iPatient in Medical Information Systems
and Future of Internet of Health

Olga Kolesnichenko
Security Analysis Bulletin
Moscow, Russia
oykolesnichenko@list.ru

Gennady Smorodin
Dell EMC External Research and Academic Alliances
St. Petersburg, Russia
Gennady.Smorodin@dell.com

Andrey Mazelis, Alexander Nikolaev, Lev Mazelis
Vladivostok State University of Economics and Service
Vladivostok, Russia
Andrey.Mazelis@vvsu.ru, sa10121992@yandex.ru
Lev.Mazelis@vvsu.ru

Alexander Martynov, Valeriy Pulit
SP.ARM
St. Petersburg, Russia
{Martynov, Valeriy.Pulit}@sparm.com

Sergey Balandin
Open Innovations Association FRUCT,
University of Information Technologies, Mechanics
and Optics (ITMO University)
St. Petersburg, Russia
Sergey.Balandin@fruct.org

Yuriy Kolesnichenko
Cybersecurity Technology
Moscow, Russia
green-apple_2000@mtu-net.ru

Abstract—**The results of Study "iHealthCare Optimization", provided by Dell EMC External Research and Academic Alliances, are presented. Big Data analytics of Medical information system qMS records was implemented using cluster analysis in Python. Software for cluster analysis was created by Andrey Mazelis (Vladivostok State University of Economics and Service). There are two directions of cluster analysis: Series treatment (number of investigation procedures for each patient) and Series time (waiting time for investigation procedures for each patient). Two models of patients management (Model A and Model B) were found, that can be used for better planning of care management. Models approach provides the new capability to implement Health Care Standard in mode aaS, using feedback after Big Data analytics. Around 80-90% of patients with Essential hypertension can get treatment in Day Hospital without hospitalization.**

I. INTRODUCTION. STUDY iHEALTHCARE OPTIMIZATION

For the first time in history people surround themselves with medical control environment. It has been named Internet of Health (IoH) [1]. Health Care undergoes total Digital transformation that includes: precise, personalized approach (personalized medicine); predictive analytics (predictive medicine); 3V (volume, velocity, variety) plus agile features (Big Data in medicine); spatiotemporal approach (biorhythmology); comprehensive approach; telemedicine [2], [3], [4], [5], [6].

Internal hospital local area network (Intranet) or Medical information systems (MIS) allow to collect patient's data and to analyze data in order to optimize the medical care. Hospital Intranet data helps to determine the socio-medical portrait of patients for each nosology. MIS provides an unique opportunity for flexible respond to patients socio-medical needs through analysis of collected Big Data. It also should be used for social mediation during Health Care reforms implementation. Big Data analytics of MIS records can make qualitative impact on key Health Care goals: improvement of population treatment; maintaining people professional longevity; prevention disability and decrease mortality rate.

There are several steps that required for MIS development in hospitals [7], [8], [9]. Data Scientist Louis Frolio listed: 1) to create a framework for data recording and collecting; 2) to organize the interaction between physicians and mathematicians / IT specialists; 3) to make MIS different records integrated with each other and available for analytics (including analytics on demand or in real time); 4) to establish training courses for Big Data analytics teams consisted of physicians and mathematicians / IT specialists; 5) to promote MIS and Big Data analytics for all hospitals; to work on standardization of Electronic Health Records.

The modern software, such as Python Integrated Development Environment, allows to analyze Big Data, collected by MIS, and to create complex picture of treatment process in hospital, simultaneously considering each patient in mathematical algorithms.

In the article the results of Study provided by Dell EMC External Research and Academic Alliances are presented. Study has been named "Case-study Big Data: iHealthCare Optimization". The records of MIS qMS collected during one year (2014-2015) from three Russian hospitals. MIS qMS

(Fig. 1) is the unique hospital Cloud-based network that collects data bringing it to the common Data Center and allows analysis of data on SaaS basis (Software as a Service).
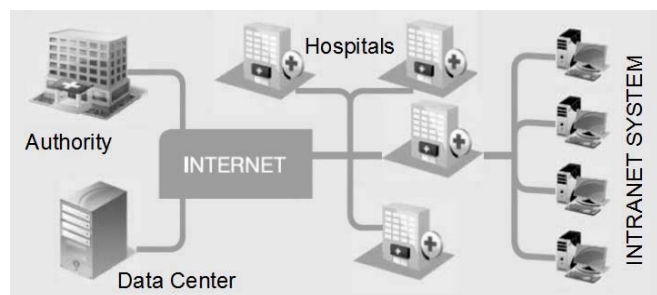

Fig. 1. Framework of Medical information system qMS. Image reprinted from SP.ARM Handbook, 2015

The goal of Study is improvement of care management process in hospital using Big Data analytics of MIS records. CodeBook (guide) as requirements specification was elaborated and included following sections: title of Case-study; the main question for answer; list of metadata for analysis; methods of analysis; hypothesis and expected findings; practical significance.

The work is based on approach of Data Scientist, Dell EMC Chief Technology Officer of Global Services Big Data Practice, Bill Schmarzo [10], [11], [12], [13]. It includes "Big Data Business Model Maturity Index" phases: Business monitoring phase; Business Insights phase; Business Optimization phase; Insights monetization phase; Business Metamorphosis phase.

During analysis two models of patients' management (Model A and Model B) were found, that can be used for better planning of treatment process and care management, to make more effective ROI (Return on Investment), and for feedback recommendation how patient should behave in daily life.
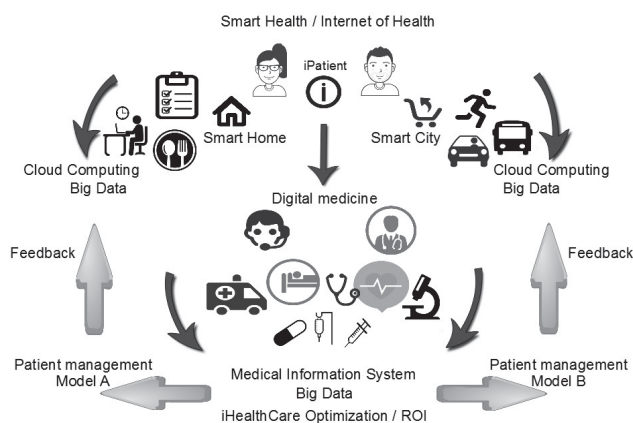

Fig. 2. Environment of Internet of Health with iPatient management models approach. Image of Study "iHealthCare Optimization", ERAA

From Internet of Health's point of view, models, which are found due to Big Data analytics of MIS records, can become the basis of personalized approach for iPatient management in environment of Smart Home and Smart City.

In all MIS qMS records of 685 patients were analyzed. They all have Hypertensive heart disease (I11) without heart failure (I11.9) by ICD-10 (10th revision of the International Statistical Classification of Diseases and Related Health Problems, World Health Organization). The average age of the patients was 56.5 years. Personal data about patients was not transferred to research group, each patient had his own encrypted code. SP.ARM processed Big Data to Microsoft Excel table with metadata. Cluster analysis in Python Integrated Development Environment was made by Department of Mathematics and Modeling in Vladivostok State University of Economics and Service. Statistical analysis was made in Microsoft Excel, including Pearson correlation coefficient.

II. BUSINESS MONITORING PHASE

A. CodeBook

The hypothesis, which has been set before study launch, was following: cluster analysis may show the difference of patients stay in hospital, that depending on time of investigation waiting and type of investigation procedures. Cluster analysis was decided to implement with full number of parameters (all investigation procedures) and with reduced number of parameters (the required minimum of investigation procedures and a little broader list of investigation procedures).

The software for cluster analysis (iPython, libraries NumPy, Pandas and Sklearn) was created by Andrey Mazelis, Associate Professor of Department of Mathematics and Modeling of Vladivostok State University of Economics and Service. Clusters based on full number of parameters (39 investigation procedures) were defined as Max Metadata clusters. Clusters based on reduced number of parameters (24 investigation procedures) were defined as Middle Metadata clusters. Clusters based on required minimum of investigation procedures (10 investigation procedures) were defined as Min Metadata clusters.

Cluster analysis was implemented with different set of clusters allocation – 2, 3 and 4 clusters. Each cluster is marked by letter "k" and has own designation. Clusters of Max Metadata are marked for set of 2 clusters allocation as Max-2 1k and Max-2 2k; for set of 3 clusters allocation as Max-3 1k, Max-3 2k, Max-3 3k; for set of 4 clusters allocation as Max-4 1k, Max-4 2k, Max-4 3k, Max-4 4k. Clusters of Middle Metadata (Middle-2, Middle-3, Middle-4) and Min Metadata (Min-2, Min-3, Min-4) have the same designation.

