UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

The relation between the evolution in blood pressure and the evoluation in renal function over time

Robert Orlowski

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Ruth NYSEN

SUPERVISOR :

Dr. Gijs VAN POTTELBERGH

2017
2018

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

The relation between the evolution in blood pressure and the evoluation in renal function over time

Robert Orlowski

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Ruth NYSEN

SUPERVISOR :

Dr. Gijs VAN POTTELBERGH

Contents

.

Abstract

The aim of this master thesis is to extend the methodology and to confirm the results obtained in the previously published paper in which the authors explored the relationship between evolution of kidney function over time and blood pressure. Specifically, it is of interest to expand the statistical methodology presented in the previous analysis by Vaes et al. (2015) and to incorporate the longitudinal structure of the data to make inference about how blood pressure measurements influence the kidney function expressed as estimated glomerular filtration rate (eGFR).

Data used in this analysis come from a large retrospective epidemiological cohort study following up patients with morbidities in Flanders, Belgium. Measurements taken between years 2002 and 2012 were extracted from the database and included in the analysis. As to the methods used in this analysis, linear mixed models for longitudinal data and generalized estimating equations (GEE) were used to answer the research questions.

A positive effect of systolic and diastolic blood pressure on the baseline eGFR values was identified. Moreover, it was concluded that the effect of BP changes as time from baseline elapses and the overall eGFR level decreases as time passes. Patients characteristics such as gender and age above 70 years and above 80 years had a significant negative effect on the eGFR value, also a positive relationship between cardiovascular mediaction and outcome was confirmed. The probability of annual rapid kidney function decline was affected by patient baseline characteristics and by baseline eGFR value with higher values contributing surprisingly to higher odds for rapid decline.

1 INTRODUCTION

The aim of this master thesis is to extend the methodology of the previously published paper in which the authors explored the relationship between evolution of kidney function over time and blood pressure. Specifically, it is of interest to expand the statistical methodology presented in the previous analysis and to incorporate the longitudinal structure of the data to make inference about how blood pressure measurements influence the kidney function expressed as estimated glomerular filtration rate (eGFR).

Kidney functioning is an important parameter reflecting general health and can be measured by among others serum creatinine. It is well documented that renal function (expressed as the behaviour of eGFR values) decreases over time, This can be consulted in Danziger et al. (2002) and in Lindeman et al. (1985). Many studies have reported up to now the connection between hypertension and cardiovascular diseases and the renal function, e.g. US Renal Data System (2013). Such studies have usually excluded patients of older age due to their existing comorbidities, especially when considering them in clinical trials setting. Analysis conducted by Vaes et al. (2015) is the first know epidemiological study to include a cohort of patients followed-up for a longer period of time and including information on existing comorbidities. Data were extracted from a huge morbidity registry containing millions of observations and measurements taken. The statistical analysis revealed that the renal functioning decline in older patients (aged 60 or above) can be attributed to the decline in blood pressure.

The main aim of this project is to confirm the results obtained in the primary analysis and to point out some flaws that failed to be eliminated from that analysis.

The thesis consists of the following sections:

- Critical description of the previous analysis conducted by Vaes et al (2015).
- Data and database description
- Description of problems frequently encountered in the analysis of highly unbalanced longitudinal data
- Description of methods and models employed in the analysis
- Results
- Discussion and comparison of results

- Limitations and suggestions for future research

2 PREVIOUS ANALYSIS BY VAES ET AL. (2015)

The previous analysis conducted by Vaes et al. (2015) included the population of patients aged 60 or older starting from year 2002. Patients' data collected between years 2002 and 2012 were included in the analysis. The authors considered a vast variety of models, each of them presenting a different approach to modelling the eGFR value or change of the latter in time.

Sadly, the results of this analysis are not reproducible due to the poor quality of data extraction description. It cannot also be confirmed (as stated by the authors themselves during one of the discussions) that the data used in this analysis were cleaned before commencing statistical modelling, e.g. it is not clear if impossible or outlying values of systolic and diastolic blood pressure were transformed (e.g. by hardcoding in the data set) or excluded from the analysis.

Apart from that, the presented analysis suffers from some shortcomings of statistical nature, some of these were apparently not even eliminated during the journal submission process. This section summarizes briefly the most flagrant statistical science malpractice which escaped the attention of the reviewers. Some of the comments are also just indicating the differences between the previous and current analysis and should not be deemed as criticism.

First of all, there is a certain difference between handling the data with regard to its longitudinal unbalanced structure. Similarly to the current analysis, Vaes et al. (2015) have chosen to present only one subject visit (one measurement of EGFR and systolic and diastolic blood pressure) per year. It certainly helps to facilitate the instances in which subjects have a cluster of visits within a short time period which can distort the analysis in a significant way. Specifically, if a subject had multiple visits during one year, it was decided to use the mean of the last two measurements as a proxy for an overall yearly measurement. This approach distorts not only the longitudinal structure but also the time point analysis meaning that the measurements (visit) times should follow a random stochastic process as they are not planned beforehand and are influenced by a variety of external factors. This approach makes it also impossible to examine the evolution of eGFR over time with regard to single time points.

In the statistical models themselves, the time variable included in the analysis was expressed

as time elapsed from baseline measurements until the last one taken on the same subject which can be considered as an error but facilitates longitudinal data handling without diverging into the realm of methods for repeated measures. The analysis did not follow according to the longitudinal data analysis principle, i.e. the models are not of a longitudinal nature as the formulation of the model does not account for multiplicity of measurements. This is a major flaw which can make the statistical inference not valid in terms of made decisions.

The authors present each and every time three versions of the same model for different age stratum (patients aged 60-69, 70-79 and over 80 years old). This approach does not allow to test if there is any statistically significant difference between the strata with regard to the parameters and outcomes of interest. The more plausible and efficient approach (presented in the current analysis) is to include variables representing the age strata in the statistical models of interest and to assess their impact and significance in the analysis. Presentation of separate models for cohorts may be viewed as inefficient and redundant. In this analysis it was of interest to examine the age effect for three different explicitly defined strata, thus age was categorized and continuous value of subject age was not used.

It should be also stressed that in the contrary to what is presented in Vaes et al. (2015) the estimate of the parameter in the linear regression analysis, let alone logistic regression, cannot be interpreted as correlation. The value of the parameter estimates show indeed a linear dependence between dependent and explanatory variable but is not a correlation in the strict sense as correlation measures the strength of association and takes values between -1 and 1. This lapse is repeated constantly and several times in the previous analysis.

