programs as preventive behaviours, which are prescribed by Hungarian law *(NM, 1997)* and *(Nemzeti Népegészségügyi Központ, 2020)*. According to this, for females we counted the participation at cervical cancer screening between 25-65 years of age, breast cancer screening between 45-65 years of age, colorectal screening between 50-70 years of age, blood pressure, blood glucose and cholesterol levels measured within a year, flu vaccination at 60+ years of age and bacterial pneumonia vaccination at 50+ years of age. For males, we counted the participation at prostate- and colorectal cancer screening between 50-70 years of age, blood pressure, blood glucose and cholesterol levels measured within a year, flu vaccination at 60+ years of age and bacterial pneumonia vaccination at 50+ years of age. To make respondents with different gender and age comparable, we calculated the preventive behaviour score (PBS) as the proportion of performed preventive behaviours compared to the maximum of preventive behaviours prescribed for a given age and gender. For example, having only blood pressure measured within a year would represent a preventive behaviour score of 0.33 for a 40-year-old man (with only blood pressure, glucose and cholesterol check recommended), while it would be a score of 0.125 for a 60-year-old woman, (for whom cervix-, breast- and colorectal

cancer screenings, blood pressure, glucose and cholesterol tests, as well as flu and bacterial pneumonia vaccinations are recommended). We also grouped respondents based on their preventive behaviour score (<50%, ≥ 50%).

II.2.3. Excluded respondents

Before data analysis, we checked the dataset for outliers and based on group consensus, we set implausible values to missing, or deleted entire records in case of potentially unreliable answer patterns. We deleted the data point if sedentary time was reported >18 hours / day. We deleted cases if the frequency of online health information seeking was reported over two categories greater than general health information seeking, response times for shorter than 4 second per item for the survey instruments (PAM-13, eHEALS), or shorter than 1 minute for the NVS instrument. One respondent was excluded due to unlikely body parameters (height 111 cm, weight 200 kg, BMI=162), and based on the PAM license owner's recommendation, we dropped individuals with a PAM-13 score of 0 and 100 (respondents who answer all strongly agree or all strongly disagree, are likely not paying attention and responding in a valid way). as well as ones with not applicable answers in more than 3 PAM-13 items *(Insignia Health, 2020)*.

II.2.4. Statistical methods

We followed the applicable COSMIN guidelines for patient-reported outcome measurement instruments when planning the methods of our study *(Mokkink et al., 2010), (Prinsen et al., 2018)* and *(Terwee et al., 2012)*. Missing data, descriptive statistics and distributional properties were assessed for all variables. The distribution of PAM-13 scores was assessed via inspection of the histogram and quantile-plot, and normality was tested via the Shapiro–Wilk test. I would also like to note that the original paper assessing normality used the Shapiro-Wilk test, as it was the standard approach in other PAM validation studies. However, while the Shapiro-Wilk test is appropriate for samples under 2,000, it is not ideal for larger samples above 50. Therefore, I also conducted the Kolmogorov-Smirnov test, although it can be affected by tied values. To ensure robustness, I further assessed normality using the Anderson-Darling test, which is considered a powerful alternative. Table 6 presents the results of these tests.

Table 6 Normality tests

Method	Target	*p* value	Conclusion
Shapiro–Wilk test for normal distribution	$p \geq 0.05$	<0.001	Deviation from normality
Shapiro–Wilk test for log-normal distribution	$p \geq 0.05$	0.811	Log-normal distribution
Kolmogorov-Smirnov test for normal distribution	$p \geq 0.05$	0.001346	Deviation from normality
Kolmogorov-Smirnov test for log-normal distribution	$p \geq 0.05$	0.0008998	Deviation from log-normal distribution
Anderson-Darling test for normal distribution	$p \geq 0.05$	<0.001	Deviation from normality
Anderson–Darling test for log-normal distribution	$p \geq 0.05$	<0.001	Deviation from log-normal distribution

Notes: p<0.05, significant; p≥0.05, not significant

Floor- and ceiling effects (frequency of the highest and lowest scores in the sample) were assessed against the threshold of 15% *(Lim et al., 2015b)*. We tabulated respondents based on their PAM-13 levels.

II.2.4.1. Reliability

We evaluated internal consistency via computing Cronbach's alpha. Test-retest reliability for PAM-13 scores was assessed by intra-class correlation coefficient of agreement using a two-way random effects model ($ICC_{agreement}$ or ICC(A,1)) *(McGraw et al., 1996)*. For categorical PAM-13 levels, we calculated weighted kappa using quadratic weights to progressively penalise greater differences between categories *(McGraw et al., 1996)*. Measurement error (standard error of measurement, SEM) was calculated using the formula $SEM = \sigma * \sqrt{1 - ICC_{agreement}}$, where σ is the pooled standard deviation of the sample from first- and repeat administrations. The smallest detectable change (SDC, the smallest change that can be detected within a single person with p<0.05 taking measurement error into consideration) was calculated via the following formula $SDC = 1.96 * \sqrt{2 * SEM}$ *(Terwee et al., 2007)*. We considered the following thresholds for good

measurement properties: ≥0.7 for $ICC_{agreement}$ and weighted kappa, and the range of 0.7-0.95 for Cronbach's alpha *(Terwee et al., 2007)*.

II.2.4.2. Validity

Content validity was assessed during the translation process, no further quantitative measurements were performed. Construct validity was assessed via confirmatory factor analysis using robust structural equitation modelling via the R package lavaan *(Rosseel, 2012)*, assuming a single underlying factor. We checked the Kaiser-Meyer-Olkin (KMO) statistic for the adequacy of sampling *(Kaiser et al., 1974)* and Bartlett-test for sphericity to check the adequacy of our data for factor analysis. Model fit was assessed by the RMSEA, the Tucker-Lewis index (TLI) and the comparative fit index (CFI), using cut-off values of ≤0.05, 0.9 and 0.9 for good fit, respectively. Convergent validity was assessed by the correlation between PAM-13 scores as well as PAM-13 levels and eHEALS scores, based on the assumption that both instruments measure advanced knowledge and are conceptually related to self-efficacy *(Hibbard et al., 2004)* and *(Norman et al., 2006)*. We expected significant positive relationship between the two measures. Discriminant validity was tested by the expectation of weak or non-significant correlation between PAM-13 scores as well as PAM-13 levels and age, education and income, based on the assumption that PAM measures qualities that cannot be explained by socio-economic status. We applied Pearson correlation between continuous measures, and polyserial correlation when categorical measures were involved.

When testing known-groups validity, our hypothesis was the following: patients practicing more preventive behaviours (preventive behaviour score ≥50%), having fewer lifestyle risk factors according to the stringent (lifestyle risk index=0) or relaxed (lifestyle risk index ≤1) criteria, those, who were more active in health information seeking, patient education, online/mobile health-information seeking- , health-related communication, disease-prevention- or disease management activities had higher mean PAM-13 scores. We defined higher activity as having at least median scores on each item, or any activity over the past year, if majority of respondents did not engage in the respective online activity. The hypotheses were tested using one-sided Welch's t-test. We also explored the same hypotheses in subgroups of patients with or without chronic disease, male or female respondents, and respondents ≥65 years of age or younger, respondents in the lowest

income group or higher and respondents with adequate (NVS≥4) or lower health literacy scores. We deleted missing values pairwise for all statistical analyses. No missing values were imputed.

II.2.4.3. Rasch analysis

Insignia Health (copyright owner of PAM-13) provided the authors with PAM-13 scores based on the Rasch model *(Hibbard et al., 2005)* and *(Rasch, 1960)* as well as values of infit and outfit indices for each participant in both administrations of the survey. These indices measure the goodness-of-fit of the model to the observed response patterns. Infit is more sensitive to deviations on items aimed at the true activation levels of the respondents, whereas outfit is more sensitive to deviations on items aimed far from their true activation levels. Several thresholds for person infit and outfit indices exist in the literature, and according to one popular system *(van Buuren et al., 2020)*, values below 0.3 imply a lack of expected variability in answers, values between 0.3 and 3.0 indicate good to high quality data, and values greater than 3 indicate unusual response patterns and thus poor-quality data.

II.2.4.4. Regression analysis

We tested the association of PAM-13 score with the preventive behaviour score (PBS), lifestyle risk index (LRI), individual lifestyle-related risk factors (BMI, smoking, alcohol, physical activity and diet), health information seeking, participation at patient education and online health-related behaviours, when controlled for eHealth-literacy (eHEALS), health literacy (NVS), age, gender, education, income, type of settlement and the health status of the respondent (MEHM items). We also replaced PAM-13 scores with PAM-13 levels. For continuous variables we applied OLS regression. Heteroskedasticity and model specification was tested via the Breusch-Pagan *(Breusch-Pagan, 1979)* and Ramsey RESET tests *(Ramsey, 1969)*, respectively. In case of heteroskedasticity, we applied robust regression. Binary risk factors were analysed via logistic regression and the ordinal health information seeking and online behaviours via ordered logit models. For logistic and ordered logit models goodness of fit was assessed via the binary and multinomial versions of the Hosmer-Lemeshow test *(Hosmer et al., 2013)* and *(Fagerland et al., 2013)*.

II.3. Results

II.3.1. Descriptive statistics

From the 900 survey respondents due to PAM-13-related quality issues we excluded 99 respondents (11.0%), and 22 respondents (2.4%) for other reasons detailed above. Altogether 779 (86.6%) individuals were included in our sample. In the repeat survey (n=100), 11 respondents (11.0%) had PAM-13-related quality issues, 4 (4.0%) other reasons for exclusion and 10 respondents were excluded in the first administration. The retest sample comprised of 75 respondents with had matching test-retest scores.

In the sample, mean age was 60.4 (SD=10.6) years, 54.0% were female, 66.5% reported to have chronic disease. Highly educated, affluent urban respondents were over-represented compared to the 40+ year-old general population. Two thirds of respondents reported to have a chronic disease (Table 7). The demographic characteristics of survey respondents and reasons for exclusion are summarized in Table 7.

Table 7 Sociodemographic characteristics and health status

		Sample		Retest sample		General population 2011[a]
		n	*%*	%	*%*	*%*
Total		779	-	75	-	-
Age group	40-49	143	*18*	14	*15*	*26*
	50-59	177	*23*	14	*21*	*28*
	60-69	306	*39*	11	*37*	*23*
	70+	153	*20*	9	*27*	*23*
Gender	Male	358	*46*	44	*49*	*44*
	Female	421	*54*	56	*51*	*56*
Education	Primary	203	*26*	62	*45*	*35*
	Secondary	288	*37*	28	*33*	*49*
	Tertiary	288	*37*	10	*21*	*16*
Region	Central Hungary	276	*35*	29	*29*	*29*
	Transdanubia	262	*34*	31	*36*	*31*
	Great Plain and North	241	*31*	40	*35*	*40*
Type of	Capital	181	*23*	17	*24*	*17*
residence	Town	447	*57*	52	*48*	*52*
	Village	151	*19*	31	*28*	*31*
Income	1st quintile	75	*11*	20	*12*	
	2nd quintile	106	*16*	20	*20*	
	3rd quintile	74	*11*	20	*13*	
	4th quintile	122	*18*	20	*17*	
	5th quintile	291	*44*	20	*38*	
	Missing	111	*14*	6	*8*	
Self-rated	Very good	39	*5*	2	*3*	
health	Good	267	*34*	25	*33*	
	Fair	397	*51*	39	*52*	
	Bad	66	*8*	8	*11*	
	Very Bad	10	*1*	1	*1*	
	Missing	0	*0*	0	*0*	
Chronic	No	253	*33*	20	*27*	
morbidity	Yes	503	*67*	54	*73*	
	Missing	23	*3*	1	*1*	
GALI	Not limited	496	*64*	40	*53*	
	Limited but not severely	243	*31*	32	*43*	
	Severely limited	37	*5*	3	*4*	
	Missing	3	*0*	0	*0*	

[a] Population census 2011 [25]

Notes: This table presents the sociodemographic characteristics and health status. "GALI" refers to activity limitations.

Mean (±SD) PAM-13, eHEALS and NVS scores were 60.6 (±10.0), 28.7 (±5.1) and 3.9 (±1.7), respectively. The number (%) of patients with PAM-13 levels 1,2,3 and 4 were 54 (6.9%), 169 (21.7%), 471 (60.5%) and 85 (10.9%), respectively. Health literacy measured by the NVS was adequate (≥4) for 491 (63.0%), possibly limited (2-3) for 197 (25.3%), and probably limited (≤1) for 91 (11.7%). Mean (±SD) lifestyle risk index was 1.2 (±1.0). Out of 779 respondents 0, 1, 2 and ≥3 lifestyle risk factors were reported by 210 (27.0%), 275 (35.3%), 207 (26.6%) and 87 (11.2%), respectively. Mean (±SD) preventive behaviour score was 0.45 (±0.25). According to the cut-off values for hypothesis testing, 466 (59.8%) respondents searched health-related information at least monthly, 171 (22.0%) participated in patient education over the past year, 443 (56.9%) performed health-related administration over the internet, 421 (54.0%) sought health-related information at least bimonthly, 196 (25.2%) communicated online over past year about health with health care professionals (HCPs), helpers or peers, 280 (35.9%) engaged in online health prevention activities 347 (44.5%) participated in online disease-management activities over the past year.

II.3.2. Classical test theory methods

The distribution plots of PAM-13 are displayed in the Appendix II.2. We summarized the psychometric properties of PAM-13 along with the applied methods, target values, and results in Table 8. Altogether, PAM-13 scores fit rather log-normal than normal distribution with no floor- or ceiling effect. Internal consistency was adequate, while test-retest reliability was moderate. The single-factor structure was confirmed, convergent validity (correlation with eHEALS scores) and discriminant validity hypotheses (low / no correlation with age, education and income) were supported. Polychoric and polyserial correlations were used for ordinal and nominal variables, respectively, while Pearson correlation was applied for continuous variables. Known groups validity tests were significant in 7 / 9 (77.8%) hypotheses. The results of known-groups hypothesis tests in various subgroups are summarized in the Appendix II.2. Generally fewer than 7 hypotheses were supported in the subgroups than in the entire sample, except for chronic

patients (7 / 9 hypothesis supported) and 18-65-year-olds (8 / 9 hypothesis supported). None of the hypotheses were supported in the >65 age group. Respondents with higher PAM-13 levels had fewer lifestyle risk factors in 9 / 10 subgroups and sought more often health information in 7 / 10 subgroups. The smallest detectable change (SDC) was 7.1 points. The associations between PAM-13 scores and normal BMI, physical activity, and a healthy diet were tested using analysis of variance (ANOVA).

Table 8 Summary of the results of classical test theory methods

Category	Property	Method	Target	Result	*p* value / [95%CI]	Comment
General	Distribution	Skewness	0.00	0.36	<0.001	Positive skew
		Kurtosis	3.00	3.24	0.165	Normal kurtosis
		Shapiro–Wilk test for normal distribution	$p \geq 0.05$	-	<0.001	Deviation from normality
		Shapiro–Wilk test for log-normal distribution	$p \geq 0.05$	-	0.811	Log-normal distribution
		Floor effect	<15%	0.13%	[0.0%-0.7%]	No floor effect
		Ceiling effect	<15%	0. 25%	[0.0%-0.9%]	No ceiling effect
Reliability	Internal consistency	Cronbach alpha	0.7-0.95	0.77	[0.74-0.79]	Adequate
	Test-retest reliability	$ICC_{agreement}$[a]	>0.7	0.62	[0.46-0.74]	Moderate
		Standard error of measurement	-	6.5	[5.4-7.8]	-
		Smallest detectable change	-	7.1	[6.4-7.7]	-
		Weighted kappa[b]	>0.7	0.46	[0.26-0.65]	Moderate
Validity	Structural validity	CFA[c] sample				Adequate
		- KMO[d]	0.5	0.84		
		- Bartlett test	$p<0.05$	-	<0.0001	
		CFA: single factor structure				Good fit
		- RMSEA[e]	<0.05	0.049	[0.041-0.057]	
		- CFI[f]	>0.90	0.947		
		- TLI[g]	>0.90	0.937		
	Convergent validity	PAM-13 – eHEALS[h] Pearson correlation	r>0.3[i]	r=0.40 -	<0.001	Supported
		PAM-13 levels – eHEALS Polyserial correlation[b]	□>0.3[i]	□=0.39 -	<0.001	Supported
	Discriminant validity	PAM-13 – age Pearson correlation	r<0.3[j]	r=0.02 -	0.524	Supported
		PAM-13 – education polyserial correlation	□ <0.3[j]	□=0.04 -	0.273	Supported
		PAM-13 – income quintiles	□ <0.3[j]	□ =0.04	0.321	Supported

	polyserial correlation				
Known groups validity	PAM-13 score difference PBS[k] ≥50% vs PBS <50%	□>0.0[i]	□□0.91	□□102	Not supported
	PAM-13 score difference Lifestyle risk index:0 vs ≥1	□>0.0[i]	□□3.87	<0.001	Supported
	PAM-13 score difference Lifestyle risk index ≤1 vs ≥2	□>0.0[i]	□□4.47	<0.001	Supported
	PAM-13 score difference Health information seeking at least monthly vs less	□>0.0[i]	□□□□41	<0.001	Supported
	PAM-13 score difference patient education over past year vs none	□>0.0[i]	□□1□88	□□□18	Supported
	PAM-13 score difference online health information seeking at least bimonthly or less	□>0.0[i]	□□0.91	□□104	Not supported
	PAM-13 score difference online health-related communication past year vs none	□>0.0[i]	□□□□61	□□025	Supported
	PAM-13 score difference online health-prevention over past year vs none	□>0.0[i]	□□□□60	□□□15	Supported
	PAM-13 score difference online disease management over past year vs none	□>0.0[i]	□□□□46	□□□44	Supported

Notes: [a]ICC: intra-class coefficient; [b]Results refer to PAM-13 levels (all other results: PAM scores); [c]CFA: confirmatory factor analysis; [d]KMO: Kaiser-Meyer-Olkin statistic; [e]RMSEA: root mean squared error of approximation; [f]CFI: comparative fit index; [g]TLI: Tucker-Lewis Index; [h]eHEALS: eHealth Literacy Scale; [i]p<0.05, significant; [j] p≥0.05, not significant; [k]PBS: preventive behaviour score

II.3.2.1. Rasch analysis

Table 9 summarizes the data quality of the Rasch model based on the person infit and outfit indices. The results indicate that the PAM-13 scores used in this paper are based on high quality, consistent data.

Table 9 Summary of Rasch data quality indices

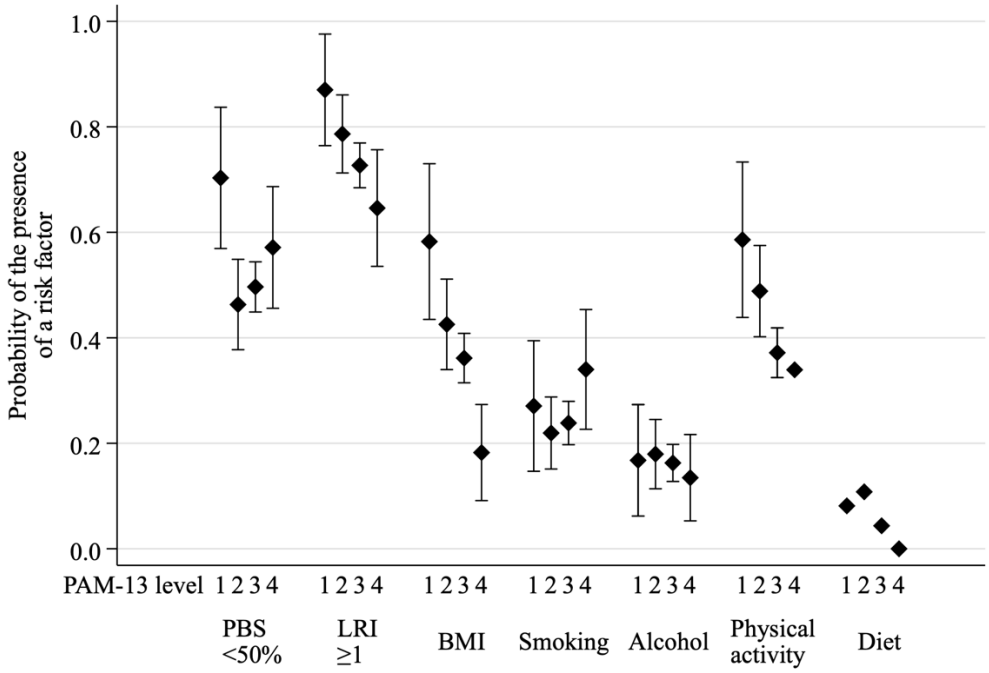
	First administration		Second administration	
	Good to high	**Poor**	**Good to high**	**Poor**
Infit	732 (94.0%)	47 (6.0%)	72 (96.4%)	3 (3.6%)
Outfit	728 (93.5%)	51 (6.5%)	72 (96.4%)	3 (3.6%)

Notes: This table presents the number and percentage of data in terms of quality.

II.3.2.2. Regression analysis

When controlled for eHealth literacy, health literacy, sociodemographic variables and respondents' health status, the association of PAM-13 scores with lifestyle-related risks remained significant. In particular, higher PAM-13 scores were associated with fewer lifestyle-related risks overall, lower probability of risky BMI (<18.5 or >30), physical inactivity or diet low in fruits and or vegetables. However, smoking and binge drinking episodes were not associated with PAM-13. Furthermore, although health information seeking, participation at patient education programs and several online health-related behaviours were associated with PAM in bivariate hypothesis tests, in multiple regression models these factors were associated with eHEALS or NVS but not with PAM-13 scores. Higher preventive behaviour scores were associated with higher education levels, but neither with patient activation nor with health literacy. The regression table is shown in the Appendix II.3. PAM-13 levels showed similar pattern to PAM scoresAppendix II.4. After controlling for the regression predictor variables, the adjusted probabilities of various lifestyle-related risks at different PAM-13 levels are depicted in Fig. 10 for the entire sample, in Fig. 11 for the subgroup with chronic morbidity and in Fig. 12 for 65+ year-old patients. While the pattern of chronic patients was similar to the entire sample, neither the overall number of risk factors, nor the level of physical activity was associated with higher PAM-13 levels in the elderly subgroup. The association of lifestyle risks with PAM-13 scores for the entire sample, chronic patients and the 65+ subgroups are depicted in Appendix II.6, II.7 and II.8 respectively.