There are two directions of cluster analysis: Series treatment (number of investigation procedures for each patient) and Series time (waiting time for investigation procedures for each patient).

B. Clusters allocation

All determined clusters are presented in Table I. Clusters were divided by criteria "average cost of a single episode of hospitalization" and "average length of stay in hospitals during single episode of hospitalization" into two groups 1 and 2 after

data visualization that is shown at Fig. 3. The simple distribution of all cases (685 patients) is presented at Fig. 4.

TABLE I. ALL CLUSTERS ALLOCATION INTO TWO GROUPS ACCORDING TO COST / LENGTH-OF-STAY PARAMETERS

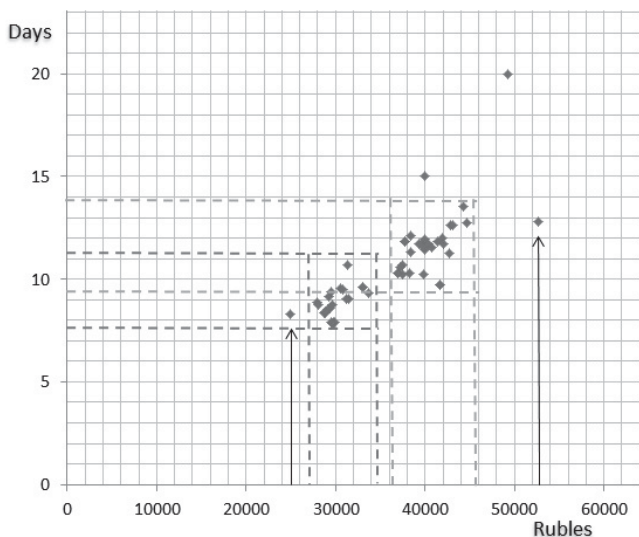| Groups	Clusters
Group 1 • 395 patients; • average cost of a single episode of hospitalization is less than 35.000 Rubles; • average length of stay in hospitals during single episode of hospitalization is 8.5 days	**Series treatment** Max-2 1k; Middle-2 1k; Min-2 1k; Max-3 1k; Middle-3 2k; Min-3 1k; Max-4 1k; Min-4 1k
	Series time Max-2 1k; Middle-2 1k; Min-2 1k; Max-3 2k; Max-3 3k; Middle-3 1k; Middle-3 2k; Min-3 1k; Max-4 2k; Max-4 4k; Middle-4 1k; Middle-4 2k; Middle-4 3k; Min-4 3k
Group 2 • 290 patients; • average cost of a single episode of hospitalization is more than 35.000 Rubles; • average length of stay in hospitals during single episode of hospitalization is 11.2 days	**Series treatment** Max-2 2k; Middle-2 2k; Min-2 2k; Max-3 2k; Max-3 3k; Middle-3 1k; Middle-3 3k; Min-3 2k; Min-3 3k; Max-4 2k; Max-4 3k; Max-4 4k; Middle-4 1k; Middle-4 2k; Middle-4 4k; Min-4 2k; Min-4 4k
	Series time Max-2 2k; Middle-2 2k; Min-2 2k; Max-3 1k; Middle-3 3k; Min-3 2k; Min-3 3k; Max-4 1k; Middle-4 4k; Min-4 1k; Min-4 4k

Fig. 3. Clusters allocation by criteria "average cost of a single episode of hospitalization" and "average length of stay in hospitals during single episode of hospitalization". Very strong Pearson correlation coefficient r = 0.8. The arrows point out the lowest cluster's average cost of a single episode of hospitalization (Middle-4 3k Series treatment) and the highest cluster's average cost of a single episode of hospitalization (Max-4 3k Series time).
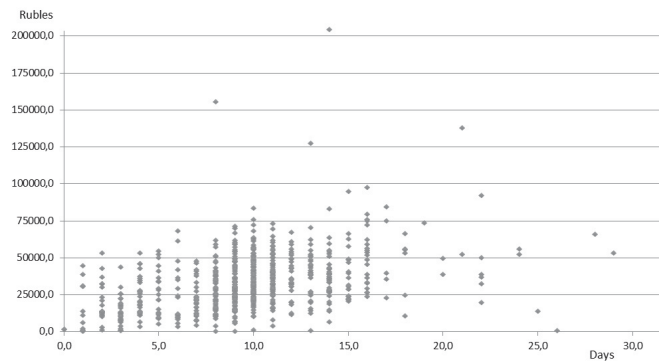

Fig. 4. The simple distribution of all cases (685 patients) by criteria "average cost of a single episode of hospitalization" and "average length of stay in hospitals during single episode of hospitalization". There is not significant Pearson correlation.

Is shown that simple statistical distribution of all cases doesn't reflect the hidden allocation into two groups. There is not significant Pearson correlation for simple statistical distribution of all cases by criteria "average cost of a single episode of hospitalization" and "average length of stay in hospitals during single episode of hospitalization", but very strong Pearson correlation coefficient was found for clusters distribution by these criteria.

C. Description of Max Metadata clusters

The cluster Max-4-1k Series treatment was found (Fig. 5, Fig. 6), it consists of 427 patients (62.3% of all patients sample) and has average cost of a single episode of hospitalization 28.000 Rubles in comparison with average cost of all patients sample 33.000 Rubles; average length of stay in hospitals (LSH) during single episode of hospitalization was 8.7 days.

Also the cluster Max-4-4k Series treatment was found, it consists of 221 patients (32.3% of all patients sample) and has average cost of a single episode of hospitalization 43.000 Rubles with average LSH during single episode of hospitalization 11.3 days. This LSH is higher than required 10 days by Health Care Standard of Essential (primary) hypertension (Ministry of Health of Russian Federation Order on November 9, 2012, No. 708n). The patients average age of this cluster is 56 years old.

Considering clusters allocation of Series time (Fig. 7, Fig. 8), the cluster Max-4 3k Series time was found, it consists of 63 patients (9.2% of all patients sample). The cluster's average cost of a single episode of hospitalization exceeds 52.000 Rubles. Cluster Max-4 3k Series time has the highest cost of treatment among all clusters in this study, with average LSH during single episode of hospitalization 12.8 days.

D. Description of Middle Metadata clusters

Cluster Middle-4 3k Series treatment has the lowest average cost of a single episode of hospitalization 24.939 Rubles. This cluster is based on reduced number of investigation procedures that strictly correspond to the Health Care Standard of Ministry of Health. The cluster's average LSH during single episode of hospitalization is 8 days.
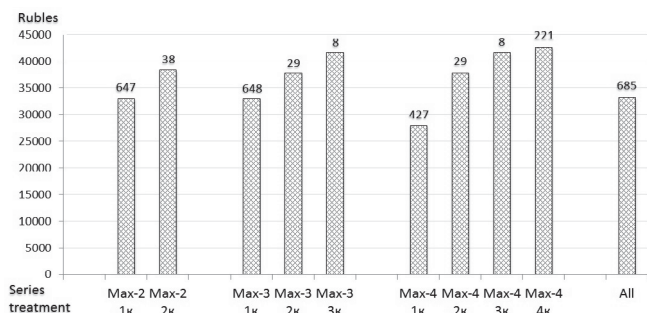
Fig. 5. Max Metadata clusters, Series treatment. Columns – average cost of a single episode of hospitalization for each of clusters; numbers – amount of patients in each of clusters.
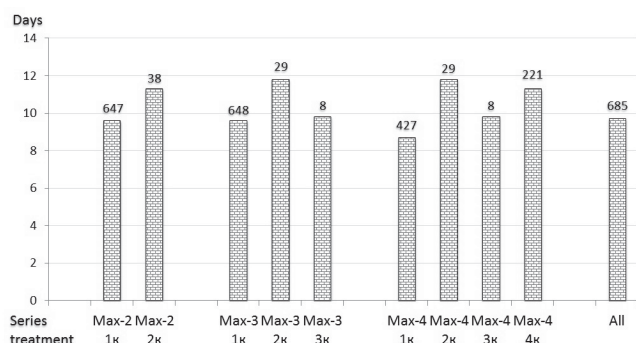

Fig. 6. Max Metadata clusters, Series treatment. Columns – average length of stay in hospitals during single episode of hospitalization for each of clusters; numbers – amount of patients in each of clusters.
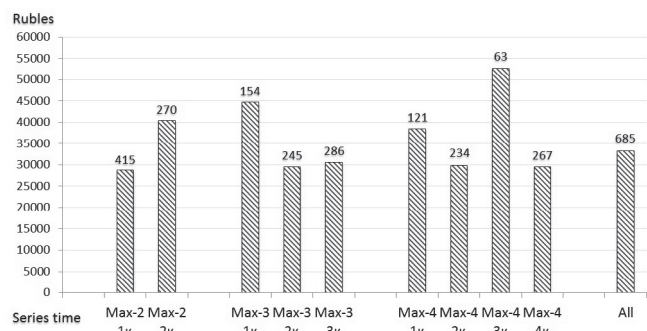

