UNIVERSITA' DEGLI STUDI DI MILANO-BICOCCA
Corso di Dottorato in Epidemiologia e Biostatistica
XXVIII Ciclo

THESIS

MANAGING CARDIOVASCULAR RISK IN HYPERTENSION: METHODOLOGICAL ISSUES IN BLOOD PRESSURE DATA ANALYSIS

Xiaoqiu Liu
Matricola 775441

Tutor: Prof.ssa Maria Grazia Valsecchi

Cotutor: Prof. Gianfranco Parati, Dr.ssa Paola Rebora
Anno Accademico 2015-2016

Acknowledgements

My deepest appreciation to Prof. Maria Grazia Valsecchi and Prof. Stefania Galimberti for the great opportunity into this wonderful research group, and to my tutor Dr. Paola Rebora for her careful and patient intuition and for being the very positive and in-depth influence in these years.

My sincere thanks to my colleagues and friends Davide Paolo Bernasconi, Ausiliatrice Lucenti, Jessica Blanco, Maria Chiara Magri and all the other wonderful people from Center of Biostatistics for Clinical Epidemiology in Monza during the last four years for their friendship and support in my study.

My greatest thanks to Istituto Auxologico Italiano for offering me a great opportunity to learn and to gain experience working with the top experts in the field of cardiovascular research. I also wish to thank Andrea Faini, Grzegorz Bilo and Davide Soranna for the discussion, idea-sharing and great friendship. I thank Fondazione Issachi Samaja very much for their support for my research program.

My special appreciation to Prof. Gianfranco Parati for his full support during my entire PhD program and in particularly in the difficult moments when I would have otherwise chosen differently if not for his continuous encouragement.

I would thank my whole family, who keep me moving forward all the time.

Contents

List of abbreviations

AFT	Accelerated Failure Time
AIC	Akaike Information Criterion
ANCOVA	Analysis of Covariance
ARV	Average Real Variability
ASV	Average Successive Variability
AUC	Area Under The Receiver Operating Characteristic Curve
BIC	Bayesian Information Criterion
BMI	Body Mass Index
BP	Blood Pressure
BPV	Blood Pressure Variability
BPVR	Blood Pressure Variability Ratio
CoV	Coefficient of Variation, as a blood pressure variability index
CV	Cardiovascular
FU	Follow-up
GEE	Generalized Estimation Equation
GLM	Generalized linear model
HR	Hazard Ratio
INI	Integrated Discrimination Improvement
KM	Kaplan Meier
LCL	Lower Confidence interval
MAR	Missing At Random
MCAR	Missing Completely At Random
MNAR	Missing Not At Random
NRI	Net Reclassification Index
OR	Odds Ratio
PH	Proportional Hazard
PO	Proportional Odds
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SE	Sensitivity
SP	Specificity
SV	Successive Variation
SVIM	Successive Variation Independent Of Mean
UCL	Upper confidence interval
VIM	Variation Independent Of Mean
wSD	Weighted Standard Deviation

Some BPV indices are not included in the list.

Research Program

Hypertension remains in 2017 a leading cause of mortality and disability worldwide. A number of issues related to the determinants of cardiovascular risk in hypertensive patients and to the strategies for better hypertension control are still pending. In such a context, aims of my research program were:

1. To investigate the contribution of blood pressure variability to the risk of cardiovascular mortality in hypertensive patients. In this setting, different methods for assessing blood pressure variability and different models exploring the link between blood pressure variability and outcome were investigated.
2. To assess the possibility that a hypertension management strategy based on hemodynamic assessment of patients through impedance cardiography might lead to a better hypertension control over 24 hours than a conventional approach only based on blood pressure measurement during clinic visits.

To these aims, this thesis summarizes data obtained by performing a). An in-depth analysis of a study conducted in the Dublin hypertensive population, including 11492 subjects, and b). The analysis of longitudinal data collected in the frame of BEAUTY (BEtter control of blood pressure in hypertensive pAtients monitored Using the hoTman® sYstem) study.

In Dublin study, the proportional hazard Cox model and accelerated failure time models have been used to estimate the additional effect of blood pressure variability on cardiovascular mortality over and above the effect of increased mean BP levels, with an attempt to identify the best threshold values for risk stratification.

On the other hand, in BEAUTY study, mixed model and generalized estimation equation are used for the longitudinal data analysis.

Key words:

Survival data analysis; Longitudinal data analysis; Clinical trial; ROC curve; Blood pressure variability; Hypertension management

Part I: Effect of Blood Pressure Variability on Cardiovascular Mortality

Cardiovascular events may be affected by both mean blood pressure and blood pressure variability together with other known or unknown risk factors, among which there is a complicated network of casual interactions. The purpose of this study was to explore the effect of blood pressure variability on cardiovascular mortality over and above the effect of mean BP in the frame of Dublin Study, a study on a large cohort of hypertensive individuals living in the city of Dublin, Ireland.

1. Introduction to blood pressure variability and related indices

The measurement of Blood pressure (BP) variability (V) estimates how much office, ambulatory or home BP change over a period of time; in this context different BPV indices may represent different perspectives in the assessment of BP variability, with different applications when considering 24h ambulatory BP recordings or BP changes over longer time intervals, such as those among days or visits. Currently, BPV attracts growing attention due to the evidence that it might predict cardiovascular (CV) events. A pending issue in this context is related to the correlation between BP mean values and BP variability, which raises the research question of how much BPV contributes to CV event, over and above the contribution given by mean BP levels.

The first step in exploring this question is to understand how BPV is categorized and calculated. The types of BPV, according to the duration of the time window over which BPV is computed, can be classified as:

1. Very short-term BPV (beat-to-beat), mainly caused by increased central sympathetic drive, reduced arterial/cardiopulmonary reflex sensitivity, humoral and rheological factors, behavioral and emotional factors, activity/sleep alternance and ventilation;

2. Short-term BPV, calculated within 24 hours. The influencing mechanisms include increased central sympathetic drive, reduced arterial/cardiopulmonary reflex sensitivity, humoral and rheological factors, behavioral and emotional factors, activity/sleep, reduced arterial compliance and improper dosing/titration of antihypertensive treatment;

3. Mid-term and long-term BPV, measured over a period of days (day-by-day), or measured from visit to visit or between seasons. The determining mechanisms include reduced arterial compliance, improper dosing/titration of anti-hypertensive treatment, reduced adherence to antihypertensive treatment, BP measurement errors, or seasonal or weather change. The difference in the responsible mechanisms may also explain how different BPV estimates could be predictors of different cardiovascular diseases.