All statistical models presented in the paper include systolic and diastolic pressure effects among the explanatory variables in all considered models, as well as pulse pressure. The fact that the pulse pressure effect is also included in each statistical model and since the value of this parameter is a difference of the two latter variables - systolic and diastolic blood pressure (and thus is their linear combination) this results in the presence of strong collinearity which eventually may cause invalidity of the estimated parameter values and inference based thereon.

In this analysis models presented in the original paper were revised and adjusted to reflect the repeated measurement structure of the data and consider the correlation between measurements taken on the same subject across time.

3 DATA DESCRIPTION

3.1 Use of registry database

The data used for the current analysis were extracted from a huge database which contains patient data gathered by 97 general practitioners active in 55 different medical centres across Flanders, Belgium [the same database was used by Vaes et al. (2015)]. The data contain not only information on systolic and diastolic blood pressure and kidney function but also on many different parameters. Details about the whole database can be consulted with Intego morbidity registration network at the Department of General Practice of the KU Leuven in Belgium. Separate data sets were provided for analysis purposes, specifically following data sets were used:

- Patients data set: data set containing patients' identification numbers and characteristics (age and sex) along with the visit details (dates)
- Measurements data set: data set containing information on the systolic and diastolic blood pressure results, as well as some other parameters which were not of interest (e.g. heart rate), all reported with dates of measurements
- Laboratory tests data set: data set containing laboratory results data along with date for a huge amount of medical tests, only creatinine serum was of interest and has been extracted
- Diagnoses data set: data set containing information on chronic diseases (comorbidities) diagnoses and date of diagnosis, only chronic diseases specified later in the next section of the thesis were extracted
- Prescription data set: data set containing information on the prescriptions for patients along with prescriptions dates, only data containing information on the use of cardiovascular medication was extracted.

All of the data sets contain millions of observations with the laboratory results data set containing over 30 million observations which made the data extraction and data set merging process computer intensive.

For the current study data for patients from the period 2002-2012 were collected which is in line with the study described by Vaes et al. (2015) Only patients of 60 years old or older were included in the analysis.

Systolic and diastolic blood pressure measurements for patients fulfilling the aforementioned conditions were extracted from the Measurement data set and subsequently cleaned to exclude observations with values considered as impossible (e.g. systolic blood pressure not larger than diastolic, systolic blood pressure equal diastolic or impossible values of pulse pressure defined as the difference between systolic and diastolic measurement). Laboratory results data for serum creatinine were extracted from the Laboratory tests data set. Only values within a normal range were included in the analysis (as instructed by the external supervisor), meaning only values between 0.6 and 1.2 mg were considered as valid.

In order to obtain the longitudinal structure of the data laboratory serum creatinine test results were merged with blood pressure measurements by data. Only instances for which merging resulted in obtaining a valid record, i.e. a record with a full date in DDMMYYYY format were included in the analysis. At a later stage patient characteristics with regard to age and sex were extracted from the Patients dataset and merged with longitudinal measurement compliant set. The kidney function parameter eGFR (estimated glomerular filtration rate) was calculated based on the value of serum creatinine using the Modification of Diet in Renal Disease (MDRD) equation (adjusting the formula for men and women and patient's characteristics).

Theoretical concepts of the equation and how it translates into the renal function can be found in Levey et al. (2007). Lower values of eGFR are associated with worsening renal function and values below 60 are deemed as a sign of kidney disorder.

Comorbidity information to be included in the model takes the form of a generic index expressed as Charlson Comorbidity Index (abbreviated as CCI) which was modified and adjusted to facilitate the current analysis. Information on the index itself can be found in Charlson et al. (1987), it should be mentioned here however that the index translates the obtained score into the number of survival years within the 10 years time interval. In order to obtain the comorbidity information needed to calculate the CCI index for patients, only diseases contributing to the index values were extracted. Table 1 presents the full frequency tabulation of comorbidities and cardiovascular medication. The limitation of this approach was however that similarly to the original analysis presented by Vaes et al. (2015) it was not possible to associate the cancer stage information with malignancy information in the existing data set. This is due to the fact that the cancer information was provided with no

further details on the stage of the disease or type of malignancy itself.

Therefore it was assumed that the value of a non-metastatic malignancy will be assigned for each instance of carcinoma (this approach is in line with what is presented in the original analysis). No information on the type of diabetes (uncomplicated or end-organ damage) and liver disease (mild or moderate to severe) could also be retrieved, therefore it was decided to assign the value for the lowest category of severity (again, a non-conservative approach). As it is impossible to retrieve and asses the connective tissue disorders based on the data from the registry, this type of comorbidity was not included in the calculation of the CCI index.

Finally, the data on cardiovascular medication at baseline were retrieved from the Prescription data set. Categories of interest were

- Beta-blockers
- ACE inhibitors and angiotensin receptor blockers
- Calcium antagonists
- Diuretics.

In the final data set information on the use of the cardiovascular (hypertensive) concomitant medication was expressed as a generic dummy variable expressing the fact of having ever used such medication (no differentiation between different classes of cardiovascular medications was made in the considered statistical models).

Table 2 gives summary statistics and category frequencies for baseline patient characteristics for patient included in the analysis. The number of subjects included was 8426.

Table 1: Baseline comorbidities and concomitant cardiovascular medication medication

	Age 60-69, n=4510	**Age 70-79, n=2759**	**Age > 80, n=1157**	**All, n=8426**
Comorbidities				
Diabetes	1216 (27.0%)	700 (25.3%)	281 (24.3%)	2197 (26.1%)
Myocardial infarction	231 (5.1%)	174 (6.3%)	94 (8.1%)	499 (5.9%)
Chronic pulmonary disease	289 (6.4%)	251 (9.1%)	96 (8.3%)	636 (7.6%)
Cerebrovascular acctdent / TIA	287 (6.7%)	321 (11.6%)	238 (20.6%)	846 (10.0%)
Dementia	52 (1.2%)	123 (4.7%)	99 (8.6%)	280 (3.3%)
Hemiplegia	34 (0.8%)	30 (1.1%)	27 (2.3%)	91 (1.1%)
Leukemia	16 (0.4%)	13 (0.5%)	4 (0.4%)	33 (0.4%)
Liver disease	356 (7.9%)	167 (6.1%)	63 (5.5%)	586 (7.0%)
History of peptic ulcer disease	333 (7.4%)	240 (8.7%)	101 (8.7%)	674 (8.0%)
Heart failure	82 (1.8%)	163 (5.9%)	142 (12.3%)	387 (4.6%)
History of cancer	659 (14.1%)	436 (15.8%)	188 (16.3%)	1283 (15.2%)
Peripheral arterial illness	252 (5.6%)	255 (9.2%)	114 (9.9%)	621 (7.4%)
Lymphoma	21 (0.5%)	25 (0.9%)	4 (0.4%)	50 (0.6%)
Cardiovascular medication				
Beta-blockers	1504 (33.4%)	1372 (49.7%)	739 (63.9%)	3615 (42.9%)
Diuretics	2563 (56.8%)	1768 (64.8%)	776 (67.1%)	5127 (60.9%)
Calcium antagonists	1397 (31.0%)	1136 (41.2%)	570 (49.3%)	3103 (36.8%)
ACE inhibitors / angiotensin receptor blockers	2206 (48.9%)	1506 (54.5%)	687 (59.4%)	4399 (52.1%)