Fig. 7. Max Metadata clusters, Series time. Columns – average cost of a single episode of hospitalization for each of clusters; numbers – amount of patients in each of clusters.
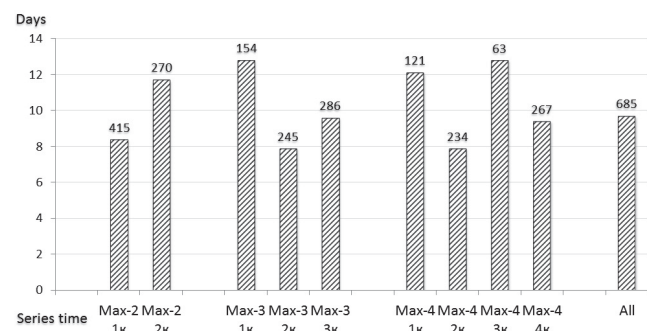

Fig. 8. Max Metadata clusters, Series time. Columns – average length of stay in hospitals during single episode of hospitalization for each of clusters; numbers – amount of patients in each of clusters.

This cluster consists of 231 patients (33.7% of all patients sample). Thus the desired low value of cost and length of stay (not going beyond the Health Care Standard) were found only for one-third part of patients, that is the reason for optimization of approaches to treatment process and care management of patients with Hypertensive heart disease.

E. Description of Min Metadata clusters

Two clusters with greater average LSH during single episode of hospitalization were found among all clusters. They are cluster Min-4 2k Series time and cluster Min-4 3k Series treatment. Both clusters are based on most often implemented and required minimum of investigation procedures and don't consider a wide range of patients investigation. Thus the greater LSH in these cases relates to list of 10 investigation and treatment procedures: electrocardiography (ECG); 24 hour Holter ECG + Blood pressure (BP) monitoring; kidney ultrasonography; renal artery ultrasonography; fundoscopy with Goldmann three-mirror lens; fingerstick test; venipuncture; intramuscular injection; intravenous injection; intravenous infusion.

Cluster Min-4 2k Series time consists of 20 patients, the patients average age of this cluster is 61 years old (it is the highest average age among all clusters in this study). The cluster's average cost of a single episode of hospitalization is 39.908 Rubles. The average LSH during single episode of hospitalization is 15 days.

Cluster Min-4 3k Series treatment includes only one patient which cost of a single episode of hospitalization was 49.256 Rubles and average LSH during single episode of hospitalization was 20 days. Cluster with one patient was not analyzed further.

More detailed analysis was carried out (Fig. 9, Fig. 10) and the reason, why patients of cluster Min-4 2k Series time stayed in hospital so long, was found. It was the need to repeat ECG. Also more number of venipuncture procedures and longer waiting time for intramuscular injection, intravenous injection and intravenous infusion were determined.

In all there were 37 patients (5.4% of all patients sample) who had ECG examinations twice and more. ECG was implemented during first day of stay in hospital in most cases. At Fig. 11 is shown the small pool of patients with repeated ECG examinations. The distribution of number of repeated ECG examinations is presented at Fig. 12, 86% of patients had two ECG examinations, other more. Repeated ECG examinations linearly depend on LSH.

Also should be mentioned the distribution of repeated fingerstick tests, which had 48 patients (7% of all patients sample), 67% of them had two fingerstick tests (Fig. 13), other more. And 342 patients (49.9% of all patients sample) received intravenous infusions, 88% of them received only one intravenous infusion (Fig. 14).

Thus the analysis showed that only one-third part of all patients had desired low value of cost and LSH. But at the same time, only around 5% of all patients needed deeper and more prolonged investigation of cardiovascular system. Most

of these patients from 5%-sample needed only two ECG examinations. Only 7% of all patients needed two fingerstick tests and around half of all patients needed one intravenous infusion. It points out to two issues:

1) more than 95% of all patients with Hypertensive heart disease without heart failure (I11.9) can get investigation and treatment procedures during short LSH with subsequent outpatient monitoring, instead found one-third part of all patients;

2) can be assumed that there is another reason to stay in hospital longer than Health Care Standard of Essential hypertension requires.

The further analysis was aimed to determine the socio-medical portrait of patients with Hypertensive heart disease and to answer the question why these patients stayed in hospitals longer than it was required.
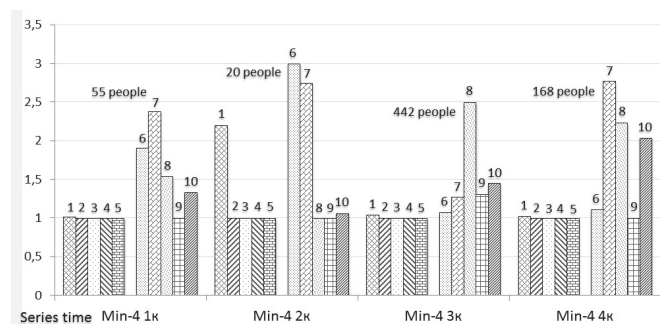

Fig. 11. Simple distribution of all cases of ECG examinations, including repeated ECG examinations (685 patients). X-axis – length of stay in hospitals during single episode of hospitalization in days; Y-axis – waiting time of ECG examination in minutes. 24 hours = 1440 minutes.

Fig. 9. Min-4 Series time clusters, amount of patients is shown above. Columns – cluster's average number of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion.
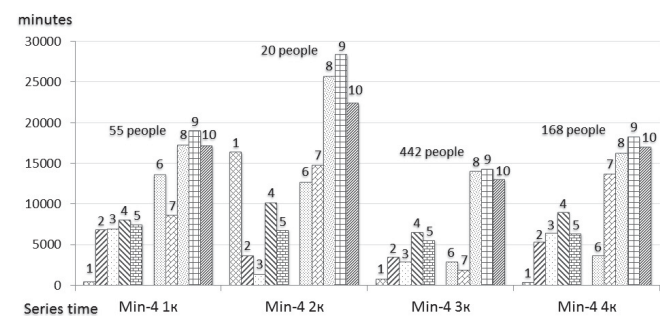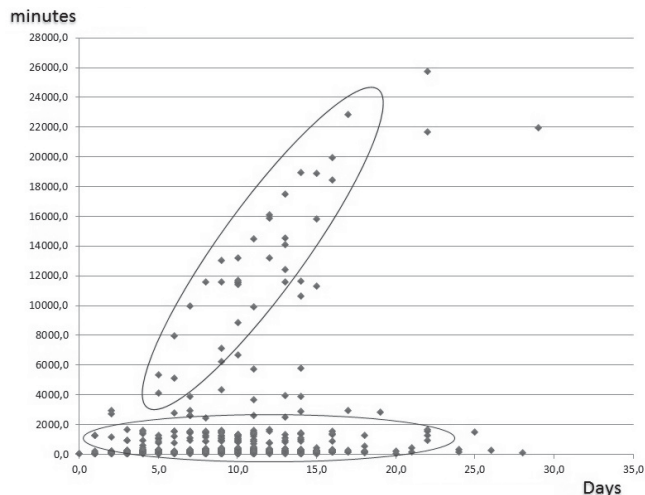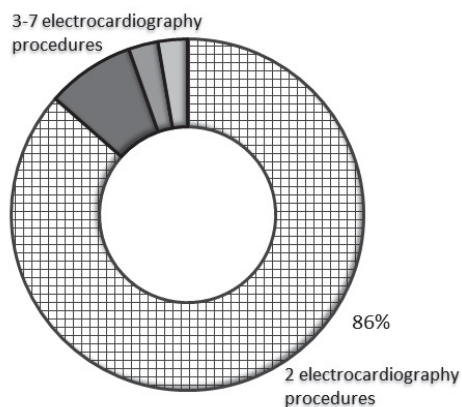

Fig. 12. Distribution of number of repeated ECG examinations

Fig. 10. Min-4 Series time clusters, amount of patients is shown above. Columns – cluster's average waiting time of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion. 24 hours = 1440 minutes.
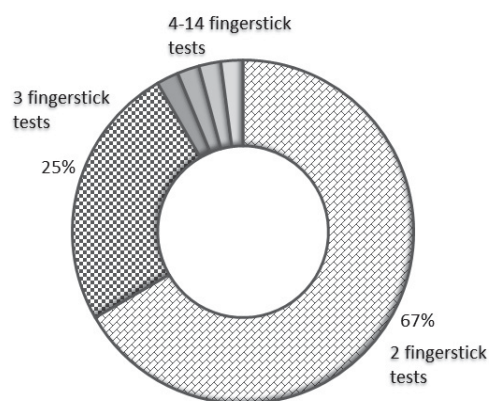

