USAID
FROM THE AMERICAN PEOPLE

CoReform
HEALTH SYSTEMS TRANSFORMATION

PREMIUM ESTIMATES FOR THE 2009 STATE HEALTH CARE BENEFIT PROGRAM FOR THE POOR

FINAL REPORT

September 2008
This publication was produced for review by the United States Agency for International Development. It was prepared under the auspices of CoReform

USAID FROM THE AMERICAN PEOPLE | CoReform HEALTH SYSTEMS TRANSFORMATION

PREMIUM ESTIMATES FOR THE 2009 STATE HEALTH CARE BENEFIT PROGRAM FOR THE POOR

FINAL REPORT

September 2008

This publication was produced for review by the United States Agency for International Development. It was prepared under the auspices of CoReform

This document was prepared with financial assistance from USAID
Contract No. GHS-1-00-03-00039-00 Task Order 800

Authors:

Guram Mirzashvili
Alexander Omanadze
Khatuna Jishiashvili
Giorgi Makharadze
Zurab Tsigroshvili

Abt Associates Inc.

CARE International

Curatio International Foundation

PREMIUM ESTIMATES FOR THE 2009 STATE HEALTH CARE BENEFIT PROGRAM FOR THE POOR

FINAL REPORT

Table of Contents

0. Executive Summary

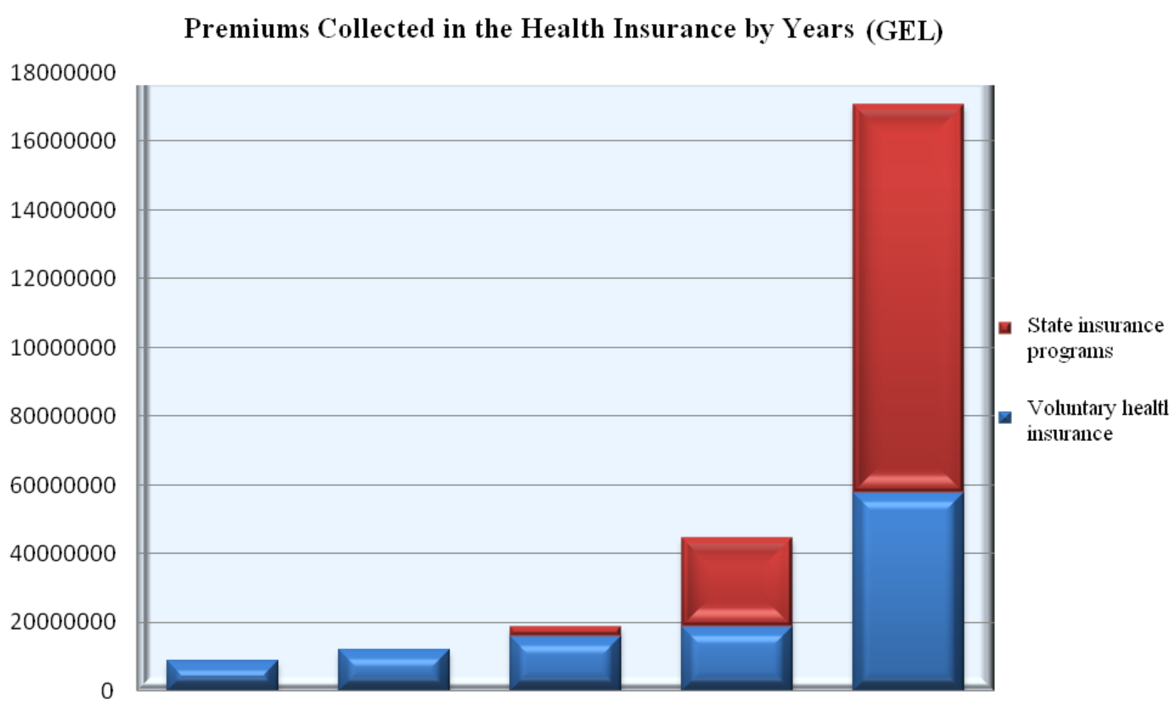
Problem Statement

The cooperation in health system transformation (CoReform) project is being implemented under financial assistance of USAID/Georgia. In the result of a technical assistance provided by the CoReform project in the field of health financing the Ministry of Labor, Health and Social Affairs of Georgia commenced implementing the poor's program on insurance-based principle in 2007. It put on the agenda the issue of elaborating actuarial estimates for a corresponding insurance premium for the state program to provide for appropriate allocations in the state budget.

Brief overview of the situation

Implementation of the pilot health insurance program for the population below the poverty line (insurance program for the needy population) started in Tbilisi and Imereti Region in 2007. In the fall 2007 the government of Georgia announced publicly that it was going to expand such programs gradually and provide a publicly funded insurance coverage to approximately 1.2 million individuals. In 2008 the number of insured will reach 800 000 individuals.

A conversion of state programs to insurance-based principles has a significant effect on health insurance in Georgia. This segment of the market which was growing at a considerable pace on account of a voluntary (mainly corporate) health insurance has acquired a leading position throughout the whole insurance industry. Even in 2007 the corresponding premium of the medical insurance program for the population below the poverty line turned out to exceed in value the whole health insurance market. In addition, as the state policy became more oriented to private insurance relationships a demand for voluntary health insurance has also increased noticeably. The following graph depicts the amount of a premium written in the health insurance market and the share of publicly funded insurance in Georgia in 2004-2008 (data for 2008 are estimates based on the results of a six-month period). The publicly funded insurance implies not only insurance programs for the poor but also similar programs funded by the state (e.g. a medical insurance program for teachers).



The graph shows clearly how the market grew due to emergence of publicly funded insurance programs. The first such case was an insurance program financed by Tbilisi Municipality in 2005. Insurance coverage for about 50 000 beneficiaries was provided by this program with the budget of 3 000 000 GEL – a moderate amount for the market of that time. In 2007 the pilot program for the poor started in Tbilisi and Imereti Region. Its budget was 17 million GEL and covered about 200 000 beneficiaries. In addition, several other programs with a design similar to the poor's program started in 2007. A total amount of a premium written under publicly funded state health insurance programs exceeded 25 million GELS while a total amount of a premium written under voluntary health insurance schemes was less than 18 million GELS. A total amount of an insurance premium to be written under the poor's programs (excluding other state-funded insurance) in 2008 is 113 million GELS which accounts for two thirds of corresponding premium income in the health insurance industry.

The health insurance is a leading line of the insurance market in Georgia: in 2007 it accounted for almost 45% of all premium income throughout the insurance market; its expected share in 2008 is 60%. The share of premiums collected only under the poor's programs will comprise 40% of the entire market in 2008.

Naturally, a portfolio of such a size exerts a great influence on a general situation on the entire market and such processes as price setting in the voluntary health insurance. For example, a price appreciation in the voluntary health insurance became apparent past several months of initiation of the pilot poor's program. One of the reasons was that at the first stage a loss pattern under the poor's program turned out to be heavier than expected and it caused immediate cost shifting to the voluntary health insurance.

In order to balance the program it is very important to plan its budget with a high precision since when a portfolio of such a size is overly profitable or unprofitable it has a negative impact for the entire insurance industry. In the first case it will deteriorate the accessibility of a voluntary insurance and in the second one it will lead to unjustifiably low premium rates.

To provide for necessary allocations in the state budget it became necessary to develop actuarial estimates for a corresponding insurance premium of the state program. To this end, the Ministry of Labour, Health and Social Affairs of Georgia asked the USAID and the Cooperation in Health System Transformation (CoReform) Project to provide a technical assistance in calculating corresponding parts of the healthcare budgets for 2008-2009.

Methodology

The project hired 5 local consultants to fulfil a technical assignment agreed at the Ministry. It implies calculating the insurance premiums for the 2008-2009 state insurance programs for the population below the poverty line and studying the systems of data collection and registration employed by the companies in the field of health insurance (the aim of the last assignment is to optimize the process of reporting of the private companies implementing state programs; the information obtained by means of this reporting is a main basis for calculating the amount of the premium).

This report presents recommendations regarding the parameters of the corresponding component of the 2009 state budget envisaged for state health insurance program for the population below the poverty line.

To fulfil the task, the working group has collected the needed statistical data and performed its initial processing and analysis. On the basis of the available statistical information respective actuarial models have been developed and the basic insurance premium for the poor's program has been calculated and recommendations on pricing coefficients have been elaborated.

Results

The basic annual insurance premium for the 2009 poor's program is 151.68 GEL; the respective monthly payment is 12.64 GEL. The premium is structured the following way:

Table 1. The structure of the monthly basic premium

Component	Amount
Inpatient service	7 GEL
Outpatient service	1.2 GEL
Medial cost appreciation	15%
Administration, acquisition, profit	20%

Along with calculating the basic insurance premium – an important parameter for the state budget – the working group has analyzed the level of losses associated with various insurance portfolios from different perspectives. The aim of such an analysis was to elaborate different price coefficients which would also facilitate a proper technical implementation of the poor's program.