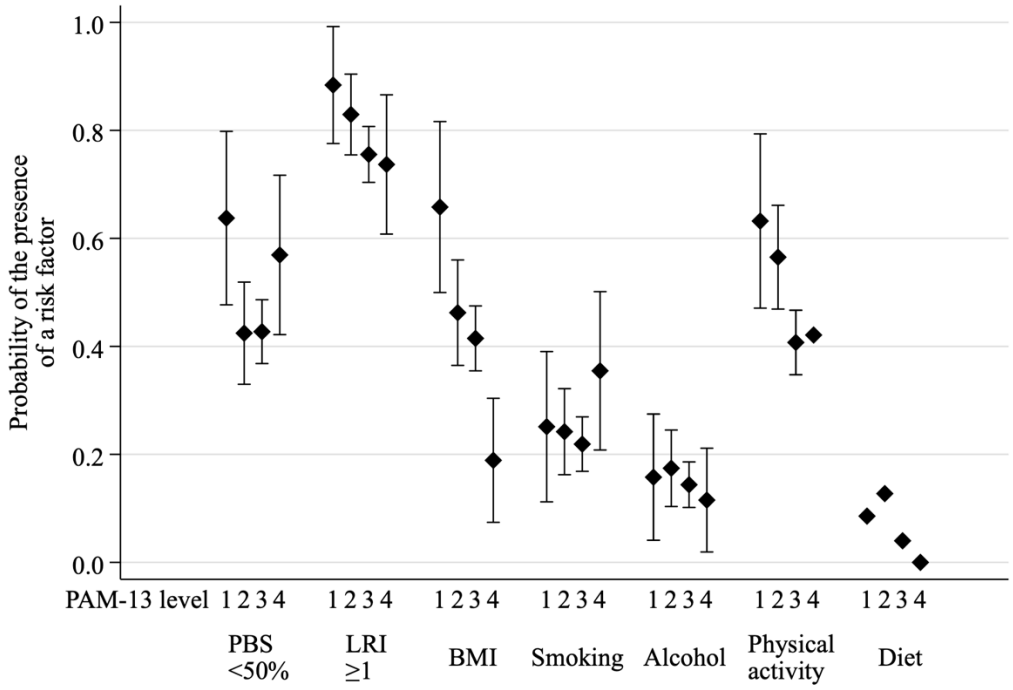
Figure 10 Adjusted probability of lifestyle-related risks at various PAM-13 levels in the entire sample



Source: (Zrubka et al., 2022)

Notes: This figure shows the adjusted probability of lifestyle-related risks at various PAM-13 levels across the entire sample. "PBS" refers to the preventive behavior score, "LRI" stands for Lifestyle Risk Index (18.5<), and "BMI" represents body mass index (<30).

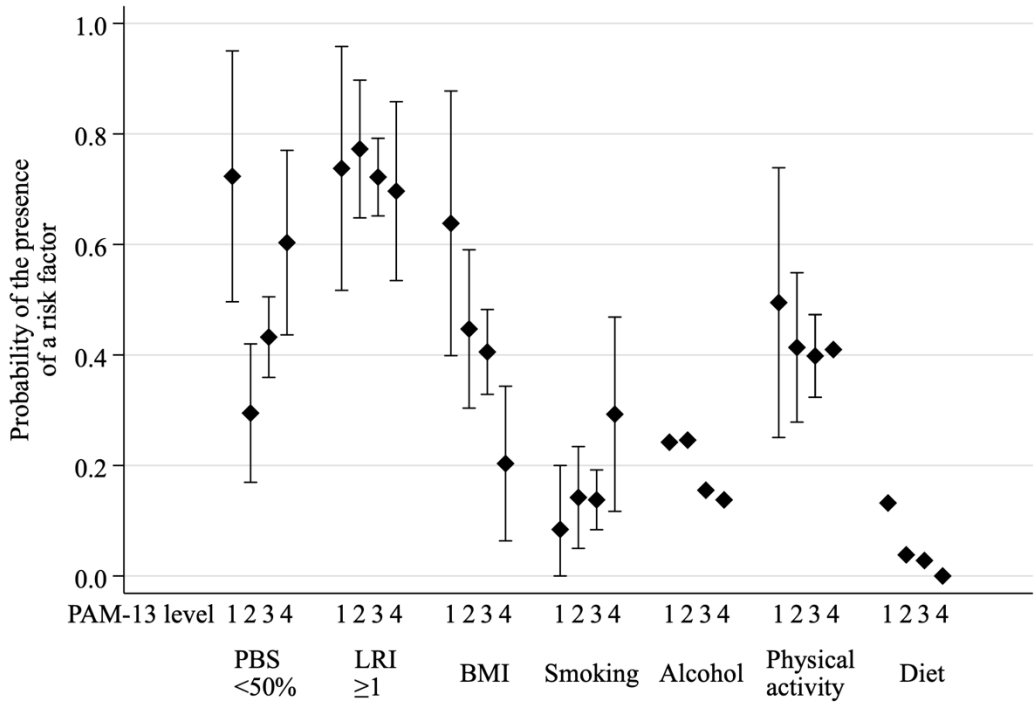
Figure 11 Adjusted probability of lifestyle-related risks at various PAM-13 levels in the subgroup with chronic morbidity



Source: (Zrubka et al., 2022)

Notes: This figure shows the adjusted probability of lifestyle-related risks at various PAM-13 in the subgroup with chronic morbidity. "PBS" refers to the preventive behavior score, "LRI" stands for Lifestyle Risk Index (18.5<), and "BMI" represents body mass index (<30).

Figure 12 Adjusted probability of lifestyle-related risks at various PAM-13 levels in the 65+ subgroup



Source: (Zrubka et al., 2022)

Notes: This figure shows the adjusted probability of lifestyle-related risks at various PAM-13 in the 65+ subgroup. "PBS" refers to the preventive behavior score, "LRI" stands for Lifestyle Risk Index (18.5<), and "BMI" represents body mass index (<30).

II.4. Discussion

In this cross-sectional online survey among the 40+ year-old Hungarian general population, we demonstrated the validity and tested the psychometric properties of the Hungarian version of the PAM-13 instrument. Following the COSMIN guidelines, we found good internal consistency, moderate test-retest reliability, a single-factor structure and good content validity consistent with majority of the predefined hypotheses. In particular, higher PAM-13 scores of PAM-13 levels were associated with fewer lifestyle related risks even when controlled for eHealth literacy, health literacy, socioeconomic characteristics and the health status of respondents. Higher PAM-13 scores were especially associated with greater probability of having normal BMI, being physically active and having a diet including vegetables and fruits. However, smoking and risky alcohol intake was not associated with PAM-13 scores.

The results of bivariate hypothesis tests were consistent with the entire sample in chronic patients and in the 40-65 years old age group. However, in the 65+ population higher PAM-13 levels were not associated with either fewer lifestyle risk factors or more health information seeking, participation at patient education or online health-related activities. Therefore, the validity of PAM-13 in the elderly population should be further studied with larger population samples and confirmed by testing alternative hypotheses more relevant to this population such as disease management, medication adherence *(Graffigna et al., 2017)* or costs of care *(Lindsay et al., 2018)*.

Our results are consistent with other validation studies. Comparisons are more relevant with surveys that were conducted among the general population, however the differences between the demographic characteristics of the samples should be considered. The PAM instruments were developed on a sample of 45+ year-old individuals from the US general population, with 34% over age of 65 *(Hibbard et al., 2004)* and *(Hibbard et al., 2005)*. Fowles et al. assessed the validity of PAM-13 in a younger age group: US employees with mean age 45 of years *(Fowles et al., 2009)*. PAM-13 was also validated in the general population of Israel *(Magnezi et al., 2014)* and the Netherlands *(Hendrikx et al., 2018)*. The average PAM-13 score varied between 60.2 and 70.7 across the four studies, our result was closest to the original development study (60.6 vs. 60.2) *(Hibbard et al., 2005)*. Cronbach alpha varied between 0.77 and 0.9 *(Hibbard et al., 2004), (Hibbard et al.,*

2005), (Magnezi et al., 2014) and *(Hendrikx et al., 2018)*. Associations between PAM-13 and demographic characteristics (age, gender, educational level) varied across the four studies, however positive association with health status (assessed by different measurement tools) was rather consistent *(Hibbard et al., 2004), (Hibbard et al., 2005), (Magnezi et al., 2014)* and *(Hendrikx et al., 2018)*.

We tested the association between patient activation and a broad set of variables concerning participants' health-related risk factors, information seeking and health management behaviour. Similarly to our findings, Hibbard et al. reported that the preventive, consumeristic and the disease-specific self-management behaviours were strongly associated with PAM-13 scores *(Hibbard et al., 2005)*. Fowles et al. found strong positive association between PAM-13 and personal healthy behaviours and health information seeking *(Fowles et al., 2009)*. In our study, the items of MEHM were used to control for the health status of respondents. However, assuming diverse health related behaviours and health states in the general population, the association between PAM-13 and health status was not tested in this study.

PAM-13 validation studies were conducted also in a series of specific patient populations including diabetes *(Kosar et al., 2019), (Laranjo et al., 2018)* and *(Zeng et al., 2019)*, metabolic syndrome *(Bahrom et al., 2020)*, multiple sclerosis *(Stepleman et al., 2010)*, neurological disorders *(Kephart et al., 2019)* and *(Packer et al., 2015)*, hypertension *(Kosar et al., 2019)* and *(Zeng et al., 2019)* , cardiac conditions *(NGooi et al., 2017)*, schizophrenia *(Melby et al., 2021)*, mental disorders *(Moljord et al., 2015)*, osteoarthritis *(Ahn et al., 2015) (Eyles et al., 2017)* and *(Eyles et al., 2020)*, rheumatic diseases *(Kosar et al., 2019)* and *(Roe et al., 2020),* HIV *(Cooper et al., 2019)* or not specified chronic diseases *(Cunha et al., 2018), (Skolasky et al., 2011)* and *(Graffigna et al., 2015)* and clinical settings in different countries *(Brenk-Franz et al., 2013), (Hung et al., 2013), (Prey et al., 2016), (Schmaderer et al., 2015)* and *(Hellstrom et al., 2019)*. In some of these studies, convergent validity was assessed using generic and disease-specific measures of health-related attitudes or behaviours such as self-efficacy *(Stepleman et al., 2010), (Kephart et al., 2019), (Roe et al., 2020)* and *(Brenk-Franz et al., 2013)*, self-esteem *(Cunha et al., 2018)*, lifestyle *(Packer et al., 2015)* or health literacy *(Prey et al., 2016), (Schmaderer et al., 2015), (Hellstrom et al., 2019)* and *(Rademakers et al., 2012)*. For the assessment of convergent validity, we chose the eHEALS scale, a widely used

measure of self-reported eHealth literacy that is strongly rooted in self-efficacy theory *(Zrubka et al., 2019)* and *(Norman et al., 2006)*. While PAM-13 and eHEALS showed moderate positive correlation, the multiple regression analyses revealed that the two instruments are associated with different facets of health-related behaviours. While higher PAM-13 scores predicted healthier lifestyles, higher eHEALS scores were associated with more frequent health information seeking and online health related behaviours.

The strength of our study is that it met the most meticulous methodological standards outlined by the COSMIN best practices guide *(Mokkink et al., 2010)*: all phases of our approach, including descriptions of the research aim, the PROM and the target population, the design requirements, the analysis of structural validity, internal consistency, measurement invariance, measurement error and reliability, criterion validity, construct validity, known-groups validity and the construct approach (hypothesis testing), fell into the highest quality category defined by these guidelines.

The limitation of our study is that PAM-13 was administered in an online population, with over-representation of highly educated, affluent urban respondents, which reflects the general sampling bias of online surveys *(Eysenbach et al., 2002)*. However, patients from rural regions or lower socio-economic status may be particularly prone to lifestyle risks and poor health status *(Pál, 2017)* and *(OECD, 2017)*, therefore the validity of the patient activation concept is of particular importance in these populations. Nevertheless, in subgroups with lowest income and inadequate health literacy levels, higher PAM-13 scores were associated with fewer lifestyle risk factors and more frequent health information seeking in our study, which supports the validity of the PAM-13 instrument in these vulnerable subgroups. Furthermore, although the validity of PAM-13 was supported among respondents with self-reported chronic morbidity, further studies are required to assess the psychometric properties of PAM in clinical populations.

Furthermore, as a potentially modifiable enhancer of patients' contribution to the health production process, the responsiveness of PAM-13 to interventions is of key interest *(Expert Panel, 2019)* and *(WHO, 2015)*. However, the responsiveness of PAM-13 is a poorly explored area. We found one validation study from Norway in which patients with chronic inflammatory arthritis attending a patient education programme were involved *(Roe et al., 2020)*. The low responsiveness was partly explained by the substantial ceiling

effect in 10 of the 13 items. Our results suggest that the moderate test-retest reliability of PAM-13 necessitates sufficiently large sample sizes in adequately powered studies to demonstrate change. We encourage further research both among the general public and patient populations to get a better insight into responsiveness of PAM-13 to interventions.

We suppose that certain properties of PAM-13 may be context specific depending on the health care setting. In Hungary, the weak primary care system acts as a gatekeeper and offers limited choice of providers *(Rotar et al., 2018)*, while health care remains hospital-centric and inefficient despite the low expenditure levels *(OECD, 2019b)*. Health-conscious patients may demand more services and therefore may cost more the system, while patients from low socioeconomic status may have poorer overall access to care *(OECD, 2017)*. Therefore, while higher PAM-13 scores were associated with lower resource use in the US health care system *(Harvey et al., 2012)* and *(Remmers et al., 2009)*, the association of PAM-13 with costs of care warrants further research in the Hungarian context.

As a conclusion, the Hungarian version of the PAM-13 demonstrated good validity in the 40-65-year-old general population as well as among people with self-reported chronic morbidity. Further research is needed to establish the validity of PAM-13 in clinical populations as well as among the elderly and its responsiveness to interventions. To provide a clearer overview of the study findings, Table 10 summarizes the key results and conclusions from the validation of the Hungarian version of the PAM-13.

Table 10 Summary Table of Main Results and Conclusions from PAM-13 Validation Study in Hungary

Category	Key Findings
Study Design	Cross-sectional online survey among Hungarian adults aged 40+
Psychometric Properties	- Good internal consistency (Cronbach's alpha) - Moderate test-retest reliability - Single-factor structure - Good content validity (confirmed most predefined hypotheses)

Category	Key Findings
Health Behavior Associations	- Higher PAM-13 scores associated with: → Normal BMI → Physical activity → Fruit & vegetable-rich diet *Tested via ANOVA, associations remained after controlling for health and socioeconomic factors*
No Associations Found With	- Smoking - Risky alcohol intake
Subgroup Findings	- Similar results in chronic patients and 40–65 age group - No significant associations in 65+ population for lifestyle factors or information-seeking behavior
eHealth Literacy (eHEALS)	- Moderate correlation with PAM-13 - PAM-13 associated with healthier lifestyle - eHEALS associated with online health information seeking
Comparison with Other Studies	- Findings consistent with US, Israeli, and Dutch validation studies - Hungarian PAM-13 mean score (60.6) close to original US score (60.2)
Limitations	- Online sample bias: overrepresentation of urban, educated, affluent respondents - PAM-13 not tested in clinical settings or with objective health status data
Strengths	- Followed COSMIN best practices for PROM validation - Strong methodological rigor in psychometric evaluation
Responsiveness	- Low responsiveness noted in previous studies (e.g., Roe et al., 2020) - Moderate test-retest reliability suggests large sample sizes needed for intervention studies
Contextual Considerations	- Hungarian health system differs from US (e.g., hospital-centric, weak primary care) - Association between PAM-13 and health care costs may vary by country and requires further study
Conclusion	- Hungarian PAM-13 is valid in general population aged 40–65 and those with chronic conditions - Further studies needed in elderly, clinical populations, and for intervention responsiveness

Part III. Use of PAM in health insurance pricing models

The Patient Activation Measure (PAM) is a pivotal instrument designed to assess the degree to which individuals possess the knowledge, skills, and confidence to effectively manage their health and health care. Since its development, PAM has emerged as a significant predictor of health care outcomes, with a substantial body of research underscoring its link to various aspects of health management, including medication adherence, utilization of health care services, and patient engagement in preventive measures. Among the most compelling areas of investigation is the relationship between patient activation levels, as measured by PAM, and health-related costs. This relationship is critical as healthcare systems around the world face rising costs and the need for models that promote efficiency and patient-centered care *(Zrubka et al., 2022)*.

A wide range of studies have demonstrated that higher levels of patient activation are associated with lower health-related costs. For instance, *Hibbard et al. (2013)* found that patients with the highest levels of activation had health care costs that were significantly lower than those with the lowest levels of activation over a one-year period. This inverse relationship holds promise for the implementation of patient-centered interventions designed to enhance activation levels and, consequently, reduce costs. Moreover, Greene and *Hibbard (2012)* identified the mechanisms through which activated patients contribute to cost savings, including reduced hospitalizations and emergency room visits, which are among the most expensive forms of health care utilization.

The financial implications of patient activation extend beyond direct health care costs to include broader economic impacts, such as productivity losses and indirect costs related to caregiver burden. As such, the exploration of PAM as a tool for identifying patients at risk of higher health care expenditures is not only pertinent but necessary for the development of targeted interventions that can mitigate these costs.

The existing literature clearly supports the notion that improving patient activation can serve as a pivot point in efforts to contain healthcare costs while improving outcomes. This paper aims to synthesize the findings from these studies, offering a comprehensive analysis of the evidence on the relationship between patient activation, as measured by PAM, and health-related costs. In doing so, it will highlight the critical role of patient

activation in the broader context of health economics and policy, underscoring its potential as a lever for achieving more sustainable health care systems.

After validating the Patient Activation Measure (PAM-13) in the Hungarian general population aged 40 years and above, I have performed this research independently. As a prerequisite of this work, the preliminary step presented in the previous chapter was particularly important: any risk indicator that is intended to meet academic standards must first be strictly validated in the target population. This ensures the reliability and applicability of the measure for assessing or predicting relevant outcomes.

With the PAM-13 validation complete, my immediate research focus shifts to exploring its potential to impact health care costs within the Hungarian system. The hypothesis driving this investigation is based on existing literature suggesting that higher patient activation levels are associated with more cost-effective health management. This inquiry is particularly pertinent given the distinct challenges and opportunities presented by Hungary's health care environment.

The methodology involves developing a Generalized Linear Model (GLM) utilizing the collected data. This approach will allow for an in-depth analysis of the association between varying patient activation levels, as delineated by PAM-13 scores, and health care-related expenditures. The GLM is tailored to account for the skewed nature of cost data, typically characterized by a majority of individuals incurring minimal expenses and a minority facing significantly higher costs due to complex health needs *(Nelder-Wedderburn, 1972).*

This analysis aims to clarify the correlation between patient activation levels and health care costs, including the utilization of hospital services, outpatient care, and medications. Beyond elucidating this relationship, the study seeks to inform health policy and practice within Hungary, advocating for strategies that could foster higher patient activation.

The implications of this work extend beyond academic circles, potentially influencing health care policy, provider practices, and insurance models in Hungary. By demonstrating a link between patient activation and cost savings, the research could offer a compelling case for integrating patient-centered strategies into the health care system.

These strategies might include targeted education, enhanced support for chronic disease management, and improved patient-provider communication.

Ultimately, this project aims to contribute significantly to the discourse on patient engagement and health care efficiency, offering insights that could lead to more sustainable health care practices. By leveraging the validated PAM-13 instrument, I hope to illuminate paths towards achieving optimal health outcomes at reduced costs, setting a precedent for health systems in Hungary and beyond.

III.1. Patient Activation and its impact: Analysis of health care costs and self-rated health outcomes

In this section, I would like to detail and compare the findings of two key articles that explore the complex relationship between Patient Activation Measure (PAM) and healthcare costs. Through a comprehensive analysis, I explore how heightened patient activation, as quantified by PAM scores, not only plays a significant role in enhancing self-care among individuals with chronic conditions is correlated with a notable reduction in health care expenditures. The comparison of these articles sheds light on the multiple benefits of patient empowerment, highlighting the dual benefit of improved health outcomes and economic savings in the healthcare landscape.

A study by Lindsay et al (2018) examined the relationship between changes in Patient Activation Measure (PAM) levels and health care costs among high-risk patients in the United States. A prospective cohort study design was used, analyzing de-identified claims, demographic data, and serial PAM scores from 2,155 patients over 3 years. Key findings from the study revealed that a higher Patient Activation Measure (PAM) score is linked to a notable decrease in follow-up costs, with each increase in PAM corresponding to an 8.3% reduction. This suggests that patient activation could potentially serve as an early indicator of the success of interventions aimed at high-cost, high-needs patients. Additionally, the research underscores the potential of PAM to act as a valuable tool for predicting changes in healthcare expenses, highlighting the critical role of patient engagement in the efficient management of healthcare costs.

A study by Tusa et al. (2020) investigated the relationship between patient activation, as measured by the Patient Activation Measure (PAM), and self-rated health (SRH) among primary care patients with chronic diseases. The study involved 605 patients from the Siilinjärvi Health Center, focusing on those monitored for hypertension, ischemic heart disease, or diabetes. The major findings of the study show a clear linear relationship between Self-Rated Health (SRH) and the Patient Activation Measure (PAM), even after controlling for factors like age, number of diseases, and depressive symptoms. This indicates that higher patient activation is linked to better self-rated health. These results suggest that strategies designed to improve patient activation could positively influence how patients perceive their health and improve their management of chronic conditions.

Both studies highlight the important role of patient activation in chronic disease management and its significant impact on healthcare costs and self-rated health. The findings suggest that increased patient activation leads to improved self-management of chronic conditions, better health outcomes and reduced health care costs. They also point to the need for targeted interventions to increase patient activation, which could help individuals manage their conditions more effectively and potentially reduce the financial burden on the healthcare system.

In addition, these studies open the door for future research into how health care systems can implement strategies to improve patient activation, leading to better outcomes and more efficient use of health care resources.