Ambulatory BP monitoring (ABPM) has been proved to play an important role in cardiovascular risk stratification, as well as in the assessment of cardiovascular protection through 24h BP control. Also information on BPV, calculated from 24h ABP recordings, has been reported to significantly affect cardiovascular health.

Different BPV indices may represent different perspectives on BP variability components. The most commonly employed methods to assess BPV, subdivided according to the BPV component explored, include:

1. Overall variability is an assessment over all the valid BP values as a whole, it reflects the absolute or relative dispersion of BP values;

2. Variability for ordered values, the calculation of which is based on not only the

absolute values but also their orders of the measurements. The calculation of ordered variability either passes from one value to the next one, or is based on a linear or non-linear regression or spectral analysis estimated from measurements in time-order, therefore, besides the absolute or relative dispersion of BP values, it introduces also the component of time, either in a direct way that the calculation included BP changes weighted by the time intervals, e.g. time rate, or ARV can also be calculated as mean of absolute BP changes weighted by time interval between consecutive readings, or the component of time is included in an indirect way, as it reflect how BP changes following time, e.g. BP fluctuation within 24h or following different physical and psychological change. It is thus a reflection of BP stability over time.

3. Extreme variability, the calculation of which is based on extreme values, such as the maximum or minimum among all the measurements; its major component is the range of BP fluctuation.

4. Drug effect variability, which accesses the efficacy of anti-hypertensive treatments on the BPV or BP fluctuation with the presence of treatment. Indices of this kind includes two components: one is the natural fluctuation of BP, and two, the fluctuation caused by antihypertensive treatment that could not assessed by the previous three categories of BPV indices. Instead, indices proposed for the purpose of assessing drug effect variability could combine those two components.

The calculation of the above mentioned variability using various indices are listed in Table 1.

Table 1. List of BPV indices

Indices	**Abbreviation and Calculation***	**Note**
Index for overall variability		
Standard deviation [1]	$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(BP_i - \overline{BP})^2}$	BP_i: i^{th} BP measurement, in the case of clinic BP measurement, it is the mean of ≥2 valid readings. $\overline{BP}$: mean of all BP measurements. N: total number of measurements
Coefficient of variation	CoV =100*SD/$\overline{BP}$	Refer to SD.
Weighted SD[2]	$wSD = \frac{SD_d \cdot T_d + SD_n \cdot T_n}{T_d + T_n}$	SD_d:Daytime SD; T_d: Daytime SD_n:Nighttime SD; T_n: Nighttime Refer to manuscript for the definition of daytime and nighttime.
Blood pressure variability ratio [3]	$BPVR = \frac{SD_{SBP24h}}{SD_{DBP24h}}$	SD_{SBP24h} and SD_{DBP24h} : SD of systolic and diastolic BP over 24h hours.
Index based on ordered BPs		
Average real variability[4]	$ARV = \frac{1}{N-1}\sum_{i=1}^{N-1}\lvert BP_{i+1} - BP_i\rvert$	Refer to note for SD.

Successive Variation[5]	$SV = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(BP_{i+1} - BP_i)^2}$	Refer to note for SD.
Time rate[6]	$TR = \frac{\sum_{i=1}^{N-1} \lvert r_i \rvert}{N-1}$	$r_i = (BP_{i+1} - BP_i)/(t_{i+1} - t_i)$, where t_{i+1}, t_i are the time of i+1th and ith measurement.
Interval weighted SD [7]	$iSD = \sqrt{\frac{1}{\sum W_i}\sum_{i=1}^{N} W_i(BP_i - \overline{BP})^2}$	Refer to note for SD. W_i: time interval between consecutive readings.
Variance of the absolute second differences between successive BP [8]	$VABS2 = \sum_{i=1}^{N-2}(\Delta d - \overline{\Delta d})^2$	$\Delta d = \lvert (BP_{i+1} - BP_i) - (BP_i - BP_{i-1}) \rvert$ $\overline{\Delta d}$: mean of Δd
Individual residual variability[9]	$Residual = \sum_{i=1}^{N}(BP_i - BP_f)^2$	BP_f : values from fitted profile based on the fast Fourier transform spectral analysis.
Variability independent of mean [10][11]	$VIM = \frac{SD}{\overline{BP}^x} * \overline{BP_P}^x$	x is estimated from non-linear regression $SD_P = a * \overline{BP_P}^x$, P: population(all subjects)
Successive variation independent of mean[12]	$SVIM = \frac{SV}{\overline{BP}^x} * \overline{BP_P}^x$	x is estimated from non-linear regression $SV_P = a * \overline{BP_P}^x$, P: population(all subjects)
Residual SD[13]	$rSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(R_i - \bar{R})^2}$	Refer to the note of SD. $\bar{R}$: mean residual estimated by simple linear regression with visit.
Sum of squared difference between each clinic BP and the trend- predicted BP [8]	$rSSR = \sum_{i=1}^{N}(BP_i - \widehat{BP})^2$	$\widehat{BP}$: Trend-predicted mean
Index on extreme values		
Range[14]	Range=Max-Min	Max: max(BP_i) Min: min(BP_i) Mean: mean(BP_i)
Peak size[12]	Peak=Max-mean	
Trough size[12]	Trough=Mean-min	
Index of assessing treatment effects on BPV		
Trough/peak ratio[15–17]	$TPR = \frac{\mathrm{Min}(BP_{v1,t} - BP_{v2,i})}{\mathrm{Max}(BP_{v1,\tau} - BP_{v2,\tau})}$	BP_t, BP_τ : tth and τ^{th} BP measurement at visit 1 without treatment ($BP_{v1,t}$) and at Visit 2 with treatment ($BP_{v2,t}$), both by 24h ABPM. Calculation requires $BP_{v1}$and BP_{v1} measured at the same hour of the day of the 24h ABPM.
Smoothness index[18]	$SI = \frac{\overline{BP_h}}{SD_{BP_h}}$	BP_h: mean of BP measured within each of the 24 hours. $\overline{BP_h}$ and SD_{BP_h} are the mean and SD of the hourly mean.
Treatment on variability index[19]	$\mathrm{TOVI} = \frac{BP_{v1} - BP_{v2}}{wSD}$	24h BP change from visit 1 without treatment (BP_{v1}) to Visit 2 with treatment (BP_{v2}); wSD: weighted SD at the visit 2 by 24h ABPM.