The results of the analysis showed that it will be necessary to increase the number of tariff classes. It is a natural and approved mechanism worldwide for balancing the risk among insurance companies participating in the program. The number of tariff classes should be increased in stages. For example, in 2009 tariff class qualifiers should comprise only age and sex, while in 2010 they should be expanded by such parameters as region and family size.

Chapter 6 of the report provides a more detailed discussion of the issue and gives different options for ratification schedules obtained by employing the above mentioned parameters.

1. Task Formulation and Employed Statistics

The main goal of the fulfilled work was to assign appropriate insurance premium for the 2009 state health insurance program for the poor. It meant determining parameters of the program's budget as well as elaborating a price schedule by risk-factors. To achieve the main goal, it proved necessary to fulfill the following intermediate tasks:

1. To obtain detailed statistical data on implementation of the 2006-2008 programs for the poor from the Health and Social Programs Agency and private insurers participating in the program.
2. To normalize the obtained information, standardize it and perform its initial processing.
3. To hold consultations with the Ministry of Labor, Health and Social Affairs of Georgia to specify the expected changes in the design (in the list of services provided for by state program), to analyze the obtained statistical information considering these expected changes.
4. To develop actuarial models and calculate the basic premium using the available statistical information on their basis.
5. To adjust the basic premium for different factors.
6. To categorize losses in the target population (beneficiaries) considering different *a priori* risk-factors as well as analyze losses within these groups.
7. To determine equalizing coefficients for different risk-groups for the basic premium.

The following data have been obtained and employed under the project:

1. Ministry of Labor, Health and Social Affairs of Georgia:

- A description of the 2006-2008 State Health Programs
- Findings of the Household Survey on Health Expenditure and Health Service Utilization (Final Report, December 2007)

2. Health and Social Programs Agency:

- Statistical and financial data on the implementation of medical assistance programs for the population below the poverty line (from July 1, 2006 to May 31, 2008)

- The pricing schedule for financing medical services as laid down in the state standards.
- The pricing schedule for financing medical services as laid down in the contracts with health service providers
- The register of citizens below the poverty line which were regarded as beneficiaries of the poor's program in 2006-2008.
- Trends in the total number of beneficiaries in 2006-2008.

3. Private insurers participating in the 2007-2008 programs for the poor:

- Statistical and financial data on the implementation of the 2007 pilot program for the poor reported in the format defined by the Health and Social Programs Agency.
- Statistical and financial data on the implementation of the 2008 program for the poor reported in the format defined by the Health and Social Programs Agency.
- The entire list of insured beneficiaries by insurance companies.

The obtained statistical data includes information on each insured accident as well as on each beneficiary of the program. To ensure privacy, these data didn't contain first and last names of beneficiaries, their ID numbers and other kinds of individually identifiable information. A unique code assigned by State Subsidy Agency was used to identify beneficiaries under this project.

2. Insurance Coverage

Government Decree #92 as of April 8, 2008 on "laying down the terms of the insurance voucher to be dispensed to the population below the poverty line for health insurance" defines the amount of the insurance coverage for the 2008 poor's' program:

The terms of the health insurance financed by the insurance voucher provide for the following healthcare costs to be covered by an insurer:

A. costs of outpatient health services not covered by the State PHC Program being in force during the insurance period and approved by the state budget of the respective year:

a. a. Emergency outpatient services;

a. b. Services provided by a family doctor, a nurse as well as specialist consultations and other medical services, including home medical care when needed;

a. c. Ultrasound and X-ray examination prescribed by a doctor, lab-testing and instrumental investigations for a planned hospitalization;

a. d. Medical investigations of the disabled needed for a medical-social expertise, particularly for granting the status of a disabled person (except for high-tech investigations – CT and MRI);

b. Costs of inpatient services:

b. a. Emergency inpatient services, including hospital admissions of patients with pregnancy complications;

b. b. Elective surgical operations; the annual insurance limit is 15 000 GEL per insured;

b. c. Co-payment expenses not covered by the other state healthcare programs being operational during the period of the insurance coverage and defined by the state budget of the respective year, namely: covering 30% co-payment provided for by the state program for cardiac surgery for the poor beneficiaries residing in Tbilisi and Imereti Region and having signed a contract with an insurance company under the 2007 state program;

b. d. Costs of chemotherapy and radiotherapy; the annual insurance limit is 12 000 GEL per insured;

c. Costs associated with labor and delivery – 400 GEL

2. According to the terms of health insurance costs of the following health services shall not be covered by the insurance voucher:

a. Costs and services covered by other healthcare programs being operational during the insurance period and provided for in the state budget or budgets of local self-governing entities;

b. Planned inpatient treatment of therapeutic patients;

c. Treatment without medical indications, treatment not prescribed by a doctor, self-treatment

d. Medical treatment abroad;
e. Sanatorium-and-spa treatment;
f. Aesthetic surgery and cosmetic treatment;
g. Costs of treatment of sexual disorders and infertility
h. Costs of treatment of AIDS and chronic Hepatitis;
i. Costs of medical services which are needed because of an inflicting self-injury, committing terrorist or criminal acts or being under the influence of drugs.
j. Costs of organ transplantation and implantation of exo-prosthesis.

As for the coverage under the similar program for 2009, some changes are expected in it, though the final decision on these changes hasn't been made by the moment of compiling this report. It is known that the main components of the coverage will remain unchanged. Hence, the core function of the program will also remain unaltered. The program beneficiaries will be eligible for emergency health services, elective surgical and obstetric care, therapeutical services for oncological diseases; financial coverage will also be provided for medical investigations and PHC services in relation with each of the above mentioned services. In addition, beneficiaries of the poors' program can make use of services provided by the other state programs except for state programs providing for financing emergency care, oncological care and cartiac surgery care services for general population.

There are two types of changes which may be implemented:
1. Beneficiaries of the poors' program are at the same time beneficiareis of the PHC state program and receive ambulatory services from the PHC program predominantly. It's inefficient approach in general (gives rise to certain technical problems and redundancy in allocating the state budget), however the state PHC program will need restructuring to transfer the entire PHC component to the poors' program. It will be possible if MoLHSA restructures state health programs by 2009.

2. Like PHC services, certain services under the state program for communicable diseases may also be transferred to the poors' program (e.g. such nosological groups as hepatitis, amoebiasis and so on). This change, in turn, will also necessitate redesigning the state programs.
Obviously, either of these changes will exert influence on the amount of the insurance premium. Preliminary estimates show that if both of these changes are introduced the premium will rise by 10-12% approximately.

Since the decision about introducing the above mentioned changes hasn't been made by the moment of compiling this report the insurance premium contained in this report matches the insurance product provided for by the 2008 poors' program. Therefore, the given premium will be adequate if the above mentioned changes are not implemented.

3. Assumptions

The assumptions based on which the actuarial calculations were made are presented below. Accordingly, all the results contained in this report will be valid only if the assumptions are met.

Assumption 1. For the part of the program dedicated to outpatient and elective inpatient services insurance companies will sign respective contracts with individual healthcare providers which will ensure provision of health services to beneficiaries. For the part of the program dedicated to emergency care respective services will be available based on the principle of geographic accessibility.

It is a common practice for either private or public insurers participating in the program implementation. The main aim of such a practice is to restrain medical prices within particular limits. If most of the health services are not provided by healthcare providers that are contracted by insurers healthcare costs will soar up. As a result, a considerable adjustment of the calculated tariff will be needed.

<u>Assumption 2.</u> Administrative and acquisition costs along with a planned profit's component in the insurance premium are defined as 20% of the gross premium.

The main goal of this project is to estimate the so-called risk-premium (i.e. net premium). In order to obtain the gross premium certain loadings - administrative and acquisition costs and a profit margin, should be added to this amount. The loading amounted to 15% of the gross premium in the poor's' programs for 2007 and 2008 and this amount turned out not to be enough to cover all expenditures. According to the insurers participating in the implementation of the poor's' program in 2008, average acquisition expenses of each of these companies were about 15% during this year, while administrative expenses turned out to be quite variable and amounted to approximately 5-10%. On the other hand, it should be taken into account that some 700 000 people signed insurance contracts over the period of 2007 and 2008 and these contracts will be extended automatically in 2009 provided that beneficiaries themselves don't express the wish to switch to another insurance company. Acquisition costs have to be borne in full also for households that were not regarded as beneficiaries of the poor's' program in 2008 and will be added to this list by 2009 (it will happen if the government increases the number of beneficiaries or because of updates in the database of the Agency for Social Subsidies).

The loading of 20% as it is given in this assumption is essentially a matter of consensus and consultations on the issue may be continued in the future as well.

<u>Assumption 3.</u> The risk premium of the program's inpatient and so-called "secondary outpatient service"components be defined as the 90th quantile of corresponding aggregate loss probability distribution function.