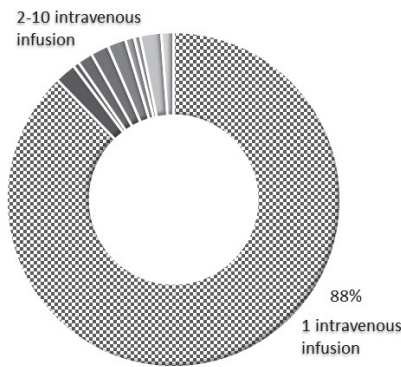Fig. 13. Distribution of number of repeated fingerstick tests

Fig. 14. Distribution of number of intravenous infusions

III. BUSINESS INSIGHTS PHASE

A. Comparison of clusters with the same cost of treatment

Clusters Max-4 2k and Max-4 4k Series time were compared. Cluster Max-4 2k Series time has average cost of a single episode of hospitalization 29.866 Rubles. Cluster Max-4 4k Series time has average cost of a single episode of hospitalization 29.543 Rubles. Cluster Max-4 2k Series time consists of 234 patients, average LSH during single episode of hospitalization was 7.9 days. Cluster Max-4 4k Series time consists of 267 patients, average LSH during single episode of hospitalization was 9.4 days. Both clusters coincide with required LSH by Health Care Standard of Essential hypertension; both have cost of a single episode of hospitalization less than average of all patients sample; both have quite equal number of patients.

The differences between two clusters are following: patients of cluster Max-4 4k Series time had more intramuscular and intravenous injections and intravenous infusions (Fig. 15, Fig. 16).
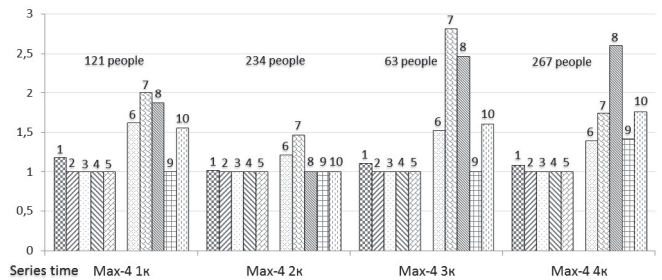

Fig. 15. Max-4 Series time clusters, amount of patients is shown above. Reduced number of parameters is presented. Columns – cluster's average number of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion.

Fig. 16. Max-4 Series time clusters, amount of patients is shown above. Reduced number of parameters is presented. Columns – cluster's average waiting time of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion. 24 hours = 1440 minutes.
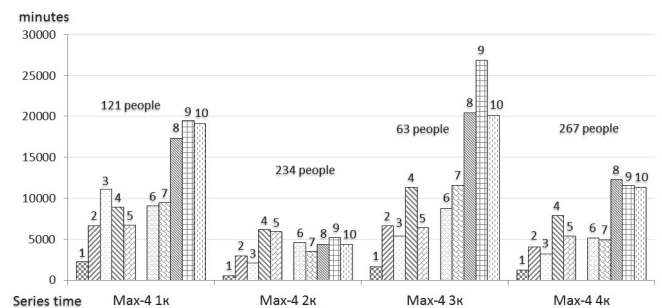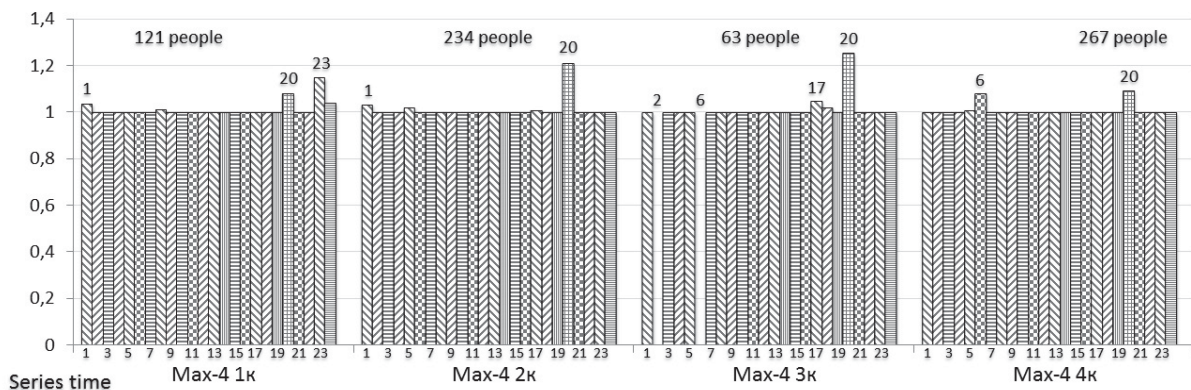

Fig. 17. Max-4 Series time clusters, amount of patients is shown above. Enlarged list of parameters is presented. Columns – cluster's average number of investigation procedures. 1 – 24 hour Holter ECG monitoring; 2 – treadmill stress test; 3 – bicycle ergometer exercise test; 4 – echocardiographic stress test; 5 – 24 hour Holter ECG + BP + reopulmonography monitoring; 6 – spirometry with bronchodilator test; 7 – respiratory investigation; 8 – triplex ultrasonography (color-flow imaging) of extracranial brachiocephalic arteries, 9 – triplex ultrasonography (color-flow imaging) of extracranial and intracranial brachiocephalic arteries; 10 – triplex ultrasonography (color-flow imaging) of intracranial brachiocephalic arteries; 11 – abdominal / renal ultrasonography; 12 – polysomnographic investigation of sleep; 13 – duplex ultrasonography of lower-extremity arteries and veins; 14 – duplex ultrasonography of lower-extremity veins with evaluation of veins function; 15 – thyroid ultrasonography; 16 – Biohit test; 17 – esophagogastroduodenoscopy (EGD); 18 – colonoscopy; 19 – breast ultrasonography; 20 – endoscopic biopsy; 21 – female pelvic ultrasonography; 22 – male pelvic ultrasonography; 23 – digital rectal examination (DRE) of prostate gland; 24 – bacteriological, cytological, hormonal investigations and polymerase chain reaction (PCR) diagnostics.

Fig. 18. Max-4 Series time clusters, amount of patients is shown above. Enlarged list of parameters is presented. Columns – cluster's average waiting time of investigation procedures. 1 – 24 hour Holter ECG monitoring; 2 – treadmill stress test; 3 – bicycle ergometer exercise test; 4 – echocardiographic stress test; 5 – 24 hour Holter ECG + BP + reopulmonography monitoring; 6 – spirometry with bronchodilator test; 7 – respiratory investigation; 8 – triplex ultrasonography (color-flow imaging) of extracranial brachiocephalic arteries, 9 – triplex ultrasonography (color-flow imaging) of extracranial and intracranial brachiocephalic arteries; 10 – triplex ultrasonography (color-flow imaging) of intracranial brachiocephalic arteries; 11 – abdominal / renal ultrasonography; 12 – polysomnographic investigation of sleep; 13 – duplex ultrasonography of lower-extremity arteries and veins; 14 – duplex ultrasonography of lower-extremity veins with evaluation of veins function; 15 – thyroid ultrasonography; 16 – Biohit test; 17 – esophagogastroduodenoscopy (EGD); 18 – colonoscopy; 19 – breast ultrasonography; 20 – endoscopic biopsy; 21 – female pelvic ultrasonography; 22 – male pelvic ultrasonography; 23 – digital rectal examination (DRE) of prostate gland; 24 – bacteriological, cytological, hormonal investigations and polymerase chain reaction (PCR) diagnostics. 24 hours = 1440 minutes.
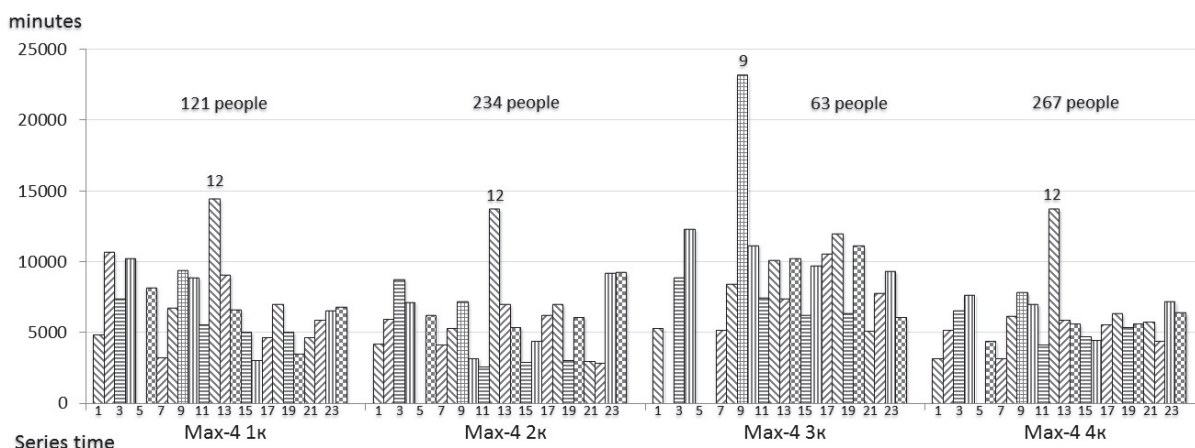
The full number of parameters, complemented reduced number of parameters, is presented at Fig. 17 and Fig. 18. Comparing clusters Max-4 2k and Max-4 4k Series time with the same cost of treatment can be mentioned that cluster Max-4 2k Series time has more endoscopic biopsies and cluster Max-4 4k Series time has more spirometries. The polysomnographic investigation of sleep is characterized by long average waiting time for both clusters. Thus two models of patients' management with Essential hypertension were found: Model A and Model B.