III.2. Utilizing Generalized Linear Models (GLM) for insurance pricing

III.2.1. Introduction

Health insurance pricing is a complex process that requires sophisticated statistical models to accurately assess risk and determine premiums. Traditional linear regression models often fall short due to their assumptions of normality and constant variance, which are rarely met in insurance data. Generalised linear models (GLMs) offer a flexible alternative that can handle the characteristics of insurance data, including skewed distributions and non-linear relationships *(Burka et al., 2021)*.

Generalized Linear Models (GLM) extend the linear modeling framework to accommodate response variables that have error distribution models other than a normal distribution. GLMs are characterized by three components:

1. Random Component: Specifies the probability distribution of the response variable (Y) (e.g., normal, Poisson, binomial, gamma), allowing for the modeling of different types of data.
2. Systematic Component: Represents the predictors ($X_1, X_2, \dots, X_n$) or independent variables that are linearly related to the response variable through a link function.
3. Link Function: Provides the relationship between the linear predictor and the mean of the distribution function. It ensures that the model predictions stay within the range that is plausible for the given distribution.

The formula for a GLM can be written as:

$$g(\mu) = \eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Where:

- μ is the expected value of the response variable.
- $g(\cdot)$ is the link function that relates the mean of the distribution of Y (i.e., μ) to the linear predictor η.
- $\beta_0, \beta_1, \dots, \beta_n$ are the regression coefficients to be estimated.
- $X_1, X_2, \dots, X_n$ are the predictor variables.

The objective in GLM is to estimate the coefficients $\beta_0, \beta_1, \dots, \beta_n$ that best predict the response variable. This is typically done through maximum likelihood estimation *(Gray-Kovács, 2001).*

This framework is highly adaptable, enabling analysts to model a wide array of response variables. In the context of health insurance pricing, GLMs can accurately model outcomes such as the number of claims (using a Poisson distribution) or the cost of claims (using a gamma distribution) *(Burka et al., 2021).*

Generalized Linear Models are particularly well-suited for health insurance pricing due to several key advantages. First, they offer flexibility in modeling skewed data, such as

when a small number of policyholders generate large claims, by using distributions like the gamma distribution for claim sizes. Second, GLMs enable more precise risk differentiation by incorporating a wide range of policyholder characteristics (such as age, medical history, and lifestyle) and external factors (like economic indicators and geographic location), resulting in fairer pricing that can also incentivize policyholders to lower their risks. Third, GLMs support regulatory compliance by providing transparent and justifiable methods, which are crucial in the highly regulated health insurance sector. Lastly, their ability to model complex relationships between policyholder attributes and claim outcomes enhances predictive accuracy, improving both risk assessment and premium setting.

The implementation of GLM in health insurance pricing involves several steps, including data preparation, model selection (distribution and link function), model fitting, and validation. Actuaries must also consider the potential for overdispersion and the need for regular updates to the model as new data becomes available *(Burka et al., 2021).*

III.2.2. Basics of insurance pricing

Insurance pricing is a critical task in the insurance industry, ensuring that premiums collected cover the claims made, administrative costs, and also yield a reasonable profit. This section provides an overview of the core components of insurance pricing, highlighting its objectives, constraints, and the regulatory framework in which it operates. Additionally, it explores the crucial role of data in informing these pricing decisions *(Banyár, 2016).*

The main objective of insurance pricing is to set premiums that accurately reflect the risk of insuring an individual or entity. This process involves several important goals. First, premiums must be adequate to cover expected claims and expenses throughout the policy period. Second, pricing should be fair, ensuring that policyholders pay according to their risk level, preventing cross-subsidization. Third, premiums must be competitive to attract and retain customers in the market. Lastly, the pricing strategy should aim for profitability to ensure the insurer's financial stability and potential for growth.

Insurance pricing is influenced by several constraints. Regulatory requirements, both at the country and EU levels, mandate adherence to minimum coverage standards, rate filing procedures, and solvency rules. Market conditions also play a role, as insurers must balance competitive pricing with maintaining financial stability. Additionally, data limitations, such as the availability, quality, and relevance of data, can affect pricing accuracy, particularly when dealing with emerging risks or limited historical data, complicating risk assessment.

Insurance is highly regulated to protect consumers and maintain the financial stability of insurers. Regulatory considerations that influence pricing include the need for rate approval, where insurers in many territories must apply to and receive approval from a regulatory body to ensure that rates are neither too high nor too low. Transparency is also crucial, with regulations often requiring insurers to disclose how rates are determined to build consumer confidence. In addition, anti-discrimination laws prevent unfair practices by ensuring that pricing is based only on risk factors directly related to the likelihood of a claim, and not on prohibited factors such as race, religion or gender. In the European Union, the most significant regulation in this area is the Gender Directive, which prohibits using gender as a factor in determining insurance pricing. However, this does not prevent actuaries from modeling gender-specific risk internally; it remains a matter of expert judgment how to translate those risk assessments into a unified final price presented to policyholders. For instance, separate reserves for men and women may still be maintained based on differing risk profiles.

Data is fundamental to insurance pricing, serving as the basis for risk assessment and pricing decisions. Insurers use various types of data, such as personal information (age, health, driving record), policy details (such as coverage limits and deductibles), claims history, and external data such as economic indicators or natural catastrophe trends. Data can be sourced from internal records, third-party providers, government agencies and, increasingly, telematics and IoT devices. To analyse this data, insurers use advanced statistical models, including generalised linear models (GLMs), to identify risk factors. This analysis helps to develop rating factors, which are then used to set individual premiums.

III.2.3. Understanding Generalized Linear Models (GLM)

This section provides an overview of the basic aspects of GLM, its key components, and the types of GLM most applicable to insurance pricing *(Gray-Kovács, 2001).*

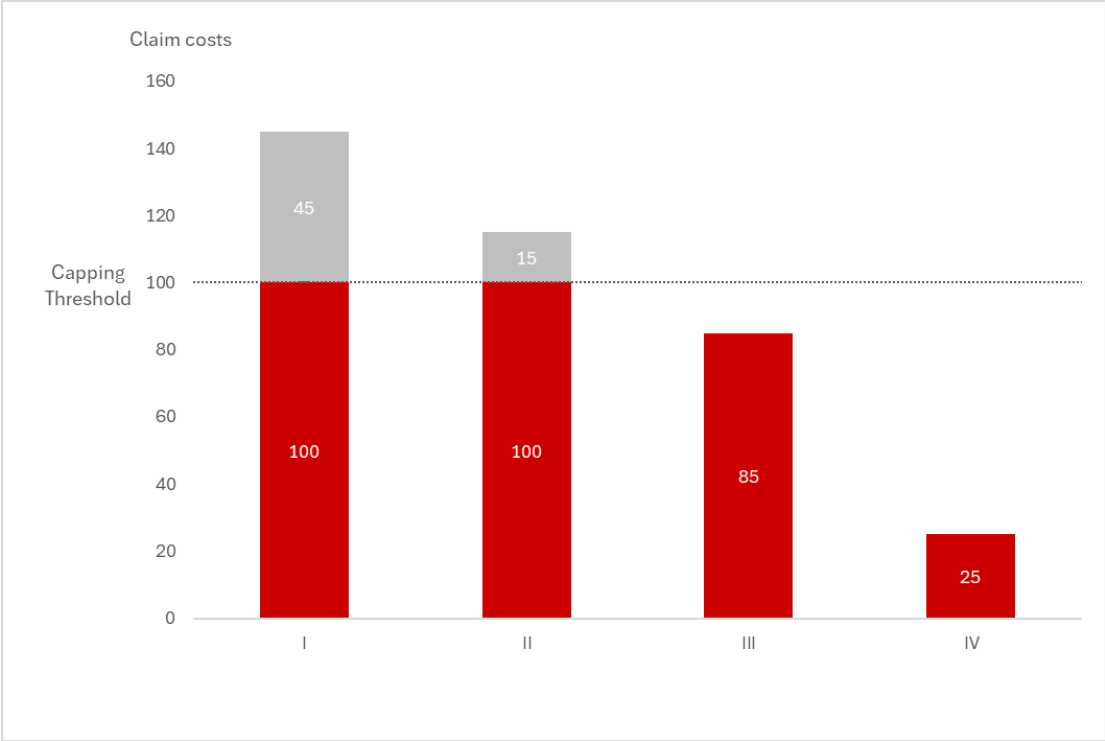
When modeling claims, the risk models in non-life insurance are typically frequency-severity models, both modeled using GLM.

The basic concept is that after obtaining the frequency and severity models, the next step is to combine them. This can be done using Emblem's built-in model combiner tool or by applying the frequency × severity formula to estimate expected claims.

To improve the accuracy of the estimation, it is advisable to filter out large claims, commonly referred to as "outliers." There are two main methods for handling these. The traditional approach involves applying a fixed capping threshold: claims exceeding this predefined level are modeled separately. Figure 13 illustrates this method. In the example, the capping threshold is set at 100, and claims I and II exceed this limit. These are modeled using two separate models: a propensity model (to estimate the likelihood of a large claim) and an average excess model (to estimate the expected amount above the threshold). Table 12 shows the properties of the propensity model.

An alternative approach is to construct a dynamic large claim model, where the capping threshold is not fixed but varies based on certain characteristics or predictors. In my personal experience, this method may offer marginal improvements in predictive accuracy, but it requires significantly more effort to implement. Therefore, I believe that the traditional fixed-threshold method is generally sufficient for most practical purposes. The example illustrated in Figure 13 demonstrates the fixed threshold version.

Figure 13 Large claim capping example



Source: own ed.

In a GLM model, weights are used for the frequency model, which represents the exposure in non-life insurance where the contract term is one year. This exposure is often expressed in 'risk years'. For severity, the weights correspond to the number of claims. Table 11 shows some key formulas:

Table 11 Key formulas

Metric	Formula
Exposure	$\frac{Number\ of\ days\ in\ calendar\ year}{Total\ days\ of\ the\ calendar\ year\ (365\ or\ 366)}$
Earned Premium (in a Year)	$Annual\ Premium\ \times\ Exposure$
Frequency (in a Year)	$\frac{Number\ of\ Claims\ incurred\ in\ a\ Year}{Exposure}$
Average Cost	$\frac{Cost\ of\ Claims\ incurred\ in\ a\ Year}{Number\ of\ Claims\ incurred\ in\ a\ Year}$
Pure Premium	$Frequency\ \times\ Average\ Cost$
Loss Ratio (Accident Year)	$\frac{Cost\ of\ Claims\ incurred\ in\ a\ Year}{Earned\ Premium}$
Rate Sum Insured	$\frac{Pure\ Premium}{Sum\ Insured}$

Source: own ed.

The Omnibus test is an important tool in evaluating the overall effectiveness of a model. Its purpose is to determine whether at least one predictor variable has a statistically significant effect on the dependent variable, thereby assessing the model's overall explanatory power rather than focusing on individual predictors. The test computes a statistic by comparing the likelihood of the data under the null model (without predictors) and the proposed model (with predictors), which follows a chi-square distribution and allows the calculation of a p-value.

A significant result (typically $p < 0.05$) indicates that the model with predictors offers a better fit to the data than the null model, meaning at least one predictor is significantly related to the dependent variable. A non-significant result suggests that the model does not improve the fit over the null model.

Omnibus tests are particularly useful in the early stages of model building to confirm that the model as a whole has explanatory value. However, it is important to note that while

an Omnibus test can indicate whether a model is overall significant, it does not specify which variables are significant. Further analysis, such as examining individual coefficient tests, is necessary to identify the specific significant predictors.

III.2.3.1 Types of GLM in insurance pricing

Several types of Generalized Linear Models (GLMs) are particularly useful for insurance pricing, each suited to specific types of data and outcomes.

Poisson regression is ideal for modeling count data, such as the number of claims filed by a policyholder, as the Poisson distribution effectively handles the discrete nature of such data.

Binomial regression is used for binary outcomes, like whether a claim occurs or not, and is useful for calculating the probability of an event happening.

Gamma regression is well-suited for modeling positive continuous data, such as claim sizes, as it can manage the skewed nature typical of claim amounts.

By accommodating the unique distributional characteristics of insurance data, GLMs enable actuaries and data scientists to develop more accurate and fair pricing models, ultimately benefiting both insurers and policyholders. The GLM model is versatile and can be used for multiple purposes, including building lapse models. Table 12 lists some models along with their corresponding parameters and main properties:

Table 12 Most important GLM risk models

Model	Response (y)	Error Structure	Link function (g)	Variance function (V)	Prior weight
Frequency	Number of Claims	Poisson	ln	μ	Exposure
Severity	Cost of Claims	Gamma	ln	μ^2	Number of claims
Pure Premium	Cost of Claims	Tweedie	ln	$\frac{1}{\lambda}\mu^p$	Exposure
Propensity	Number of large claims	Binomial(μ,n)	$logit$	$\mu(1-\mu)$	Number of claims(n)
Lapses	Number of lapses	Binomial(μ,n)	$logit$	$\mu(1-\mu)$	Number of policies(n)

Source: *Gray-Kovács (2001)*

III.2.4: Application of GLM in insurance pricing

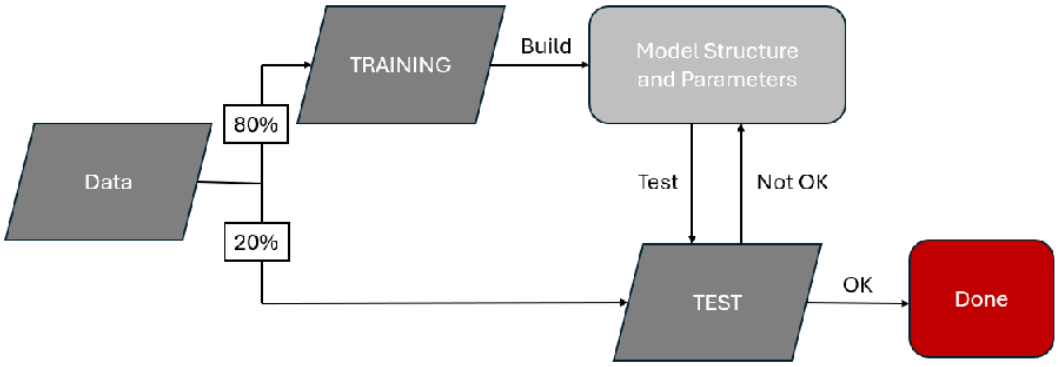
This section explores the application of GLM in insurance pricing, detailing the step-by-step process from data preparation to model validation *(Gray-Kovács, 2001)*.

III.2.4.1 Data preparation

The accuracy of a Generalized Linear Model (GLM) heavily depends on the quality of the input data. The pricing process begins with data collection, where information is gathered from both internal and external sources, such as policyholder demographics, claims history, and external factors like economic indicators that affect risk. Next is data cleaning, which involves resolving issues like missing values, outliers, and inconsistencies to maintain the dataset's integrity. Variable selection follows, where relevant predictors are identified based on statistical significance and industry expertise, such as age, gender, or driving record, depending on the type of insurance. Finally, data transformation is applied to ensure the data aligns with the assumptions of the chosen GLM, including adjustments like logarithmic transformations for skewed data.

To avoid overfitting, which is a significant risk, split sample validation may be applied, i.e., splitting the dataset into a training and a test dataset using simple random sampling. A common solution is to use 80% of the dataset for training and 20% for testing. This way, the models can ensure that they measure the performance of their models on a dataset that was not used for training, thereby avoiding overfitting. However, sampling error can be significant, and more sophisticated approaches such as crossvalidation can be used to remedy this. Figure 14 illustrates the concept of split sample validation.

Figure 14 Splitting sample process



Source: own ed.

III.2.4.2. Model selection

Selecting the appropriate Generalized Linear Model (GLM) requires choosing the right distribution for the response variable and a suitable link function to describe the relationship between the predictors and the response. Distribution selection involves picking a distribution from the exponential family, such as Poisson for count data, binomial for binary data, or gamma for positive continuous data, based on the nature of the response variable.

The link function must then be chosen to ensure that the linear predictor appropriately connects to the mean of the response variable, with options like the log, identity, or probit link functions depending on the model.

III.2.4.3. Model fitting

Once the data is prepared and the model is specified, the next step is fitting the Generalized Linear Model (GLM) to the data. Parameter estimation is performed using statistical software, typically through maximum likelihood estimation or similar methods. Model refinement follows, where the significance of predictors and their interactions is assessed. Non-significant variables are removed to enhance the model's fit and prevent overfitting, ensuring the model is both accurate and generalizable.

III.2.4.4 Model validation

Validating the Generalized Linear Model (GLM) is essential to ensure its predictive accuracy and reliability. Diagnostic checks are performed to assess goodness-of-fit, including analyzing residuals to detect any patterns that may indicate violations of model assumptions. Cross-validation techniques, such as k-fold cross-validation, are used to evaluate the model's predictive performance on unseen data, ensuring its generalizability. Additionally, model comparison is conducted by evaluating the GLM's performance against other models or benchmarks to determine its effectiveness in predicting outcomes accurately.

The use of GLMs in insurance pricing is a methodical process that involves accurate data preparation, careful model selection and rigorous validation. When done correctly, GLMs provide a robust framework for developing pricing models that accurately reflect risk, benefiting both insurers and policyholders through fair and transparent premium setting.

In my experience, the best and most commonly used software by insurance companies around the world for working with GLM in an insurance company at the moment is Willis Towers Watson's software called Emblem. This software is very practical and includes many built-in tests. For instance, it offers a time interaction test, allowing one to observe trends in the data over time. This feature is practical for identifying changes in risk profiles or claim patterns, which are crucial for adjusting premiums and understanding risk dynamics. Emblem's user-friendly interface and comprehensive analytics tools make it particularly appealing for actuaries and data analysts working in the insurance sector. Emblem is currently the leading software in the market, widely adopted by insurance companies across the globe. As I did not have access to a license for Emblem, I conducted my calculations using IBM’s SPSS, a commonly used alternative. While SPSS includes several built-in features, it requires more manual steps to complete analyses and verify results. R could also be a viable option; however, due to its complexity, I would not personally recommend it. To put the efficiency into perspective, a few hours of work in Emblem can take several days in SPSS and over a week in R.
The "Tests of Model Effects" section of the SPSS output provides an analysis of each variable within the model. By calculating statistics such as the Wald chi-square, degrees of freedom and associated p-values for each predictor, it is possible to identify which

variables significantly affect the dependent variable.

One of the strengths of this test lies in its ability to handle complex models with multiple predictors, including interactions between variables. It helps in refining the model by pinpointing which predictors have a statistically significant effect, thereby guiding decisions on model simplification or the need for additional data. The results from the "Tests of Model Effects" can directly inform the iterative process of model building, where variables may be added or removed based on their contributions to explaining variance in the dependent variable.

III.2.5. Further benefits of using GLM in insurance pricing

GLMs are adaptable to different data types, unlike traditional models that assume a normal distribution. By accommodating Poisson, binomial, or gamma distributions, GLMs allow for more precise modeling of claims frequency, severity, and likelihood. Additionally, GLMs handle non-linear relationships between predictors and the response variable through the use of link functions, making it easier to capture complex patterns within the data.

By accurately modeling the relationship between risk factors and insurance claims, GLMs enable insurers to segment risks more effectively. This segmentation leads to fairer pricing, as premiums more closely reflect the actual risk posed by each policyholder.

GLMs offer flexibility in distribution selection, allowing the model to fit the nature of the response variable, whether the data is skewed, kurtic, or zero-inflated. This adaptability ensures more accurate modeling of complex data patterns. Additionally, GLMs improve the estimation of extremes by better modeling the tails of the distribution, which is crucial for setting reserves and assessing the insurer's liability in extreme events.

It is worth noting that other Hungarian researchers have also examined the use of Generalized Linear Models (GLMs) in insurance premium calculation. For example, Burka, Szepesváry and Kovács (2021) analyzed the application of GLMs in Motor Third Party Liability (MTPL) claim modeling. They compared GLMs with more advanced techniques such as Generalized Additive Models (GAMs), Random Forests, and Neural

Networks. Their findings—consistent with my own experience—highlight that while these more complex models can provide improved predictive performance, this often comes at the cost of interpretability, transparency, and increased model complexity. Notably, their final model was a hybrid approach combining several base methods, suggesting that while GLMs alone may not always offer the best predictive power, their clarity and ease of implementation remain valuable—especially in regulated environments like insurance.

Kovács (2021) further explored the computational aspects of these modeling approaches. He compared the runtime performance of GLMs and various GAM implementations, finding that with modern, optimized algorithms such as HGHS, the runtime of GAMs has become significantly more efficient—approaching that of GLMs—without sacrificing much in terms of accuracy. While GAMs remain methodologically more complex, these results suggest they are increasingly viable for real-world applications, particularly where nuanced modeling is required and computational resources permit.

III.2.6. Challenges and limitations of using GLM in insurance pricing

While Generalized Linear Models have revolutionized insurance pricing with their flexibility and predictive power, their application is not without challenges and limitations. Understanding these potential hurdles is crucial for actuaries and data scientists to effectively navigate and mitigate issues that may arise during model development and implementation. This section explores the most significant challenges and limitations associated with using GLM in insurance pricing.

The accuracy and reliability of GLM outcomes are highly dependent on the quality and availability of data. Insufficient, inaccurate, or biased data can result in misleading conclusions and potentially unfair pricing.

Limited historical data, especially for new products or coverage areas, can hinder the ability to train and validate the model effectively. Data quality issues, such as incorrect or incomplete data, can distort model predictions, leading to inaccurate premium calculations and risk assessments. Ensuring robust and reliable data is crucial for the fairness and accuracy of GLM-based pricing decisions.

The complexity of GLMs can present challenges in model development, interpretation, and communication with stakeholders.

Model overfitting is a common risk, where incorporating too many variables or complex interactions may lead to a model that performs well on training data but fails on new, unseen data. Additionally, the intricate nature of GLMs, especially with numerous variables and interactions, can make it challenging for non-technical stakeholders to understand and accept the model's findings. Clear communication and careful model simplification are key to addressing these issues.

The application of Generalized Linear Models in insurance pricing must carefully address regulatory requirements and ethical concerns, particularly regarding data privacy and the risk of discrimination.