Table 2: Summary statistics for patient baseline characteristics

Variable	Age 60-69, n=4510	Age 70-79, n=2759	Age > 80, n=1157	All, n=8426
No. of women	2179 (48.3%)	1476 (53.3%)	715 (61.8%)	4369 (51.9%)
No. of men	2332 (51.7%)	1283 (46.5%)	442 (38.2%)	4057 (48.2%)
SBP	134 ± 16.2	137.45 ± 15.3	139 ± 16.0	139 ± 16.0
SBP <120 mmHg	394 (8.7%)	156 (5.7%)	48 (4.2%)	598 (7.1%)
SBP 120-129 mmHg	1033 (22.9%)	450 (15.3%)	191 (16.5%)	1674 (19.9%)
SBP 130-139 mmHg	1203 (26.7%)	765 (27.7%)	272 (23.5%)	2240 (26.6%)
SBP 140-149 mmHg	1067 (23.7%)	764 (27.7%)	322 (27.8%)	2153 (25.6%)
SBP >150 mmHg	813 (18.0%)	624 (22.6%)	324 (28.0%)	1761 (20.9%)
DBP	82 ± 9.3	81 ± 8.2	80 ± 8.1	81 ± 8.9
DBP <70 mmHg	92 (2.0%)	49 (1.8%)	40 (3.5%)	181 (2.2%)
DBP 70-79 mmHg	891 (19.8%)	590 (21.4%)	294 (25.4%)	1775 (21.1%)
DBP 80-89 mmHg	2387 (52.9%)	1580 (57.3%)	632 (54.6%)	4599 (54.9%)
DBP >90 mmHg	1140 (25.3%)	540 (19.6%)	191 (16.6%)	1871 (22.2%)
eGFR	79 ± 13.8	75 ± 13.5	71 ± 13.9	77 ± 14.07
eGFR 45-59	260 (5.8%)	389 (14.1%)	262 (22.6%)	912 (10.8%)
eGFR >60	4250 (94.2%)	2370 (85.9%)	894 (77.4%)	7514 (89.2%)

3.2 Missing data

After conducting the data set merge across all aforementioned domains it was discovered that no missing data with regard to the renal function parameter (eGFR) and both systolic and diastolic blood pressure were revealed. This came unexpected given the size of the data sets. It should be stressed that no criteria on missingness were specified when merging the data sets. Missingness occurred only when calculating the Charlson Comorbidity Index and retrieving data on cardiovascular medication for patients included in the analysis and fulfilling criteria of the cohort of interest. This can be however interpreted as not having a comorbidity and not being on hypertensive treatment, thus no further action is required and taken to deal with this kind of missingness. In such cases subjects were assigned a CCI score for an individual without comorbidities,

4 THE NATURE OF THE UNBALANCED LONGITUDINAL DATA

4.1 Unbalanced data in longitudinal studies

The data collected in epidemiological studies and clinical trials usually are unbalanced. This holds not only for data whose structure is longitudinal in their nature but also for cross-sectional observations. It should be stressed that the lack of balance in the data may be caused by several factors ranging from the presence of missing data (this often occurs in the case of randomized clinical trials testing drugs' efficacy and safety but as revealed in section 3.2 this does not have any effect on the current analysis), different numbers of observations collected on the same subject (patient) or different time intervals between measurements or inpatient and outpatient visits.

When data originating from longitudinal cohort studies are considered, no matter if prospective or retrospective, e.g. from general practitioner registries like in this analysis, it should come as no surprise that such data are highly unbalanced. There are multiple reasons for this phenomenon. First of all, patients present in the registries vary in terms of their health status. Some of them may suffer from recurring conditions which require constant or frequent observation or intervention of the doctor. Therefore, patients with better general health condition will generally have less observations than patients dealing with a chronic disorder such as for instance hypertension or renal function decline. It can be argued on the other hand that the number of visits for a patient in the registry may be considered as a random stochastic process as it cannot be said or foreseen when visits take place, examples are provided in Cote and Stein (2007). It might be the case for healthy individuals who show up for a check-up every year but even in this case one can never be sure if an urgent condition will not force him or her to conduct some medical tests and examinations. At the same time subjects who experience some (severe) conditions tend to be examined on regular basis and within shorter time periods. In the latter case patients may experience any kind of exacerbations which cause their frequent presence at hospitals or general practitioners' practices. Those frequent visits may contribute to clusters on the timeline of visits for a single patient in the sense that there will be a vast amount of observations taken within a short period followed by a longer break when the condition is cured or in remission. This can cause serious problems within the framework of statistical modelling of e.g. longitudinal data. To facilitate this, it was decided to take only one measurement per subject per year to cater for more or less equal spacing between visits and to avoid clustering of visits (measurements) during one year.

In case of more than one measurements taken within one day, the second one was extracted from the database following the simple reasoning that there must have been a reason to repeat measurement e.g. because of error during the first attempt. It is possible that models for longitudinal data would converge if this was not done but usually the Hessian is not positive definite in such cases (when considering the mixed models).

All of the above mentioned patterns (as described earlier with underlying random process) have a huge impact on the structure of the data collected in studies such as this one. In this retrospective longitudinal study patients were followed up without any specific pattern. A single visit can be described as a result of a random process and the spacing between visits varies between patients. The number of visits per subject between years 2002 and 2012 (only observations from this period and cohort were extracted) varies between 2 and 8 and the longest break (disruption) between observations was 4 years.

Unbalanced data of this kind require also different handling of the baseline value. As it is impossible from definition (especially in retrospective epidemiological cohort studies) to register the first observation at the same time, the time variable has to be adjusted so that it represents the time of inclusion for the particular subject. In this analysis the baseline time value was adjusted as zero for each and every individual registered in the database, not taking into account that in fact the baseline measurements for all patients were not collected at a joint 'moment zero' resulting in a more 'artificial' definition of baseline. In other words, the clock starts ticking from the first measurement taken in the first registered year of being under observation. This approach is however acceptable as the individuals are independent of one another.