Model A is characterized by more parenteral therapy and cardiorespiratory system investigation and Model B is characterized by more diagnostics of comorbidities of digestive system.

B. Description of cluster with highest cost of treatment

As was mentioned above the cluster Max-4 3k Series time has the highest cost of treatment among all clusters in this study. This cluster is characterized by more venipunctures (more blood tests), more intramuscular injections and longer parenteral intravenous therapy courses.

Cluster has more esophagogastroduodenoscopies (EGD) and more endoscopic biopsies. Also cluster is characterized by long average waiting time for triplex ultrasonography (color-flow imaging) of extracranial and intracranial brachiocephalic arteries.
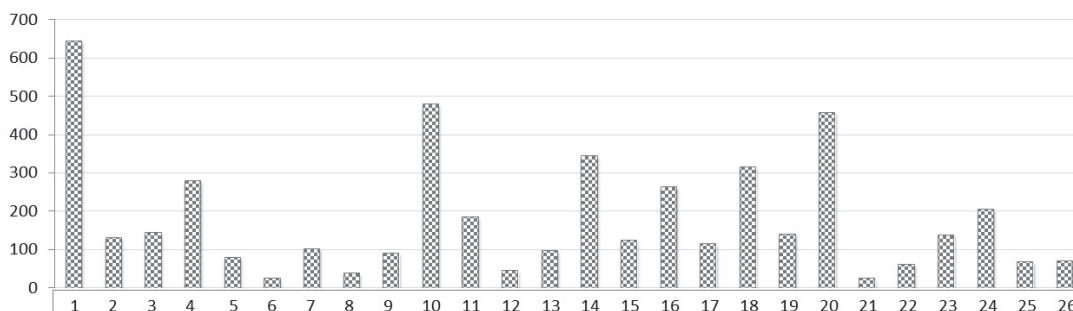

Fig. 19. The distribution of all investigation procedures for all patients sample (685 patients). Columns – number of investigation procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG monitoring; 3 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 4 – 24 hour Holter ECG + BP + reopulmonography monitoring; 5 – treadmill stress test; 6 – bicycle ergometer exercise test; 7 – echocardiographic stress test; 8 – spirometry with bronchodilator test; 9 – respiratory investigation; 10 – triplex ultrasonography (color-flow imaging) of brachiocephalic arteries; 11 – fundoscopy with Goldmann three-mirror lens; 12 – polysomnographic investigation of sleep; 13 – duplex ultrasonography of lower-extremity arteries and veins; 14 – thyroid ultrasonography; 15 – breast ultrasonography; 16 – endoscopic biopsy; 17 – Biohit test; 18 – esophagogastroduodenoscopy (EGD); 19 – colonoscopy; 20 – abdominal / renal ultrasonography; 21 – kidney ultrasonography; 22 – renal artery ultrasonography; 23 – female pelvic ultrasonography; 24 – male pelvic ultrasonography; 25 – digital rectal examination (DRE) of prostate gland; 26 – bacteriological, cytological, hormonal investigations and polymerase chain reaction (PCR) diagnostics.

C. Distribution of all investigation procedures

The distribution of all investigation procedures, that were implemented for all patients sample (685 patients), is presented at Fig. 19. Electrocardiography (ECG) took the first place (more often implemented). Triplex ultrasonography of brachiocephalic arteries took second place. Abdominal / renal ultrasonography took third place.

Percentage distribution of all investigation procedures, that were implemented for all patients sample, is presented at Fig. 20. The heart investigation procedures dominate, but also the prepotency of diagnostics of comorbidities of digestive system should be emphasized.
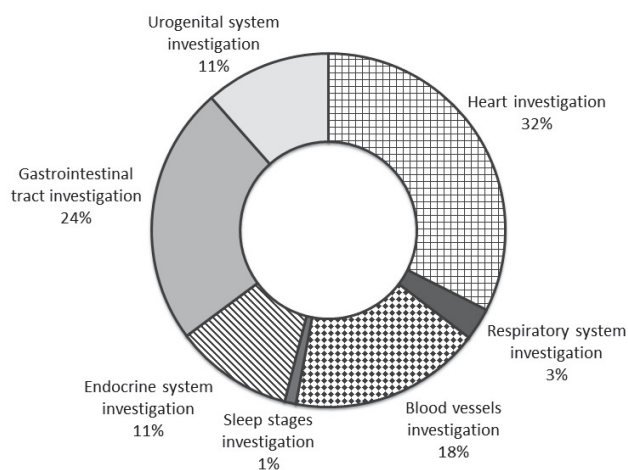

Fig. 20. Percentage distribution of all investigation procedures that were implemented for all patients sample (685 patients)

D. Description of cluster with lowest cost of treatment

As was mentioned above cluster Middle-4 3k Series treatment has the lowest average cost of a single episode of hospitalization. This cluster is characterized by short average waiting time for intravenous infusions (Fig. 21 and Fig. 22).
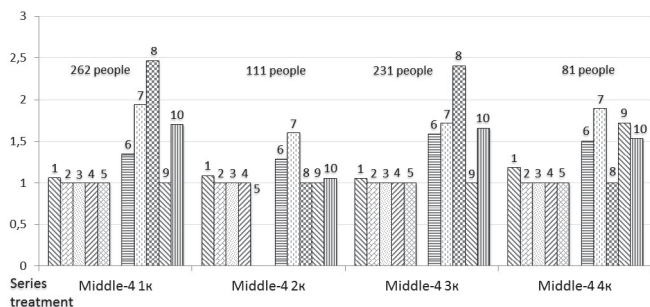

Fig. 21. Middle-4 Series treatment clusters, amount of patients is shown above. Reduced number of parameters is presented. Columns – cluster's average number of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion.

Fig. 22. Middle-4 Series treatment clusters, amount of patients is shown above. Reduced number of parameters is presented. Columns – cluster's average waiting time of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion. 24 hours = 1440 minutes.
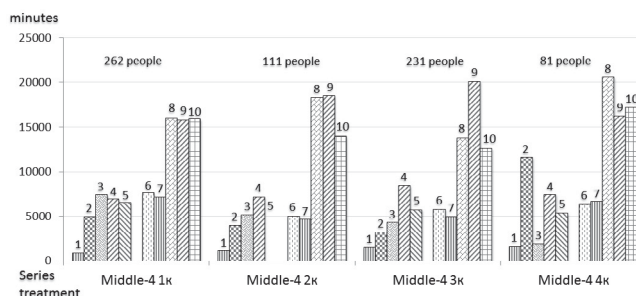
Considering all investigation procedures of Middle Metadata (is not presented at figures) can be mentioned that clusters have prepotency either spirometries or 24 hour Holter ECG monitoring. Polysomnographic investigation of sleep is characterized by longest waiting time, it points out to implementing investigation of sleep in the last turn.

E. Comparison of the two groups / models

The data of MIS qMS records include both cases of health insurance in Russia: the compulsory medical insurance known as OMC, and privately purchased medical insurance known as DMC. In both cases of medical insurance an additional investigation of patient exceeding Health Care Standards is allowed, if it is required in concrete situation (Ministry of Health of Russian Federation Newsletter №14-3/10/2-11668, November 24, 2011). Comorbidities are treated in accordance with related Health Care Standards of Ministry of Health of Russian Federation.