Compliance with local and international regulations governing data use, privacy, and fairness in pricing is essential. Additionally, there is a risk that GLMs might inadvertently include variables that act as proxies for prohibited factors like race or gender, leading to discriminatory pricing practices. Ensuring that models are both compliant and ethically sound is critical to maintaining fairness and avoiding bias.

Generalized Linear Models, like all statistical models, rely on a set of assumptions about the data and the relationships between variables. Incorrect assumptions can compromise the validity of the model.

The selection of the appropriate distribution for the response variable and the link function is crucial; if these choices do not accurately reflect the underlying data, the model's performance can suffer. While GLMs provide significant advantages in insurance pricing, successfully navigating their complexities requires diligent data management, thorough model testing, and strict adherence to ethical and regulatory standards. By recognizing and addressing these limitations, insurers can harness GLMs to create fair, accurate, and efficient pricing models.

III.2.7. Integrating Patient Activation Measure (PAM) into GLM for health insurance modeling

In this section, I construct five Generalized Linear Models (GLM) to predict medical expenses, comparing their outcomes to select the most effective model. The primary objective is to design a sickness insurance product by incorporating the validated Patient Activation Measure (PAM) data from Part II into the insurance framework established in Part I. The analysis primarily uses data collected before the Covid-19 pandemic, intentionally excluding pandemic years because of their unusual characteristics. However, available data from the pre-pandemic period were relatively limited. Medical costs were estimated using survey responses, focusing on reported average healthcare expenditures and out-of-pocket expenses. To enhance precision, these estimates were further verified with official healthcare expenditure data obtained from NEAK's 2019 records, including general practitioner visits and hospital services. Medical costs are calculated as follows:

$$cost = gp \cdot avg\ gp\ cost + spec \cdot avg\ spec\ cost + hosp \cdot avg\ length \cdot daily\ cost + oop + avg\ med\ cost \cdot med - avg\ med\ oop\ cost.$$

Where

$$avg\ gp\ cost = 1276\ HUF$$
$$avg\ spec\ cost = 2006\ HUF$$
$$avg\ lenght = 7.9\ day$$
$$daily\ cost = 110100\ HUF$$
$$avg\ med\ cost = 13695\ HUF$$
$$avg\ med\ oop\ cost = 12897\ HUF$$

Here, 'gp' represents the number of GP visits, 'spec' stands for specialist visits, 'hosp' denotes the number of hospitalizations, 'oop' is out-of-pocket payments, and 'med' refers to money spent on prescription medicine.

The PAM database contains 779 observations and 147 variables. Overall, I use 11 variables in the 5 GLM models. I chose these variables because many of them are commonly used in other GLM models, such as gender, age, income, and territorial data. The others are risk indicators, and it is interesting to see whether they can be used in a

risk model. Also, note that territorial variables are not commonly used in health and life insurances in Hungary. However, I will demonstrate later on that these variables were significant in my data. The list of predictors is as follows:

- Gender (sex)
- Age group (agegr)
- Family status (fam)
- Education group (edugr)
- Income (inc)
- Type of settlement (stype)
- Region (rg)
- EQ-5D-5L index
- eHEALS quartiles (ehgr)
- Health literacy level (nvsgr)
- PAM levels (pamgr)

Of the 11 variables, 9 are categorical. Table 13 summarizes all categories of these detailing the counts and percentages for each. For the PAM group, I combined levels 2 and 3 due to their almost identical coefficients in the models, creating a new variable named 'pamgr2'. As mentioned in the data preparation section, I randomly split the database into 80% for training and 20% for testing. Table 13 displays the frequency distributions of the categorical predictor variables in the training dataset.

Table 13 Training sample distribution

		N	Percent
sex	1 (Female)	343	54,4%
	2 (Male)	287	45,6%
	Total	630	100,0%
agegr	4 (70+ years old)	120	19,0%
	3 (60-69 years old)	242	38,4%
	2 (50-59 years old)	145	23,0%
	1 (40-49 years old)	123	19,5%
	Total	630	100,0%
fam	6 (Other)	4	0,6%
	5 (Widowed)	68	10,8%

	4 (Divorced)	87	13,8%
	3 (Domestic partnership)	77	12,2%
	2 (Married)	307	48,7%
	1 (Single)	87	13,8%
	Total	630	100,0%
edugr	3 (Tertiary)	236	37,5%
	2 (Secondary)	234	37,1%
	1 (Primary)	160	25,4%
	Total	630	100,0%
rg	7 (Dél-Dunántúl)	70	11,1%
	6 (Nyugat-Dunántúl)	60	9,5%
	5 (Közép-Dunántúl)	66	10,5%
	4 (Dél-Alföld)	69	11,0%
	3 (Észak-Alföld)	68	10,8%
	2 (Észak-Magyarország)	67	10,6%
	1 (Közép-Magyarország)	230	36,5%
	Total	630	100,0%
ehgr	4 (4th quartile)	174	27,6%
	3 (3rd quartile)	164	26,0%
	2 (2nd quartile)	158	25,1%
	1 (1st quartile)	134	21,3%
	Total	630	100,0%
nvsgr	3 (Adequate)	399	63,3%
	2 (Possibly limited)	156	24,8%
	1 (Probably limited)	75	11,9%
	Total	630	100,0%
pamgr2	4 (High PAM score)	69	11,0%
	2 &3 (Medium PAM score)	458	72,7%
	1 (Low PAM score)	103	16,3%
	Total	630	100,0%
stype	3 (Village)	118	18,7%
	2 (Town)	359	57,0%
	1 (Capital)	153	24,3%
	Total	630	100,0%

Source: own ed.

Notes: This table presents the distribution of the training sample. "Sex" refers to gender, "agegr" indicates age group, "fam" represents family status, "edugr" stands for education group, "stype" means type of settlement, "rg" represents region, "ehgr" indicates eHEALS quartiles, "nvsgr" represents health literacy level, and "pamgr2" refers to PAM levels.

Table 14 presents information on continuous variables. It includes two predictor variables as well as the target variable, cost.

Table 14 Training sample continuous variables properties

Continuous Variable Information

		N	Minimum	Maximum	Mean
Dependent Variable	cost	630	1	8850932	121302,93
Covariate	i5	630	-0,482	1	0,854
	inc	630	0	614110	127460

Source: own ed.

Notes: This table presents the continuous variables properties of the training sample. "cost" refers to Medical costs, "inc" refers to income and "i5" denotes the EQ-5D-5L index

For weights, I use the 'wgt' variable, which represents the population weight of the person listed in that row of the table.

In order to fully present the sample of 779 respondents, here are the test sample characteristics:

Table 15 Test sample distribution

		N	Percent
sex	1 (Female)	78	52,35%
	2 (Male)	71	47,65%
	Total	149	100,00%
agegr	4 (70+ years old)	33	22,15%
	3 (60-69 years old)	64	42,95%
	2 (50-59 years old)	32	21,48%
	1 (40-49 years old)	20	13,42%
	Total	149	100,00%
fam	6 (Other)	0	0,00%
	5 (Widowed)	14	9,40%
	4 (Divorced)	24	16,11%
	3 (Domestic partnership)	17	11,41%
	2 (Married)	77	51,68%
	1 (Single)	17	11,41%
	Total	149	100,00%
edugr	3 (Tertiary)	52	34,90%

	2 (Secondary)	54	36,24%
	1 (Primary)	43	28,86%
	Total	149	100,00%
rg	7 (Dél-Dunántúl)	19	12,75%
	6 (Nyugat-Dunántúl)	9	6,04%
	5 (Közép-Dunántúl)	17	11,41%
	4 (Dél-Alföld)	29	19,46%
	3 (Észak-Alföld)	17	11,41%
	2 (Észak-Magyarország)	12	8,05%
	1 (Közép-Magyarország)	46	30,87%
	Total	149	100,00%
ehgr	4 (4th quartile)	54	36,24%
	3 (3rd quartile)	35	23,49%
	2 (2nd quartile)	37	24,83%
	1 (1st quartile)	23	15,44%
	Total	149	100,00%
nvsgr	3 (Adequate)	92	61,74%
	2 (Possibly limited)	41	27,52%
	1 (Probably limited)	16	10,74%
	Total	149	100,00%
pamgr2	4 (High PAM score)	15	10,07%
	2 &3 (Medium PAM score)	113	75,84%
	1 (Low PAM score)	21	14,09%
	Total	149	100,00%
stype	3 (Village)	33	22,15%
	2 (Town)	88	59,06%
	1 (Capital)	28	18,79%
	Total	149	100,00%

Source: own ed.

Notes: This table presents the distribution of the test sample. "Sex" refers to gender, "agegr" indicates age group, "fam" represents family status, "edugr" stands for education group, "stype" means type of settlement, "rg" represents region, "ehgr" indicates eHEALS quartiles, "nvsgr" represents health literacy level, and "pamgr2" refers to PAM levels.

Table 16 Test sample continuous variables properties

Continuous Variable Information

		N	Minimum	Maximum	Mean
Dependent Variable	cost	149	1	1807736	87104
Covariate	i5	149	-0,329	1	0,830
	inc	149	0	325078	121174,2

Source: own ed.

Notes: This table presents the continuous variables properties of the test sample. "cost" refers to Medical costs, "inc" refers to income and "i5" denotes the EQ-5D-5L index

III.2.7.1 Description of the five GLM models

In this subsection, I describe the 5 GLM models created using the IBM SPSS program. IBM SPSS Statistics is a statistical software platform that offers advanced features for data analysis, including Generalized Linear Models (GLM).

Using the GLM in IBM SPSS, you can specify the distribution of their dependent variable (e.g., normal, binomial, Poisson) and choose an appropriate link function to model the relationship between the dependent variable and one or more independent variables. SPSS allows for the analysis of both continuous and categorical data, accommodating scenarios such as predicting outcomes (regression), analyzing count data, or modeling binary outcomes (logistic regression).

III.2.7.1.1 Model #1: The outlier-filtered version

In this version, I first calculated the Cook's Distance in the model using SPSS. After obtaining Cook's Distance in the training dataset, I filtered out the cases where Cook's Distance exceeded the threshold of 4/n *(Kovács, 2011).*

The dependent variable in our case is medical costs in all 5 models. The built model's distribution is Gamma with a Log link function. The predictors used are: gender, age group, family status, education group, region, income, eHEALS group, NVS group, EQ-5D-5L, and PAM group 2 (Table 18 below).

Incomplete data can compromise the model's integrity. SPSS provides mechanisms to

handle missing values effectively, ensuring that analyses are based on complete and accurate information. In my case, no data was missing.

Table 17 Model #1 missing data summary

Case Processing Summary

	N	Percent
Included	598	100,0%
Excluded	0	0,0%
Total	598	100,0%

Source: own ed.

This model incorporates an additional filter compared to the other four, resulting in different distributions among the categorical and continuous variables. The tables below illustrate this case, while the others are as previously shown.

Table 18 Model 1# data distribution

		N	Percent
sex	1 (Female)	270	45,2%
	2 (Male)	328	54,8%
	Total	598	100,0%
agegr	1 (40-49 years old)	116	19,4%
	2 (50-59 years old)	137	22,9%
	3 (60-69 years old)	230	38,5%
	4 (70+ years old)	115	19,2%
	Total	598	100,0%
fam	1 (Single)	84	14,0%
	2 (Married)	291	48,7%
	3 (Domestic partnership)	72	12,0%
	4 (Divorced)	82	13,7%
	5 (Widowed)	67	11,2%
	6 (Other)	2	0,3%
	Total	598	100,0%
edugr	1 (Primary)	150	25,1%
	2 (Secondary)	225	37,6%
	3 (Tertiary)	223	37,3%
	Total	598	100,0%

rg	1 (Közép-Magyarország)	220	36,8%
	2 (Észak-Magyarország)	63	10,5%
	3 (Észak-Alföld)	65	10,9%
	4 (Dél-Alföld)	66	11,0%
	5 (Közép-Dunántúl)	62	10,4%
	6 (Nyugat-Dunántúl)	54	9,0%
	7 (Dél-Dunántúl)	68	11,4%
	Total	598	100,0%
ehgr	1 (1st quartile)	127	21,2%
	2 (2nd quartile)	151	25,3%
	3 (3rd quartile)	155	25,9%
	4 (4th quartile)	165	27,6%
	Total	598	100,0%
nvsgr	1 (Probably limited)	69	11,5%
	2 (Possibly limited)	149	24,9%
	3 (Adequate)	380	63,5%
	Total	598	100,0%
pamgr2	1 (Low PAM score)	96	16,1%
	2&3 (Medium PAM score)	437	73,1%
	4 (High PAM score)	65	10,9%
	Total	598	100,0%

Source: own ed.

Notes: This table presents the distribution of the training sample. "Sex" refers to gender, "agegr" indicates age group, "fam" represents family status, "edugr" stands for education group, "rg" represents region, "ehgr" indicates eHEALS quartiles, "nvsgr" represents health literacy level, and "pamgr2" refers to PAM levels.

Table 19 Model #1 Continuous variable properties

Continuous Variable Information

		N	Minimum	Maximum	Mean
Dependent Variable	cost	598	1	2685808	54533,94
Covariate	i5	598	-0,482	1	0,855
	inc	598	0	1857	385,902

Source: own ed.

Notes: This table presents the continuous variables properties of the training sample. "cost" refers to Medical costs, "inc" refers to income and "i5" denotes the EQ-5D-5L index

I employed the Omnibus test to assess the significance of my GLM model, ensuring its overall explanatory power and validity. Specifically, in the context of regression analysis, it tests the null hypothesis that all the regression coefficients in the model are equal to zero, implying that none of the independent variables in the model have a significant impact on the dependent variable.

Table 20 Model #1 Omnibus test

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
125,137	26	,000

Source: own ed.

Notes: a. Compares the fitted model against the intercept-only model. Where "df" represents degrees of freedom and "Sig." provides the p-value for the omnibus test. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

In the analysis of Generalized Linear Models (GLM) within SPSS, the "Tests of Model Effects" plays a crucial role in evaluating the significance of individual variables. This feature is specifically designed to provide insights into how each predictor contributes to the model, offering a detailed examination beyond the overall model fit assessed by omnibus tests.

The results are shown in Table 21.

Table 21 Model #1 Tests of Model Effects

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	1381,810	1	,000
gender	,485	1	,486
age group	9,764	3	,021
family status	10,199	5	,070

education group	,891	2	,641
region	4,591	6	,597
e-heals group	1,160	3	,763
nvs group	1,173	2	,556
pam group	2,289	2	,318
EQ-5D index	44,393	1	,000
income	,381	1	,537

Source: own ed.

Notes: Where "df" represents degrees of freedom and "Sig." provides the p-value for the model effects tests. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

Based on Table 21 only the age group, family status, and EQ-5D-5L are significant. In this study, my goal is to demonstrate the effect of the PAM group. Therefore, I will focus primarily on that variable. Table 22 shows the parameter estimates.

Table 22 Model #1 Parameter Estimates

Parameter Estimates

Variable	Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		Exp(B)
				Lower	Upper	Wald Chi-Square	Sig.	
	(Intercept)	11,713	0,4103	10,909	12,518	814,91	0	122190
sex	1 (Female)	-0,088	0,1261	-0,335	0,159	0,485	0,486	0,916
	2 (Male)	0[a]	.	.	.	.	.	1
agegr	4 (70+ years old)	0,65	0,2128	0,233	1,067	9,333	0,002	1,916
	3 (60-69 years old)	0,384	0,1675	0,055	0,712	5,244	0,022	1,468
	2 (50-59 years old)	0,268	0,1807	-0,086	0,622	2,202	0,138	1,307
	1 (40-49 years old)	0[a]	.	.	.	.	.	1
fam	6 (Other)	0,157	1,0109	-1,824	2,139	0,024	0,876	1,17
	5 (Widowed)	0,284	0,2502	-0,207	0,774	1,287	0,257	1,328
	4 (Divorced)	0,546	0,2263	0,102	0,989	5,816	0,016	1,726
	3 (Domestic partnership)	0,059	0,2306	-0,393	0,511	0,066	0,797	1,061
	2 (Married)	0,429	0,1818	0,073	0,785	5,577	0,018	1,536
	1 (Single)	0[a]	.	.	.	.	.	1
edugr	3 (Tertiary)	0,081	0,1717	-0,256	0,417	0,221	0,639	1,084
	2 (Secondary)	-0,05	0,1579	-0,359	0,26	0,099	0,752	0,951

	1 (Primary)	0[a]	.	.	.	.	.	1
rg	7 (Dél-Dunántúl)	-0,017	0,1996	-0,408	0,375	0,007	0,934	0,984
	6 (Nyugat-Dunántúl)	0,067	0,2185	-0,361	0,495	0,094	0,759	1,069
	5 (Közép-Dunántúl)	-0,079	0,2124	-0,496	0,337	0,14	0,708	0,924
	4 (Dél-Alföld)	0,376	0,2099	-0,035	0,787	3,209	0,073	1,456
	3 (Észak-Alföld)	-0,079	0,2034	-0,477	0,32	0,15	0,698	0,924
	2 (Észak-Magyarország)	0,08	0,2123	-0,336	0,497	0,143	0,705	1,084
	1 (Közép-Magyarország)	0[a]	.	.	.	.	.	1
ehgr	4 (4th quartile)	0,135	0,1875	-0,233	0,502	0,516	0,472	1,144
	3 (3rd quartile)	0,086	0,1787	-0,264	0,437	0,234	0,629	1,09
	2 (2nd quartile)	-0,035	0,1724	-0,373	0,303	0,041	0,839	0,966
	1 (1st quartile)	0[a]	.	.	.	.	.	1
nvsgr	3 (Adequate)	0,21	0,1955	-0,174	0,593	1,15	0,284	1,233
	2 (Possibly limited)	0,153	0,2142	-0,266	0,573	0,514	0,474	1,166
	1 (Probably limited)	0[a]	.	.	.	.	.	1
pamgr2	4 (High PAM score)	-0,352	0,2517	-0,845	0,141	1,956	0,162	0,703
	2 &3 (Medium PAM score)	-0,223	0,169	-0,554	0,108	1,739	0,187	0,8
	1 (Low PAM score)	0[a]	.	.	.	.	.	1
EQ-5D index	i5	-2,046	0,3071	-2,648	-1,444	44,393	0	0,129
Income	inc	0	0,0002	0	0,001	0,381	0,537	1
	(Scale)	1,902[b]	0,0912	1,731	2,089			

Source: own ed.

Notes: Where "B" represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable, while holding all other variables constant. "Std. Error" refers to the Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard deviation of the sampling distribution of the coefficient. "Sig." stands for Significance or p-value, and it indicates the probability that the observed relationship between the independent and dependent variables occurred by chance. "Exp(B)" denotes the Exponentiated Coefficient.

III.2.7.1.2 Model #2: without outlier filter and with region

The database contains only 779 responses, and for a GLM, a larger database typically results in stronger models. The aim is to demonstrate how GLM functions in insurance

pricing and to assess whether PAM can be used as a predictor. Naturally, with 10,000 responses, I could build a model with even stronger explanatory power and a better level of significance. What we observe is that filtering out more data from the training database results in fewer significant predictors. For this reason, in subsequent models, I will not filter out the outliers but rather keep them in the model to see how it affects the results.

The dependent variable in our case is medical costs in all 5 models. The built model's distribution is Gamma with a Log link function. The predictors used are: gender, age group, family status, education group, region, income, eHEALS group, NVS group, EQ-5D-5L, and PAM group 2 (see the table 13 above).

The handling of missing data is the same as in model #1. In my case, no data was missing. Table 23 is the same for the remaining models, so I will not include it in the document later on.

Table 23 Model #2-5 Missing data summary

Case Processing Summary

	N	Percent
Included	630	100,0%
Excluded	0	0,0%
Total	630	100,0%

Source: own ed.

As mentioned before, the distribution of the categorical and continuous variables is the same as what was described earlier at the beginning of this section.

The result of the Omnibus test is as follows:

Table 24 Model #2 Omnibus test

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
163,109	26	,000

Source: own ed.

Notes: a. Compares the fitted model against the intercept-only model. Where "df" represents degrees of freedom and "Sig." provides the p-value for the omnibus test. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

Table 25 displays the significance levels of the variables.

Table 25 Model #2 Tests of Model Effects

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	1303,221	1	,000
sex	,778	1	,378
agegr	,905	3	,824
fam	12,059	5	,034
edugr	18,086	2	,000
rg	35,866	6	,000
ehgr	8,589	3	,035
nvsgr	1,997	2	,368
pamgr2	20,122	2	,000
i5	23,521	1	,000
inc	1,255	1	,263

Source: own ed.