Finally, statistical considerations of more theoretical nature cannot be forgotten. Another drawback which arises when modelling of repeated measurements is of interest is that statistical models for longitudinal data require suitable approach to variance-covariance matrix choice. As most of the variance-covariance matrices are suitable mainly or only for equally spaced longitudinal data, the choice between different types of underlying covariance structure becomes narrower and more cumbersome. As the maximum number of observations per subject may be high, it might be not possible to estimate the parameters in more complex covariance structures due to the high number of parameters to be estimated. This happens for instance in the case of the unstructured covariance matrix. In unbalanced longitudinal data setting models can quickly become overparametrized resulting in poor quality of

estimators. It might also happen that convergence criteria might not be met. The choice of parsimonious structures in this case is limited and the models require some rigorous assumptions. This is however the price to pay for more efficient estimation process. In the case of the present analysis where maximum number of measurements equals 9 the use of unstructured covariance matrix would entail estimation of 45 additional parameters.

Table 3 summarizes the frequency of the number of observations taken per subject and shows that data are structured in a highly unbalanced way.

Table 3: Number of measurements taken per subject (no. of visits per subject)

No. of Visits	Frequency	Percent
3	2958	35.1 %
4	1778	21.1 %
5	1389	16.5 %
6	1016	12.1 %
7	734	8.7 %
8	398	4.7 %
9	153	1.8 %

4.2 Covariance structure considerations

As mentioned in the previous sections, modelling of highly unbalanced longitudinal data requires modified approach to the choice of the covariance structure which expresses how measurements taken at different time points are correlated with each other. First the fact that the data are unequally spaced, i.e. there are different intervals between measurements has to be taken into account. What is more, subjects vary in terms of the number of available observations, meaning that some may have only 3 registered measurements, some 5 an some up to 9 measurements taken. This fact alone implies that the assumption of all subject having identical variance-covariance structure is impossible, not even only when its dimensions are considered.

It has been already mentioned that most of the popular covariance structures are either not feasible with regard to the estimation process (e.g. unstructured covariance matrix) or are not suitable for longitudinal data which are not characterized by the property of having all measurements equally spaced. This sole condition already puts aside such covariance structures as among others Toeplitz, AR(1) [autoregressive of first order] and banded. The

reasoning behind the choice of the structure for unbalanced data with different measurement times for all subjects should follow the logic. It goes without saying that the most intuitive thinking should be considered as the most appropriate one. Thus, measurements taken at time points adjacent to each other should be more correlated with each other than with other measurements taken within more distant interval.

This property can be achieved by generalizing the autoregressive AR(1) structure to more general one, in which measurements can be unequally spaced. Such covariance structure is expressed as

$$\mathbf{\Sigma} = \mathrm{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \begin{pmatrix} 1 & \rho^{d_{12}} & \cdots & \rho^{d_{1n}} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{d_{n1}} & \rho^{d_{n2}} & \cdots & 1 \end{pmatrix}$$

and is usually referred to as spatial covariance power structure because of the fact that it implements the distances between observations in the exponent. It still holds even if the measurement times are different as covariance within a subject observations depend on the distance between registered visits. This structure is sometimes referred to in the literature as Markov structure [e.g. Diggle (2013)].

Another possible structure might be the compound symmetry (CS) variance-covariance structure as it does not depend on the values of the times. This structure is given by

$$\mathbf{\Sigma}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

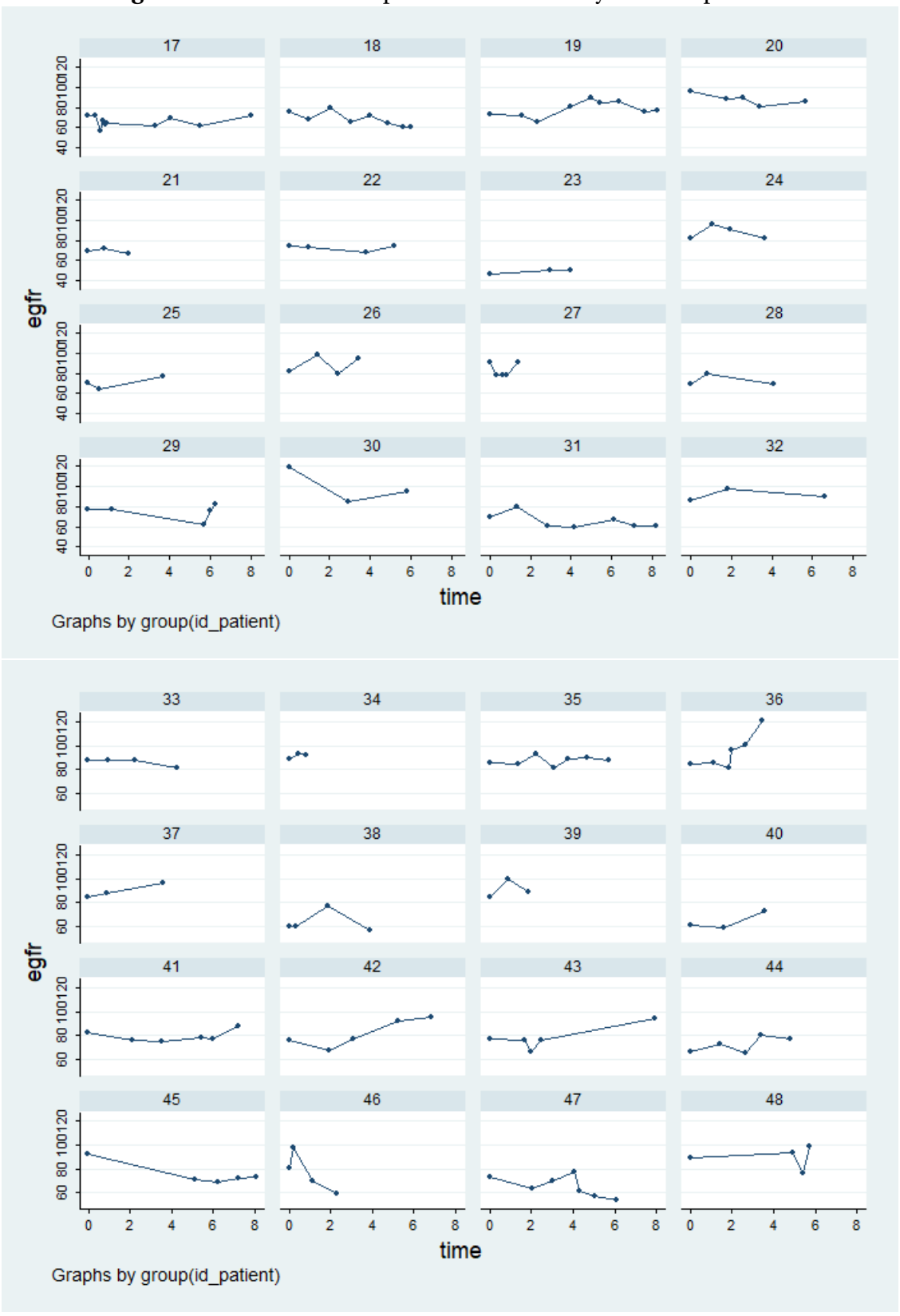
This approach is however limited by the rigorous condition that all correlation between all observations remain the same. It should be also noted that the unstructured variance-covariance matrix has another drawback in the situation of modelling measurements taken at different time points. Subjects do not have common covariance parameters, so the values in the matrix should depend on subject specific quantities.