*All the BPV indices are calculated for an individual subject. Only the calculation of VIM depends on the BP of population level.

BP_i: One BP measurement. For major information, please refer to the original papers.

Comments on the BPV indices

Indices for overall variability

SD and CoV are the classic BPV indices, while SD is an absolute variation or dispersion, CoV is a unitless, standardized measure of relative dispersion, represents a simple way to normalize BPV for mean BP levels, and thus is more suitable to provide a direct comparison between BPV of day and night, between systolic and diastolic BPV.

When focusing on 24h BPV, estimates of short term BPV separately computed by excluding the day-night BP changes are of particular clinical relevance. This is because mean night BP is in general lower than mean daytime BP, and nocturnal BP dipping, which quantitatively contributes to overall 24h BPV, has a favorable prognostic impact on CV outcome. Indices of short term BPV, not affected by day – night BP changes, include: the weighted 24h SD (wSD), ARV, individual residual BPV (i.e. the power of BP spectral components computed after excluding the first two harmonics over 24h tracing), or calculation of daytime and nighttime BPV, separately considered.

Indices based on ordered BPs

The calculation of BPV indices based ordered BPs either passes from one BP value to the next one, or is based on a linear or non-linear regression estimated from measurements in time-order. In general, BPV indices based on ordered values are better applied for short-term BPV measured by 24h ABPM than for BPV over longer duration. This is because that time interval, i.e. how fast the BP changes, is an important factor to be considered over a relatively short period of time, however, it becomes less meaningful when the measurement interval is long, for example, BP change over 6 months or 1 years may not truly reflect BP fluctuation. Moreover, BP fluctuates within 24h reflect the instant BP values while visit-to-visit BP are normally measured in office with standard procedure, and thus the average BP values fluctuate within a limited range. Adding time interval into the calculation of long-term BPV may thus be an overcorrection.

Indices on extreme values

BPV based on the extreme values excludes all the other useful values that reflect real BP profile and carries the risk of including artefactual outlies. Due to these defects, it is not often recommended. If to be used, it may fit only the long-term BPV estimation where the BP is usually the average of 3 measurements taken in a standard office condition with on-time control by medical personals.

Indices to assess treatment effects on BPV

When the study aim includes effect of antihypertensive treatment, BPV indices designed for such purpose is required. Smoothness index is more reproducible than trough/peak ratios, furthermore, SI provides more reliable and more clinically relevant information on the effect of antihypertensive treatment on 24h ambulatory blood

pressure. TOVI reflects another perspective in BPV quantification, its numerator is the BP change over long term, which reflects the overall fluctuation, while the denominator is the weighted SD over 24h, which reflects the fluctuation due to the patient's own physiological feature. In such way, the true fluctuation caused by treatment effect can be accessed. Higher smoothness index or TOVI was attributed to stronger BP-lowering effect and longer duration of drug action[20].

All the indices used for long-term BPV assessment need to be considered with caution if the patient changes therapy during the follow-up duration, since in such case the BP fluctuations are composed of spontaneous BP fluctuations related to cardiovascular control mechanisms, change induced by drugs, and change induced by change of drugs. The BP fluctuations reflecting BP reduction by a change in treatment may obviously represent a protective feature, and cannot be combined with the degree of spontaneous BPV reflecting the interaction between environmental stimulations and the response by cardiovascular regulatory mechanisms, because an increase in the latter component may represent a risk factor for cardiovascular system.

Major characteristics and applicable BP types are summarized for each index in Box 1.

BOX 1. Characteristics of all the previously proposed BPV indices

Indices	BP Type	Characteristics
SD[1]	ABPM HBPM OBPM	- It reflects the dispersion of values around their mean without accounting for the BP measurements order - It is sensitive to low sampling frequency of ABPM
CoV	ABPM HBPM OBPM	- It is a standardized, unit-less measure of dispersion
wSD[2]	ABPM	- Evaluates 24h BPV by excluding the "protective" component represented by of day/night changes - It focuses on short term BP fluctuations
BPVR[3]	ABPM	- May be independent of mean arterial pressure[3]
ARV[4]	ABPM	- It reflects the time series of ABPM data, accounting for the BP change between day and night.
SV[5]	ABPM	- Highly correlated with ARV
TR[6]	ABPM HBPM OBPM	- When the measurement interval of BP is fixed, the time rate is determined predominantly by the magnitude of each BP difference, particularly in ABPM[6].
iSD[7]	ABPM	- May also fit HBPM or OBPM, but interval may be overestimated.
VABS2[8]	OBPM	- It is related to the penalty for smooth function estimation (eg, in splines, spectral functions, and wavelets)[8].
Residual[9]	ABPM	- It represents the fast BP fluctuations that remain after exclusion of the slower components of the 24 h BP profile through spectral analysis
VIM[10]	HBPM	- A strong correlation is noted between VIM and other BPV indices[14].