Services provided for by the program may be divided in two groups according to the financing methodology used by insurers. Primary outpatient services financed on the basis of capitation or some other similar principle will fall into the first group implying transferring the risk in full or in part to a service provider. The rest of services (inpatient and "secondary outpatient" services) financed through the classical principle of insurance reimbursement will fall into the second group. In financial terms the second group comprises a large part of the insurance premim and this paper is mostly dedicated to its calculation. A mathematical model is developed and a probability distribution of losses is determined for this part. Premiums for 2007 and 2008 poors' programes were calculated exactly on the basis of 90^{th} quantile principle.

<u>Assumption 4.</u> Each company participating in the 2009 Poor's' Program will obtain "sufficiently large" insurance portfolio.

The size of an insurance premium is largely dependent on changes in the number of insured. The number of beneficiaries of the 2008 Poor's' Program is up to 800 000. If the portfolio of this size is used for calculating the premium it will ensure a well balanced virtual portfolio for the entire insurance market. Nevertheless, at the same time some companies participating in the program may still face substantial losses based on the size of the portfolio at their disposal. The insurance premium presented in this paper ensures loss rate stability if a virtual portfolio is split up into moderate chunks. A "sufficiently large" portfolio in this assumption means a portfolio of up to 10 000 insured person. If a company owns a smaller portfolio purchasing of an adequate reinsurance is recommended.

4. Basic Insurance Premium

This chapter presents a brief description of methods used for calculating a basic insurance premium and obtained results.
When calculating an insurance premium its following traditional structure is considered: a premium consists of two major components – a risk-premium (net premium) and loadings. The risk-premium serves for covering insurance liabilities, while loadings account for various direct and indirect expenses and a profit.

Essentially, the goal of actuarial calculations was to estimate the size of the risk-premium. As for the second part of the premium – the loading, its amount is generally contingent upon a tariff policy of an individual insurer and, in our case, depends on the government's position as well. A standard assumption about the size of the loading (20% of the gross premium) and its justification was given in chapter 3.

The insurance premium was calculated based on a one-year-long insurance period.

Essentially, determining an insurance premium means calculating a so-called insurance fund which has to highly probable to be sufficient for covering annual total losses of a given insurance portfolio. A state program's budget is determined based on this amount exactly. To make it convenient and draw comparisons with the budget of the similar program in 2007, obtained size of the insurance fund can be divided by the number of beneficiaries which gives us a basic premium i.e. the program's budget per beneficiary.

To this end, it became necessary to break up the insurance coverage in two components. For convenience we denote them as an *inpatient care* and an *outpatient care*. These titles are conventional. The exact meaning of the coverage components is the following:
Inpatient care – includes chemotherapy, radiotherapy, and surgical operations as well as all related laboratory and instrumental investigations (pre- and post-operational lab-tests and investigations to make a precise diagnosis). Statistical data pertinent for this component was provided in a disaggregated form by both the HESPA and private insurers – i.e. separate records were available for every case.

Outpatient care – implies so-called primary outpatient services; this component is basically an addition to the state PHC program. Its share in the amount of insurance losses is relatively small (approximately 15%). Insurers use a quasi-capitation method of financing to reimburse such services. It implies almost entire transfer of an insurance risk to a service provider. Because of this, in general, an insurer doesn't require a detailed report on a fulfilled work from a service provider, alternatively such a report is not filed in an electronic form (detailed report means a full description of individual cases).

Such decomposition of insurance coverage was performed because of peculiarities of the available statistical data rather than with the intention to analyze prices of components.
The annual basic premium calculated this way amounts to 151.68 GEL (12.64 GEL monthly). The structure of the premium by components is as follows:

Component	Amount
Inpatient care	7 GEL
Outpatient Care	1.2 GEL
Appreciation of medical costs	15%
Administration, acquisition, profit	20%

Here is the formula to calculate the gross premium:
$P_G = (P_I + P_o) * (1+i)/(1-e)$
Where,
P_G is the basic gross premium
P_I is the risk-premium for inpatient care
P_o is the risk-premium for outpatient care
i is the rate of appreciation of medical costs
e is the share of administrative and acquisition costs and a planned profit in the premium

The risk-premium for inpatient services was calculated using a stochastic model. To this effect, incidence rates of insurance claims and total losses per insurance policy have been analyzed individually. The incidence rate of claims means here the number of insured persons who experienced at least one insured accident during the insurance period; the average incidence rate was 18%, which is a rather higher than in previous years. A binomial distribution was used for modeling incidence rate. The average incidence rate and number of insured were picked as parameters for the distribution (the premium is estimated for an insurance portfolio of 10000 insured to account for the risk resulting from a variation of a number of insured in different insurance companies. See the assumption on “sufficiently large portfolios” in chapter 3). As for total losses per insurance policy, their conditional mean was (given that the loss took place) 350 GEL with a standard deviation of 913 GEL. Data were divided into several groups to model this component. A mix of shifted lognormal distributions was chosen as a model of its probability distribution. A computer simulation was used for modeling total losses. The insurance fund corresponding to this component of the premium is the 90^{th} quintile of the distribution function of total annual losses.

Since the available statistical data don’t allow performing a stochastic analysis a deterministic model was used to calculate ***risk premium for outpatient services***. Data on payments made by insurance companies to primary healthcare providers served as a main basis to calculate this component. The nominal amount of this component haven’t changed compared to the similar data in 2008, however it was adjusted for inflation as in case of the inpatient component.

The rate of appreciation of medical costs is not an inflation rate of medical costs throughout the country but rather it is data about an increase of medical costs for the organizations implementing the Poor’s program. The absence of a unified medical classifier in the country makes it very difficult to calculate these data. It required picking the most common and expensive services from the statistics of losses and estimate the inflation rate on their basis using the Laspeyres formula. Selected data comprises approximately 30% of the total expenditure – the fact that obviously lowers its validity. The data was adjusted for inflation for the period of 1.5 years since the data available to the working group covers the first half of 2008 and the insurance premium for the 2009 poor’s program will remain in force to be used for one-year insurance contracts to be signed mainly in the second and third quarters of 2009.

The assumption about **administrative and acquisition costs and a loading for a planned profit margin** was given in chapter 3. It is increased by 5% compared to poor’s programs implemented in previous years.

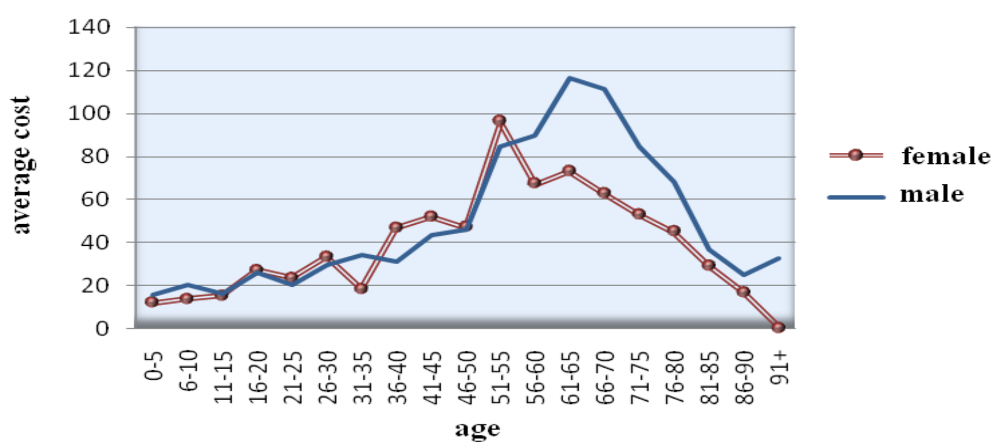
5. Analysis of Risk-Factors

To expand the basic premium into tariff classes the working group has performed an empirical analysis of the available statistical data by different categories. This chapter provides the main findings of the analysis and recommendations on the tariff classes. A primary goal for working out a tariff schedule is to assign an adequate insurance premium to certain risk-groups. As it is evidenced by the ongoing poor’s program the absence of tariff classes exerts considerable influence on the acquisition policy pursued by the companies participating in the program. For example, the fact that there are no diversified tariffs determined for Tbilisi has made the insurance companies to compete with each other more intensely in the regions (as the statistical data shows the amount of losses is higher in Tbilisi than in the rest of Georgia).

The data for 6 months of 2008 which are extracted from statistical reports on the inpatient component of the pilot poor’s program and cover only Tbilisi and Imereti region have been used for analysis.

First of all, the pattern of losses by age and sex is a matter of interest. Under the current terms and conditions of the program the premium is defined for two age-groups; these groups are: 0-64 years and more than 65 years old people. Based on the picture of losses it is desired to increase the number of these groups. The figure below shows the average cost per insured by age-group and sex.