Notes: Where "df" represents degrees of freedom and "Sig." provides the p-value for the model effects tests. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

The parameter estimates are as follows:

Table 26 Model #2 Parameter Estimates

Parameter Estimates

Variable	Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		Exp(B)
				Lower	Upper	Wald Chi-Square	Sig.	
	(Intercept)	12,572	0,4621	11,666	13,477	740,281	0	288299,2
sex	1 (Female)	-0,128	0,1451	-0,412	0,156	0,778	0,378	0,88
	2 (Male)	0[a]	.	.	.	.	.	1
agegr	4 (70+ years old)	0,071	0,2546	-0,428	0,57	0,078	0,78	1,074
	3 (60-69 years old)	0,023	0,2184	-0,405	0,451	0,011	0,916	1,023
	2 (50-59 years old)	-0,124	0,2302	-0,576	0,327	0,292	0,589	0,883
	1 (40-49 years old)	0[a]	.	.	.	.	.	1
fam	6 (Other)	-0,317	0,8416	-1,967	1,332	0,142	0,706	0,728
	5 (Widowed)	0,142	0,3177	-0,48	0,765	0,201	0,654	1,153
	4 (Divorced)	0,666	0,2877	0,102	1,23	5,363	0,021	1,947
	3 (Domestic partnership)	-0,086	0,2755	-0,626	0,454	0,097	0,755	0,918
	2 (Married)	0,435	0,2406	-0,037	0,906	3,27	0,071	1,545
	1 (Single)	0[a]	.	.	.	.	.	1
edugr	3 (Tertiary)	-0,125	0,1983	-0,513	0,264	0,395	0,53	0,883
	2 (Secondary)	-0,681	0,1855	-1,044	-0,317	13,467	0	0,506
	1 (Primary)	0[a]	.	.	.	.	.	1
rg	7 (Dél-Dunántúl)	-0,12	0,233	-0,577	0,336	0,267	0,605	0,886
	6 (Nyugat-Dunántúl)	0,944	0,2556	0,443	1,445	13,64	0	2,57
	5 (Közép-Dunántúl)	0,945	0,2476	0,46	1,43	14,571	0	2,573
	4 (Dél-Alföld)	0,538	0,2403	0,067	1,009	5,019	0,025	1,713

	3 (Észak-Alföld)	-0,024	0,2443	-0,503	0,454	0,01	0,92	0,976
	2 (Észak-Magyarország)	0,721	0,2532	0,225	1,217	8,11	0,004	2,057
	1 (Közép-Magyarország)	0[a]	.	.	.	.	.	1
ehgr	4 (4th quartile)	0,587	0,2238	0,148	1,026	6,874	0,009	1,798
	3 (3rd quartile)	0,261	0,2065	-0,144	0,666	1,598	0,206	1,298
	2 (2nd quartile)	0,095	0,2085	-0,314	0,504	0,208	0,649	1,1
	1 (1st quartile)	0[a]	.	.	.	.	.	1
nvsgr	3 (Adequate)	0,304	0,2209	-0,129	0,737	1,89	0,169	1,355
	2 (Possibly limited)	0,301	0,2439	-0,176	0,779	1,528	0,216	1,352
	1 (Probably limited)	0[a]	.	.	.	.	.	1
pamgr2	4 (High PAM score)	-1,288	0,2925	-1,862	-0,715	19,395	0	0,276
	2 &3 (Medium PAM score)	-0,679	0,197	-1,065	-0,293	11,885	0,001	0,507
	1 (Low PAM score)	0[a]	.	.	.	.	.	1
EQ-5D index	i5	-1,722	0,355	-2,418	-1,026	23,521	0	0,179
Income	inc	0	0,0003	0	0,001	1,255	0,263	1
	(Scale)	2,515[b]	0,1146	2,3	2,75			

Source: own ed.

Notes: Where "B" represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable, while holding all other variables constant. "Std. Error" refers to the Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard deviation of the sampling distribution of the coefficient. "Sig." stands for Significance or p-value, and it indicates the probability that the observed relationship between the independent and dependent variables occurred by chance. "Exp(B)" denotes the Exponentiated Coefficient.

III.2.7.1.3 Model #3: without outlier filter and with type of settlement

In this model, the most significant change is the replacement of the Region variable with the Type of Settlement variable. The dependent variable in our case is medical costs in all 5 models. The built model's distribution is Gamma with a Log link function. The predictors used are: gender, age group, family status, education group, type of settlement, income, eHEALS group, NVS group, EQ-5D-5L, and PAM group 2 (see Table 13 above).

The result of the Omnibus test is as follows:

Table 27 Model #3 Omnibus Test

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
133,399	22	,000

Source: own ed.

Notes: a. Compares the fitted model against the intercept-only model. Where "df" represents degrees of freedom and "Sig." provides the p-value for the omnibus test. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

Table 28 displays the significance levels of the variables.

Table 28 Model #3 Tests of model Effects

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	1197,893	1	,000
fam	9,938	5	,077
edugr	18,756	2	,000
ehgr	10,066	3	,018
pamgr2	22,831	2	,000
i5	29,267	1	,000
sex	2,454	1	,117

agegr	1,454	3	,693
nvsgr	2,696	2	,260
inc	,208	1	,649
stype	6,270	2	,043

Source: own ed.

Notes: Where "df" represents degrees of freedom and "Sig." provides the p-value for the model effects tests. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

The parameter estimates are as follows:

Table 29 Model #3 Parameter Estimates

Parameter Estimates

Variable	Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		Exp(B)
				Lower	Upper	Wald Chi-Square	Sig.	
	(Intercept)	13,01	0,4776	12,074	13,946	742,004	0	446877
fam	6 (Other)	-0,447	0,8545	-2,121	1,228	0,273	0,601	0,64
	5 (Widowed)	0,172	0,3337	-0,482	0,826	0,265	0,607	1,187
	4 (Divorced)	0,652	0,2892	0,085	1,219	5,076	0,024	1,919
	3 (Domestic partnership)	-0,086	0,2795	-0,634	0,461	0,096	0,757	0,917
	2 (Married)	0,342	0,2426	-0,134	0,817	1,986	0,159	1,408
	1 (Single)	0[a]	.	.	.	.	.	1
edugr	3 (Tertiary)	-0,137	0,1919	-0,513	0,239	0,512	0,474	0,872
	2 (Secondary)	-0,719	0,1869	-1,085	-0,352	14,786	0	0,487
	1 (Primary)	0[a]	.	.	.	.	.	1
ehgr	4 (4th quartile)	0,638	0,2207	0,205	1,07	8,353	0,004	1,892
	3 (3rd quartile)	0,328	0,2027	-0,069	0,725	2,619	0,106	1,388

	2 (2nd quartile)	0,145	0,2078	-0,263	0,552	0,484	0,487	1,155
	1 (1st quartile)	0[a]	.	.	.	.	.	1
pamgr2	4 (High PAM score)	-1,357	0,2844	-1,915	-0,8	22,775	0	0,257
	2 &3 (Medium PAM score)	-0,633	0,1984	-1,022	-0,244	10,188	0,001	0,531
	1 (Low PAM score)	0[a]	.	.	.	.	.	1
EQ-5D index	i5	-1,974	0,3649	-2,689	-1,259	29,267	0	0,139
sex	1 (Female)	-0,239	0,1525	-0,538	0,06	2,454	0,117	0,788
	2 (Male)	0[a]	.	.	.	.	.	1
agegr	4 (70+ years old)	0,232	0,2575	-0,272	0,737	0,813	0,367	1,261
	3 (60-69 years old)	0,072	0,227	-0,373	0,516	0,099	0,753	1,074
	2 (50-59 years old)	-0,028	0,2363	-0,491	0,436	0,014	0,907	0,973
	1 (40-49 years old)	0[a]	.	.	.	.	.	1
nvsgr	3 (Adequate)	0,308	0,2275	-0,138	0,754	1,83	0,176	1,36
	2 (Possibly limited)	0,105	0,2503	-0,385	0,596	0,177	0,674	1,111
	1 (Probably limited)	0[a]	.	.	.	.	.	1
Income	inc	0	0,0003	0	0,001	0,208	0,649	1
stype	3 (Village)	0,552	0,2205	0,119	0,984	6,256	0,012	1,736
	2 (Town)	0,254	0,1775	-0,094	0,602	2,043	0,153	1,289
	1 (Capital)	0[a]	.	.	.	.	.	1
	(Scale)	2,593[b]	0,1178	2,372	2,835			

Source: own ed.

Notes: Where "B" represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable,

while holding all other variables constant. "Std. Error" refers to the Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard deviation of the sampling distribution of the coefficient. "Sig." stands for Significance or p-value, and it indicates the probability that the observed relationship between the independent and dependent variables occurred by chance. "Exp(B)" denotes the Exponentiated Coefficient.

III.2.7.1.4 Model #4: without outlier filter, including region, and with reduced variables

This model is based on model #2, but it excludes the non-significant variables. These variables were: Gender, Age group, Income, and NVS group. The dependent variable in our case is medical costs in all 5 models. The built model's distribution is Gamma with a Log link function. The predictors used are: family status, education group, region, eHEALS group, EQ-5D-5L, and PAM group 2 (see the table 13 above).

The result of the Omnibus test is as follows:

Table 30 Model #4 Omnibus test

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
157,761	19	,000

Source: own ed.

Notes: a. Compares the fitted model against the intercept-only model. Where "df" represents degrees of freedom and "Sig." provides the p-value for the omnibus test. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

Table 31 displays the significance levels of the variables.

Table 31 Model #4 Tests of Model Effects

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	1407,182	1	,000
fam	16,181	5	,006
edugr	18,013	2	,000
rg	39,563	6	,000
ehgr	10,459	3	,015
pamgr2	22,089	2	,000
i5	21,950	1	,000

Source: own ed.

Notes: Where "df" represents degrees of freedom and "Sig." provides the p-value for the model effects tests. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

The parameter estimates are as follows:

Table 32 Model #4 Parameter Estimates

Parameter Estimates

Variable	Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	Hypothesis Test	Exp(B)
				Lower	Upper	Wald Chi-Square	Sig.	
	(Intercept)	12,887	0,4066	12,09	13,684	1004,513	0	395169,2
fam	6 (Other)	-0,442	0,8249	-2,058	1,175	0,287	0,592	0,643
	5 (Widowed)	0,153	0,2822	-0,4	0,706	0,293	0,588	1,165
	4 (Divorced)	0,684	0,2676	0,16	1,209	6,538	0,011	1,982
	3 (Domestic partnership)	-0,171	0,2652	-0,691	0,349	0,415	0,519	0,843
	2 (Married)	0,425	0,2176	-0,002	0,851	3,81	0,051	1,529
	1 (Single)	0[a]	.	.	.	.	.	1
edugr	3 (Tertiary)	-0,019	0,1844	-0,38	0,343	0,01	0,919	0,981

	2 (Secondary)	-0,618	0,1799	-0,97	-0,265	11,782	0,001	0,539
	1 (Primary)	0[a]	.	.	.	.	.	1
rg	7 (Dél-Dunántúl)	-0,164	0,2266	-0,608	0,28	0,525	0,469	0,849
	6 (Nyugat-Dunántúl)	0,89	0,2474	0,405	1,374	12,931	0	2,434
	5 (Közép-Dunántúl)	0,944	0,2323	0,489	1,399	16,51	0	2,57
	4 (Dél-Alföld)	0,481	0,2352	0,02	0,942	4,191	0,041	1,618
	3 (Észak-Alföld)	-0,103	0,2363	-0,566	0,361	0,189	0,664	0,902
	2 (Észak-Magyarország)	0,661	0,2454	0,18	1,142	7,263	0,007	1,937
	1 (Közép-Magyarország)	0[a]	.	.	.	.	.	1
ehgr	4 (4th quartile)	0,597	0,2195	0,167	1,027	7,401	0,007	1,817
	3 (3rd quartile)	0,22	0,2023	-0,177	0,617	1,182	0,277	1,246
	2 (2nd quartile)	0,026	0,2024	-0,371	0,423	0,017	0,898	1,026
	1 (1st quartile)	0[a]	.	.	.	.	.	1
pamgr2	4 (High PAM score)	-1,323	0,2892	-1,89	-0,756	20,921	0	0,266
	2 &3 (Medium PAM score)	-0,717	0,1933	-1,095	-0,338	13,743	0	0,488
	1 (Low PAM score)	0[a]	.	.	.	.	.	1
EQ-5D index	i5	-1,666	0,3556	-2,363	-0,969	21,95	0	0,189
	(Scale)	2,529[b]	0,1151	2,313	2,765			

Source: own ed.

Notes: Where "B" represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable, while holding all other variables constant. "Std. Error" refers to the Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard

deviation of the sampling distribution of the coefficient. "Sig." stands for Significance or p-value, and it indicates the probability that the observed relationship between the independent and dependent variables occurred by chance. "Exp(B)" denotes the Exponentiated Coefficient.

III.2.7.1.5 Model #5: without outlier filter, including type of settlement, and with reduced variables

This model builds on model #3, excluding the non-significant variables: Gender, Age group, Income, and NVS group. Additionally, it incorporates Age as a continuous variable. The dependent variable in our case is medical costs in all 5 models. The built model's distribution is Gamma with a Log link function. The predictors used are: age, family status, education group, type of settlement, eHEALS group, EQ-5D-5L, and PAM group 2 (see the table 13 above).

The result of the Omnibus test is as follows:

Table 33 Model #5 Omnibus test

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
129,871	16	,000

Source: own ed.

Notes: a. Compares the fitted model against the intercept-only model. Where "df" represents degrees of freedom and "Sig." provides the p-value for the omnibus test. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

Table 34 displays the significance levels of the variables.

Table 34 Model #5 Tests of Model Effects

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	457,033	1	,000
fam	9,677	5	,085
edugr	19,744	2	,000
ehgr	11,493	3	,009
pamgr2	26,673	2	,000
i5	28,883	1	,000
stype	9,718	2	,008
age	4,447	1	,035

Source: own ed.

Notes: Where "df" represents degrees of freedom and "Sig." provides the p-value for the model effects tests. If this p-value is below your chosen significance level (e.g., 0.05), you can reject the null hypothesis, indicating that the model explains a significant amount of variance in the dependent variable(s).

The parameter estimates are as follows:

Table 35 Model #5 Parameter estimates

Parameter Estimates

Variable	Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	Hypothesis Test	Exp(B)
				Lower	Upper	Wald Chi-Square	Sig.	
	(Intercept)	12,402	0,5441	11,336	13,468	519,494	0	243281,3
fam	6 (Other)	-0,825	0,847	-2,485	0,835	0,948	0,33	0,438
	5 (Widowed)	-0,22	0,3313	-0,87	0,429	0,442	0,506	0,802
	4 (Divorced)	0,412	0,2837	-0,144	0,968	2,109	0,146	1,51
	3 (Domestic partnership)	-0,255	0,2676	-0,78	0,269	0,912	0,34	0,775
	2 (Married)	0,091	0,2458	-0,391	0,573	0,137	0,711	1,095
	1 (Single)	0[a]	.	.	.	.	.	1
edugr	3 (Tertiary)	0,001	0,1786	-0,349	0,351	0	0,997	1,001

	2 (Secondary)	-0,654	0,1805	-1,007	-0,3	13,113	0	0,52
	1 (Primary)	0[a]	.	.	.	.	.	1
ehgr	4 (4th quartile)	0,603	0,2156	0,18	1,025	7,817	0,005	1,827
	3 (3rd quartile)	0,248	0,1979	-0,14	0,636	1,573	0,21	1,282
	2 (2nd quartile)	0,016	0,1995	-0,375	0,407	0,007	0,936	1,016
	1 (1st quartile)	0[a]	.	.	.	.	.	1
pamgr2	4 (High PAM score)	-1,429	0,2768	-1,972	-0,887	26,664	0	0,239
	2 &3 (Medium PAM score)	-0,632	0,1932	-1,011	-0,253	10,693	0,001	0,532
	1 (Low PAM score)	0[a]	.	.	.	.	.	1
EQ-5D index	i5	-1,986	0,3696	-2,711	-1,262	28,883	0	0,137
stype	3 (Village)	0,643	0,2086	0,234	1,052	9,496	0,002	1,902
	2 (Town)	0,322	0,1633	0,002	0,642	3,888	0,049	1,38
	1 (Capital)	0[a]	.	.	.	.	.	1
	age	0,017	0,0078	0,001	0,032	4,447	0,035	1,017
	(Scale)	2,603[b]	0,1182	2,381	2,845			

Source: own ed.

Notes: Where "B" represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable, while holding all other variables constant. "Std. Error" refers to the Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard deviation of the sampling distribution of the coefficient. "Sig." stands for Significance or p-value, and it indicates the probability that the observed relationship between the independent and dependent variables occurred by chance. "Exp(B)" denotes the Exponentiated Coefficient.

III.2.7.2 Reliability and Validity testing

To ensure the robustness and validity of the findings derived from the five GLM models previously built using SPSS (with a Gamma distribution), additional modeling was conducted using Emblem—an industry-standard actuarial software widely used in insurance pricing. The objective of this step was to replicate and validate the core results in a different modeling environment, using both the Gamma and Tweedie distributions, which are commonly applied to skewed, positive-valued cost data.

While Emblem offers a user-friendly interface and efficient implementation for insurance-related modeling, it is important to acknowledge a key limitation: the software can only handle a maximum of 255 unique categorical and numerical levels per variable. As a result, the dataset required preprocessing before modeling. Specifically, the income variable—which originally had more than 255 unique values—had to be banded into broader income categories. This manual categorization ensured that Emblem could process the variable without losing relevant information. Due to a similar issue, the EQ-5D health-related quality of life variable, which also exceeded the category limit, was excluded from the Emblem-based models.

Further variable exclusions were based on statistical and structural considerations. The territorial information (region or location data) was excluded due to its high standard errors, which were likely caused by the relatively limited sample size (N = 779) and the even smaller number of observations within each territorial category. Similarly, the education variable did not yield stable estimates and was also omitted. The family status variable was removed due to its strong correlation with age groups, which created issues of multicollinearity and interpretability.

As a result, the final Emblem models included the following variables:

- Age group
- E-HEALS group
- Gender
- PAM-13 group

Both the Gamma and Tweedie models in Emblem were constructed using the same set of

variables, identical banding, and consistent groupings. The modeling goal was to generate stable parameter estimates with reasonable standard errors, suitable for interpretation and actuarial decision-making. The fitted parameter tables show (Table 36-37) consistent and interpretable coefficients for the included variables.

Table 36 Gamma model parameters

Name		Value	Standard Error	Standard Error (%)	Weight	Weight (%)	Exp(Value)
Mean		1.9886	0.03631	1.8	779	100.0	7.3055
Gender	2 (Male)	-0.0606	0.04009	66.1	358	46.0	0.9412
	1 (Female)				421	54.0	
PAM group	1 (Low PAM score)	0.1382	0.05654	40.9	124	15.9	1.1482
	2 &3 (Medium PAM score)				571	73.3	
	4 (High PAM score)	-0.1069	0.06477	60.6	84	10.8	0.8986
E-Heals group	1st quartile	-0.1600	0.05374	33.6	157	20.2	0.8521
	2nd quartile	-0.0763	0.04843	63.5	195	25.0	0.9266
	3rd and 4th quartiles				427	54.8	
Age group	1 (40-49 years old)	-0.2684	0.05207	19.4	143	18.4	0.7646
	2-3 (50-69 years old)				483	62.0	
	4 (70+ years old)	0.1688	0.05142	30.5	153	19.6	1.1839

Source: own ed.

Notes: Where "Value"(=B) represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable, while holding all other variables constant. The Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard deviation of the sampling distribution of the coefficient. Number colors in the Emblem Beta tab indicate proximity to statistical significance: green denotes results clearly significant, grey indicates borderline or non-significant results, and red highlights results moving away from significance. "Exp(Value)" denotes the Exponentiated Coefficient.

Table 37 Tweedie model parameters

Name		Value	Standard Error	Standard Error (%)	Weight	Weight (%)	Exp(Value)
Mean		1.9854	0.03612	1.8	779	100	7.2819
Gender	2 (Male)	-0.0576	0.04036	70	358	46	0.9440
	1 (Female)				421	54.0	
PAM group	1 (Low PAM score)	0.1334	0.05606	42	124	15.9	1.1427
	2 &3 (Medium PAM score)				571	73.3	
	4 (High PAM score)	-0.1029	0.06602	64.2	84	10.8	0.9023
E-Heals group	1st quartile	-0.1473	0.05475	37.2	157	20.2	0.8630
	2nd quartile	-0.0796	0.0487	61.2	195	25.0	0.9235
	3rd and 4th quartiles				427	54.8	
Age group	1 (40-49 years old)	-0.2625	0.05504	21	143	18.4	0.7691
	2-3 (50-69 years old)				483	62.0	
	4 (70+ years old)	0.1664	0.05016	30.1	153	19.6	1.1811

Source: own ed.

Notes: Where "Value"(=B) represents the Unstandardized Coefficient. The B value indicates the change in the dependent variable for a one-unit change in the independent variable, while holding all other variables constant. The Standard Error, which measures the accuracy of the B coefficient estimate by representing the standard deviation of the sampling distribution of the coefficient. Number colors in the Emblem Beta tab indicate proximity to statistical significance: green denotes results clearly significant, grey indicates borderline or non-significant results, and red highlights results moving away from significance. "Exp(Value)" denotes the Exponentiated Coefficient.

Table 38 shows a direct model comparison between the Gamma and Tweedie distributions in Emblem. Both models converged successfully, and the Tweedie model demonstrated a lower Akaike Information Criterion (AIC), indicating a slightly better model fit.

Table 38 Model comparison

Metric	Current Model	Reference Model	Difference
Model Label	Tweedie	Gamma	
Fitted Parameters	8	8	(e)
Deviance	975.59	227.69	747.90
AICc	4,492.59	13,514.65	-9,022.06
Fitting Result	Converged OK	Converged OK	

Source: own ed.

The overall findings from the Emblem models strongly aligned with those obtained in SPSS. Most notably, the Patient Activation Measure (PAM-13) groupings remained a statistically significant predictor of health-related costs. Higher PAM levels were consistently associated with lower expected medical costs, reinforcing the key conclusions drawn from the SPSS-based analysis. This convergence of results across different tools and distributional assumptions enhances confidence in the reliability and generalizability of the model's insights.

III.2.7.3 Results and overall model overview

In the exploration of five Generalized Linear Models (GLM) designed to evaluate the significance of the Patient Activation Measure (PAM) variable alongside other predictors, an intriguing pattern emerged. Remarkably, in four out of the five models, the PAM variable demonstrated significant predictive power. This consistency underscores the potential relevance of patient activation levels, as captured by PAM, in forecasting medical costs within the insurance domain.

The models were further assessed using Omnibus tests, which indicated that all five models were statistically significant. This suggests that the models were correctly specified and that the included variables contribute meaningfully to explaining variations in medical costs. In addition to the Omnibus tests, other model fit statistics provided by SPSS were also examined, including Akaike’s Information Criterion (AIC), the finite sample corrected AIC (AICC), the Bayesian Information Criterion (BIC), and the consistent AIC (CAIC). These criteria evaluate model fit while accounting for model complexity, with lower values indicating a better balance between fit and parsimony. The results from these indicators were consistent with the Omnibus tests and supported the selected model specifications. Table 39 summarizes the results. As a result, based on all four criteria (AIC, AICC, BIC, and CAIC), Model #1 demonstrates the best overall fit.

Table 39 Goodness of Fit

	Goodness of Fit				
	#1	#2	#3	#4	#5
Akaike's Information Criterion (AIC)	13823,75	15076,48	15098,19	15067,83	15089,72
Finite Sample Corrected AIC (AICC)	13826,61	15079,18	15100,17	15069,35	15090,84
Bayesian Information Criterion (BIC)	13946,77	15200,96	15204,89	15161,19	15169,74
Consistent AIC (CAIC)	13974,77	15228,96	15228,89	15182,19	15187,74

Source: own ed.