5 METHODS

5.1 Linear mixed models

In order to answer the research question, so to explore the association between kidney function over time and blood pressure (expressed as systolic and diastolic blood pressure separately) the linear mixed model with random intercept was employed. The random intercept was added to reflect the variation between baseline eGFR values across different subjects. The trend can be approximated by the linear function as it can be seen on the profile plots on Figure 1. The plot contains observed eGFR value evolution across time for randomly selected 32 subjects from the cohort which was included in the current analysis. It can be easily seen that there is a need for random effect on subject level as starting eGFR values differ to some extent from patient to patient. Such model contains two sources of variation: first random intercept (variation in baseline values across individuals) and random component (variability not explained by the model). The model was estimated in SAS 9.4 software using the PROC MIXED procedure from SAS STAT 14.3 product. The program employed the restricted maximum likelihood method (REML) to obtain the parameter estimates for fixed effects and to calculate the variance of the random intercept effect. Contrary to the classical maximum likelihood estimators, REML provides unbiased estimates of the variance and covariance components.

Figure 1: EGFR evolution profiles for randomly selected patients



As the interest lies in the population effect, the inference will be made on the fixed effects and no subject-specific values of the random effects will be shown. It should be stressed though that the mixed models follows taking into account the value of subject-specific random effect (intercept in this case).

To test the effect of the variance-covariance structure on the models of interest, both models were run in two versions: first with the compound symmetry structure and second with the spatial power (Markov) structure. The assumptions that follow in all presented linear models is that the random intercept and random component follow normal distribution with itself specific variance.

5.1.1 Model specification

General remark for all models: for controlling variables Age, Sex and Use of cardiovascular medication, age group of subjects between 60 and 69 years old, women and no cardiovascular medication used respectively were selected as reference categories.

5.1.2 Model 1.1 and 1.2

These models explore the impact of systolic and diastolic blood pressure on the value and evolution (through the inclusion of effect interactions with time variable) of estimated glomerular filtration rate in time controlling for baseline characteristics. The dependent variable is the eGFR for i-th subject at time point t. The explanatory variables are:

- Systolic blood pressure for i-th subject at time point t
- Diastolic blood pressure for i-th subject at time point t
- Interaction of systolic blood pressure for i-th subject at timepoint t and time
- Interaction of dystolic blood pressure for i-th subject at timepoint t and time
- Sex of i-th subject
- Binary variable representing age between 70 and 79 years old for i-th subject at baseline
- Binary variable representing age 80 and more years old for i-th subject at baseline
- Binary variable representing fact of using cardiovascular medication i-th subject at baseline

- CCI index score for i-th subject at baseline
- Random intercept effect representing change in eGFR due to i-th subject itself.

5.1.3 Model 2.1 and 2.2

Models 2.1 and 2.2 explores the impact of the value of the baseline blood pressure measurements and baseline eGFR measurement on the subsequent measurements in time using the value of eGFR registered after baseline as dependent variable. As it is of interest to examine the effect of the baseline measurement on the subsequent values of eGFR, the dependent variable is eGFR for i-th subject at time point t, t>0, thus the number of observations used in this case is decreased by the number of subjects (as the baseline value is not used as a dependent eGFR value but rather as an explanatory). The full list of explanatory variables contain:

- Baseline systolic blood pressure for i-th subject
- Baseline diastolic blood pressure for i-th subject
- Interaction of baseline systolic blood pressure for i-th subject and time
- Interaction of baseline diastolic blood pressure for i-th subject and time
- Sex of i-th subject
- Binary variable representing age between 70 and 79 years old for i-th subject at baseline
- Binary variable representing age 80 and more years old for i-th subject at baseline
- Binary variable representing fact of using cardiovascular medication i-th subject at baseline
- CCI index score for i-th subject at baseline
- Random intercept effect representing change in eGFR due to i-th subject itself.

5.2 Generalized Estimating Equations (GEE)

In order to assess the impact of blood pressure controlled for patient baseline characteristics on the probability of rapid annual decline of kidney function the dataset was first adjusted in order to implement the probability evolution modelling. New binary variable 'rapid annual

eGFR decline' has been defined and served as an outcome variable in the analysis [annual decline in eGFR of at least 3 mL/min/1.73m^2, value of change reported by Rifkin et al.(2013)]. Because of the fact that annual change is being modelled only observations taken within subsequent years (and with interval of approximately one year - ranging from 11 to 13 months) were extracted from the data set used during modelling of continuous outcome. This resulted in a dramatic decrease in the number of observations and subjects. Data for 5277 subjects with measurements taken in subsequent years are available and used to estimate the model parameters (with the minimum number of observations (measurements) per subject being 2 and maximum 8).

The modelling process followed using the generalized estimating equations (GEE). GEE is a class of non-likelihood (quasi-likelihood) methods for correlated data and follows by the choice of the correlation working structure. The choice of this structure is not of crucial importance as the correlation between measurements taken on the same subject is treated rather as a nuisance. For instance under the independence working correlation assumption parameters are equivalent to those obtained by classical maximum likelihood methods with adjusted standard errors. The peculiar thing about GEE methods is that the estimates of model parameters are valid even if the working correlation structure has been misspecified. In this case the estimation followed using the independence assumption (other structures were also considered but e.g. unstructured failed to produce estimates in the SAS PROC GENMOD program - this structure was considered as even though it might not be plausible, the inference and parameter estimates would still be valid as described in section 5.2). It is also worth noticing that due to the fact that GEE does not use maximum likelihood methods but rather quasi-likelihood some researchers argue that it is an estimation method rather than a model as mentioned in Agresti (2007).