	OBPM	- Most of the studies support the independence of VIM on mean BP - It is highly sample-specific and may not be compared across populations - It is inappropriate in 24h ABPM because of the cardinal and diurnal variation and the highly diverse pattern among individuals
SVIM[12]	HBPM OBPM	- Similar to VIM
rSD[13]	HBPM OBPM	- The linear assumption of BP values over time should be satisfied. - Does not fit for ABPM.
rSSR[8]	HBPM OBPM	- It represents the direct estimate of residual (error) variability and does not show a variance-mean relationship[8] - Subjects should have the same no. of values.
Range[14]	OBPM	- Easily influenced by outliers - Ignore most of the useful BP values
Peak[12]	OBPM	
Trough[12]	OBPM	
TPR[15–17]	ABPM	- Easily influenced by outliers - Pre-post design required - A measurement of extreme values
SI[18]	ABPM	- Calculates the homogeneity of BP reduction by treatment throughout the 24 hours, by focusing on treatment induced hourly BP changes. - A measurement of overall BPV
TOVI[19]	OBPM+ ABPM	- Combines information on the reduction of 24-h average BP and on 24h BPV after treatment (assessed by wSD), measures the effects of antihypertensive treatment on both mean BP levels and BP variability; - Pre-post design required. Both ABPM and OBPM are required. - A measurement of overall BPV

Use of different indices representing different BPV components has been suggested[21]: SD can be calculated also for all BPV types, including the beat-to-beat variability, while 24h wSD and ARV are better suited for short-term BPV, which is often measured in 24 hours. For long-term BPV, SD, CoV, VIM and residual are preferred. For different types of BPV, the methods of BP Measurement, advantages and disadvantages, choice of BPV indices as well as proposed mechanisms have been well summarized (Table 2). In this report, a brief summary of the preference and use of those indices is in BOX 2.

It has been suggested that a suitable BPV index should be easily measurable and applicable in clinical practice, should be reproducible; should have defined normalcy and interventional thresholds, should be independent contributes to CV risk, should be modifiable by treatment; and patients' prognosis is improved when additional treatment targets are set for BPV beyond those for average BP[22]. However, at present, evidence for most of these questions is missing, and therefore BPV remains a challenging research issue deserving thorough investigation.

Table 2. Types of BPV: Methods of measurement, prognostic relevance, and proposed mechanisms (Cited from Parati, Nat. Rev. Cardiol 2013[21])

Table 1 | Types of BPV: methods of measurement, prognostic relevance, and proposed mechanisms

Characteristic	Very short-term BPV (beat-by-beat)	Short-term BPV (within 24 h)	Long-term BPV (day-by-day)	Long-term BPV (visit-to-visit)
Method of BP measurement	Continuous BP recordings in a laboratory setting or under ambulatory conditions	ABPM	ABPM over ≥48 h HBPM	ABPM OBPM HBPM
Measurement intervals	Beat-to-beat over variable recording periods (1 min to 24 h)	Every 15–20 min over 24 h	Day-by-day, over several days, weeks, or months	Spaced by visit over weeks, months, and years
Advantages	Assessment of indices of autonomic cardiovascular modulation	Extensive information on 24 h BP profile Identification of patterns of circadian BP variation	Appropriate for long-term monitoring	Appropriate for long-term monitoring
Disadvantages	Stability of measurements might not be guaranteed outside the laboratory setting	Cannot be repeated frequently	Patient training and involvement is required for HBPM ABPM over 48 h is neither always well tolerated or accepted by patients	OBPM and HBPM provide limited information on BP profiles
Indices of BPV	SD Indices of autonomic modulation can be calculated (that is, fluctuations in very low, low, and high frequency bands [spectral analysis])	24 h, daytime, and night-time SD and CV 24 h weighted SD Day-to-night BP changes ARV	SD CV	SD CV
Proposed mechanisms	Increased central sympathetic drive Reduced arterial/cardiopulmonary reflex Humoral and rheological factors Behavioural and emotional factors Activity/sleep Ventilation	Increased central sympathetic drive Reduced arterial/cardiopulmonary reflex Humoral and rheological factors Behavioural and emotional factors Activity/sleep Reduced arterial compliance Improper dosing/titration of AHT	Reduced arterial compliance Improper dosing/titration of AHT Reduced adherence to AHT BP measurement errors	Improper dosing/titration of AHT Reduced adherence to AHT BP measurement errors Seasonal change

Abbreviations: ABPM, ambulatory blood pressure monitoring; AHT, antihypertensive treatment; ARV, average real variability; BP, blood pressure; BPV, blood pressure variability; CV, coefficient of variation; HBPM, home blood pressure monitoring; OBPM, office blood pressure measurement.

BOX 2. Application note of BPV indices for different BP types: Applicable (✔), applicable but not recommended (▼) and Recommended (R)

BPV type	Very short-term	Short-term			Mid-term	Long-term
	(Beat-by-beat)	Day	Night	24 Hours	(Day-by-day)	(Visit-to-visit)
BP type	-	ABPM			HBPM	OBPM/HBPM
SD[1]	✔ R	✔ R	✔ R	✔	✔ R	✔ R
CoV	✔	✔	✔	✔	✔	✔
wSD[2]				✔ R		
BPVR[3]	▼	✔	▼	▼	▼	▼
ARV[4]	✔ R	✔ R	✔ R	✔ R	▼	▼
SV[5]	✔	✔	✔	✔	▼	▼
TR[6]	✔	✔	✔	✔		
iSD [7]	✔	✔	✔	✔	▼	▼
VABS2[8]	✔	✔	✔	✔	✔	✔
Residual[9]				✔ R		
VIM[10][11]					✔ R	✔ R
SVIM[12]					✔	✔
rSD[13]					✔	✔
rSSR[8]					✔	✔
Range[14]					▼	▼
Peak size[12]					▼	▼
Trough size[12]					▼	▼
TPR[15–17]				✔		
SI[18]				✔ R		
TOVI[19]*				✔ *R		

2. Review of literature on the role of blood pressure variability on the risk of cardiovascular events

2.1. Discrepancies in previous studies

The previously published papers are of various designs including different study populations (general population/in-hospitalized patients; differences in ethnicity, age range, disease situation); they focused on different CV outcomes (all-cause mortality, all CV events or other diseases like ischemic/hemorrhagic stroke, peripheral arterial disease, coronary event, cardiac events, intima-media thickness and so on); with different follow-up duration (6 months -12 years or even longer), adjusting different covariates in the statistic models. After all, the choices of BPV indices are very different, and in different studies they were taken from ABPM, office, home BP self-measured by patients; home BP by nurses. Some studies used only baseline BPV while others used BPV quantified during a period of follow-up (which may cause immortal bias[23]). Moreover, calculation of BPV in different studies was based on very different numbers of BP values measured at different time interval (every 3/6 months, every 1/2year).