As a final step in the evaluation process, an examination of the models' predictive accuracy was conducted by comparing the estimated costs derived from the models against the actual medical costs on record. This comparison aimed to provide a concrete measure of the models' practical utility in real-world settings. Correlations between actual and predicted logarithmic costs in the test dataset, detailed in Table 40, serve as critical indicators of model performance, offering insights into the precision and reliability of the GLM approach in capturing the complexities of insurance pricing.

Table 40 Model correlations

Correlations						
		ln_pred_cost_1	ln_pred_cost_2	ln_pred_cost_3	ln_pred_cost_4	ln_pred_cost_5
ln_cost	Pearson Correlation	0.263	0.201	0.179	0.168	0.2
	Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00
	N	779	779	779	779	779

Source: own ed.

Table 40 offers a valuable perspective on the predictive performance of various models, measured by the Pearson correlation between the actual and predicted logarithmic costs

in the test dataset. Although the correlation coefficients are modest, with the highest being 0.263 in model #1 and notable performances of 0.201 and 0.2 in models #2 and #5 respectively, it is important to consider these results in context.

A correlation in the range of 0.2 is indicative of a positive but weak linear relationship. However, it is critical to understand that in the complex field of insurance cost prediction, such correlations may not fully capture the models' usefulness or the significance of variables like PAM. The relatively lower correlations do not necessarily imply that PAM is not a valuable predictor. Instead, it suggests that while PAM contributes to the model, it is one of many factors, and its impact may be more nuanced.

Furthermore, the breadth and depth of data play a crucial role in improving model accuracy. With the current dataset of 779 responses, the models provide a glimpse into the relationships between variables and costs. However, for a more precise and accurate prediction, a larger dataset is typically more desirable. More data would offer a better opportunity to capture the complexities and variations inherent in medical costs, thereby potentially increasing the correlation coefficients and the models' predictive powers.

Model #2 showcases the highest correlation among those where PAM was significant, followed closely by model #5. This finding is encouraging and suggests that with additional data, the predictive power of these models could improve. The significance of PAM in these models supports its inclusion as a predictor and opens the door for further exploration. As data accumulates and more sophisticated modeling techniques are employed, the predictive accuracy is likely to improve, affirming the importance of patient activation measures in predicting medical costs.

Table 41 displays the estimated multipliers for the PAM categories across the five GLM models.

Table 41 PAM GLM estimates in Models #1-5

Parameter	Model 1			Model 2			Model 3			Model 4			Model 5		
	Exp(B)	95% Wald Confidence Interval for Exp(B)		Exp(B)	95% Wald Confidence Interval for Exp(B)		Exp(B)	95% Wald Confidence Interval for Exp(B)		Exp(B)	95% Wald Confidence Interval for Exp(B)		Exp(B)	95% Wald Confidence Interval for Exp(B)	
		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
[pamgr2=4]	0,703	0,429	1,152	0,276	0,155	0,489	0,257	0,147	0,449	0,266	0,151	0,47	0,239	0,139	0,412
[pamgr2=2 &3]	0,800	0,575	1,114	0,507	0,345	0,746	0,531	0,36	0,783	0,488	0,334	0,713	0,532	0,364	0,776
[pamgr2=1]	1	.	.	1	.	.	1	.	.	1	.	.	1	.	.

Source: own ed.

Notes: Where "Exp(B)" denotes the Exponentiated Coefficient.

The output from the Generalized Linear Models offers nsights into the Patient Activation Measure (PAM) as a predictor within the context of insurance risk assessment. Across all models, we observe a consistent trend: the highest PAM group, indicated as [pamgr2=4], is associated with lower Exp(B) values, while the baseline group, [pamgr2=1], has an Exp(B) set to 1. This pattern aligns with the underlying hypothesis that individuals with higher patient activation, reflecting a greater engagement in managing their health, are likely to be less risky and thus have a lower multiplier on their predicted costs.

In more specific terms, the Exp(B) value, often referred to as the risk multiplier in the context of insurance modeling, inversely correlates with the level of patient activation. For example, model 2 shows an Exp(B) value of 0.276 for the highest PAM group ([pamgr2=4]), which can be interpreted as a 72.4% decrease in the risk multiplier compared to the baseline group. This suggests that those with the highest activation levels are deemed significantly less likely to incur high medical costs.

Similarly, individuals in the [pamgr2=2&3] category, who have a moderate level of patient activation, also exhibit lower multipliers than the baseline but are not as low as the highest PAM group. This monotone effect, where each step up in patient activation is associated with a decrease in risk multiplier, provides a quantifiable validation of the PAM scale's effectiveness in differentiating risk levels among patients.

The consistency of this trend across all five models further substantiates the reliability of

PAM as a valuable predictor in this setting. Not only does this trend make sense intuitively—engaged and proactive patients often take better care of their health, potentially reducing the frequency and severity of claims—but it also has a clear manifestation in the model outputs. Insurance companies can leverage this information to refine their risk assessments, tailoring premiums and policies to more accurately reflect the varying levels of health engagement among their insured population.

In conclusion, the results of these models resonate with the principle that higher patient activation is indicative of lower risk, justifying the use of PAM as an influential factor in predicting health care costs in insurance modeling.

The significant role of the PAM variable across multiple models highlights the importance of considering patient activation in health care cost predictions. Meanwhile, the Omnibus test results and the comparison of estimated versus actual costs collectively affirm the models' validity and applicability in insurance analytics.

III.2.8. Future of GLM in insurance pricing

The landscape of insurance pricing is continuously evolving, driven by advancements in technology, data analytics, and regulatory changes. While Generalized Linear Models (GLM) have been a cornerstone in insurance pricing for decades, the future promises further innovation and adaptation of these models. This section explores the emerging trends and potential challenges that may shape the future of GLM in insurance pricing, offering insights into what lies ahead for actuaries and data scientists.

III.2.8.1 Integration with Machine Learning and Artificial Intelligence

The integration of Generalized Linear Models (GLMs) with machine learning (ML) and artificial intelligence (AI) is a key development in insurance pricing. These technologies enhance GLMs by boosting predictive accuracy, managing complex interactions, and enabling the analysis of unstructured data.

Hybrid models combine GLMs with machine learning algorithms like random forests or neural networks, utilizing the interpretability of GLMs alongside the advanced pattern recognition capabilities of ML. Additionally, AI enables automated feature selection,

helping to identify relevant predictors and interactions more efficiently, thereby reducing the time and complexity required for model development. Combined, they provide a strong foundation for more accurate, scalable insurance pricing.

III.2.8.2 Big Data and Advanced Analytics

The rise of big data presents new opportunities to refine and enhance insurance pricing models. Advanced analytics can process large datasets from various sources, offering deeper insights into risk factors.

Telematics and Internet of Things (IoT) data can be integrated into GLMs to provide more detailed risk assessments based on real-time behavior and environmental factors. Additionally, data from social media and other alternative sources may reveal new predictors of risk. However, the use of such data introduces ethical and privacy concerns that must be carefully managed to ensure responsible and fair usage.

III.2.8.3 Regulatory and Ethical Considerations

As Generalized Linear Models (GLMs) and other advanced models grow more complex, regulatory bodies are expected to increase their scrutiny of insurance pricing practices. Ensuring fairness, transparency, and privacy will become essential.

The demand for explainability and transparency may drive the adoption of more interpretable machine learning techniques that complement GLMs, ensuring models remain understandable for both regulators and policyholders. Additionally, as big data and AI become more integrated into pricing models, insurers will face greater ethical and privacy challenges. Striking a balance between detailed risk assessments and safeguarding individual privacy will be crucial for maintaining trust and regulatory compliance.

III.2.8.4 Challenges and Opportunities

The future of GLM in insurance pricing is not without challenges. Keeping pace with technological advancements, managing data privacy concerns, and meeting regulatory requirements will be ongoing issues. However, these challenges also present opportunities for innovation, improved risk assessment, and more personalized insurance

products.

The future of GLM in insurance pricing is poised at the intersection of tradition and innovation. By embracing new technologies and data sources while adhering to ethical and regulatory standards, the insurance industry can enhance the accuracy and fairness of pricing models, benefiting both insurers and insured parties.

Bibliography

ÁNTSZ (2020) ’A Nemzeti Népegészségügyi Központ módszertani levele a 2020. évi védőoltásokról’, ÁNTSZ [National Public Health and Medical Officer Service]: Online.

Actuarial Standards Board (2019) 'Actuarial Standard of Practice No. 49: Using Health Status Based Risk Adjustment Methodologies in Actuarial Work'.

Ágoston, K. Cs., Burka, D., Kovács, E., Vasközi, Á., & Vékás, P. (2019) 'Klaszterelemzési eljárások halandósági adatokra', Statisztikai Szemle, 97(7), pp. 629–655. Available at: https://doi.org/10.20311/stat2019.7.hu0629.

Ahn, Y.H. et al. (2015) 'Psychometric Properties of the Korean Version of the "Patient Activation Measure 13" (PAM13-K) in Patients With Osteoarthritis', Evaluation & the Health Professions, 38(2), pp. 255–264.

Asztalos, L. Gy. (1997) Biztosítási kézikönyv. Budapest: Biztosítási Oktatási Intézet.

Ayuso, M., Guillen, M. and Alcañiz, M. (2019) 'Telematics and Gender Discrimination: Some Usage-Based Evidence on Whether Men's Risk of Accidents Differs from Women's', Risks, 7(2), p. 63.

Aziz, A., Reynolds, R. & Ansari, A. (2019) PROCESS AND SYSTEMS: ’A population-based model of care for people with inflammatory bowel disease - patient-reported outcomes.’ Future Healthc J, 6(1), pp. 30–35.

Bahrom, N.H. et al. (2020) 'Validity and reliability of the Patient Activation Measure(R) (PAM(R))-13 Malay version among patients with Metabolic Syndrome in primary care', Malays Fam Physician, 15(3), pp. 22–34.

Baji, P. & Gulácsi, L. (2010) 'Beteg önrész” – a lakosság fizetési hajlandósága az egészségügyi szolgáltatásokért', Esély, 21(4), pp. 106–114.

Baji, P. & Gulácsi, L. (2012a) 'A helyzet változatlan – Egy reprezentatív kérdőíves felmérés eredményei a magyar lakosság hálapénz-fizetési szokásairól és a hálapénz megítéléséről', Egészségügyi Gazdasági Szemle, 50(4), pp. 30–36.

Baji, P. & Gulácsi, L. (2012b) 'A magyar lakosság fizetési hajlandósága az egészségügyi ellátásokért – egy reprezentatív kérdőíves felmérés eredményei', Biztosítási Szemle, 15 November, pp. 1–20.

Bandeira, R., Bahia Filho, J. A., Neto, V. C. L., & others. (2025) Maximizing Value in Public Healthcare: Assessing the Financial Return on Investment in Surgical Infection Prevention. Open Forum Infectious Diseases. Available at: https://doi.org/10.1093/ofid/ofae631.558.

Bandura, A. (1977) ’Self-efficacy: Toward a unifying theory of behavioral change.’ Psychological Review, 84(2), pp. 191–215.

Banyár, J. & Vékás, P. (2016) 'A pénzügyi termékek ára', Közgazdasági Szemle, 63(4), pp. 380–406. Available at: https://doi.org/10.18414/KSZ.2016.4.380.

Banyár, J. (2016) Életbiztosítás. Budapest: Budapesti Corvinus Egyetem Operációkutatás és Aktuáriustudományok Tanszék.

Bircher, J. & Kuruvilla, S. (2014) ’Defining health by addressing individual social and environmental determinants: new opportunities for health care and public health.’ J Public Health Policy, 35(3), pp. 363–386.

Boncz, I. & Sebestyén, A. (2015) Egészségbiztosítási ismeretek. Budapest: Medicina Könyvkiadó.

Brenk-Franz, K. et al. (2013) 'Validation of the German version of the patient activation measure 13 (PAM13-D) in an international multicentre study of primary care patients', PLoS One, 8(9), p. e74786.

Breusch, T.S. and Pagan, A.R. (1979) 'A Simple Test for Heteroscedasticity and Random Coefficient Variation', Econometrica, 47(5).

Burka, D., Kovács, L., & Szepesváry, L. (2021). Modelling MTPL insurance claim events: Can machine learning methods overperform the traditional GLM approach?. Hungarian Statistical Review, 4(2). Available at: 10.35618/hsr2021.02.en034

Carroll, J.K. et al. (2019) '"Get Ready and Empowered About Treatment" (GREAT) Study: a Pragmatic Randomized Controlled Trial of Activation in Persons Living with HIV.' J Gen Intern Med, 34(9), pp. 1782–1789.

Coleman, K. et al. (2009) 'Evidence on the Chronic Care Model in the new millennium. Health Aff (Millwood)', 28(1), pp. 75–85.

Cooper, V. et al. (2019) 'Measuring empowerment among people living with HIV: a systematic review of available measures and their properties', AIDS Care, 31(7), pp. 798–802.

Cox, B. et al. (2009) 'The reliability of the Minimum European Health Module', International Journal of Public Health, 54(2), pp. 55–60.

Cummins, J.D. and Weiss, M.A. (2001) 'The Global Market for Reinsurance: Consolidation Capacity and Efficiency', Brookings-Wharton Papers on Financial Services, 2001(1), pp. 159-222.

Cunha, C.M. et al. (2018) 'Cultural adaptation and validation of the Brazilian Version of the Patient Activation Measure-22 items', Rev Bras Enferm, 71(4), pp. 1891–1898.

Dash, A. & Grimshaw, D. (1993) 'Dread Disease cover. An actuarial perspective', Journal of the Staple Inn Actuarial Society, 33, pp. 149–193. Available at: https://doi.org/10.1017/S2049929900010564.

Deloitte (2021) 2021 Insurance Industry Outlook.

Denuit, M. and Lang, S. (2004) 'Non-life Rate-Making with Bayesian Generalized Linear Models', Insurance: Mathematics and Economics, 34(3), pp. 513–531.

Denuit, M., Maréchal, X., Pitrebois, S. and Walhin, J.-F. (2007) Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. Wiley.

Di Angelantonio, E. et al. (2016) 'Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents', The Lancet, 388(10046), pp. 776–786.

Dobson, A.J. and Barnett, A.G. (2008) An Introduction to Generalized Linear Models. 3rd edn. CRC Press.

Doherty, N.A. and Dionne, G. (1993) 'Insurance with Undiversifiable Risk: Contract Structure and Organizational Form of Insurance Firms', Journal of Risk and Uncertainty, 6(2), pp. 187–203.

European Central Bank Eurosystem (2020) 'Euro foreign exchange reference rates'. Available at: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-huf.en.html (Accessed: 1 April 2020).

Eurostat (2023) Eurostat database. Available at: https://ec.europa.eu/eurostat/data/database (Accessed: 26 January 2023).

Expert Panel on Effective Ways of Investing in Health (EXPH) (2019) Opinion on defining value in “value-based healthcare”. European Commission: Brussels.

Eyles, J.P. et al. (2017) 'Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties', Osteoarthritis and Cartilage, 25(8), pp. 1210–1222.

Eyles, J.P. et al. (2020) 'Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis', Health and Quality of Life Outcomes, 18(1), p. 121.

Eysenbach, G. and Wyatt, J. (2002) 'Using the Internet for surveys and health research', J Med Internet Res, 4(2), p. E13.

Fagerland, M. and Hosmer, D. (2013) 'A goodness-of-fit test for the proportional odds regression model', Statistics in Medicine, 32, pp. 2235–2249.

Faragó, J. (2012) Aktuáriusi modellek az egészségbiztosításban. Eötvös Loránd Tudományegyetem Biztosítási és pénzügyi matematika mesterszak szakdolgozat, pp. 1-53.

Fónagy, F. (2016) A long-term care biztosítás aktuáriusi modellezése. Budapesti Corvinus Egyetem Biztosítási és pénzügyi matematika mesterszak szakdolgozat, pp.1-56.

Fowles, J.B. et al. (2009) 'Measuring self-management of patients' and employees' health: further validation of the Patient Activation Measure (PAM) based on its relation to employee characteristics', Patient Education and Counseling, 77(1), pp. 116–122.

Frees, E.W. (2010) Regression Modeling with Actuarial and Financial Applications. Cambridge University Press.

Frees, E.W. and Valdez, E.A. (1998) 'Understanding Relationships Using Copulas', North American Actuarial Journal, 2(1), pp. 1–25.

Frees, E.W., Derrig, R.A. and Meyers, G. (2014) 'Predictive Modeling in Actuarial Science', Variance, 8(1), pp. 1–20.

Frees, E.W., Derrig, R.A. and Meyers, G. (2014) 'Predictive Modeling in Actuarial Science', Variance, 8(1), pp. 1–20.

Gogola, J. & Vékás, P. (2020) 'Élettartam-kockázat Csehországban és Magyarországon', Biztosítás és kockázat, 7(3–4), pp. 14–26. Available at: https://doi.org/10.18530/BK.2020.3-4.14.

Graffigna, G. et al. (2015) 'Measuring patient activation in Italy: Translation adaptation and validation of the Italian version of the patient activation measure 13 (PAM13-I)', BMC Med Inform Decis Mak, 15, p. 109.

Graffigna, G., Barello, S. & Bonanomi, A. (2017) The role of Patient Health Engagement Model (PHE-model) in affecting patient activation and medication adherence: A structural equation model. PLoS One, 12(6), p. e0179865.

Greene, J. and Hibbard, J.H. (2012) 'Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes', Journal of General Internal Medicine, 27(5), pp. 520–526.

Haberman, S. & Pitacco, E. (1999) Actuarial models for disability insurance. London: Chapman & Hall/CRC.

Haberman, S. and Renshaw, A.E. (1996) 'Generalized Linear Models and Actuarial Science', The Statistician, 45(4), pp. 407–436.

Harvey, L. et al. (2012) ’When activation changes what else changes? The relationship between change in patient activation measure (PAM) and employees' health status and health behaviors.’ Patient Educ Couns, 88(2), pp. 338–343.

Hellstrom, A. et al. (2019) 'Validation of the patient activation measure in patients at discharge from hospitals and at distance from hospital care in Sweden', BMC Public Health, 19(1), p. 1701.

Hendrikx, R.J.P. et al. (2018) 'How to Measure Population Health: An Exploration Toward an Integration of Valid and Reliable Instruments', Population Health Management, 21(4), pp. 323–330.

Hibbard, J.H. et al. (2004) ’Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers.’ Health Serv Res, 39(4 Pt 1), pp. 1005–1026.

Hibbard, J.H. et al. (2005) 'Development and testing of a short form of the patient activation measure', Health Serv Res, 40(6 Pt 1), pp. 1918–1930.

Hibbard, J.H., Greene, J. and Overton, V. (2013) 'Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'', Health Affairs, 32(2), pp. 216–222.

Holdroyd, I., Lunn, A. D., Diamond, P., & Harasgama, S. (2024) What works: Payment mechanisms to improve prevention spending in health care settings. HEEC. Available at: https://www.heec.co.uk/wp-content/uploads/2024/12/What-works-Payment-mechanisms.pdf.

Hosmer, D.J., Lemeshow, S. and Sturdivant, R. (2013) Applied Logistic Regression. 3rd edn. Hoboken, NJ: Wiley.

Hung, M. et al. (2013) 'Psychometric assessment of the patient activation measure short form (PAM-13) in rural settings', Qual Life Res, 22(3), pp. 521–529.

Insignia Health (2019) The Science of the PAM survey. Available at: https://www.insigniahealth.com/research/science.

Insignia Health (2020) 'Best Practices for Analysing PAM data', Insignia Health Editor. Available online: Insignia Health.

Institute of Medicine (2005) Crossing the quality chasm: A new Health System for the 21 century. Washington DC.

Insurance Information Institute (2023) Issues Update: Insurance Pricing.

Jean, J. (2004) 'Accident insurance', in Teugels, J. L. & Sundt, B. (eds.) Encyclopedia of Actuarial Science, vol. 1. Hoboken: John Wiley & Sons, pp. 1–4. Available at: https://doi.org/10.1002/9780470012505.taa001.

Kaiser, H. and Rice, J. (1974) 'Little jiffy mark IV', Educational and Psychological Measurement, 34(1), pp. 111–117.

Kar, L. B. (2016) Klinikai vizsgálatok aktuáriusi kapcsolatai. Eötvös Loránd Tudományegyetem Biztosítási és pénzügyi matematika mesterszak szakdolgozat, pp.1-56.

Kephart, G. et al. (2019) 'The structural and convergent validity of three commonly used measures of self-management in persons with neurological conditions', Qual Life Res, 28(2), pp. 545–556.

Kincses, Gy. (1999) Egészség-gazdaság? Budapest: Praxis Server Egészségügyi Tanácsadó Kft.

Kincses, Gy. (2006) 'A co-payment alkalmazási lehetősége a korszerű egészségpolitikában', Informatika és Menedzsment az Egészségügyben, 5(5), pp. 14–21.
Koltai, J. and Kun, E. (2016) 'The practical measurement of health literacy in Hungary and in international comparison', Orvosi Hetilap, 157(50), pp. 2002–2006.

Korm. határozat az „Egészséges Magyarország 2014–2020” Egészségügyi Ágazati Stratégia 2017–2018 évekre vonatkozó cselekvési tervéről in 1886/2016. (XII. 28.) (2017) Egészségügyi Közlöny, LXVI. évfolyam 1.

Kosar, C. and Besen, D.B. (2019) 'Adaptation of a patient activation measure (PAM) into Turkish: reliability and validity test', Afr Health Sci, 19(1), pp. 1811–1820.

Kovács, E. (2011) 'Pénzügyi adatok statisztikai elemzése', Budapest: Tanszék Kft.

Kovács, E. & Vékás, P. (2017) 'Mortality and Longevity Risk', in Pompella, M. & Scordis, N. A. (eds.) The Palgrave Handbook of Unconventional Risk Transfer. New York: Palgrave Macmillan, pp. 269–297. Available at: https://doi.org/10.1007/978-3-319-59297-8_9.

Kovács, E., Rétallér, O. & Vékás, P. (2015) 'Modellpontok szerepe a nyugdíj-hatásvizsgálatban', Közgazdasági Szemle, 62(12), pp. 1328–1342. Available at: https://doi.org/10.18414/KSZ.2015.12.1328.