Explanatory variables include

- Systolic blood pressure for i-th subject at time point t (measurement date)
- Diastolic blood pressure for i-th subject at time point t (measurement date)
- Interaction of systolic blood pressure for i-th subject at time point t (measurement date) and time
- Interaction of diastolic blood pressure for i-th subject at time point t (measurement date) and time

- Sex of i-th subject
- Binary variable representing age between 70 and 79 years old for i-th subject at baseline (baseline defined as the first measurement in the year-to-year sequence)
- Binary variable representing age 80 and more years old for i-th subject at baseline
- Binary variable representing fact of using cardiovascular medication i-th subject at baseline
- CCI index score for i-th subject at baseline

6 RESULTS

All models specified earlier in the Methods section reached convergence criteria and the parameters estimates were obtained. Tables 4 and 5 present fixed effects estimates for models 1.1 and 1.2 respectively and Table 7 and 8 present model estimates for models 2.1 and 2.2.

As the number of parameters to be estimated in the variance-covariance matrix in both cases (spatial power covariance structure and compound symmetry covariance structure) equals 2, it is impossible to compare models with these structures using the standard likelihood ratio approach as models are not nested within the parameter space. Instead, model with the lowest AIC value will be chosen and the inference will follow using the estimates for such model. However, output for both models is given as it might be of interest to assess the impact of the choice of the covariance structure on obtained estimates.

Table 10 contains parameter estimates for the GEE model.

Table 4: Linear mixed model 1.1, CS covariance structure, dependent variable: eGFR value at time point t (including baseline)

	Estimate	Standard error	$Pr > \|t\|$
Intercept	62.4707	1.0050	<.0.001
Diastolic	0.0663	0.0127	<.0.001
Systolic	0.0586	0.0071	<.0.001
Diastolic× Time	-0.0133	0.0031	<.0.001
Systolic× Time	-0.0050	0.0018	0.0043
Time	2.4310	0.2290	<.0.001
Sex	9.7698	0.2590	<.0.001
Age (at the entry)			
70 - 79	-4.2632	0.3100	<.0.001
80 and more	-8.0038	0.4606	<.0.001
Drugs	-2.3609	0.3387	<.0.001
CCI	0.1337	0.099	0.1810

Table 5: Linear mixed model 1.2, spatial power covariance structure, dependent variable: eGFR value at time point t (including baseline)

	Estimate	Standard error	$Pr > \|t\|$
Intercept	64.8580	1.0111	<.0.001
Diastolic	0.0524	0.0127	<.0.001
Systolic	0.0486	0.0071	<.0.001
Diastolic× Time	-0.0071	0.0032	0.0289
Systolic× Time	-0.0042	0.0018	0.0220
Time	1.8223	0.2406	<.0.001
Sex	9.7615	0.2574	<.0.001
Age (at the entry)			
70 - 79	-4.2388	0.3081	<.0.001
80 and more	-8.0074	0.4583	<.0.001
Drugs	-2.3054	0.3364	<.0.001
CCI	0.1352	0.0994	0.1739

Table 6: Linear mixed model 1.3, spatial power covariance structure, dependent variable: eGFR value at time point t (including baseline), non-significant effect of CCI excluded

	Estimate	Standard error	$Pr > \|t\|$
Intercept	65.2108	0.9773	<.0001
Diastolic	0.05203	0.01265	<.0001
Systolic	0.04874	0.007106	<.0001
Diastolic× Time	-0.00702	0.003228	0.0296
Systolic× Time	-0.00418	0.001817	0.0215
Time	1.8220	0.2406	<.0001
Sex	9.7999	0.2559	<.0001
Age (at the entry)			
70 - 79	-4.0782	0.2846	<.0001
80 and more	-7.6847	0.3921	<.0001
Drugs	-2.2407	0.3331	<.0001

Table 7: Linear mixed model 2.1, CS covariance structure, dependent variable: eGFR value at time point t (excluding baseline)

	Estimate	Standard error	$Pr > \|t\|$
Intercept	24.477	1.6183	<.0.001
eGFR(baseline)	0.6658	0.0107	<.0.001
eGFR(baseline)× Time	-0.0160	0.0020	<.0.001
Diastolic (baseline)	-0.0187	0.0208	0.3681
Systolic (baseline)	0.0186	0.0114	0.1027
Diastolic (baseline)× Time	-0.0037	0.0040	0.3578
Systolic (baseline)× Time	-0.0154	0.0022	<.0.001
Time	4.2667	0.3093	<.0.001
Sex	3.8886	0.2300	<.0.001
Age (at the entry)			
70 - 79	-1.8045	0.2607	<.0.001
80 and more	-3.5206	0.3919	<.0.001
Drugs	-0.7448	0.2898	0.0102
CCI	0.0135	0.0835	0.8717

Table 8: Linear mixed model 2.2, spatial power covariance structure, dependent variable: eGFR value at time point t (excluding baseline)

	Estimate	Standard error	$Pr > \|t\|$
Intercept	24.1886	1.7291	<.0.001
eGFR (baseline)	0.6735	0.0114	<.0.001
eGFR (baseline)× Time	-0.0178	0.0023	<.0.001
Diastolic (baseline)	-0.0221	0.0223	0.3204
Systolic (baseline)	0.0183	0.0122	0.1324
Diastolic× Time	-0.0021	0.0046	0.6455
Systolic× Time	-0.0156	0.0025	<.0.001
Time	4.3125	0.3557	<.0.001
Sex	3.8737	0.2293	<.0.001
Age (at the entry)			
70 - 79	-1.7866	0.2597	<.0.001
80 and more	-3.5180	0.3909	<.0.001
Drugs	-0.7377	0.2888	0.0107
CCI	0.0147	0.0832	0.8595

Table 9: Linear mixed model 2.3, spatial power covariance structure, dependent variable: eGFR value at time point t (excluding baseline), non-significant effect of CCI excluded

	Estimate	Standard error	$Pr > \|t\|$
Intercept	24.2299	1.7132	<.0001
eGFR (baseline)	0.6735	0.01144	<.0001
eGFR (baseline)× Time	-0.01782	0.002321	<.0001
Diastolic (baseline)	-0.02232	0.02223	0.3154
Systolic (baseline)	0.01836	0.01214	0.1307
Diastolic× Time	-0.00212	0.004613	0.6461
Systolic× Time	-0.01563	0.002500	<.0001
Time	4.3124	0.3557	<.0001
Sex	3.8778	0.2282	<.0001
Age (at the entry)			
70 - 79	-1.7694	0.2410	<.0001
80 and more	-3.4835	0.3389	<.0001
Drugs	-0.7306	0.2860	0.0106

Table 10: GEE model estimation results, independence working correlation assumption, dependent variable: dummy representing annual eGFR value decline of at least 3 mL/min/1.73m^2