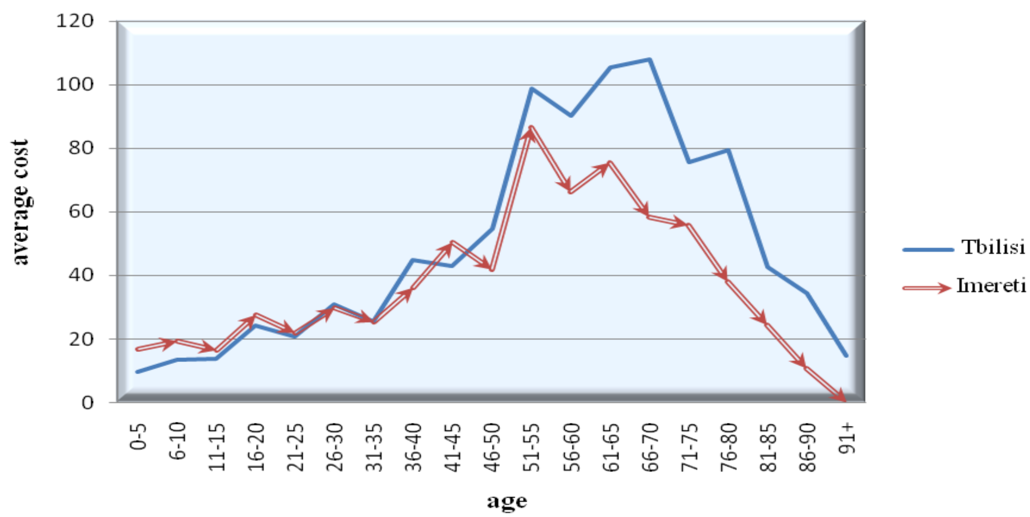
Average cost per insured by age and sex



Based on such a pattern we feel it rational that there are 3 tariff classes by age-groups: 0-50y, 50-75y and above 75y. A decrease in average expenditures at older ages is due to the characteristics of the program's design – the most part of losses are attributed to surgical operations, which are, indeed, rare in higher age-groups. As for expenditures by sex, they almost equal in groups of 0-50 and 75+ years old people, but differ substantially in the most costly class of 50-75 years old.

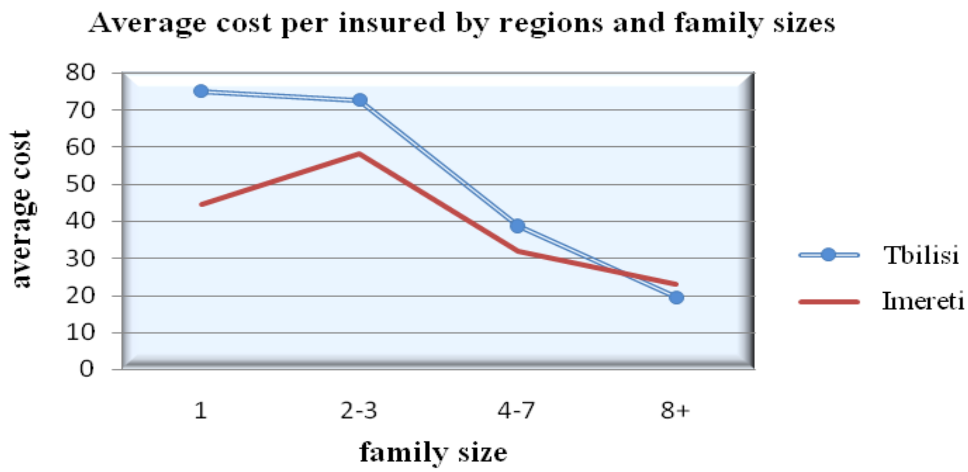
As for regional differences, they are perfectly shown on the following graph:

Average cost per insured by regions and age

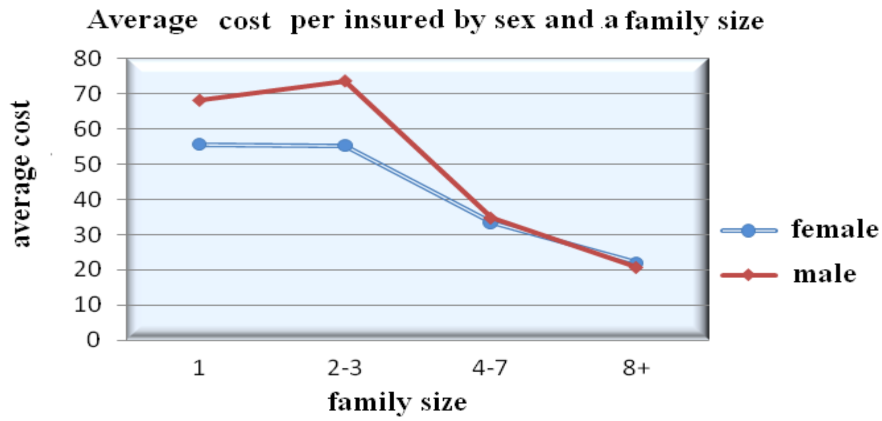


Considerable differences among regions arise in the group of 50-75 years old; differences are also noticeable in the third age-group.

The comparison of losses by a family size yielded interesting results. Due to the fact that the poor's program provides for a family insurance and doesn't allow for insuring individual family members, emergence of tariff classes on the basis of family characteristics may have significant impact on the acquisition strategies employed by insurance companies. For example, if we consider the amount of losses by a family size (number of family members) and regions (Tbilisi, Imereti) we obtain the following picture:



It's also an interesting picture in the distribution of losses by sex and a family size:



On the basis of this empirical analysis several options for tariff parameters have been determined and used to develop tariff groups. Respective equalizing insurance premiums are presented in the next chapter.

Average expenditure per hospital case by age

The main problem with calculating this indicator was a definition of an "inpatient case" associated with the specificities of a health insurance.

In terms of insurance one hospital case frequently yields several records falling into the claims database. It means that codes of several diseases associated with a given surgical intervention may be assigned to one case of hospitalization. Below are given examples, where two records were made for individual patients:

1. Hydatid disease (liver, lung, pancreas – surgical treatment)
2. Internal standard (exceptional cases)

1. Tendon injury (plastic surgery)
2. Nerve injury (plastic surgery)

It is also possible that a patient will need laboratory investigations before (or after) a surgical operation and the case is registered under a separate service code. It is not clearly defined anywhere whether such cases are considered to comprise one case of a hospital care.

Therefore, the working group had to make several assumptions while calculating a mean cost of an "inpatient case":

Assumption 1:
An instance of a hospital admission of an individual has been regarded as one hospitalization and all disease codes and associated costs have been assigned to that single case of an inpatient treatment.

On the other hand, the statistical data available to the working group covered different time periods and the reporting was also different depending on which organization was in charge of managing the program.

The situation was further complicated by the unavailability of a unified medical classifier, which made it difficult to recognize the overall picture in medical terms. Even the data provided by the Health and Social Program Agency was recorded different ways in different years since they updated their disease classifier several times.

The Agency data referring to the period from 2006 / VI to 2007 / III was chosen as a basis to work with. This information was elaborated in details in medical terms by the working group earlier and case categories for each case were readily available there.

All inpatient cases (planned hospitalization, oncological surgery, urgent hospitalization, cases of critical care, obstetrics and so on) and all outpatient cases associated with a hospitalization (oncological diagnostics, ambulatory lab investigations associated with a planned hospitalization and so on) have been selected from this database.

Data on categories of inpatient care (without categories of outpatient care) have been considered to define a case of a hospital care. Despite such a selection even this database required additional refinement since it was common that a patient received services under 3 or 4 disease codes. For example:

1. Hematemesis
2. Brain infarction (critical state)
3. Internal standard (exceptional cases)
4. Hypovolemic shock (hemorrhagic)
5. Internal standard (exceptional cases)

1. Unstable stenocardia
2. Left cardiac insufficiency (acute)
3. Cholelithiasis with acute cholecystitis
4. Acute and sub acute hepatic insufficiency (encephalopathy)

Even the fact that an individual was recorded 10 times with the database with different dates of hospital admission and discharge has been discovered.

Therefore, it become necessary to make an assumption on after what period a case could be regarded as a new one rather than a continuation of the previous admission.

Opinions of medical experts were sought to deal with the problem. According to them even clinically exacerbated cases are resolved within an interval from two weeks to a month. Cardiac and chronic cases are characterized by a markedly prolonged period of a hospital stay.

Thus, a grouping of the database at hand was performed in terms of intervals between a date of discharge in one case of hospitalization and a date of admission in another case of inpatient treatment.

Finally, we've got the following picture:

Number of days	0	7	14	21	28	35	42	60	350
Continuation of an old case	**1307**	**1601**	**1791**	**1929**	**2059**	**2192**	**2280**	**2498**	**2917**
New case	**18952**	**18658**	**18468**	**18330**	**18200**	**18067**	**17979**	**17761**	**17342**

Out of these the interval in 20 days was regarded as the most acceptable one and overall data were grouped according to such a definition of a "new" hospital case (typically, different disease codes which were assigned to a patient at the same day have been attributed to the same inpatient case).