Kovács, L. (2022). 'Változószelekciós algoritmusok vizsgálata általánosított additív modellekben – Egy új, hibrid metaheurisztika elemzése.', Budapesti Corvinus Egyetem Közgazdasági és Gazdaságinformatikai Doktori Iskola PhD Disszertáció, pp. 1-155. Available at: 10.14267/phd.2022058

Kruppa, J. and Schwender, H. (2015) 'Prediction of Insurance Claims Using Machine Learning Techniques', Insurance: Mathematics and Economics, 65, pp. 14–22.

KSH (2011) Population census 2011. Hungarian Central Statistical Office: Online.

KSH (2020) Demográfiai évkönyv 2019. Budapest: Központi Statisztikai Hivatal.

KSH (2023a) KSH Morbiditási adatgyűjtés. Available at: https://www.ksh.hu/elef2019 (Accessed: 26 January 2023).

KSH (2023b) KSH Továbvezetett népesség adattábla. Available at: https://www.ksh.hu/apps/meta.objektum?p_lang=HU&p_menu_id=110&p_ot_id=100&p_obj_id=AAAB (Accessed: 26 January 2023).

KSH (2023c) KSH database. Available at: http://statinfo.ksh.hu/Statinfo/ (Accessed: 26 January 2023).

Lantos, Cs. (2015) 'Állami és magáfinanszírozás az egészségügyben', VI. MABISZ Nemzetközi Biztosítási Konferencia.

Laranjo, L. et al. (2018) 'Translation and Validation of the Patient Activation Measure in Portuguese People with Type 2 Diabetes Mellitus', Acta Med Port, 31(7-8), pp. 382–390.

Lee, R. D. & Carter, L. R. (1992) 'Modeling and forecasting U.S. mortality', Journal of the American Statistical Association, 87, pp. 659–671. Available at: http://dx.doi.org/10.2307/2290201.

Lim, C.R. et al. (2015) 'Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set', BMJ Open, 5(7), p. e007765.

Lim, S.S. et al. (2015) 'Validation of a new predictive risk model: measuring the impact of the major modifiable risks of death for patients and populations', Popul Health Metr, 13, p. 27.

Lindsay, A., Hibbard, J.H., Boothroyd, D.B., Glaseroff, A. and Asch, S.M. (2018) 'Patient Activation Changes as a Potential Signal for Changes in Health Care Costs: Cohort Study of US High-Cost Patients', Journal of General Internal Medicine, 33(12), pp. 2106–2112. Available at: https://doi.org/10.1007/s11606-018-4657-6.

Lorig, K. (1993) 'Self-management of chronic illness: a model for the future.' Generations, 17(3), pp. 11–14.

Löfdahl, B. (2013) 'Stochastic modelling of disability insurance in a multi-period framework', KTH Royal Institute of Technology, 2(9), pp. 1–14. Available at: https://doi.org/10.1080/03461238.2013.779594.

MABISZ (2023) Magyar biztosítók évkönyve. Budapest: Magyar Biztosítók Szövetsége.
Magnezi, R. and Glasser, S. (2014) 'Psychometric properties of the Hebrew translation of the patient activation measure (PAM-13)', PLoS One, 9(11), p. e113391.

Mansfield, E.D. et al. (2018) 'Canadian adaptation of the Newest Vital Sign(c) a health literacy assessment tool', Public Health Nutrition, 21(11), pp. 2038–2045.

Marano, P. and Siri, M. (2020) 'Regulating Big Data: The Guidelines of the European Insurance and Occupational Pensions Authority', European Journal of Risk Regulation, 11(2), pp. 389–404.

McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models. 2nd edn. Chapman and Hall/CRC.

McGorrian, C. et al. (2011) 'Estimating modifiable coronary heart disease risk in multiple regions of the world: the INTERHEART Modifiable Risk Score', European Heart Journal, 32(5), pp. 581–589.

McGraw, K.O. and Wong, S.P. (1996) 'Forming inferences about some intraclass correlation coefficients', Psychological Methods, 1(1), pp. 30–46.

Melby, K. et al. (2021) 'Test-Retest Reliability of the Patient Activation Measure-13 in Adults with Substance Use Disorders and Schizophrenia Spectrum Disorders', International Journal of Environmental Research and Public Health, 18(3).

Mokkink, L.B. et al. (2010) 'The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study', Quality of Life Research, 19(4), pp. 539–549.

Moljord, I.E. et al. (2015) 'Psychometric properties of the Patient Activation Measure-13 among out-patients waiting for mental health treatment: A validation study in Norway', Patient Education and Counseling, 98(11), pp. 1410–1417.

Móri, T. (2011) Élettartamadatok elemzése. Budapest: Typotex Kft.

Morton, K. et al. (2017) ’Using digital interventions for self-management of chronic physical health conditions: A meta-ethnography review of published studies.’ Patient Educ Couns, 100(4), pp. 616–635.

Mosen, D. et al. (2007) ’Is Patient Activation Associated With Outcomes of Care for Adults With Chronic Conditions?’ Journal of Ambulatory Care Management, 30(1), pp. 21–29.

Nelder, J.A. and Wedderburn, R.W.M. (1972) 'Generalized Linear Models', Journal of the Royal Statistical Society. Series A (General), 135(3), pp. 370–384.

Németh, A. O., Németh, P. & Vékás, P. (2020a) 'Demographics Labour Market and Pension Sustainability in Hungary', Society and Economy, 42(2), pp. 146–171. Available at: https://doi.org/10.1556/204.2019.015.

Németh, A. O., Németh, P. & Vékás, P. (2020b) 'Childbearing and Pensions in the V4 Countries', Köz-gazdaság, 15(2), pp. 120–129. Available at: https://doi.org/10.14267/RETP2020.02.16.

Németh, P. (2016) ‘Egészségbiztosítások árazási folyamata és kihívásai egy társadalombiztosítási rendszerben.’ Eötvös Loránd Tudományegyetem Biztosítási és pénzügyi matematika mesterszak szakdolgozat, pp. 1-55.

Németh, P. (2024) ‘Az egészségbiztosítások árazási módszerei, folyamatai és kihívásai.’ Statisztikai Szemle, 102(4), pp. 307–325. Available at: https://doi.org/10.20311/stat2024.04.hu0307.

Nemzeti Rákregiszter (2023) Nemzeti Rákregiszter. Available at: http://www.onkol.hu/hu/nemzeti_rakregiszter (Accessed: 26 January 2023).

Neter, E. and Brainin, E. (2017) 'Perceived and Performed eHealth Literacy: Survey and Simulated Performance Test', JMIR Human Factors, 4(1), p. e2.

Ng, R. et al. (2020) 'Smoking drinking diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease', International Journal of Epidemiology, 49(1), pp. 113–130.

Ngooi, B.X. et al. (2017) 'Validation of the Patient Activation Measure (PAM-13) among adults with cardiac conditions in Singapore', Quality of Life Research, 26(4), pp. 1071–1080.

NIAAA (2004) 'NIAAA Newsletter', NIAAA Newsletter, Office of Research Translation and Communications NIAAA NIH: Online, p. 3.

NICE (2019) Evidence standards framework for digital health technologies. London.

NM rendelet a kötelező egészségbiztosítás keretében igénybe vehető betegségek megelőzését és korai felismerését szolgáló egészségügyi szolgáltatásokról és a szűrővizsgálatok igazolásáról in 51/1997. (XII. 18.) (1997) Magyar Közlöny, 1997/114. Hungary.

Norman, C.D. and Skinner, H.A. (2006) 'eHEALS: The eHealth Literacy Scale', Journal of Medical Internet Research, 8(4), p. e27

OECD (2004) Proposal for a taxonomy of health insurance. Technical report, OECD Private Health Insurance Study. Available at: https://doi.org/10.1787/9789264007451.

OECD (2015) Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives. Paris: OECD Publishing.

OECD (2019a) Health at a Glance 2019: OECD Indicators. Paris: OECD Publishing. Available at: https://doi.org/10.1787/4dd50c09-en.

OECD (2019b) Hungary: Country Health Profile 2019. European Observatory on Health Systems and Policies.

OECD (2023) OECD Database. Available at: http://stats.oecd.org/ (Accessed 26 January 2023).

OECD/European Observatory on Health Systems and Policies (2017) Hungary: Country Health Profile 2017. OECD Publishing/European Observatory on Health Systems and Policies: Paris/Brussels.

Ohlsson, E. and Johansson, B. (2010) Non-life Insurance Pricing with Generalized Linear Models. Springer.

Packer, T.L. et al. (2015) 'The Patient Activation Measure: a validation study in a neurological population', Quality of Life Research, 24(7), pp. 1587–1596.

Pál, V. (2017) 'Egészségünk földrajza - területi különbségek a hazai egészségi állapotban', Magyar Tudomány, 178(2017/3), pp. 311–321.

Parker, R.N. and Fenwick, R. (1983) 'The Pareto Curve and Its Utility for Open-Ended Income Distributions in Survey Research', Social Forces, 61(3).

Pasaribu, U.S., Husinah, H., Sari, R.R.K.N. & Yanti, A.R. (2019) ’Pricing Critical Illness Insurance Premiums Using Multiple State Continous Markov Chain Model.’ Journal of Physics: Conference Series, 1366(1), p. 012112. Available at: https://doi.org/10.1088/1742-6596/1366/1/012112.

Pirie, K. et al. (2013) 'The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK', The Lancet, 381(9861), pp. 133–141.

Pitacco, E. (2014) Health Insurance – Basic Actuarial Models. Springer. Available at: https://doi.org/10.1007/978-3-319-12235-9.

Prey, J.E. et al. (2016) 'Reliability and validity of the patient activation measure in hospitalized patients', Patient Education and Counseling, 99(12), pp. 2026–2033.

Prinsen, C.A.C. et al. (2018) 'COSMIN guideline for systematic reviews of patient-reported outcome measures', Quality of Life Research, 27(5), pp. 1147–1157.

Prochaska, J., Redding, C. & Evers, K. (2008) The transtheorethical model and stages of change in Health Behaviour and Health Education Theory Research and Practice, 4th Edition, Glanz, K., Rimer, B. & Viswanath, K. (Editors), Jossey-Bass: San Francisco CA.

Rademakers, J. et al. (2012) 'Measuring patient activation in The Netherlands: translation and validation of the American short form Patient Activation Measure (PAM13)', BMC Public Health, 12, p. 577.

Ramsey, J. (1969) 'Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis', Journal of the Royal Statistical Society. Series B, 31(2), pp. 350–371.

Rasch, G. (1960) Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danmarks Paedagogiske Institut.

Remmers, C. et al. (2009) ’Is patient activation associated with future health outcomes and healthcare utilization among patients with diabetes?’ J Ambul Care Manage, 32(4), pp. 320–327.

Røe, R., Grønning, K., Eriksson, L. R., & Zangi, H. A. (2020). ’Outcomes in patient education programmes for people with rheumatic diseases: Results from a Delphi process and a study of feasibility and responsiveness.’ Musculoskeletal care, 18(2), 195–203. https://doi.org/10.1002/msc.1456

Roger Gray and Kovács Erzsébet (2001) 'Módszertani tanulmányok', Statisztikai Szemle, 79(8). Available at: https://www.ksh.hu/statszemle_archive/2001/2001_08/2001_08_689.pdf

Rosseel, Y. (2012) 'lavaan: An R Package for Structural Equation Modeling', Journal of Statistical Software, 48(2).

Rotar, A.M. et al. (2018) 'Shared decision making between patient and GP about referrals from primary care: Does gatekeeping make a difference?', PLoS One, 13(6), p. e0198729.

Schmaderer, M. et al. (2015) 'Psychometric Properties of the Patient Activation Measure in Multimorbid Hospitalized Patients', Journal of Nursing Measurement, 23(3), pp. 128–141.

Schnock, K.O. et al. (2019) ’Acute Care Patient Portal Intervention: Portal Use and Patient Activation.’ J Med Internet Res, 21(7), p. e13336.

Shao, A.W., Sherris, M. & Fong, J.H. (2015) ’Product pricing and solvency capital requirements for long-term care insurance.’ Scandinavian Actuarial Journal, (February 27). Available at: https://doi.org/10.2139/ssrn.2574533.

Skolasky, R.L. et al. (2009) 'Psychometric properties of the Patient Activation Measure among individuals presenting for elective lumbar spine surgery', Quality of Life Research, 18(10), pp. 1357–1366.

Skolasky, R.L. et al. (2011) 'Psychometric properties of the patient activation measure among multimorbid older adults', Health Services Research, 46(2), pp. 457–478.

Spedicato, G.A., Dutang, C. and Petrini, L. (2018) 'Machine Learning Methods and Generalized Linear Models', North American Actuarial Journal, 22(3), pp. 452–477.

Stamatakis, E. et al. (2019) 'Sitting Time Physical Activity and Risk of Mortality in Adults', Journal of the American College of Cardiology, 73(16), pp. 2062–2072.

Stein, C.J. and Colditz, G.A. (2004) 'Modifiable risk factors for cancer', British Journal of Cancer, 90(2), pp. 299–303.

Stepleman, L. et al. (2010) 'Validation of the patient activation measure in a multiple sclerosis clinic sample and implications for care', Disability and Rehabilitation, 32(19), pp. 1558–1567.

Stiglitz, J.E. (2000) A kormányzati szektor gazdaságtana. Budapest: KJK Kerszöv.

Szabó, A. & Gilly, Gy. (2003) ’Több biztosító jelenléte az egészségbiztosításban az egészségügyi piac kudarca tükrében.’ Egészségügyi Gazdasági Szemle, 42(1), pp. 51–61.

Szentkereszti, G. & Vékás, P. (2022) ’Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal.’ Statisztikai Szemle, 100(10), pp. 905–922. Available at: https://doi.org/10.20311/stat2022.10.hu0905.

Taghizadeh, N., Vonk, J.M. and Boezen, H.M. (2016) 'Lifetime Smoking History and Cause-Specific Mortality in a Cohort Study with 43 Years of Follow-Up', PLoS One, 11(4), p. e0153310.

Terwee, C.B. et al. (2007) 'Quality criteria were proposed for measurement properties of health status questionnaires', Journal of Clinical Epidemiology, 60(1), pp. 34–42.

Terwee, C.B. et al. (2012) 'Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist', Quality of Life Research, 21(4), pp. 651–657.

Tőkey, B. (2015) ’Kitekintés a magán-egészségbiztosítások legelterjedtebb szabályozási megoldásaira.’ Biztosítás és Kockázat, 2(4), pp. 49–61. Available at: https://doi.org/10.18530/BK.2015.4.48.

Tusa, N., Kautiainen, H., Elfving, P., Sinikallio, S. and Mäntyselkä, P. (2020) 'Relationship between patient activation measurement and self-rated health in patients with chronic diseases', BMC Family Practice, 21, p. 225. Available at: https://doi.org/10.1186/s12875-020-01301-y.

Tusnády, G., Gaudi, I., Rejtő, L., Kásler, M. & Szentirmay, Z. (2008) ’A magyar daganatos betegek túlélési esélye a Nemzeti Rákregiszter adatai alapján.’ Magyar Onkológia, 52(4), pp. 339–349. Available at: https://doi.org/10.1556/MOnkol.52.2008.4.2.

Vallyon, A. (2011) A kiegészítő biztosítások szerepe az egészségügyben az Európai Unióban és Magyarországon. Doktori (PhD) értekezés, pp. 1–196.

van Buuren, S. and Eekhout, I. (2020) Child Development with the D-score: Turning Milestones into Measurement. Available at: D-score.org.

van der Vaart, R. et al. (2011) 'Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations', Journal of Medical Internet Research, 13(4), p. e86.

Vékás, P. (2019a) Az élettartam-kockázat modellezése. Budapesti Corvinus Egyetem, Budapest.

Vékás, P. (2019b) ’Változások a halandóságjavulás mintázatában Magyarországon.’ Biztosítás és kockázat, 5(3), pp. 34–47. Available at: https://doi.org/10.18530/BK.2018.3.34.

Vékás, P. (2020) ’Rotation of the age pattern of mortality improvements in the European Union.’ Central European Journal of Operations Research, 28(3), pp. 1031–1048. Available at: https://doi.org/10.1007/s10100-019-00617-0.

Wallston, K.A., Wallston, B.S. & DeVellis, R. (1978) ’Development of the Multidimensional Health Locus of Control (MHLC) Scales.’ Health Educ Monogr, 6(2), pp. 160–170.

Wang, X. et al. (2014) 'Fruit and vegetable consumption and mortality from all causes cardiovascular disease and cancer: systematic review and dose-response meta-analysis of prospective cohort studies', BMJ, 349, p. g4490.

Weiss, B.D. et al. (2005) 'Quick assessment of literacy in primary care: the newest vital sign', Annals of Family Medicine, 3(6), pp. 514–522.

Werner, G. and Modlin, C. (2009) Property and Casualty Insurance. LOMA.

WHO (2014) Global Status Report About Noncommunicable Diseases. Geneva: WHO.
WHO (2015) WHO global strategy on people-centred and integrated health services - Interim report. Geneva: WHO.

WHO (2016) Action plan for the prevention and control of noncommunicable diseases in the WHO European Region. Copenhagen: WHO Regional Office for Europe.

WHO (2019a) Chronic diseases and health promotion. Available at: https://www.who.int/chp/about/integrated_cd/en/ (accessed 1 September 2019).

WHO (2019b) Healthy prosperous lives for all: the European Health Equity Status Report. Copenhagen: World Health Organization Regional Office for Europe.

WHO (2019c) 'Process of translation and adaptation of instruments'. Available at: https://www.who.int/substance_abuse/research_tools/translation/en/

WHO (2023) WHO Database. Available at: https://gco.iarc.fr/overtime/en (accessed 26 January 2023).

Xi, B. et al. (2017) 'Relationship of Alcohol Consumption to All-Cause Cardiovascular and Cancer-Related Mortality in U.S. Adults', Journal of the American College of Cardiology, 70(8), pp. 913–922.

Zeng, H. et al. (2019) 'Measuring patient activation in Chinese patients with hypertension and/or diabetes: reliability and validity of the PAM13', Journal of International Medical Research, 47(12), pp. 5967–5976.

Zrubka, Z. et al. (2019) 'Psychometric properties of the Hungarian version of the eHealth Literacy Scale', European Journal of Health Economics, 20(Suppl 1), pp. 57–69.

Zrubka, Zs., Vékás, P., Németh, P., Dobos, Á., Hajdu, O., Kovács, L., Gulácsi, L., Hibbard, J. & Péntek, M. (2022) ’Validation of the PAM-13 instrument in the Hungarian general population 40 years old and above.’ European Journal of Health Economics, 23(8), pp. 1341–1355. Available at: https://doi.org/10.1007/s10198-022-01434-0.

Appendix I. Supplementary materials for Part I

This is the data used for Critical Illness Insurance pricing.
Population:
Male:

Year\Age	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
2015	232 190	253 720	246 847	268 291	320 840	315 549	325 315	422 519	379 670	337 041
2016	233 839	249 806	247 659	259 846	315 964	317 405	317 281	402 056	399 110	348 842
2017	237 334	243 160	249 863	253 732	306 587	320 288	311 254	381 187	414 419	356 343
2018	238 908	240 436	250 647	251 512	297 716	323 196	310 577	362 588	426 192	359 950
2019	240 537	235 885	253 532	250 546	289 541	325 446	315 030	345 806	433 920	362 642

Year\Age	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-	Total
2015	289 208	325 782	319 087	236 822	183 367	115 965	75 629	47 937	4 695 779
2016	286 957	307 308	327 130	248 517	182 871	119 533	74 746	49 649	4 688 519
2017	289 295	292 174	330 778	256 411	183 521	123 968	73 906	51 071	4 675 291
2018	301 229	281 779	327 001	258 224	186 801	129 257	73 400	52 189	4 671 602
2019	315 080	275 799	314 966	265 009	191 850	132 807	73 952	53 473	4 675 821

Female:

Year\Age	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
2015	219 975	240 831	233 623	253 806	304 171	297 793	317 782	412 167	372 604	337 681
2016	221 029	238 095	234 020	245 347	298 973	298 819	307 420	393 107	390 669	347 874
2017	224 405	231 542	236 561	239 337	290 073	300 275	299 055	373 952	404 889	354 496
2018	226 448	228 036	237 441	237 828	280 966	302 341	296 234	354 107	415 650	356 469
2019	228 068	223 294	240 423	236 508	272 262	303 908	296 687	336 011	422 530	357 079

Year\Age	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-	Total
2015	305 190	370 174	389 483	315 096	277 783	216 904	162 291	132 438	5 159 792
2016	299 994	347 381	397 975	329 127	274 329	221 271	160 685	135 851	5 141 966
2017	299 848	327 892	400 596	341 457	273 652	223 932	161 214	139 094	5 122 270
2018	310 149	314 105	394 901	344 354	276 272	229 126	160 896	141 446	5 106 769
2019	322 650	304 742	379 905	352 328	283 333	231 599	161 883	143 725	5 096 935

Number of cases:
Male:

Year\Age	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
2015	90	49	63	106	188	276	455	805	1 024	1 707
2016	98	39	73	138	204	299	446	808	1 130	1 797
2017	85	39	56	98	217	297	429	729	1 215	1 699
2018	61	29	36	93	176	254	437	680	1 191	1 593
2019	83	44	41	89	157	270	377	602	1 209	1 634

Year\Age	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-	Total
2015	2 664	5 118	7 714	8 007	7 828	5 534	3 738	2 230	47 596
2016	2 625	4 776	8 309	8 543	7 773	5 879	3 764	2 333	49 034
2017	2 631	4 696	8 124	8 820	8 226	6 202	3 957	2 487	50 007
2018	2 605	4 126	7 541	8 750	8 014	6 233	3 625	2 305	47 749
2019	2 557	4 105	7 123	8 807	8 088	6 651	3 502	2 306	47 645

Female:

Year\Age	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
2015	98	39	47	113	216	410	787	1 406	1 958	2 589
2016	71	41	44	126	199	457	700	1 411	2 025	2 763
2017	58	39	42	87	187	383	769	1 379	2 145	2 854
2018	69	29	35	89	195	376	716	1 266	2 133	2 526
2019	61	32	54	72	177	372	695	1 212	2 283	2 645

Year\Age	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-	Total
2015	2 966	5 008	7 221	6 949	6 848	6 072	4 665	3 433	50 825
2016	3 081	4 692	7 241	7 404	7 223	6 308	4 735	3 765	52 286
2017	3 112	4 372	7 472	7 580	7 262	6 692	4 907	3 800	53 140
2018	3 129	4 124	7 095	7 573	7 354	6 502	4 718	3 697	51 626
2019	3 154	4 012	6 392	7 886	7 142	6 848	4 534	3 538	51 109

Appendix II. Supplementary materials for Part II

Electronic Supplementary Material 1.