	Estimate	Standard error	$Pr > \|t\|$
Intercept	-2.3414	0.3298	<.0.001
Diastolic	0.0033	0.0043	0.4426
Systolic	-0.0022	0.0025	0.3727
Diastolic× Time	-0.0015	0.0011	0.1591
Systolic× Time	0.0003	0.0006	0.6000
eGFR (baseline)	0.0203	0.0011	<.0.001
Time	0.0820	0.0790	0.2992
Sex	-0.2049	0.0285	<.0.001
Age (at the entry)			
70 - 79	0.0515	0.0318	0.1055
80 and more	0.1547	0.0489	0.0016
Drugs	0.1348	0.0356	0.0002
CCI	0.0095	0.0105	0.3665

7 DISCUSSION OF RESULTS

7.1 Models 1.1 and 1.2

The AIC values for models 1.1 and 1.2 are 289 719.1 and 288 238.1 respectively. Thus the inference will follow using model 1.2 (spatial power covariance structure). The variance of the random effects is estimated as 106.12 and the residual variance is given by 87.98. It should be stressed stressed that inference in models containing random effects should follow given the value of the random effects. Main effects of systolic and diastolic blood pressure as well as the effect of interaction of the latter variables with time are highly significant. It should be stressed that the main effect should not be interpreted alone and without taking interaction terms into account. In this case however, the main effect may be interpreted as the effect of systolic and diastolic blood pressure on the baseline value of eGFR (when the time variable takes value 0 and thus the interaction term cancels out). It means that the value of both systolic and diastolic blood pressure have a significant impact on the baseline eGFR value as well as on the evolution of eGFR in time. Men have significantly higher value of the estimated glomerular filtration rate by almost 10 mL/min/1.73m^2 when compared to women. Age groups associated with higher age are also a strong predictor of lowered eGFR values. The effect of the concomitant cardiovascular medication is also highly significant and contributes to the lower values of eGFR. Comorbidity expressed as Charlson Comorbidity Index (CCI) does not influence the outcome.

Systolic and diastolic blood pressure have a significant contribution to the baseline eGFR value and this effect is modified at subsequent measurements by the highly significant interaction with time. For the baseline measurement there is a positive relation between blood pressure and eGFR, meaning that lower values of blood pressure translate into lowered eGFR when holding other effects constant. The estimate of both interactions (for systolic and diastolic blood pressure and time) are negative which implies that along with higher time values the effect is modified and that the eGFR values decrease as time passes. It is in line with the fact that the values of the outcome variable tend to decrease in older groups of patients. The joint effect of blood pressure for the diastolic measurement is for example expressed as

$$0.0524 \times DBP - 0.0071 \times DBP \times time.$$

By solving the simple inequality we arrive at the conclusion that after approximately 7.4 years the contribution becomes negative irrespective of the value of the diastolic blood pressure. For systolic blood pressure this happens after more than 11 years.

7.2 Models 2.1 and 2.2

The AIC values for models 1.1 and 1.2 are 222 346.3 and 221 430.5 respectively. Thus the inference will follow using model 2.2 (spatial power covariance structure). The variance of the random effects is estimated as 58.07 and the residual variance is given by 83.78.

The model reveals that the subsequent observations (the ones taken after baseline) depend significantly on the baseline value of the eGFR. It should be stressed, however, that the main effects should not be interpreted without taking the value interaction effects into account (time does not take value of 0 in this model as only post-baseline measurements are being modelled). This effect is modified negatively by the interaction with time meaning that the values become lower as time from baseline elapses. This confirms the inference made in model 1.2. In this model no statistically significant main effect was revealed between eGFR values and systolic and diastolic blood pressure. However, the interaction of systolic blood pressure baseline value with time was statistically significant effect modifier in time with negative parameter estimate translating into lower eGFR values along with time. Controlling effects of sex and age group were highly significant (just like in the previously considered model 1.2) and no significant effect of comorbidities expressed as CCI index was observed.

Tables 6 and 9 present the revised models 1.2 and 2.2 (both with spatial power covariance matrix) with the insignificant effect of CCI excluded. In the latter case it was decided to include the non-significant blood pressure effect and effect modifiers as one of the interactions (SBP×time) was significant in model 2.2. As expected, in the case of these models inference about the connection between blood pressure values and eGFR does not change.

7.3 GEE

When modelling the probability of the annual rapid kidney function decline no statistically significant association between the instances of rapid decline and blood pressure values were found (both systolic and diastolic). The analysis confirmed solely that the probability of annual rapid kidney function decline is influenced significantly and by the value of baseline eGFR measurement and by patient characteristics such as gender (odds ratio 0.81 for men), group of patients aged 80 years or older (odds ratio 1.17) and the fact of being on cardiovascular medication (odds ratio 1.14).

However, a reverse relation between baseline eGFR was found when compared to the results

obtained in the linear models, suggesting that higher baseline eGFR values are associated with higher probability of annual rapid kidney function decline. This might be due to loss of information in the data when excluding the baseline values to calculate the annual change.

7.4 Comparison of results

The results obtained after reanalyzing this large retrospective cohort longitudinal study are mostly in line with the results obtained in the previous analysis conducted by Vaes et al. (2015) if, of course, we assume that the results of that analysis are valid. Methods used in this thesis confirmed that there is an association between kidney function and blood pressure, both systolic and diastolic as well as values of both mentioned parameters influence the change of the eGFR over time.

Specifically it was confirmed that lower values of systolic and diastolic blood pressure cause lower values of eGFR (in the presence of a constant significant contribution to the model). Applying models for longitudinal data enabled to make inference on the change of eGFR when time from baseline elapses. This was not possible in the previous analysis as the models were not longitudinal in their nature, thus only constant associations were found and examined using the valid statistical methods. Baseline patients characteristics such as age group and cardiovascular medication were confirmed to have a statistically significant impact on the value of estimated glomerular filtration rate. In the current analysis there was no sufficient evidence in the data that the comorbidity influences the eGFR level.

Increasing age and time elapsing from baseline were identified as a factor which contributes to the lower eGFR values. This relationship was also previously noticed by Shilpak et al (2005). On the contrary to the previous analysis the relation between the probability of annual rapid kidney function decrease turned out to have different direction. In the past, only baseline blood pressure was identified as factor having an influence on the probability of rapid annual kidney function decline in studies that used data from another registry (as reported by Ritkin et al.).

7.5 Limitations of the current analysis and drawbacks of presented methods

Even though much effort has been put into the data extraction, it was not possible to extract the same data as it was done by Vaes et al. (2015) in the previous analysis. Due to the fact that

the current study used the data from the same registry all of limitations identified by Vaes et al. (2015) still hold. Specifically there is no data and information whatsoever on the mortality and censoring. For instance, nothing is known about subjects after their last visit reported in the data in terms of discontinuing the study (being out of registry because of for example change of general practitioner). Patients lost to follow-up cannot be determined as well.