Previously published papers exploring BPV effects independent of the mean have mainly followed these approaches:

1. Cox proportional hazard model including BPV index categorized into a few groups (halves[24], quartiles[25–28], quintile[10,29], deciles[10] or per SD increase[13]) for the estimation of hazard ratio (HR) for CV outcomes, setting the smallest BPV group as reference. Mean BP level as a continuous variable, BP change over time or the mean BP categorized into quartile/quintiles levels were adjusted in the model, in situation when no significant correlation between BPV and BP were found[24]. In a study by Hastie[28], mean time-weighted SBP and DBP were included in the model with BPV index (Visit-to-Visit ARV) even when the mean BP is associated with BPV index (rho<0.5), however the sample size was big enough to allow proper estimation of the regression coefficients (n=14522). Another approach is to categorize BPV indices after splitting the population into groups according to subjects' mean BP level, and then assess the impact of differences in BPV on outcome for any given mean BP level[30].
2. Cox proportional hazard model including BPV index as continuous variable for the prediction of CV events or time to event, in some cases after testing the correlation between BPV and mean BP[31–33] while in some other cases without clear indication of correlation tests [34]. In a study by FF Wei[35], multiple regression

analysis included both BP level and BPV (beat-to-beat, reading-to-reading, day-to-day, VIM, ARV, Range) index and then the variance inflation factor were further calculated to assess to what extent parameter estimates for BP level and variability were affected by collinearity in fully adjusted regression models. In all of these analyses the variance inflation factor was <1.31.

3. Finding BPV indices that are likely to be independent of the average BP level, i.e. coefficient of variation, SD independent of mean (VIM) as well as SVIM, ARVIM proposed initially by Rothwell[10,12] in 2010. The calculation of them are presented in Table 1. After being proposed, VIM has become very popular and well accepted, being used in various papers focusing on either home BPV[26,31,36] or office BPV[8,35–37]. Most of the studies support the independence of VIM on mean BP, in fact, a study by FF Wei et al[35] indicated a nice independence of VIM with mean BP measured in beat-to-beat, reading-to-reading, day-to-day. It should be noticed that some of the studies estimated the impact of VIM on the CV events without adjusting for the mean BP in the Cox model[8,26,31] while some did[10]. Not all the studies are in agreement, for example, R Schutte et al[31] found VIM decrease with mean BP within a single visit but showed no dependence on overall BP obtained in two visits. Note that the calculation of VIM depends on the mean of BP from all the subjects of a study, so it is very sample-specific. Furthermore, VIM is considered inappropriate in 24h ABPM, according to its proposer P. Rothwell, as he replied my question by email: '*It is difficult to apply the VIM concept to ABPM data because of the considerable diurnal variation in BP – and the fact that the diurnal pattern varies so much between individuals. It is better to use SD - calculated separately for daytime and night and to then simply adjust for mean BP in any modelling.*'

4. Some studies have proposed BPV cutoff threshold using arbitrary selection based on clinical judgement, for example, the median value, the upper 75% or 99%[38] percentile values. Optimal BPV cutoff points were also determined using outcome-oriented approaches, including the receiver operating characteristic (ROC) curve[39] or the log-rank test statistic[7,40]. In their context, an optimal cutoff point was defined as the value of the continuous covariate that best separated, using statistical criteria, low- and high-risk patients with respect to a cardiovascular outcome event. The ROC curve based methods resulted in an area under the AUC curve (AUC) close to 0.5 and optimal value in SD in systolic visit-to-visit BPV of 8.3-8.4 mmHg. Palatini et al[7] found that a nighttime SBP SD of ≥12.2 mmHg, and a nighttime DBP SD of ≥7.9 mm Hg were associated with a greater risk of cardiovascular events and of all-cause mortality as compared with lower SD values.

5. Applying various statistical methods to assess the extra impact of BPV on CV outcome. For example, generalized R^2 statistics were used in the studies by K Asayama[26], R Schutte [31] and LJ Mena[41], In a paper by Palatini[7], the net reclassification improvement (NRI) and the category-free integrated discrimination

improvement (IDI) are calculated. The paper by Chambless[42] evaluated improvement in risk prediction models and compared the effectiveness of NRI, IDI, c-statistic, the population attributable risk and the ratio of predicted risk in the top quintile to the corresponding variables in the bottom quintile.

As regard to how many readings within 24h ABPM are sufficient to estimate BPV without losing prognostic information, Mena et al suggested 48 measures of BP as a minimum number of BP readings to assess ARV without meaningful loss of prognostic information [41].

Due to those discrepancies, a meta-analysis or pooled analysis is difficult to conduct. Using "Blood pressure variability" and ("Meta-analysis" or "pooled analysis") as key words in PubMed (31/03/2017), 9 Meta-analyses[43–50] on the effect of BPV on CV outcomes were found, among which there were 5 on Visit-to-Visit BPV[43,46,47,49,50], all aimed at CV events/mortality and all-cause mortality, and the included studies overlapped mostly. There were 4 papers on the effect of short term (not necessarily 24h) BPV[44–46,48], but with different outcomes of interest, among which, only one paper [46] has pooled studies of home BPV on the effect of cardiovascular events/mortality and all-cause mortality. One paper concentrated on the heterogeneity on the existing literature exploring the value of 24-hour BP variability as a prognostic index[48], and the finding indicates a vast diversity among included studies. Among 40 included studies, 36 different measures of BPV (systolic or diastolic BP, BPV indices) and 13 definitions of night- and day-time periods were used. The interpretation and use of 24h BPV in clinical practice, as an important prognostic indicator of CV events, is hampered by insufficient evidence and divergent methodologies.