To clarify the picture, 500 cases have been sampled randomly from the dataset grouped the way described earlier. Each case has been analyzed in clinical terms to find out whether or not cases which were regarded as "new cases of hospitalization" for the same patient were actually continuations of old cases of hospitalization.

5 types of cases have been found from the analysis:

1. a new case — Cases differing by a disease code, a type (urgent / planned) or a clinic

2. continuation of an old case — The same / similar disease codes or successive cases where treatment was continued at the same clinic (e.g. urgent + critical state)

3. hard to tell apart; generally, a continuation of an old case is suspected — Disease code indicates to a critical state or an exceptional case. Therefore, it's difficult to discern whether or not it is a continuation of the initial case

4. Cardiac and 5. asthmatic cases — Disease codes corresponding to cardiac and asthmatic cases were extremely common; acute insufficiency or stenocardia and a following operation were the most typical under cardiac disease category.

All five types of cases comprised 27..20% of the selected 500 records.

#	**Type of a case**	**Number of cases (instance)**	**% rate**	**Time passed from the first case of hospitalization**
1.	New case	46	9.20%	Typically more than a month
2.	Continuation of an old case	22	4.40%	The same month, extends to the following month ≈ 20%
3.	Hard to distinguish; generally, a continuation of an old case is suspected	16	3.20%	The same month, extends to the following month ≈ 12%
4.	Cardiac cases	40	8.00%	Generally, 2-3 months
5.	Asthma	12	2.40%	Generally the same month
	Sum	136	27.20 %	
	Total	500	100%	

To ensure greater validity "new inpatient cases" that where 30 days apart from each other were analyzed and the same 500 records were examined by hand. We didn't get much difference. There were only 2-3 discrepancies found.

This is pretty apparent from the column "Time passed from the first case of hospitalization" of the table:

– Time between “new” and preceding cases of hospitalization was already more than a month. Thus increasing the interval by 10 days couldn’t have any effect on this category;
– Out of suspected and old cases only several of them fell into the increased interval;
– Cardiac cases were marked by greater length of stay right from the beginning; and
– There was a small number of asthma cases in this sample right from the beginning;

In absolute numbers the difference between 20- and 30-days-long intervals was 191 cases out of 20260, which was a rather small share.

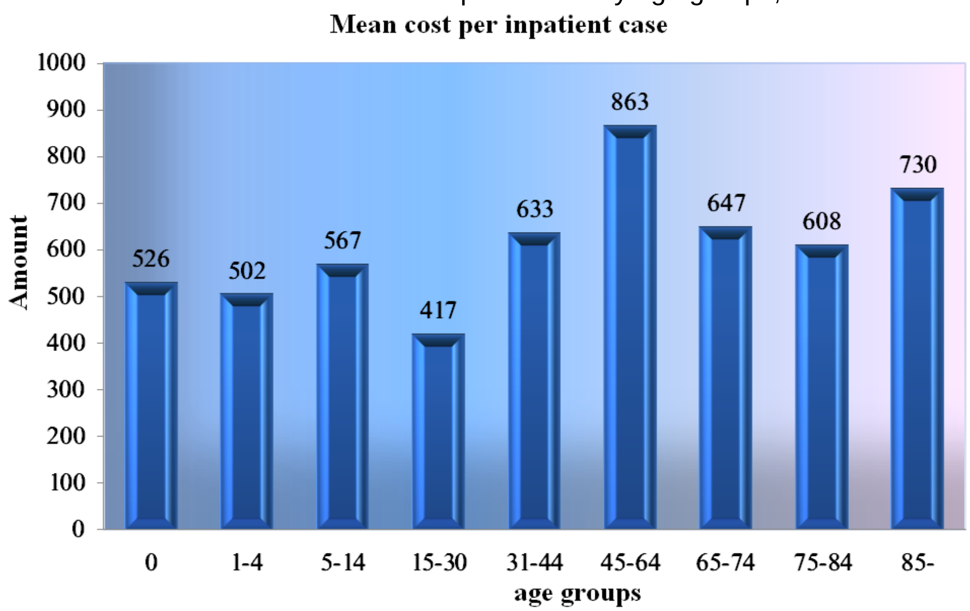
Hence, it was necessary to make the third assumption:

<u>Assumption 3:</u>
A case which was 20 days apart from the date when the previous case of hospitalization ended was taken as a new case of hospitalization.

To reconstruct the whole picture the hospital data defined in such a way was combined with the outpatient data of the respective period.

In the result we’ve got that a mean cost per case was 63639 GEL.

As for the mean cost of a case of hospitalization by age groups, it took the following form:



However, for better precision two extreme cases of recognizing a case as an “inpatient” one have been also analyzed:

1. Lower limit – any hospital admission was regarded as a separate hospital case (i.e. if there were 5 disease records for a patient then 5 hospital cases were recorded)
2. Upper limit – minimal – one individual registered with the database was regarded as one case of hospitalization despite the number of his or her admissions to the hospital.

Respectively, a maximal number of hospital cases were recorded in the first case. Since the total amount of losses hasn’t changed a per-case cost was small here. On the contrary, in the other case the mean per-case cost was big:

#	Age-groups	Upper limit	Lower limit
1	0	551.44	526.37
2	1-4	580.37	497.78
3	5-14	618.76	557.30
4	15-30	434.98	697.49
5	31-44	687.43	695.58
6	45-64	986.79	816.01
7	65-74	737.15	615.51
8	75-84	677.44	590.08
9	85-	813.63	701.11

6. Insurance Premium by Tariff Classes

Based on the analysis of the risk-factors presented in the previous chapter the working group developed several recommendations on factors of tariffication. These factors are: age, sex, family size and region. There are only two tariffication factors in the 2008 poor's program: for the insured under 65 years of age and above 65 years of age. In general, it's clear that the number of tariff classes should be increased, since the inflexible tariff structure creates an apparent imbalance in a risk distribution. Existence of a diversified and equitable insurance premium is one of the important principles of private insurance and its violation leads to creating practical problems. On the other hand, by changing the format of tariff classes one can exert a great influence on the course of the program.

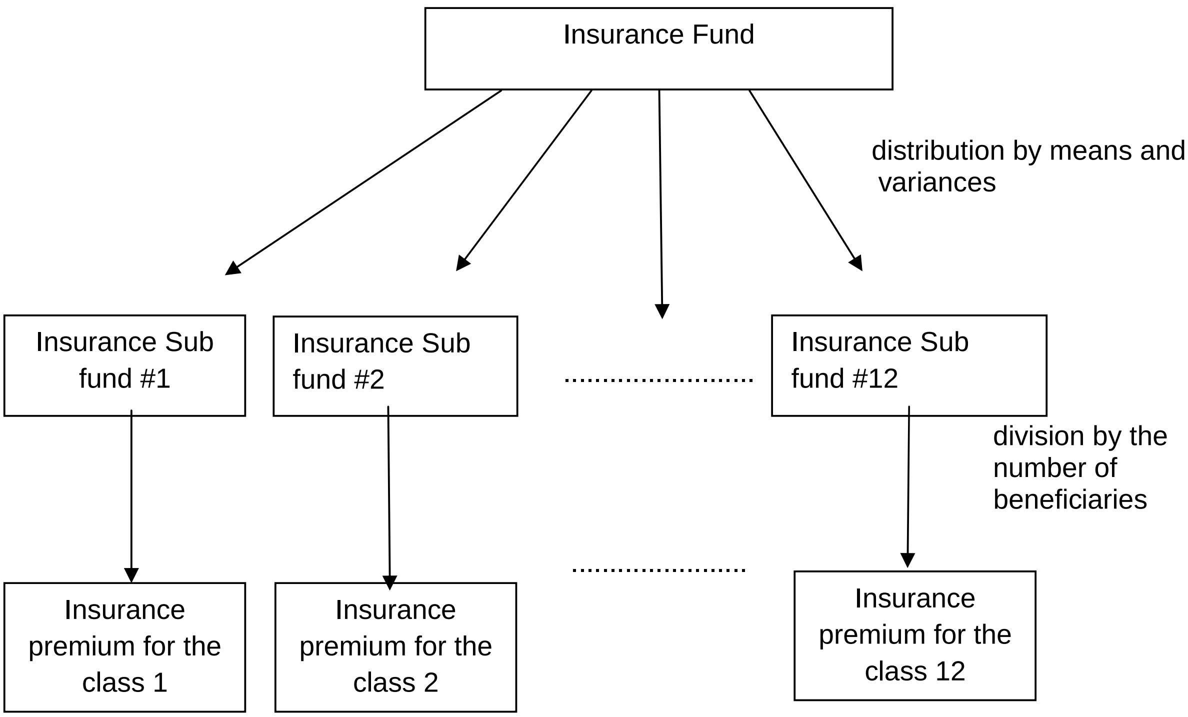
For example, emergence of regional tariff coefficients implies the existence of a diversified premium for Tbilisi (or for Tbilisi and other big cities) which, for sure, will have an influence on the acquisition activities of insurers in the regions. Its absence leads to an increased activity of the insurance companies in the regions and their inactivity in Tbilisi. It is clearly seen from the statistics, that losses are greater with the urban population than with the rural population. It is not a specific characteristic of Georgia. The same picture emerges also from any other country's data. On the one hand it is a positive development that insurance companies intensify regional activities, however, on the other hand it can lead to the situation when some companies may find themselves in a disadvantageous position compared to the other companies.