Due to the fact that data were collected in different centres, the laboratory results and measurements may vary because of centre effect (different measurement appliances and general practitioner clustering effect).

It was also not possible to extract time varying patient characteristics variables like use of cardiovascular medication at the given time point.

The analysis suffers also from all drawbacks that can be identified for studies in which highly unbalanced longitudinal data are modelled. Most of the issues has been already mentioned in earlier sections of this analysis.

7.6 Ideas for future research

Given the size of the database, it is possible to explore much more about how renal functioning changes across time in the cohort of patients aged 60 and over 60 years old. In the future research it might be of interest for example to explore the random process generating the number of visits per patient by exploring and linking subject's health status. It might be also possible to meta-analyse the results of studies conducted in older cohort of patients to obtain the generic overall across study estimate of effects. As the study is retrospective in its nature it is rather impossible to conduct any kind of time-to-event analysis which might definitely have been of interest to estimate e.g. time to kidney functioning decline (defined as eGFR ≤ 60) in the subjects included in the analysis. In order to make the time-to-event analysis possible, subject should be followed-up starting from the same time point and the observation time should overlap and should not constitute a disjoint collection of observation time space.

In terms of the relation between probability of annual rapid renal functioning decline, it might be of interest to study the relation between the year-to-year eGFR values and the baseline value to explain the relation found in this analysis.

8 CONCLUSION

In this analysis a positive effect of systolic and diastolic blood pressure on the baseline eGFR values was identified for the cohort of patients from a large multi-centre registry of patients in Flanders, Belgium. Moreover, it was concluded that the effect of BP changes as time from baseline elapses and the overall eGFR level decreases as time passes. Patients characteristics such as gender and age above 70 years and above 80 years had a significant negative effect on the eGFR value, also a positive relationship between cardiovascular mediaction and outcome was confirmed. The probability of annual rapid kidney function decline was affected by patient baseline characteristics and by baseline eGFR value with higher values contributing surprisingly to higher odds for rapid decline.

.

REFERENCES

Agresti, A. (2007). *An introduction to categorical data analysis.* Hoboken, NJ: Wiley-Interscience.

Charlson, M., Pompei, P., Ales, K. and MacKenzie, C. (1987). *A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation.* Journal of Chronic Diseases, 40(5), pp.373-383.

Cote, M. and Stein, W. (2007). *A stochastic model for a visit to the doctorâĂŹs office.* Mathematical and Computer Modelling, 45(3-4), pp.309-323.

Danziger, R., Tobin, J., Becker, L., Lakatta, E. and Fleg, J. (1990). *The Age-Associated Decline in Glomerular Filtration in Healthy Normotensive Volunteers.* Journal of the American Geriatrics Society, 38(10), pp.1127-1132.

Diggle, P. (2013). *Analysis of longitudinal data.* Oxford: Oxford University Press.

Levey, A., Coresh, J., Greene, T., Marsh, J., Stevens, L., Kusek, J. and Van Lente, F. (2007). *Expressing the Modification of Diet in Renal Disease Study Equation for Estimating Glomerular Filtration Rate with Standardized Serum Creatinine Values. Clinical Chemistry*, 53(4), pp.766-772.

Lindeman, R., Tobin, J. and Shock, N. (1985). *Longitudinal Studies on the Rate of Decline in Renal Function with Age.* Journal of the American Geriatrics Society, 33(4), pp.278-285.

Rifkin, D., Katz, R., Chonchol, M., Shlipak, M., Sarnak, M., Fried, L., Newman, A., Siscovick, D. and Peralta, C. (2013). *Blood Pressure Components and Decline in Kidney Function in Community-Living Older Adults: The Cardiovascular Health Study.* American Journal of Hypertension, 26(8), pp.1037-1044.

Shilpak, M., Sarnak, M. and Katz, R. (2005). *Cystatin C and the Risk of Death and Cardiovascular Events Among Elderly Persons.* ACC Current Journal Review, 14(9), pp.23-24.

U.S. Renal Data System, USRDS 2013 Annual Data Report: Atlas of chronic kidney disease and end-stage renal disease in the United States. Bethesda, MD: National Institutes of Health,

National Institute of Diabetes and Digestive and Kidney Diseases, 2013

Vaes, B., Beke, E., Truyers, C., Elli, S., Buntinx, F., Verbakel, J., Goderis, G. and Van Pottelbergh, G. (2015). *The correlation between blood pressure and kidney function decline in older people: a registry-based cohort study.* BMJ Open, 5(6).

9 APPENDIX

SAS codes for the presented models (1.1, 1.2, 2.1, 2.3, GEE)

```
>Model 1.1
proc mixed data=final;
class id_patient age_entry_70(ref='0') age_entry_80(ref='0') drugs(ref='0');
model egfr=diastolic systolic diastolic*time systolic*time time sex
age_entry_70 age_entry_80 drugs cci / solution;
random intercept / subject=id_patient;
repeated / subject=id_patient type=cs;
run;

>Model 1.2
proc mixed data=final;
class id_patient age_entry_70(ref='0') age_entry_80(ref='0') drugs(ref='0');
model egfr=diastolic systolic diastolic*time systolic*time time sex
age_entry_70 age_entry_80 drugs cci / solution;
random intercept / subject=id_patient;
repeated / subject=id_patient type=sp(pow)(time);
run;

>Model 2.1
proc mixed data=baseline;
class id_patient age_entry_70(ref='0') age_entry_80(ref='0') drugs(ref='0');
model egfr=egfr_baseline egfr_baseline*time systolic_baseline diastolic_baseline
systolic_baseline*time diastolic_baseline*time time sex age_entry_70
age_entry_80 drugs cci / solution;
random intercept / subject=id_patient;
repeated / subject=id_patient type=cs;
run;

>Model 2.2
proc mixed data=baseline;
class id_patient age_entry_70(ref='0') age_entry_80(ref='0') drugs(ref='0');
model egfr=egfr_baseline egfr_baseline*time systolic_baseline diastolic_baseline
```

```
systolic_baseline*time diastolic_baseline*time time sex age_entry_70
age_entry_80 drugs cci / solution;
random intercept / subject=id_patient;
repeated / subject=id_patient type=sp(pow)(time);
run;

>GEE
proc genmod data=gee descend;
class id_patient age_entry_70(ref='0') age_entry_80(ref='0') drugs(ref='0');
model rapid_egfr=diastolic systolic diastolic*time systolic*time
egfr_baseline time sex age_entry_70 age_entry_80 drugs cci/ dist=bin link=logit;
repeated subject=id_patient;
run;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
The relation between the evolution in blood pressure and the evoluation in renal function over time

Richting: **Master of Statistics-Biostatistics**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Orlowski, Robert

Datum: **22/01/2018**