Patient from the point of view of Sociology of Medicine interacts not only with the Health Care system but broader with Social system that surrounds people [14]. Patient becomes part of Information Society and can be defined as iPatient (Intranet Patient) which has specific characteristics in MIS Big Data. iPatient provides more powerful feedback for analysis due to Big Data analytics than patient from previous time with only personal feedback "patient – physician". MIS Big Data reveals the new information about patient in addition to Anamnesis morbi, Anamnesis vitae and data of physical examination and investigation procedures. This new information can be called socio-medical portrait of patients for each nosology. It is also very important for patients planning of treatment process and care management.

The two models of patients care management with Essential hypertension were detected. Models are based on two groups which are described in Table II and which are shown at Fig. 23. Model A "Cardiovascular" is based on group 1 (395 patients or 58% of all patients sample), Model B "Comorbidities, digestive" is based on group 2 (290 patients or 42% of all patients sample).

TABLE II. TWO GROUPS ACCORDING TO COST / LENGTH-OF-STAY
PARAMETERS AND ITS CHARACTERISTICS
Absolute number / Standardized rates per 100 patients

	Number of investigation procedures			
Groups	ECG	Holter ECG	Holter + BP	Holter + reo
1	364 / 92.2	63 / 15.9	90 / 22.8	144 / 36.5
2	281 / 96.9	69 / 23.8	55 / 18.9	137 / 47.2

Groups	Treadmill	Bicycle	EchoCG	Spirometry
1	47 / 11.9	4 / 1	32 / 8.1	18 / 4.6
2	33 / 11.3	21 / 7.2	71 / 24.5	20 / 6.9

Groups	Respiratory	BCA US	Fundoscopy	PSG
1	28 / 7.1	226 / 57.2	67 / 16.9	5 / 1.3
2	63 / 21.7	254 / 87.6	118 / 40.7	40 / 13.8

Groups	LE US	Thyroid US	Breast US	Biopsy
1	34 / 8.6	147 / 37.2	45 / 11.4	80 / 20.3
2	64 / 22.1	198 / 68.3	80 / 27.6	184 / 63.4

Groups	Biohit	EGD	Colonoscopy	A/R US
1	30 / 7.6	116 / 29.4	30 / 7.6	239 / 60.5
2	85 / 29.3	201 / 69.3	111 / 38.3	219 / 75.5

Groups	Kidney US	Renal a. US	Pelvic/f US	Pelvic/m US
1	17 / 4.3	37 / 9.4	55 / 13.9	84 / 21.3
2	8 / 2.8	25 / 8.6	84 / 28.9	122 / 42.1

	Number of investigation / treatment procedures				
Groups	DRE	BCH/PCR	IM inj	IV inj	IV inf
1	17 / 4.3	23 / 5.8	48 / 12.2	19 / 4.8	256 / 65
2	51 / 17.6	47 / 16.2	120 / 41	22 / 7.6	290 / 100

Contracted terms in Table II:

ECG – electrocardiography; Holter ECG – 24 hour Holter ECG monitoring; Holter + BP – 24 hour Holter ECG + Blood pressure (BP) monitoring; Holter + reo – 24 hour Holter ECG + BP + reopulmonography monitoring; Treadmill – treadmill stress test; Bicycle – bicycle ergometer exercise test; EchoCG – echocardiographic stress test; Spirometry – spirometry with bronchodilator test; Respiratory – respiratory investigation; BCA US – triplex ultrasonography (color-flow imaging) of brachiocephalic arteries; Fundoscopy – fundoscopy with Goldmann three-mirror lens; PSG – polysomnographic investigation of sleep; LE US – duplex ultrasonography of lower-extremity arteries and veins; Thyroid US – thyroid ultrasonography; Breast US – breast ultrasonography; Biopsy – endoscopic biopsy; Biohit – Biohit test; EGD – esophagogastroduodenoscopy (EGD); Colonoscopy – colonoscopy investigation; A/R US – abdominal / renal ultrasonography; Kidney US – kidney ultrasonography; Renal a. US – renal artery ultrasonography; Pelvic/f US – female pelvic ultrasonography; Pelvic/m US – male pelvic ultrasonography; DRE – digital rectal examination (DRE) of prostate gland; BCH/PCR – bacteriological, cytological, hormonal investigations and polymerase chain reaction (PCR) diagnostics; IM inj – intramuscular injection; IV inj – intravenous injection; IV inf – intravenous infusion.
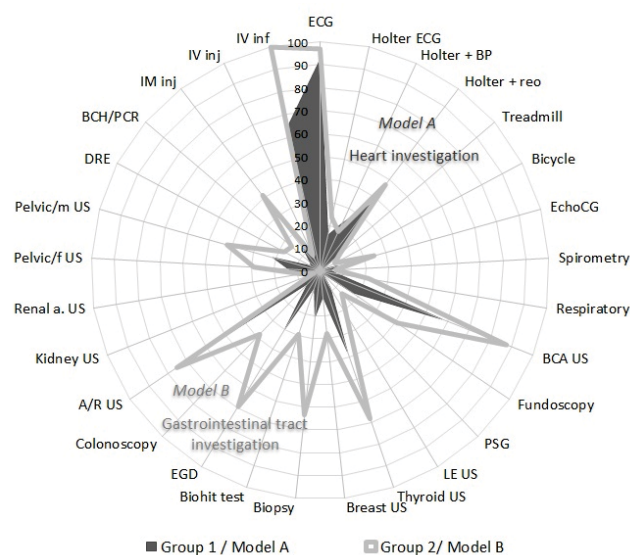

Fig. 23. Two models visualization. Number of investigation / treatment procedures, standardized rates per 100 patients. Contracted terms see below Table II.

For Model A (group 1) "Cardiovascular" average cost of a single episode of hospitalization is less than 35.000 Rubles; average LSH during single episode of hospitalization is 8.5 days. For Model B "Comorbidities, digestive" (group 2) average cost of a single episode of hospitalization is more than 35.000 Rubles; average LSH during single episode of hospitalization is 11.2 days.

Two models are differed, Model A has less number of all investigation and treatment procedures, and more concentrates on electrocardiography, Holter ECG and Blood pressure monitoring, treadmill stress test, ultrasonography of brachiocephalic arteries; is characterized by accent on parenteral therapy and cardiorespiratory system. Thus Model A reflects the targeted, in a short time investigation of the cardiovascular system of patients with Essential hypertension in accordance with Health Care Standards.

Model B has a lot more number of all investigation and treatment procedures and is characterized by accent on gastrointestinal tract investigation: abdominal ultrasonography, esophagogastroduodenoscopy, colonoscopy investigation, Biohit test, endoscopic biopsy. Thus Model B reflects the protracted, expensive, exceeding the Health Care Standards for the underlying disease (Essential hypertension) care management of patients with comorbidities digestive diseases.

Can be described the socio-medical portrait of patients (or iPatients) with Hypertensive heart disease without heart failure (I11.9, ICD-10). Mostly they need 1 ECG examinations, half of them need intravenous infusion (mostly one infusion), equal attention is paid to heart investigation and gastrointestinal tract investigation.

The very important parameter registered by MIS qMS is average waiting time (AWT) of the latest analysis. It has moderate positive linear relationship with average LSH during

single episode of hospitalization (Fig. 24) and thus it affects to LSH. Venipuncture AWT has very strong relationship with average LSH and was determined as the most influencing parameters on LSH.

For clusters of Min Metadata, that have only required minimum of investigation procedures and were purified from most of investigation procedures, the very strong relationship was found between intravenous infusion AWT and LSH (r=0.99), between venipuncture AWT and LSH (r=0.94), and between intravenous injection AWT and LSH (r=0.89). It points out that blood tests and parenteral therapy are the most crucial among shortest list of procedures during episode of hospitalization in terms of impact on LSH.
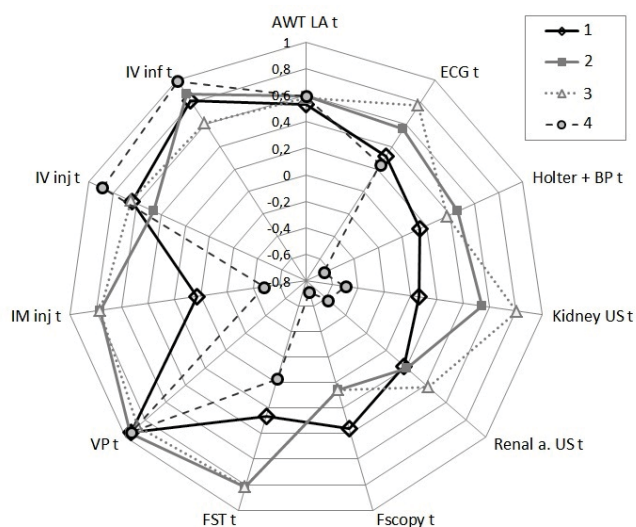

Fig. 24. Visualization of Pearson correlation coefficients between average length of stay in hospitals during single episode of hospitalization and: AWT LA t – average waiting time of the latest analysis; ECG t – electrocardiography AWT; Holter + BP t – 24 hour Holter ECG + Blood pressure monitoring AWT; Kidney US t – kidney ultrasonography AWT; Renal a. US t – renal artery ultrasonography AWT; Fscopy t – fundoscopy with Goldmann three-mirror lens AWT; FST t – fingerstick test AWT; VP t – venipuncture AWT; IM inj t – intramuscular injection AWT; IV inj t – intravenous injection AWT; IV inf t – intravenous infusion AWT. For set of data of 1 – all 24 clusters of 4 clusters allocation Series treatment and Series time; 2 – 8 clusters of Max Metadata of 4 clusters allocation Series treatment and Series time; 3 – 8 clusters of Middle Metadata of 4 clusters allocation Series treatment and Series time; 4 – 8 clusters of Min Metadata 4 clusters allocation Series treatment and Series time.