Existence of diversified premiums for different risk-groups is a common practice in countries whose health care systems are based on private insurance companies (e.g. Netherlands). However, in most cases it is a mandatory insurance which ensures availability of advance payment schemes and pooling mechanisms. In such a case, the assignment of a diversified premium is associated with much greater difficulties, since individuals who find themselves in high-risk groups will experience difficulties with paying respective insurance premiums (in a system of mandatory insurance, indeed, policy holders themselves are paying for insurance). It requires creating so-called "risk equalizing schemes" which operate based on the following principle: an insurance premium which citizens pay is not diversified by risk-groups, however this diversification takes place after individuals get insured. Risk equalizing scheme ensures equitable distribution of the premium among insurance companies depending on what kind of portfolio they end up with. I.e. despite the fact that individuals pertaining to different risk-groups pay equal premiums insurance companies get diversified premiums corresponding to actual risk profiles.

The situation is much simpler with the poor's program. There is no need in creating complex instruments of risk equalization since only the state pays for insurance. In such a situation it is possible to assign diversified premiums directly, not to create troubles for program beneficiaries and fully meet the principle of equitable insurance premium.

It is necessary that a proper consideration be given to the importance of the availability of a pricing schedule and its gradual elaboration. It is recommended to introduce tariff classes for age-groups and sex in 2009 and to add extra factors such as region and family size in 2010. As the quality and adequacy of data improves it will be necessary to revise and update tariff classes periodically during the first several years. After a while an optimal pricing schedule can be introduced ultimately. Some other parameters, such as a chronic disease existence can be added to the pricing schedule later.

Four options for a pricing schedule with varying number of tariff classes are given below. Mean annual aggregate loss and its standard deviation were computed for each of these groups. Corresponding insurance funds have been calculated in proportion to these amounts based on the following principle: the total insurance fund has been decreased by corresponding amounts of the mean annual losses. The remaining part of the fund was distributed among classes in proportion to respective standard deviations of the annual aggregate losses in these classes. To derive an insurance premium from the resultant insurance sub funds these funds have been divided by the number of insured falling into each of these classes. Schematically the process looks as follows:



This principle was used for the component of inpatient care. As for the component of outpatient care due to the lack of the appropriate statistical data the findings of the household health expenditure survey were used to distribute the insurance premium among classes. Administrative and acquisition costs along with a profit margin have been distributed uniformly among the insured.

The following tables of insurance premiums have been developed using this method:

Pricing schedule 1 – factors: age and sex

Tariff class	*Male Age 0-49y*	*Male Age 50-74y*	*Male Age 75 y and above*	*Female Age 0-49y*	*Female Age 50-74y*	*Female Age 75 y and above*
Monthly insurance premium	8.92	21.94	19.12	9.14	15.55	12.55

Pricing schedule 2 – factor: age

Tariff class	*Age 0-49y.*	*Age 50-74y,*	*Age 75y. and above*
Monthly insurance premium	9.04	17.86	14.72

Pricing schedule 3 – factors: age, sex, family size

Tariff class	*Male Age 0-49y 1-3 family members*	*Male Age 50-74y 1-3 family members*	*Male Age 75 y and above 1-3 family members*	*Male Age 0-49y 4 and more family members*	*Male Age 50-74y 4 and more family members*	*Male Age 75 y and above 4 and more family members*
Monthly insurance premium	12.66	23.63	20.33	7.22	19.45	39.99
Tariff class	*Female Age 0-49y 1-3 family members*	*Female Age 50-74y 1-3 family members*	*Female Age 75 y and above 1-3 family members*	*Female Age 0-49y 4 and more family members*	*Female Age 50-74y 4 and more family members*	*Female Age 75 y and above 4 and more family members*
Monthly insurance premium	13.78	14.57	12.05	7.64	15.86	16.39

Pricing schedule 4 – factors: age, sex, region

Tariff class	*Male Age 0-49y Tbilisi*	*Male Age 50-74y Tbilisi*	*Male Age 75 y and above Tbilisi*	*Male Age 0-49y Outside Tbilisi*	*Male Age 50-74y Outside Tbilisi*	*Male Age 75 y and above Outside Tbilisi*
Monthly insurance premium	9.21	28.66	36.78	8.05	19.24	19.25
Tariff class	*Female Age 0-49y Tbilisi*	*Female Age 50-74y Tbilisi*	*Female Age 75 y and above Tbilisi*	*Female Age 0-49y Outside Tbilisi*	*Female Age 50-74y Outside Tbilisi*	*Female Age 75 y and above Outside Tbilisi*
Monthly insurance premium	9.14	18.31	19.11	8.76	12.95	10.22

Annex 1. Mathematical Description of the Model

I. Statistical data analysis

As we've noted above, insurance industry has already amassed quite a bit of the data on medical insurance of the poor and therefore the following analysis, unlike previous reports, is grounded exactly on such type of data. In particular, we've analyzed amount of expenditures on a medical assistance to 70000 insured persons during the last year. According to these data, in total there were 11457 insurance claims and overall spending amounted to 3 076 055 GEL.
Despite the fact that costs (a "loss") per individual claims are rather non-uniform and fall in a sufficiently broad range (from 2 to 1878 GEL) it wasn't our main goal to estimate loss distribution by uniform classes of claims at this stage (as we did it before); on this occasion we decided to analyze the amount of incurred expenditures per insured person (rather than per claim) as our main goal.

It's obvious that some insured didn't file a claim at all; there are people who addressed a company only once in a year, however there are customers who had to (and will have to) apply more than once. Therefore it's interesting to analyze a number of claims and total annual expenditures per insured person. To this end, we sort out individual claims out of their total number (11 457) and formed more or less uniform groups based just on this criterion. In the result we obtained the picture presented in the table below (see Table 1). Te table shows that the first group – insured persons who filed a claim only once – is the most sizeable, followed by the group of insured who applied to the company twice and so on. There are almost no customers who required (will need) assistance more than 6 times per year.

Table 1. Annual number of claims filed by insured

Number of claims per year *N*	**Number of insured**	**% of insured**	**Number of claims**	**% of insured**
1	6738	76.68%	6738	58.81%
2	1555	17.70%	3110	27.14%
3	390	4.44%	1170	10.21%
4	84	0.96%	336	2.93%
5	17	0.19%	85	0.74%
6	3	0.03%	18	0.16%
Total	8787	100%	11457	100%

According to this table, the average number of claims was μ_N = 1.304, with a standard deviation σ_N = 0.6227. We tried to apply an appropriate probability model and it's worth noticing that due to a rather big overall volume of data, 8787, χ^2 criterion refused to comply with data of traditional, poisson, binomial and negative binomial models. The admissible model has been found among a so-called "replicable" class of distributions of the following type:

$$p_n \equiv P\{N=n\} = \begin{cases} 1-\Gamma(t;\alpha;\beta), & n=1 \\ \Gamma(t;(n-1)\cdot\alpha;\beta)-\Gamma(t;n\cdot\alpha;\beta), & n\geq 2 \end{cases} \qquad (1)$$

Where positive t, α and β parameters assumed the following values:

$$t = 0.1,\ \alpha = 0.219 \text{ და } \beta = 100.35. \qquad (2)$$

Here $\Gamma(t;\alpha;\beta)$ is a gamma distribution function of the form:

$$\Gamma(t;\alpha;\beta) = \frac{\int_0^{t/\beta} u^{\alpha-1}\cdot e^{-u}du}{\int_0^{\infty} u^{\alpha-1}\cdot e^{-u}du}. \qquad (3)$$

Let's recall that statistics of the □ criterion has the following form:

$$X_n^2(k) = n \cdot \sum_{i=1}^{k} \frac{(f_i - p_i)^2}{p_i}, \qquad (4)$$

Where n is a total number of observations (here n = 8787), k is a number of categories (summands) (in our case, k = 6); as for f_1, f_2,..., f_k – they are relative frequencies of observations falling into these categories (in our case, vector of numbers in the third column of table 1) and p_1, p_2,..., p_k are respective (theoretical) probabilities of falling into these categories (here p_1, p_2,..., p_5 are the probabilities computed by the formula (1): and $p_6 = 1 - (p_1 + p_2 + ... + p_5)$). As it is known, if a model of the random variable is appropriate, (4) should have χ^2-distribution with $d = k - 1 - m$ degrees of freedom, where m is a number of parameters under evaluation (in our case, m = 3). The procedure of computing χ^2 statistics using the formula (4) is presented in the following table:

Table 2. Computation of *chi-squared* statistics

Number of claims per customer per year ***i***	**Relative frequencies of customers** f_i	**Probabilities of the model (1)** p_i	**Summands of χ^2 statistics**
1	0,7668	0,7590	0,00008
2	0,1770	0,1863	0,00047
3	0,0444	0,0429	0,00005
4	0,0096	0,0094	0,00000
5	0,0019	0,0020	0,00000
6	0,0003	0,0005	0,00005
Total	1,000	1,000	8787*0.00065 = 5.73

As we see, the value of □ statistics is 5.73, which doesn't exceed the upper quantile of □ -distribution for 0.95 with $d = k - 1 - m = 6 - 1 - 3 = 2$ degrees of freedom the value for which is 5.99 according to statistical tables for χ^2-distribution. From here we infer that the model (1) is consistent with the claims data in Table 1.