Zrubka Z, Vékás P, Németh P, Dobos Á, Hajdu O, Kovács L, Gulácsi L, Péntek M, *Validation of the PAM-13 instrument in the Hungarian general population*. European Journal of Health Economics 2021.

13 kérdésből álló Betegaktiváció Kérdőív® (PAM-13)

Az alábbiakban néhány olyan állítás szerepel, melyeket az emberek az egészségükkel kapcsolatosan szoktak mondani. Kérjük, jelölje be, hogy mennyire ért egyet - vagy nem ért egyet - azzal, hogy Önre jellemzőek ezek az állítások. Válaszai tükrözzék azt, amit önmagára nézve igaznak tart, és ne azt, amiről azt gondolja, hogy mások elvárnának Öntől.

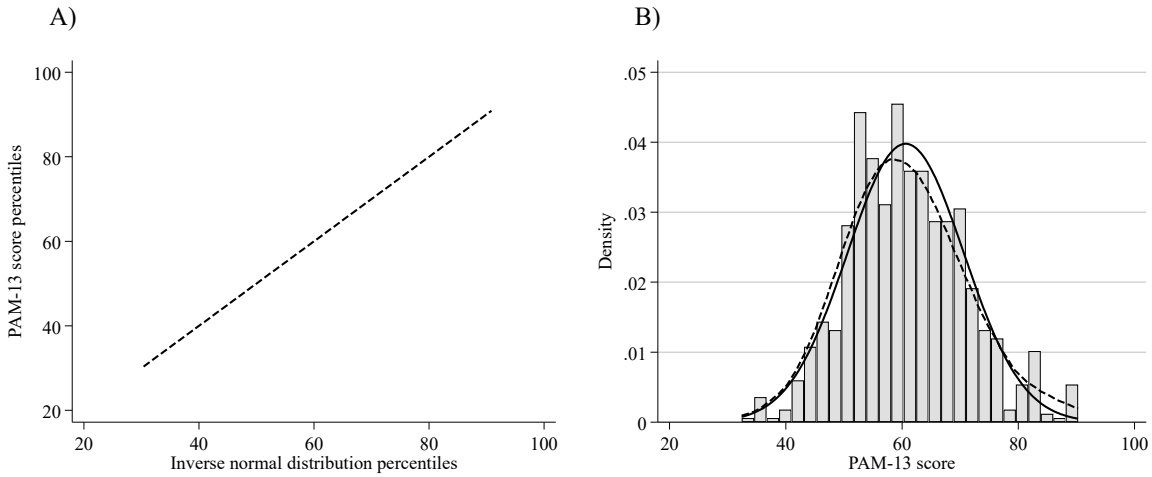
Amelyik állítás nem alkalmazható Önre, annál jelölje meg a "nem jellemző" lehetőséget.

1	Összességében az én felelősségem, hogy vigyázzak a saját egészségemre	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
2	Az egészségemet leginkább az befolyásolja, hogy aktívan foglalkozom vele.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
3	Biztos vagyok benne, hogy tudok segíteni az egészségemmel kapcsolatos problémák megelőzésében vagy csökkentésében.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
4	Minden felírt gyógyszeremről tudom, hogy mi a hatása.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
5	Biztos vagyok benne, hogy meg tudom állapítani, hogy egy egészségi problémával orvoshoz kell fordulnom, vagy magam is meg tudom azt oldani.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
6	Biztos vagyok benne, hogy el tudom mondani az orvosnak az aggályaimat akkor is, ha ő nem kérdezi.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
7	Biztos vagyok benne, hogy ha szükségem van rá, el tudom végezni az otthonra előírt kezeléseket.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
8	Értem az egészségi problémáimat és azok lehetséges kiváltó okait.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
9	Tudom, hogy egészségi problémáimra milyen kezelési lehetőségek állnak rendelkezésre.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
10	Kitartó tudtam maradni, amikor életmódot változtattam (pl. helyes táplálkozás vagy testmozgás).	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
11	Tudom, hogyan előzzem meg az egészségi problémáimat.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző
12	Biztos vagyok benne, hogy találok megoldást, ha új egészségi problémáim merülnek fel.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző

13	Biztos vagyok benne, hogy ha életmódot változtatok (pl. helyes táplálkozás vagy testmozgás), akkor még stresszes időszakokban is kitartó tudok maradni.	Egyáltalán nem értek egyet	Inkább nem értek egyet	Inkább egyetértek	Teljesen egyetértek	Nem jellemző

Electronic Supplementary Material 2.
Zrubka Z, Vékás P, Németh P, Dobos Á, Hajdu O, Kovács L, Gulácsi L, Péntek M, *Validation of the PAM-13 instrument in the Hungarian general population*. European Journal of Health Economics 2021.

Distribution of PAM-13 scores



A) Quantile-plot vs normal distribution, B) histogram with kernel-density plot (dashed line) and normal curve (solid line)

Electronic Supplementary Material 3.

Zrubka Z, Vékás P, Németh P, Dobos Á, Hajdu O, Kovács L, Gulácsi L, Péntek M, *Validation of the PAM-13 instrument in the Hungarian general population*. European Journal of Health Economics 2021.

Bivariate hypothesis tests for known-groups validity

			Subgroups										
Variable	**Category**	Total	Male	Female	No chronic disease	Chronic disease	40-65 years old	65+ years old	2-5th Income quintile	1st Income quintile	Adequate health literacy[a]	Not adequate health literacy	Supported hypo-theses
Preventive behaviours	Preventive behaviour score ≥50%	61.09	59.86	62.04	63.68	60.30	61.59	61.02	61.09	60.96	60.43	62.30	
	Preventive behaviour score <50%	60.17	59.29	60.99	61.98	59.48	59.76	60.43	60.16	60.29	59.74	60.86	1 / 10
	p value	*0.102*	*0.288*	*0.149*	*0.090*	*0.185*	*0.022*	*0.316*	*0.104*	*0.402*	*0.215*	*0.121*	
Lifestyle risks stringent criteria	Lifestyle risk index=0	63.45	62.94	63.78	64.17	63.56	64.40	61.95	63.26	65.47	62.84	64.51	
	Lifestyle risk index ≥1	59.57	58.53	60.55	61.85	58.85	59.19	60.21	59.64	58.97	59.05	60.45	9 / 10
	p value	*<0.00 1*	*<0.00 1*	*0.002*	*0.040*	*<0.00 1*	*<0.00 1*	*0.109*	*<0.00 1*	*0.013*	*<0.00 1*	*0.002*	
Lifestyle risks relaxed criteria	Lifestyle risk index ≤1	62.30	60.81	63.40	63.62	61.93	62.93	61.36	62.20	63.28	61.66	63.43	
	Lifestyle risk index ≥2	57.84	57.86	57.81	60.31	57.11	57.00	59.41	58.02	56.17	57.41	58.53	9 / 10
	p value	*<0.00 1*	*0.002*	*<0.00 1*	*0.005*	*<0.00 1*	*<0.00 1*	*0.056*	*<0.00 1*	*0.006*	*<0.00 1*	*<0.00 1*	
Health information seeking	At least montly	61.58	61.13	61.88	63.48	61.13	61.85	61.17	61.53	62.02	60.70	63.05	
	Less often than monthly	59.18	57.89	60.80	61.62	57.96	58.76	59.92	59.36	56.71	59.19	59.16	7 / 9
	p value	*<0.00 1*	*0.001*	*0.158*	*0.065*	*<0.00 1*	*<0.00 1*	*0.153*	*0.002*	*0.041*	*0.045*	*0.001*	
Patient education	Over past year	62.09	60.29	63.14	65.25	61.40	62.51	61.35	61.94	63.26	61.57	62.87	
	None	60.20	59.40	60.97	62.20	59.41	60.01	60.52	60.27	59.61	59.69	61.11	6 / 9
	p value	*0.018*	*0.266*	*0.031*	*0.047*	*0.030*	*0.013*	*0.290*	*0.035*	*0.139*	*0.044*	*0.124*	
Online health information seeking	At least bimonthly	61.04	60.41	61.67	64.05	60.16	61.28	60.79	61.03	61.13	60.27	62.45	
	Less often than bimonthly	60.13	58.84	61.43	61.53	59.60	59.68	60.60	60.17	59.73	59.85	60.55	2 / 9
	p value	*0.104*	*0.062*	*0.409*	*0.018*	*0.271*	*0.039*	*0.437*	*0.125*	*0.306*	*0.316*	*0.062*	
Online health-related communication	Over past year	61.83	61.26	62.18	63.95	61.45	61.79	61.87	61.81	61.95	61.44	62.32	
	None	60.21	59.10	61.26	62.36	59.29	60.20	60.23	60.23	60.08	59.70	61.19	4 / 9
	p value	*0.025*	*0.040*	*0.205*	*0.160*	*0.014*	*0.063*	*0.114*	*0.033*	*0.263*	*0.046*	*0.203*	
Online health prevention activity	Over past year	61.64	60.68	62.23	62.99	61.42	62.21	60.76	61.45	63.51	61.04	62.40	
	None	60.04	59.08	61.03	62.44	59.11	59.68	60.65	60.16	58.95	59.64	60.86	3 / 9
	p value	*0.015*	*0.068*	*0.119*	*0.328*	*0.006*	*0.003*	*0.465*	*0.045*	*0.058*	*0.069*	*0.105*	
Online disease management activity	Over past year	61.43	59.94	62.32	64.00	60.90	62.24	61.03	61.23	63.06	60.74	62.47	
	None	59.97	59.33	60.68	61.90	58.99	59.42	60.36	60.15	58.07	59.60	60.66	4 / 9
	p value	*0.022*	*0.278*	*0.051*	*0.049*	*0.017*	*0.001*	*0.289*	*0.075*	*0.030*	*0.103*	*0.070*	

N		779	358	421	253	503	483	296	704	75	491	288	
Supported hypotheses		7 / 9	4 / 9	3 / 9	5 / 9	7 / 9	8 / 9	0 / 9	6 / 9	4 / 9	5 / 9	3 / 9	

[a] NVS

Electronic Supplementary Material 4.

Zrubka Z, Vékás P, Németh P, Dobos Á, Hajdu O, Kovács L, Gulácsi L, Péntek M, *Validation of the PAM-13 instrument in the Hungarian general population*. European Journal of Health Economics 2021.

Regression analyses using PAM-13 scores as predictor

		PBS[i]	LRJ[j]	BMI[k]	Smoking[l]	Alcohol[m]	Physical activity[n]	Diet[o]	HIS[p]	PPE[q]	OHA[r]	OHIS[s]	OHC[t]	OHP[u]	ODA[v]
Model		OLS	Robust regression	Logistic	Logistic	Logistic	Logistic	Logistic	Ordered logit	Ordered logit	Ordered logit	Ordered logit	Ordered logit	Ordered logit	Ordered logit
PAM-13 score		0.00	-0.02***	-0.05***	0.00	-0.01	-0.04***	-0.08**	0.00	0.01	0.00	-0.01	0.00	0.00	-0.01
eHEALS[a]		0.00	0.00	0.03	-0.01	0.01	-0.01	-0.05	0.16***	0.04	0.05**	0.14***	0.07***	0.09***	0.11***
NVS[b]		0.00	-0.04	0.00	-0.09	-0.12	0.00	-0.16	0.00	-0.14*	-0.07	-0.02	-0.20***	-0.17**	-0.09
Age		0.00	-0.01*	0.01	-0.04***	0.00	-0.01	-0.04*	0.01	-0.02	0.01	-0.01	0.01	0.00	0.02**
Gender		-0.01	-0.09	-0.11	0.07	-0.82***	0.23	-0.43	0.65***	0.41*	0.23	0.61***	0.35	0.39*	0.46**
Education[c]	Secondary	0.05*	-0.03	-0.01	-0.56*	0.27	0.23	0.05	-0.03	-0.05	-0.31	-0.09	-0.38	-0.11	-0.13
	Tertiary	0.10***	-0.20	-0.24	-0.99***	0.37	0.08	-1.03	0.35	0.39	0.28	0.53*	0.40	0.15	0.16
Income[d]	2nd quintile	0.00	0.16	0.00	0.40	-0.38	0.76*	-0.32	-0.51	0.06	-0.36	-0.18	0.13	0.19	-0.43
	3rd quintile	0.00	0.03	-0.34	0.38	-0.34	0.42	-0.11	-0.63*	-0.25	-0.53	-0.51	0.16	0.13	-0.64
	4th quintile	0.06	-0.08	-0.14	0.10	-0.53	0.24	-0.94	-0.24	-0.09	-0.17	0.10	0.55	0.26	-0.04
	5th quintile	0.04	0.16	0.13	0.61	-0.16	0.47	-0.62	-0.36	-0.07	-0.30	-0.34	0.03	-0.07	-0.63*
Chronic morbidity[e]	Yes	0.05*	0.13	0.59***	-0.13	-0.79***	0.55*	0.02	-0.08	0.65**	0.51**	0.14	0.58*	-0.14	0.39
Self-rated health[f]	Bad	-0.01	0.50	2.26*	0.36	-1.22	0.02	12.39	-0.67	-0.35	-0.86	-0.14	-1.25	-1.50*	-1.13
	Fair	-0.04	0.36	1.91	0.11	-0.22	0.09	10.91	-	-0.78	-1.22	-0.66	-0.90	-	-0.99

	Good	-0.06	0.11	1.69	-0.06	-1.13	-0.21	11.01	-1.43*	-1.08	-1.38*	-0.93	-0.91	-1.52*	-0.97
	Very good	-0.04	-0.12	1.33	-2.04	-2.59	0.39	11.00	-1.68*	-0.72	-2.32**	-1.36	-0.77	-1.36	-0.92
GALI[g] (Limited due to health problems)	Not severely	0.03	-0.01	-0.36	-0.12	0.14	0.36	-0.21	0.58**	0.17	0.23	0.65***	0.79***	0.34	0.17
	Severely	0.10	-0.25	-0.28	-0.17	-1.03	0.31	-15.48	0.39	0.35	0.92*	0.67	0.62	0.63	0.16
Settlement[h]	Town	-0.04	-0.06	0.08	-0.29	0.48	-0.38	-0.14	0.16	0.03	-0.19	0.07	-0.17	-0.36	-0.06
	Village	-0.09**	-0.02	0.26	-0.16	0.41	-0.44	0.09	-0.14	-0.85*	-0.54*	-0.37	-0.54	-0.46	-0.31
Constant		0.38**	3.07***	-1.11	2.37	0.45	1.9	-4.32							
N		648	648	648	648	648	648	648	648	648	648	648	648	648	648
Breusch-Pagan test	*p* value	0.214	-	-	-	-	-	-	-	-	-	-	-	-	-
Ramsey RESET test	*p* value	0.188	0.705	-	-	-	-	-	-	-	-	-	-	-	-
Goodness of Fit test	*p* value	-	-	0.32	0.019	0.344	0.256	0.997	0.284	0.928	0.329	0.12	0.412	0.176	0.728

*p < 0.05 ; **p<0.01; ***p<0.001

[a] eHealth Literacy Scale;[b]Newest Vital Sign; [c]base: Primary; [d]base: 1st quintile ; [e]base: No chronic morbidity; [f]base: Very bad, [g]Global activity limitation indicator, base: no limitation; [h]base: Capital; [i]Preventive behaviour score; [j]Lifestyle risk index; [k]18.5<BMI(body mass index)<30; [l]Current smoker; [m]Binge drinking ≥1 per week; [n]Sedentary behaviour ≥8 hours per day with <150 min exercise per week or no exercise at all; [o]no fruit and / or vegetable intake; [p]Health information seeking (general); [q]Participation in patient education;[r]Online health administration; [s]Online health information seeking; [t]Online health-related communication; [u]Online health prevention; [v]Online disease management activity

Electronic Supplementary Material 5.

Zrubka Z, Vékás P, Németh P, Dobos Á, Hajdu O, Kovács L, Gulácsi L, Péntek M, *Validation of the PAM-13 instrument in the Hungarian general population*. European Journal of Health Economics 2021.

Regression analyses using PAM-13 levels as predictor

		PBS[i]	LRJ[j]	BMI[k]	Smoking[l]	Alcohol[m]	Physical activity[n]	Diet[o]	HIS[p]	PPE[q]	OHA[r]	OHIS[s]	OHC[t]	OHP[u]	ODA[v]
Model		OLS	Robust regression	Logistic	Logistic	Logistic	Logistic	Logistic	Ordered logit	Ordered logit	Ordered logit	Ordered logit	Ordered logit	Ordered logit	Ordered logit
PAM level	2	0.07	-0.30	-0.67	-0.31	0.08	-0.42	0.34	-0.05	-0.74	0.21	-0.40	0.43	0.53	0.29
	3	0.06	-0.56***	-0.95**	-0.21	-0.08	-0.93**	-0.77	-0.13	-0.34	0.01	-0.47	0.24	0.55	0.17
	4	0.04	-0.75***	-1.92***	0.32	-0.39	-1.07*	-17.18	-0.13	-0.05	-0.02	-0.55	0.04	0.06	-0.18
eHEALS[a]		0.00	0.00	0.02	-0.02	0.00	-0.02	-0.06	0.16***	0.04*	0.05**	0.14***	0.08***	0.09***	0.11***
NVS[b]		0.00	-0.04	0.00	-0.09	-0.12	0.00	-0.17	0.00	-0.14*	-0.07	-0.01	-0.21***	-0.16**	-0.09
Age		0.00	-0.01*	0.01	-0.04***	0.00	-0.01	-0.04*	-0.01	-0.02	0.01	-0.01	0.01	0.00	0.02**
Gender		-0.01	-0.10	-0.13	0.07	-0.82***	0.20	-0.51	0.66***	0.44*	0.22	0.61***	0.35	0.39*	0.46**
Education[c]	Secondary	0.06*	-0.03	-0.01	-0.58*	0.28	0.23	0.14	-0.03	-0.12	-0.30	-0.10	-0.33	-0.08	-0.11
	Tertiary	0.10***	-0.19	-0.23	-0.98***	0.36	0.11	-1.12	0.36	0.37	0.29	0.54*	0.43	0.13	0.16
Income[d]	2nd quintile	-0.01	0.18	-0.02	0.49	-0.40	0.79*	-0.60	-0.51	0.13	-0.37	-0.15	0.09	0.10	-0.47
	3rd quintile	0.00	0.04	-0.35	0.41	-0.34	0.43	-0.19	-0.63*	-0.19	-0.54	-0.50	0.11	0.07	-0.66
	4th quintile	0.06	-0.05	-0.15	0.17	-0.53	0.30	-1.03	-0.23	-0.05	-0.17	0.11	0.51	0.18	-0.07
	5th quintile	0.03	0.18	0.10	0.67	-0.18	0.49	-0.73	-0.36	0.01	-0.31	-0.33	-0.03	-0.16	-0.67*
Chronic morbidity[e]	Yes	0.05*	0.12	0.61**	-0.13	-0.80**	0.53*	-0.01	-0.09	0.68*	0.50**	0.13	0.56*	-0.12	0.40
Self-rated health[f]	Bad	-0.01	0.49	2.22	0.41	-1.23	0.03	13.96	-0.66	-0.37	-0.87	-0.15	-1.22	-1.53*	-1.15
	Fair	-0.04	0.36	1.84	0.19	-0.26	0.11	12.35	-1.42*	-0.75	-1.23	-0.67	-0.90	-1.63*	-1.05

	Good	-0.07	0.08	1.58	0.01	-1.19	-0.24	12.28	-1.82*	-1.03	-1.40*	-0.95	-0.90	-1.38	-1.05
	Very good	-0.04	-0.18	1.23	-2.04	-2.64	0.26	12.39	-1.67*	-0.67	-2.34**	-1.39	-0.74	-1.45	-0.98
GALI[g] (Limited due to	Not severely	0.03	-0.01	-0.37	-0.09	0.13	0.36	-0.25	0.58**	0.18	0.23	0.67***	0.78***	0.32	0.17
health problems)	Severely	0.09	-0.24	-0.36	-0.09	-1.06	0.33	-17.65	0.40	0.38	0.92*	0.69	0.60	0.55	0.11
Settlement[h]	Town	-0.04	-0.07	0.07	-0.30	0.48	-0.40	-0.09	0.16	0.05	-0.19	0.06	-0.17	-0.36	-0.06
	Village	-0.09**	-0.03	0.23	-0.18	0.40	-0.45	0.30	-0.14	-0.86*	-0.52*	-0.38	-0.52	-0.48	-0.31
Constant		0.36**	2.36***	-2.57	2.77*	-0.04	0.64	-9.22							
N		648	648	648	648	648	648	612	648	648	648	648	648	648	648
Breusch-Pagan test	*p* value	0.431	-	-	-	-	-	-	-	-	-	-	-	-	-
Ramsey RESET test	*p* value	0.308	0.734	-	-	-	-	-	-	-	-	-	-	-	-
Goodness of Fit test	*p* value	-	-	0.307	0.031	0.342	0.261	1	0.314	0.523	0.199	0.014	0.366	0.825	0.691

*p < 0.05 ; **p<0.01; ***p<0.001

[a] eHealth Literacy Scale;[b]Newest Vital Sign; [c]base: Primary; [d]base: 1st quintile ; [e]base: No chronic morbidity; [f]base: Very bad, [g]Global activity limitation indicator, base: no limitation; [h]base: Capital; [i]Preventive behaviour score; [j]Lifestyle risk index; [k]18.5<BMI(body mass index)<30; [l]Current smoker; [m]Binge drinking ≥1 per week; [n]Sedentary behaviour ≥8 hours per day with <150 min exercise per week or no exercise at all; [o]no fruit and / or vegetable intake; [p]Health information seeking (general); [q]Participation in patient education;[r]Online health administration; [s]Online health information seeking; [t]Online health-related communication; [u]Online health prevention; [v]Online disease management acti