2.2. Results of the peer studies

Let alone the heterogeneity in previous studies, most of them found a significant independent predictive value of BPV on CV outcomes, after adjusting for the impact of mean BP. The publication bias cannot be excluded, however, and more studies are still needed to define its real clinical importance, on the background of the achieved statistical significance. Yet, a relevant predictive effect of BPV independent of the means level is not supported by all the studies, i.e., JA. Staessen et al [26,31,35,51,52] found only small independent effect of BPV on cardiovascular events in a pooled analysis using data from different countries. This study analyzed 24h ABP recordings based on different methodological application of 24h ABPM, which may have biased the different estimates of BPV obtained from different populations. In 2014, Diaz KM et al published a meta-analysis on the relation between visit-to-visit BPV and CV disease and all-cause mortality[37]. This meta-analysis included 37 studies, representing 41 separate cohorts. Across studies, systolic BPV and diastolic BPV showed significant associations with outcomes in 181 of 312 (58.0%) and 61 of 188 (32.4%) analyses, respectively. Few studies provided sufficient data for pooling risk

estimates. For each 5 mm Hg higher SD of systolic BP, the pooled hazard ratio for stroke across 7 cohorts was 1.17 (95% confidence interval [CI], 1.07-1.28), for coronary heart disease across 4 cohorts it was 1.27 (95% CI, 1.07-1.51), for CV Disease across 5 cohorts it was 1.12 (95% CI, 0.98-1.28), for CV mortality across 5 cohorts it was 1.22 (95% CI, 1.09-1.35), and for all-cause mortality across 4 cohorts it was 1.20 (95% CI, 1.05-1.36). In 2016, SL Stevens et al have published a meta-analysis in which the association long term (clinic), mid-term (home), and short term (ambulatory) BPV, independent of mean BP, with cardiovascular disease events and mortality has been reviewed separately using standardized hazard ratio in prospective cohort studies and clinical trials. Increased long term systolic BPV associated with risk of all-cause mortality (hazard ratio 1.15, 95% confidence interval 1.09 to 1.22), cardiovascular disease mortality (1.18, 1.09-1.28), cardiovascular disease events (1.18, 1.07-1.30), coronary heart disease (1.10, 1.04-1.16), and stroke (1.15, 1.04-1.27). Increased mid-term and short term daytime systolic BPV were also associated with all-cause mortality (1.15, 1.06-1.26 and 1.10, 1.04- 1.16, respectively).

All these meta-analysis focusing on the effect of BPV on CV events are summarized in Table 3.

As regard to the possible threshold of BPV index, most of the studies used data-based approach, including an arbitrary value or quartile/tertile/95% percentile, the results of which thus are in heaving dependence of their subject feature. There are only few papers using outcome-based approach to best separate subjects into lower/higher CV risks. The ROC curve based methods resulted in an area under the AUC curve (AUC) close to 0.5 and optimal value in SD in systolic visit-to-visit BPV of 8.3-8.4 mmHg[39]. In study by P. Palatini[7], they suggested 12.2 mm Hg (Contal'sq=4.15) and 7.9 mm Hg (Contal'sq=2.71) for night-time SBP and DBP SDs using log-rank test statistic[40]. To my best knowledge, no study has yet reposted estimation of BPV cutoff values using ROC approach extended for survival data.

In conclusion, review of literature does not lead to a definitive answer to the question whether BPV can predict CV event independent of BP levels, and if so the predictive power. It is not surprising that the current interpretation and use of 24h BPV in clinical practice, as an important prognostic indicator of CV events, is still matter of debate, being hampered by insufficient evidence and divergent methodologies.

Table 3. Previously published meta-analysis of BPV effect on CV events

Author [ref. no]	Included studies	Events	BP type	Risk evaluation and BPV index	Results
K.M Diaz 2014[37]	37 studies representing 41 cohorts.	CV diseases (stroke, CHD, CVD), CVD mortality and all-cause mortality	Sys and Dias OBPM	Pooled HR per 5mmHg increase of systolic SD	All systolic HR>1 except for CVD No enough data for diastolic BPV evaluation
K.S. Taylor 2014[38]	40 cohorts including 36 different measures of BPV and 13 definitions of night- and day-time periods	all-cause mortality, CV mortality, all CV events, stroke and coronary heart disease	Sys and Dias ABPM	RR for 5mmHg increase for SD, CoV, for 1mmHg dipping and for 10mmHg increase for morning surge	Night dipping ~ low risk of CV events
L.S.Manning 2015[39]	7 prospective, observational cohorts	functional post-stroke outcome (death or disability)	Sys and Dias OBPM	Pooled OR per 10-mm Hg increment in Systolic SD or CoV	Systolic BPV ~ poor functional outcome
C. Tai 2015 [40]	13 cohorts	CV events and all-cause mortality	Systolic VVV	Pooled HR per 1 mmHg increase in SD or 1% increase in CoV.	Systolic VVV is a predictor of CV and all-cause mortality and stroke
S.L.Stevens 2016 [41]	23 separate analyses (observational cohort or clinical trials)	All-cause mortality and CV diseases and mortality	OBPM HBPM ABPM	Standardized HRs for increase in BPV (SD, CoV, VIM, and ARV).	Increased systolic VVV ~ all-cause, CV mortality and CV diseases; Increased mid- short-term daytime systolic BPV ~ all-cause mortality
J.M. Madden 2016 [42]	12 studies (case-control or cross-sectional)	left-ventricular mass index (LVMI)	ABPM	Correlation coefficient of LVMI withARV, SD, wSD, CoV across 24 h/day/night periods	Weak correlation between LVMI and systolic 24h SD, 24h ARV, wSD, day SD
J. Wang 2017[43]	23 observational cohort studies	CV diseases (stroke, CHD, CVD), CV mortality and all-cause mortality	OBPM	RR for increase in SD and CoV	Systolic VVV ~ all-cause mortality, CV mortality, CHD incidence and stroke incidence
Alastair J.S. Webb[44]	14 large randomized controlled trials	new atrial fibrillation	OBPM	OR for increase in ratio of SD^2 between treatment/control groups	Effects of randomized treatment on variability in BP are unrelated to risk of new-onset AF.

Note: CHD: Coronary heart disease; CVD: Cardiovascular disease; HR: Hazard ratio; RR: relative risk or risk ratio; OR: Odds ratio; ~ : associated with

3. Dublin study -a registry dataset

3.1. Introduction to Dublin Study

The Blood Pressure Unit (formerly located at the Charitable Infirmary and now based at Beaumont Hospital in Dublin) has been receiving patients who were referred to the unit by their family doctors because of an elevated clinic BP; 11291 such patients entered into Dublin outcome study during the study period (June 1, 1980 to September 30, 2002). More than one paper have been published from the dataset of Dublin study[53,54]. To be eligible for this study, patients had to be either untreated at baseline, or had all antihypertensive drugs discontinued for 1 week before their baseline visit to the unit; demographic details and cardiovascular risk factors (sex, age, body mass index, smoking status, presence of diabetes mellitus, and history of previous cardiovascular events) had to be recorded; The Hospital Ethics Committee approved the study.