Now, let's move on to a cost analysis. Let's start with the fact that there are only 20 observations in the last two categories of claimants (group 5 and group 6) in table 1. Therefore, we, naturally, combined these categories with the previous one and named the resultant group as "≥ 4", now including 104 observations. In the result, table 1 will change as follows:

Table 3. Number of claims by groups

Number of claims per year *N*	**Number of insured**	**% of insured**	**Number of claims**	**% of insured**
1	6738	76.68%	6738	58.81%
2	1555	17.70%	3110	27.14%
3	390	4.44%	1170	10.21%
≥ 4	104	1.18%	439	3.83%
Total	8787	100%	11457	100%

Let's begin analyzing groups 1, 2, 3 and ≥ 4 presented in table 3 by considering numerical characteristics of expenditures in these groups as it is shown in the following table:

Table 4. Numerical caracteristics of annual per patient expenses by groups

Group characteristic	**1**	**2**	**3**	**≥ 4**
Number	6 738	1 555	390	104
Total	1 676 388	869 169	331 371	199 127
Min	2	4	12	25
Max	10 883	12 000	10 360	22 044
Mean	249	559	850	1 915
Standard deviation	714	1094	1276	2 971
Variation Coeff.	2.9	2.0	1.5	1.6
Skewness	6.9	4.5	3.6	4.1
Kurtosis	59.4	27.7	17.8	22.2

Before turning to describing the statistical analysis of individual groups we note that the number of insured is diminishing across these groups and a pattern of this reduction is determined exactly by the model (1) described above. As for the other characteristics, their relation with the numerical values is best illustrated by the following table of correlation and determination coefficients (squared percentage values of determination coefficients):

Table 5. Correlation and determination coefficients for group characteristics and numerical values

Group characteristic	**Coefficient of**	
	correlation	**determination**
Total	0,970782	94%
Min	-0,697131	49%
Max	-0,451409	20%
Mean	-0,710488	50%
Standard deviation	-0,645328	42%
Variation Coeff.	0,989894	98%
Skewness	0,984371	97%
Kurtosis	0,991091	98%

As we see, for the characteristics which are positively correlated with numbers (total, variation, kurotsis and Skewness) determination coefficients are sufficiently high. In terms of regression terminology, say for the total, it can be understood the following way: 94% of total expenditures may be explained by the number of summands comprising the total. In mathematical terms it means that there exist some positive numbers *a* and *b* such that:

$$S_n = a \cdot n + b + E, \quad (5)$$

where n is a number of data items in a group, S_n _ is a total expenditure and E is a (normally distributed) random "error" with a mean equaling 0, $\mu_E = 0$, and unknown standard deviation σ_E. Of course if we know the latter and the numbers a and b it will be possible to determine a conficence interval based on a number of data items in the group with a high likelihood that a total expenditure will fall in this interval. It is possible to estimate these three numerical characteristics using methods of regression analysis, though due to a small number of data, in our case small number of groups (there are only four of them), these estimated won't be reliable. For example, in our case we might obtain that $a = 210.6$, $b = 306\ 383$, and $\sigma_E = 405.15$ which would lead to the following point estimates of total expenditures for the groups 1, 2, 3 da ≥ 4:

1 725 391, 633 863, 388 516, 328 285.

If we compare these figures with real expenditures

1 676 388, 869 169, 331 371, 199 127

we may feel the projection is not so bad, though as we've already mentioned it these estimates cannot be reliable due to a small number of data; besides there is no way to verify the main assumption of the model (5) that the variable *E* is normally distributed.

Hence, let's turn to statistical analysis of individual groups. Note that a variation is much higher in the first group compared to the others indicating that the first group is rather heterogeneous. It is due to the fact that we look at summary values (rather than individual summands) in the other groups leading to relatively "stable statistical behaviour" of the values. Therefore, despite a considerable dispersion of the data we didn't experience difficulties in finding probability models matching to these three groups (2, 3 and ≥ 4) (see. below): distribution is lognormal in all three groups with the parameters μ and σ presented in the table:

Table 6. μ and σ parameters of lognormal distributins in groups 2, 3 and ≥ 4

Group	2	3	≥ 4
μ	5,369	6,075	6,947
σ	1,487	1,297	1,263

Let's recall that a variable X is considered to follow a lognormal distribution with parameters μ and σ, if its distribution function is given by the formula:

$$F_X(x) = P\{X \le x\} = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \quad x > 0, \tag{6}$$

where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{z} e^{-u^2/2} du \tag{7}$$

is a standard normal distribution function.

As for the first group, due to a big number of data included in this group (it includes of all kinds of expenditures on all types of medical services) along with quite a large heterogeneity of these data we didn't succeed in fitting any uniform model to it. Therefore we generated "uniformity" in these data artificially by breaking the range into three parts: expenses up to 1000 GEL (6380 observations), expenses in the range from 1000 to 5000 GEL (300 observations) and expenditures exceeding 5000 GEL (58 observations). We managed to perform cost modeling in these subgroups employing the same lognormal models, though now these were so-called three-parameter shifted log-normal models with parameters d, μ and σ, which follow the following distribution function:

$$F_X(x) = P\{X \le x\} = \Phi\left(\frac{\ln(x+d) - \mu}{\sigma}\right), \quad x + d > 0. \tag{8}$$

Now let's discuss the application of models to the above-mentioned data. Like the discrete case described earlier (we imply here the model (1)) it once again means to build statistics of □ criterion

$$X_n^2(k) = n \cdot \sum_{i=1}^{k} \frac{(f_i - p_i)^2}{p_i},$$

so that it's value will be less than the admissible tabular value. In this case, frequencies $f_1, f_2, \ldots, f_k$ are now relative frequencies of $x_1, x_2, \ldots, x_n$ expenditures observed in a given group falling within disjoint intervals $[z_0; z_1), [z_1; z_2), \ldots, [z_{k-1}; z_k)$ on the number line, while $p_1, p_2, \ldots, p_k$ are respective (theoretical) probabilities of falling into these intervals, i.e. if a sought-for (fitting) distribution function is $F(z)$, then

$$p_i = F(z_i) - F(z_{i-1}),\ i = 1,2,\ldots, k. \tag{9}$$

Of course, the values that frequencies assume depend on chosen points $z_0, z_1, \ldots, z_k$, while probabilities depend not only on these points but also on a $F(z)$ distribution being selected. It's therefore clear that a smallness of □ statistics is dependent on appropriate picking $z_0, z_1, \ldots, z_k$ points and a $F(z)$ distribution and the goal of fitting is exactly to make such a choice that the value of this statistics will be less than the admissible tabular value.

As we've mentioned before, such a choice in groups 2, 3 da ≥ 4 is a log-normal distribution with parameters □ and □ given in table 6. As for the first group these are three-parameter lognormal distribution of the form (8) for the above-mentioned three subgroups with the parameters presented in the following table:

Table 7. d, μ and σ parameters of three subgroups of the group 1

Subgroup	---1000	1000---5000	5000---
d	670	-825	-1937
□	6.465	6.560	8.365
□	0.393	0,750	0.228

Finally, the distribution function for the group 1 is given as a weighted sum of these three subgroups where the weights are relative shares of data falling in a given subgroup, i.e:

$$F_1(x) = w_1 \cdot \Phi\left(\frac{\ln(x+d_{11}) - \mu_{11}}{\sigma_{11}}\right) + w_2 \cdot \Phi\left(\frac{\ln(x+d_{12}) - \mu_{12}}{\sigma_{12}}\right) + w_3 \cdot \Phi\left(\frac{\ln(x+d_{13}) - \mu_{13}}{\sigma_{13}}\right), \quad (10)$$

Where

w_1 = 6380/6738 = 0.946869, w_2 = 300/6738 = 0.044524 da w_3 = 58/6738 = 0.008608. (11)
and as it can be easily understood, subscripts of the other parameters indicate a group and subgroup number.