For clusters of Middle Metadata, that have reduced number of parameters and exclude investigation of comorbidities digestive diseases, the very strong relationship was found between venipuncture AWT and LSH (r=0.88), between fingerstick test AWT and LSH (r=0.82), between kidney ultrasonography AWT and LSH (r=0.81); and the strong relationship was found between electrocardiography AWT and LSH (r=0.77), between intramuscular injection AWT and LSH (r=0.77). These findings were confirmed by founded clusters allocation Middle Metadata Series time with cluster with shortest average LSH during single episode of hospitalization (Fig. 25). Cluster Middle-4 1k Series time has 7.8 days average LSH that is shortest among all clusters in this study. Shortest cluster is characterized by shortest electrocardiography AWT, shortest kidney ultrasonography

AWT, shortest venipuncture AWT, shortest intramuscular injection AWT. Electrocardiography and kidney ultrasonography are the two important investigation procedures, which waiting time influences to the LSH.
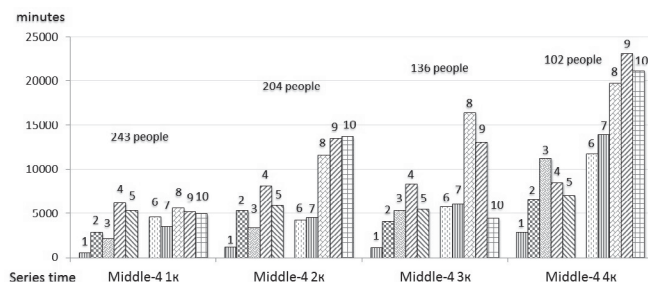

Fig. 25. Middle-4 Series time clusters, amount of patients is shown above. Reduced number of parameters is presented. Columns – cluster's average waiting time of investigation procedures and treatment procedures. 1 – electrocardiography (ECG); 2 – 24 hour Holter ECG + Blood pressure (BP) monitoring; 3 – kidney ultrasonography; 4 – renal artery ultrasonography; 5 – fundoscopy with Goldmann three-mirror lens; 6 – fingerstick test; 7 – venipuncture; 8 – intramuscular injection; 9 – intravenous injection; 10 – intravenous infusion. 24 hours = 1440 minutes.

For clusters of Max Metadata the very strong relationship was found between fingerstick test AWT and LSH (r=0.82), between venipuncture AWT and LSH (r=0.95), between intramuscular injection AWT and LSH (r=0.77), between intravenous infusion AWT and LSH (r=0.87). The blood tests and parenteral therapy show the importance for determining of LSH. The same results show all 24 clusters of 4 clusters allocation Series treatment and Series time – blood tests and parenteral therapy are the most crucial for LSH.

Considering correlation for all set of investigation procedures the very strong relationship was found between abdominal / renal ultrasonography AWT and LSH (r=0.86) and strong relationship was found between echocardiographic stress test AWT and LSH (r=0.71) for 16 clusters of 4 clusters allocation Max Metadata and Middle Metadata Series treatment and Series time.

Separately for 8 clusters of Max Metadata very strong relationship was found between abdominal / renal ultrasonography AWT and LSH (r=0.96), and strong relationship was found between echocardiographic stress test AWT and LSH (r=0.71), thyroid ultrasonography AWT and LSH (r=0.74), breast ultrasonography AWT and LSH (r=0.70), male pelvic ultrasonography AWT and LSH (r=0.77), colonoscopy AWT and LSH (r=0.73).

Separately for 8 clusters of Middle Metadata strong relationship was found between abdominal / renal ultrasonography AWT and LSH (r=0.74), between echocardiographic stress test AWT and LSH (r=0.76), and treadmill stress test AWT and LSH (r=0.75).

Thus some procedures are the most influencing on LSH: blood tests, parenteral therapy, ultrasonography (kidney, abdominal, echocardiographic, thyroid, breast, male pelvic), electrocardiography, treadmill stress test and colonoscopy. These investigation procedures should be taken into account for care management improving through decreasing waiting

time. Parenteral therapy is the unchangeable factor of LSH, because if patient needs long-term treatment, it cannot be cut. The ultrasonography examination is the changeable factor, that requires the better accessibility to ultrasonography examination for patients in hospital (i.e. more numbers of ultrasound machines, more physicians – doctors of ultrasound diagnostics, or more outpatient ultrasonography examinations before hospitalization).

IV. BUSINESS OPTIMIZATION PHASE

Digital transformation of Health Care allows to make medicine precise, personalized and predictive. Cluster analysis in Python of MIS data showed the proportion of patients with lower and higher cost of treatment, and we can understand the common trends, why cost increases or decreases (Fig. 26).
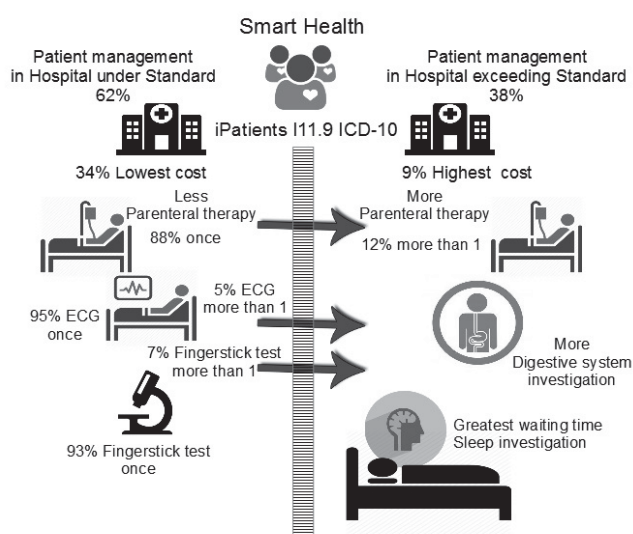

Fig. 26. Environment of Internet of Health / Smart Health with iPatient management models approach. Clusters main insights. Image of Study "iHealthCare Optimization", ERAA.

The small part of patients, that needed deeper investigation with higher cost, was found. They are characterized by repeated ECG examinations, in most cases they additionally got monitoring of cardiovascular / cardiorespiratory parameters (Fig. 27), and predominantly they got only one additional complex examination. They also had more parenteral therapy. The MIS can determine this kind of patients at early stage of a single episode of hospitalization, if physician would schedule retry of ECG (ECG control in dynamics).

Hospital Intranet data helps to determine the socio-medical portrait of patients with Hypertensive heart disease without heart failure: mostly they need 1 ECG examination, half of them need intravenous infusions (mostly one infusion), equal attention is paid to heart investigation (to underlying disease) and gastrointestinal tract investigation (to comorbidities digestive diseases).

Most of the patients were treated under Health Care Standard, only one-third part of patients had the desired low value of cost and length of stay. Around 38% of patients

exceeded the Health Care Standard. All of this is the reason for optimization of care management.
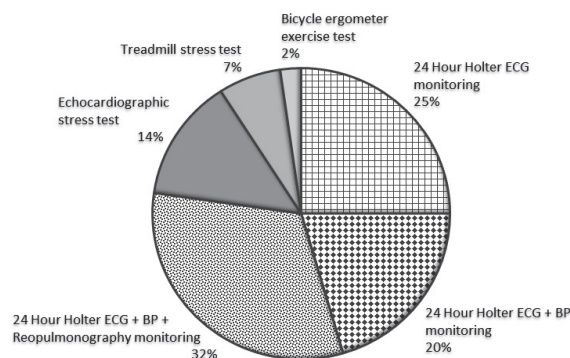

Fig. 27. Percentage distribution of investigation procedures of cardiovascular system among patients who had more than one ECG examinations (see Fig. 12).