ABPM was performed by validated oscillometric SpaceLabs 90202 and 90207 monitors (SpaceLabs Inc, USA) set to measure BP every 30 mins throughout the 24-hour period. Mortality outcome was ascertained by 30 September 2002 by searching a national computerized register of deaths for each individual whose name appeared in the *dabl*$^{@}$ BP database (ECF Medical Ltd, Blackrock, Co. Dublin, Ireland). The death certificate of each individual was examined and the cause of death was coded according to the World Health Organization's International Classification of Diseases, 9th Revision (ICD-9). Cardiac mortality included myocardial infarction (ICD-9, 4100 to 4109), heart failure (4280 to 4289), sudden death (7980 to 7989), and chronic coronary heart disease (4140 to 4149). Cerebrovascular mortality included stroke (4300 to 4389). Cardiovascular mortality was a composite of cardiac mortality, stroke, and other vascular deaths.

3.2. Data management and quality control

Data quality control mainly focuses on ABPM readings based on the following steps: SBP<50mmHg or >300mmHg, DBP<40mmHg or >150mmHg, SBP-DBP<10mmHg or >150mmHg were considered as abnormal BP readings and were deleted. As the actual bedtime and awakening hours were not available for individual subjects, Daytime was defined as the time period between hours [10:00, 22:00) and nighttime between hours [24:00, 06:00) (narrow fixed time intervals approach, excluding transition times between wake and sleep). BP readings that were registered after 22:00 of the second day were not considered. After excluding abnormal BP readings, patients (n=759) who had less than 20/7 readings for the day/night, or <70% of expected readings in 24h, were excluded from this study, following the ESH position paper[55].

Among the remaining subjects, 40 subjects under 18 years at the time of enrollment were excluded from this study.

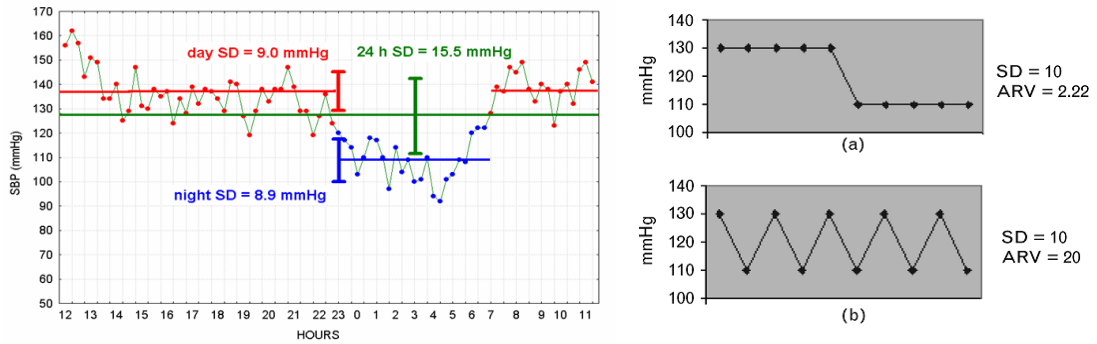
In the end, 10492 patients were included. Threshold for hypertension diagnose based on ABPM are ≥130/80 mmHg for 24h average, ≥135/85 mmHg for daytime average and ≥120/70 mmHg for night-time average. Patients with one or more average (24h, daytime, night-time) above the threshold value(s) are defined as systolic and/or diastolic hypertensive.

As in Dublin study the BP data are limited to readings obtained by 24h ABPM without antihypertensive treatment, the BPV index selection is concentrated on those of overall variability and variability based on ordered BPs. Variability on extreme values are not considered as it carries the risk of including artefactual outlies. Three measures of short term BPV, namely CoV, ARV and wSD for 24h, SD of daytime and nighttime BP separately considered. CoV is chosen because it is unitless and it is a standardized measurement of the dispersion. wSD provides a weighted average of the daytime and nighttime BP SDs, excluding the influence of day-night changes (Figure 1, Left panel), while ARV quantifies successive variations from measure to measure. This is exemplified in Figure 1, Right panel, where the example (a) has same SD but significantly lower ARV than the other. Thus, ARV may able to better reflect the actual short term BP variability, and may be less sensitive to the relative low sampling frequency of the ambulatory blood pressure monitoring devices. The calculation of ARV is not majorly influenced by the night fall in BP, and thus it is an index nicely suited to quantify short term 24h BPV. In this report, CoV, ARV, wSD for 24h BPV, SD for day and night time BPV, separately considered, are used.

Multiple regression analysis are used to explore the potential influencing factors of BPV. Proportional Cox model and Weibull models are used to assess the effect of different BPV indices on the CV mortality. Based on their results, cutoff values are estimated for significant BPV indices using receiver extended operating characteristic (ROC) based methods. At last, the model fitness is evaluated using Brier score and ROC.

In this report, SAS 9.3 and R 3.3.2 are used for data management and/or analysis.

Figure 1. Example data for the calculation of wSD (Left) and ARV (Right)
Figures cited from [2] (Left) and [4] (Right)



3.3. Data analysis /results

3.3.1. Baseline information

Out of 11291 subjects in the dataset, 10492 were included into analysis set, among whom there were 4914(46.9%) males, 2144(20.4%) smokers, 792(7.5%) individuals with diabetes, 1314(12.5%) with previous CV diseases. Mean age was 54.1±14.4 years and BMI 27.4 ±4.7 kg/cm^2. For systolic and diastolic BP, the threshold values are BP>135/85mmHg at day or >120/70mmHg at night or >130/80mmHg over 24h. Only both systolic and diastolic BP at day, night and 24h BP are all below threshold values are considered as normotensive. 1649 subjects were normotensive, 8843 were hypertensive (including 1435 systolic-only hypertensive, 682 diastolic-only hypertensive, and 6726 were hypertensive for both systolic and diastolic pressure). After a median follow-up of 5.69 years, 502 patients died from CV diseases, including 130 stroke, 319 cardiac mortality, 53 other CV mortality. Baseline BP and BPV information are in Table 4. Compared with patients who died, the censored patients were younger, slightly heavier, less males, with less smoking, less diabetes, less previous CV diseases, lower BP levels and lower BPV levels.