II. Description of cost distribution models and simulations

II.1. Distribution of annual expenditures per insured person

In order to describe distribution $C(x)$ of annual expenditures per insured let's use a total probability formula. It's obvious that:

$$C(x) \equiv P\{X \le x\} = P\{X \le x \mid A\} \cdot P(A) + P\{X \le x \mid \overline{A}\} \cdot P(\overline{A}) = F(x) \cdot P(A) + P(\overline{A}), \quad (12)$$

where X is the annual expenditure on a randomly selected insured, A is an event that he or she will file at least one claim to a company and $F(x)$ is a conditional distribution of annual expenditures per insured given that he or she applies to the company at least once, i.e. $F(x)$ is annual expenditures on individual "patient". Once again, it's obvious from the total probability formula, that:

$$F(x) = \sum_{n=1}^{3} P\{X \le x \mid N = n\} \cdot P\{N = n\} + P\{X \le x \mid N \ge 4\} \cdot P\{N \ge 4\} =$$

$$= p_1 \cdot F_1(x) + p_2 \cdot F_2(x) + p_3 \cdot F_3(x) + p_{\ge 4} \cdot F_{\ge 4}(x)$$

i.e. finally:

$$F(x) = p_1 \cdot F_1(x) + p_2 \cdot \Phi\left(\frac{\ln x - \mu_2}{\sigma_2}\right) + p_3 \cdot \Phi\left(\frac{\ln x - \mu_3}{\sigma_3}\right) + p_{\ge 4} \cdot \Phi\left(\frac{\ln x - \mu_4}{\sigma_4}\right), \quad (13)$$

Where $F_1(x)$ is a distribution explained by equation (10) and the values of □ and □ parameters are given by table 6, and p_1, p_2, p_3, $p_{\ge 4}$ probabilities are determined according to table 2,

p_1 = 0.7590, p_2 = 0.1863, p_3 = 0.0429, $p_{\ge 4}$ = 0.0119. (14)

Let's denote a probability of the event A by q. It's evident that it can be estimated as a ratio of a number of "patients" to a number of insured, i.e. in our case q = 8787/70000 = 0.12553.

So, of we summarize (10), (12) and (13) for the cost distribution $C(x)$ per insured we will get:

$$C(x) = 1 - q \cdot \left(1 - p_1 \cdot F_1(x) - p_2 \cdot \Phi\left(\frac{\ln x - \mu_2}{\sigma_2}\right) - p_3 \cdot \Phi\left(\frac{\ln x - \mu_3}{\sigma_3}\right) - p_{\ge 4} \cdot \Phi\left(\frac{\ln x - \mu_4}{\sigma_4}\right)\right), \quad (15)$$

where

$$F_1(x) = w_1 \cdot \Phi\left(\frac{\ln(x+d_{11}) - \mu_{11}}{\sigma_{11}}\right) + w_2 \cdot \Phi\left(\frac{\ln(x+d_{12}) - \mu_{12}}{\sigma_{12}}\right) + w_3 \cdot \Phi\left(\frac{\ln(x+d_{13}) - \mu_{13}}{\sigma_{13}}\right).$$

II.2. Aggregate claims distribution

It's clear, that the distribution of total claims, i.e. aggregate expenditure on all insured, is a sum of independent variables, with a distribution given by (15). Thus, the aggregate claims distribution cost will be a M-fold convolution of the distribution $C(x)$:

$$C_S(x) = C^{M^*}(x), \quad (16)$$

where M is a total number of insured (in our case, M = 70000).

Indeed, this fact though theoretically correct is of a less practical value, because we won't be able to compute not only a 70000 fold but even a two-fold convolution of a distribution (15). Therefore, we have the following way out here: simulate 70000 independent variables which follow the distribution $C(x)$ and summ them up; then simulate these sums many times, say 1000, and consider obtained results as realizations of independent variables following the distribution $C_S(x)$ and find an admissible model of $C_S(x)$ distribution using the matching procedure described above. However, even this way has ist major drawback: modelling a random variable which follows the disribution $C(x)$ requires finding an inverse function of $C(x)$ and this is yet quite a difficult task given the equation (15).

We can avoid such difficulties the following way. Since each of the M insured may apply to a company at least once or not apply at all independently from each other during a year the number of "patients" is thus a binomially distributed random variable with the parameters M and q. Therefore for a given q (as we remember q = 0.12553) and M (this parameter will be different for individual companies depending on various factors which we won't discuss here but we consider M = 5000; 10000; 20000 (in addition to M = 70000 as an exercise) let's firstly simulate the number n of future "patients". If we multiply the obtained number by the probabilities p_1, p_2, p_3, $p_{\geq 4}$ given by (14) we thus automatically simulate numbers n_1, n_2, n_3, $n_{\geq 4}$ of insured falling in groups 1, 2, 3 and ≥ 4 for which the equation $n = n_1 + n_2 + n_3 + n_{\geq 4}$ holds. Similarly, if we multiply n_1 by weights w_1, w_2 da w_3 given by (11) we thus automatically simulate n_{11}, n_{12}, n_{13} – numbers of "patients" falling in three subgroups of the group 1 – which obviously satisfy the equation $n_1 = n_{11} + n_{12} + n_{13}$. Thus, as we've elaborated the numbers of independent summands with their own probability distributions, let's simulate variables which follow all of the six kinds of distributions, considering exactly those numbers of variables that we've mentioned above.

i.e. simulate n_{11} variables with the distribution $\Phi\left(\frac{\ln(x+d_{11})-\mu_{11}}{\sigma_{11}}\right)$, n_{12} variables with the distribution $\Phi\left(\frac{\ln(x+d_{12})-\mu_{12}}{\sigma_{12}}\right)$, n_{13} variables with the distribution $\Phi\left(\frac{\ln(x+d_{13})-\mu_{13}}{\sigma_{13}}\right)$, n_2 variables with the distribution $\Phi\left(\frac{\ln x-\mu_2}{\sigma_2}\right)$, n_2 variables with the distribution $\Phi\left(\frac{\ln x-\mu_3}{\sigma_3}\right)$ and $n_{\geq 4}$ variables following the distribution $\Phi\left(\frac{\ln x-\mu_4}{\sigma_4}\right)$ and sum up all the obtained realizations. It's obvious that in the result we will get a realization of a variable following the distribution $C_S(x)$, since if we summed up each of the variables given above the total would have yielded realization of a variable with the distribution $C(x)$.

As for the simulation procedure itself, since in every case we deal with a shifted lognormal distribution (the last three cases are particular ones with d = 0), therefore to find inverses of distribution functions the following equation should be solved for x in each of the cases:

$$\Phi\left(\frac{\ln(x+d)-\mu}{\sigma}\right)=u\ ,$$

This is apparently not a problem at all. The solution has the following form:

$$x=\exp\left(\mu+\sigma\cdot\Phi^{-1}(u)\right)-d\ . \quad (17)$$

Hence, if we substitute the realization of the variable *U* which is obtained using the command "=RAND()" in an *xls* editor and is evenly distributed on the interval [0;1] for the argument *u* in the right side of the last equation, in the result we will get the realization of variable *X* with the following distribution:

$$X = \exp\left(\mu + \sigma \cdot \Phi^{-1}(U)\right) - d \,. \qquad (18)$$

The inverse function $\Phi^{-1}(u)$ of the standard normal distribution participating in the equations (17) and (18) is described by the command "=NORMSINV ()" in an xls editor.

As for the binomial variable to be simulated in the first place and to have parameters *M* and *q,* the simulation is also very simple here: it's sufficient to simulate *M* realizations of Bernoulli variables with a probability of success *q* and sum them up. This Bernoulli variable is simulated by the command „=IF(RAND() < *q*; 1; 0)".

III. Results and implications of the simulation

As we've mentioned above, we simulated models for four different scenarios, *M* = 5000; 10000; 20000; 70000 (1000 realizations for each case), which corresponds to portfolios of very small, small, middle-sized and large companies. It's interesting observation that in all of the four cases a shifted log-normal distribution proved to be a good model of $C_S(x)$ distribution. The parameters are given in the following table:

Table 8. *d*, μ and σ parameters of simulated log-normal distributions by portfolio size.

	Portfolio size			
Parameter	**5000**	**10000**	**20000**	**70000**
D	183000	316250	0	90000
μ	11.9729	12.8136	14.12961	15.014
σ	0.2398	0.140528	0.057113	0.033

If we don't consider the parameter which is only an indicator of level of shifting for a log-normal distribution we can easily see interesting dynamics for the main parameters of the model as a portfolio becomes larger: values of the parameter increases while the values of the parameter falls.

For practical purposes, indeed, characteristics of total expenditure (i.e. 5 variables) are of the most interest and are presented in the following table:

Table 9. Characteristics of total expenditures by portfolio sizes

Parameters	**Size of a portfolio**			
	5000	**10000**	**20000**	**70000**
μ_S	163 024	370 820	1 371 262	3 316 911
σ_S	39 662	52 369	78 381	109 488
X	$C_S^{-1}(x)$	$C_S^{-1}(x)$	$C_S^{-1}(x)$	$C_S^{-1}(x)$
0.9	32 392	123 381	1 472 989	3 368 312
0.95	51 999	146 409	1 503 872	3 410 023
0.99	93 719	192 909	1 563 560	3 489 628
0.999	149 348	250 607	1 633 285	3 581 011