V. INSIGHTS MONETIZATION PHASE

More accurate financial planning can be implemented having clear picture of what type of clinical / economic model refers to patient. In this study the models approach has been proposed, that helps to make more effective ROI. Two models of iPatient management were determined due to cluster analysis in Python. Simple statistical distribution of all cases doesn't reflect the hidden value: the hidden allocation into two groups, which are the basis for two models. Model A is "Cardiovascular" and Model B is "Comorbidities, digestive" (Fig. 28).

Model A (58% of patients) reflects the targeted, in a short time investigation of the cardiovascular system of patients with Essential hypertension. The more severe, small subgroup of them had highest cost of treatment, that relates to unchangeable factor for cost saving because of needed period of parenteral therapy.
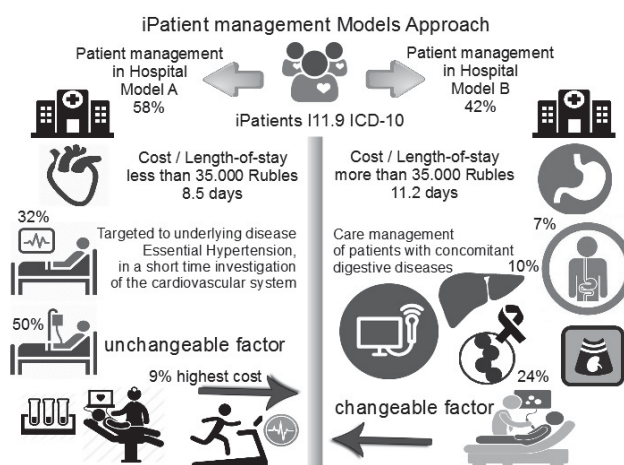

Fig. 28. Environment of Internet of Health / Smart Health with iPatient management models approach. Image of Study "iHealthCare Optimization", ERAA.

Model B (42% of patients) had a lot more number of all investigation and treatment procedures and is characterized by

accent on gastrointestinal tract investigation, on comorbidities digestive diseases. In most cases this model is expensive, exceeding the Health Care Standards for the underlying disease Essential hypertension. In Russia the digestive diseases are widespread, the normal parameters deviations are detected even among early school teens [15]. There are many ultrasonography procedures in this model. The ultrasonography examination is the changeable factor in terms of cost saving, it is the most influencing factor on length of stay in hospitals. Thus MIS can show the reminder mark to stick to the terms of the Health Care Standard, if physician would schedule abdominal ultrasonography and other ultrasonography examinations. This study demonstrates, that based on scheduled appointment procedures, MIS can assign patients to two clusters (Model A and Model B) with relevant forecast for cost and length of stay in hospitals, giving opportunity to reduce cost and length of stay.

VI. CONCLUSION. BUSINESS METAMORPHOSIS PHASE

The models approach provides the new capability to implement Health Care Standard in mode aaS, using feedback after Big Data Analytics. It shows on the collected data that a sub-group of patients with Essential hypertension also have a digestive diseases comorbidity, which should be considered in the Health Care Standard as flexible respond to patients socio-medical needs.
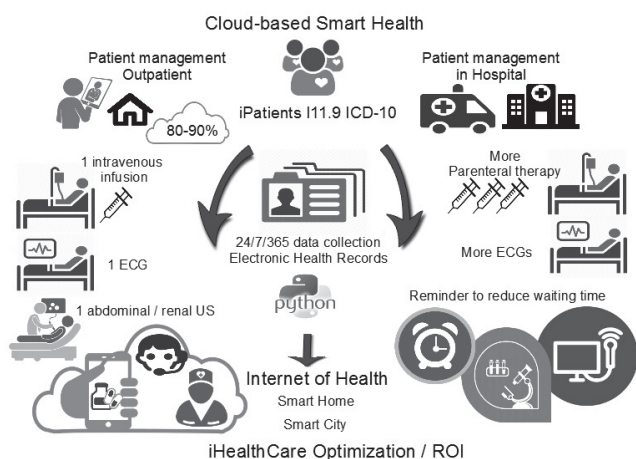

Fig. 29. Environment of Cloud-based Internet of Health / Smart Health. Image of Study "iHealthCare Optimization", ERAA.

The cluster analysis revealed that around 80-90% of patients with Hypertensive heart disease without heart failure can get investigation and treatment procedures during outpatient monitoring (including Day Hospital), without hospitalization. The road map for patient with model A or B determination can be created by MIS based on scheduled appointment procedures. Patient can get the most of investigation procedures and short-term parenteral therapy in Day Hospital (Fig. 29).

Thus the crucial metamorphosis can be achieved in the framework of Cloud-based Health Care. The goal to develop Cloud-based Health Care (telemedicine) has been set in Russia by Minister of Health Veronika Skvortsova [16], [17], [18]. This goal will significantly reduce budget spending, and will be the engine to build Internet of Health environment in Russia with Cloud-based Health Care in mode 24/7/365.

REFERENCES

[1] O. Vermesan, P. Friess, *"Digitising the Industry. Internet of Things Connecting the Physical, Digital and Virtual Worlds"*. River Publishers Series in Communications, Denmark-Netherlands, vol. 49, 2016.

[2] A. Reyss and S. Balandin, "Healthcare, medical support and consultancy applications and services for mobile devices", *in Proc. IEEE SIBIRCON Conf.* July 2010, pp. 300-305.

[3] D.G. Korzun, A.V. Borodin, I.A. Timofeev, I.V. Paramonov and S.I. Balandin, "Digital assistance services for emergency situations in personalized mobile healthcare: Smart space based approach" *in Proc. International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON),* IEEE. 2015, pp. 62-67.

[4] D.G. Korzun, A.V. Borodin, A.V. Paramonov, A.M. Vasilyev and S.I. Balandin, "Smart spaces enabled mobile healthcare services in Internet of Things environments", *International Journal of Embedded and Real-Time Communication Systems*, vol. 6 (1), 2015, pp. 1-27.

[5] E. Balandina, S. Balandin, Y. Koucheryavy and D. Mouromtsev, "IoT Use Cases in Healthcare and Tourism", *in Proc. of the 17th IEEE Conference on Business Informatics (CBI).* Lisbon, Portugal, July 13-16, 2015, pp. 37-44.

[6] D. Laure, O. Medvedev, S. Balandin, K. Lagutina, "Mobile apps for stimulating healthy life: Walky doggy reference example*", in Proc. of 17th FRUCT Conf.* 2015, pp. 110-115.

[7] L. Frolio, "Big Data Insights in Healthcare, Part I: Great Ideas Transcend Time", *Dell EMC In Focus.* Hopkinton MA, USA, September, 2015.

[8] L. Frolio, "Big Data Insights in Healthcare, Part II. A Perspective on Challenges to Adoption", *Dell EMC In Focus.* Hopkinton MA, USA, October 2015.

[9] L. Frolio, "Big Data Insights in Healthcare, Part III: Mitigating Challenges to Adoption and What Will Follow", *Dell EMC In Focus.* Hopkinton MA, USA, November 2015.

[10] B. Schmarzo, *Big Data: Understanding How Data Powers Big Business.* USA: John Wiley & Sons. Inc., 2013.

[11] B. Schmarzo, "Big Ideas Big Data Business Model Maturity Index", *Dell EMC In Focus.* Hopkinton MA, USA, 2014.

[12] B. Schmarzo, "Thinking Like a Data Scientist (Parts I—III)". *Dell EMC In Focus.* Hopkinton MA, USA, April-June 2015.

[13] B. Schmarzo, *Big Data MBA: Driving Business Strategies with Data Science*. USA: John Wiley & Sons. Inc., 2016.

[14] A.V. Reshetnikov, S.A. Efimenko, *The Sociology of the Patient.* Moscow: Zdorov'e i obshhestvo, 2008.

[15] Yu.Yu. Kolesnichenko, *Some aspects of Ultrasonography in pediatrics in primary health care.* Moscow: Folium, 2008.

[16] V.I. Skvortsova, Speech by the Minister of Health of the Russian Federation at the Board Meeting of the Ministry of Health of the Russian Federation. Moscow, April 20, 2016.

[17] V.I. Skvortsova, Speech by the Minister of Health of the Russian Federation at the Meeting of Government Council on Strategic Development and Priority Projects of Russian Federation. Moscow, July 13, 2016.

[18] V.I. Skvortsova, Speech by the Minister of Health of the Russian Federation at the Meeting of Government Council on Strategic Development and Priority Projects of Russian Federation. Moscow, August 31, 2016.