Table 4. Baseline information

Variables	**Systolic** Mean (SD)	**Diastolic** Mean (SD)
24h Mean (mmHg)	139.4(17.3)	82.8(11.3)
Day Mean (mmHg)	145.3(18.1)	87.4(12.1)
Night Mean (mmHg)	128.2(18.7)	74.2(12.0)
24h ARV (mmHg)	10.2(2.6)	8.1(2.0)
24h wSD (mmHg)	11.7(3.4)	8.9(2.4)
24h CoV	8.4 (2.2)	10.8(2.9)
Day ARV (mmHg)	10.4(3.2)	8.2(2.6)
Day SD (mmHg)	12.3(4.1)	9.1(2.9)
Day CoV	8.4(2.6)	10.6(3.3)
Night ARV (mmHg)	9.4(3.5)	7.7(2.9)
Night SD (mmHg)	10.5(4.3)	8.4(3.2)
Night CoV	8.2(3.2)	11.4(4.4)

In general, all the systolic BPV indices have greater values than diastolic BPV, except for CoV, which is a standardized measure of dispersion of BPV. Both systolic and diastolic BPV following a trend that day BPV is slightly higher than 24h BPV, 24h BPV is higher than night BPV, except for diastolic CoV at night time that is higher than day and 24h BPV, this could be caused by the very small night diastolic BP.

3.3.2. Variables associated with BPV

Linear regressions are performed to identify variables which could be potentially associated with BPV, including age, BMI, gender, previous CV disease, diabetes and smoking. After identification of age, BMI, gender and previous CV disease as covariate, generalized linear regression is applied to check the association of mean BP with BPV, adjusting for the above covariates (Table 5). Although the systolic and diastolic mean BP values were all strongly correlated with the corresponding BPV, the coefficients were always very small. Almost all the BPV values increased with age, average of corresponding BP levels and smoking, except for the opposite trend or inconsistency in CoV; all the BPV values estimated for 24h or daytime seem to be higher in females, but higher in males when estimated for nighttime. Effect of BMI and previous CV diseases are not consistent.

Table 5. The coefficients from multivariable linear regression

		Mean SBP‡	Mean DBP‡	Age‡	Smoke	BMI	Gender	Previous CV
Systolic	24h wSD	0.07		0.06	0.46‡	-0.02‡	-0.32‡	-0.08
	24h ARV	0.05		0.05	0.19†	-0.01†	-0.27‡	0.15†
	24h CoV	-0.01		0.05‡	0.36‡	-0.01†	-0.19‡	-0.05
	Day SD	0.07		0.08	0.19	-0.04‡	-0.47‡	-0.03
	Day ARV	0.05		0.06	0.06	-0.04‡	-0.44‡	0.17
	Day CoV	-0.01		0.06‡	0.14*	-0.03‡	-0.30‡	0.03
	Night SD	0.06		0.04‡	1.06‡	-0.02*	0.03	-0.32*
	Night ARV	0.04		0.04‡	0.35‡	0.03‡	0.17*	-0.10
	Night CoV	-0.02		0.03‡	0.86‡	0.02†	0.05	-0.25†
Diastolic	24h wSD		0.05	0.02‡	0.43‡	0.02†	-0.03	-0.12
	24h ARV		0.04	0.01‡	0.25‡	0.03‡	-0.13†	0.09
	24h CoV		-0.06	-0.02‡	0.53‡	0.03‡	-0.08	-0.14
	Day SD		0.05	0.02‡	0.28†	-0.00	-0.18†	0.09
	Day ARV		0.05	0.01‡	0.20†	0.01*	-0.40‡	0.11
	Day CoV		-0.06	0.02‡	0.31‡	0.00	-0.20†	-0.10
	Night SD		0.05	0.00	0.75‡	0.05‡	0.51‡	-0.21*
	Night ARV		0.04	0.00	0.30‡	0.05†	0.45‡	0.07
	Night CoV		-0.08	-0.00	1.01‡	0.07‡	0.71‡	-0.28*

Gender: Female=0, Male=1; *p<=0.05; †p<=0.01 ‡:p<=0.001.

Self-regulation in human body allows BP to vary relatively freely within a threshold, beyond which the BP becomes hard to increase or decrease. e.g. within the frame of, say, 120-160 mmHg, the BP may vary in a large range, but it is a lot more difficult for the very low BP values to decrease further or for the very high BP values to increase further. Thus, it is expected that the BP values of a subject with mean BP within the frame could be considered as normal, a subject with mean BP below the frame would have his BP values left skewed (mean<mode) and a subject above with BP values right

skewed (mean>mode). However, what about the BPV estimates in patients with lower or higher mean BP levels? Are they also skewed or normally distributed? Are they smaller or bigger than the BPV estimates obtained in patients with BP around average level? Currently, among the large number of studies which have been designed to explore the effects of BPV on targeted outcomes, some explored the relationship between mean BP and BPV[31–33], while others did not pay much attention to this aspect[34].

In order to better explore this issue, the mean 24h SBP and DBP were divided into quartiles: 1st quartile (≤127/75 mmHg), 2nd quartile (127-137/75-82), 3rd quartiles (137-150/82-89 mmHg), and 4th quartile (≥150/89 mmHg), respectively. The number of patients in each quartile is 2623. The distribution of 24h wSD, ARV and CoV are presented in Table 6.

BP variability represented by all the above indices increased with mean BP levels, and its variance and interquartile range also increased with the mean BP. All indices showed increasing trends with BP quartiles except for diastolic CoV. The variance, e.g. the standard deviation of the BPV increases not only with the increase in mean BP, but also with the increase in the BPV. wSD and ARV among mean BP subgroups show right skewness however without any significant trend, as they could all be considered as normally distributed.

Table 6. Distribution of mean BP and BPV in the quartiles of 24h BP

		SBP									DBP								
Variable	**rank**	**Mean**	**SD**	**Range**	**Min**	**Max**	**IQR**	**Q1**	**Q3**	**Skew**	**Mean**	**SD**	**Range**	**Min**	**Max**	**IQR**	**Q1**	**Q3**	**Skew**
mean24h	**1**	119.3	6.3	39.3	88.0	127.3	8.3	115.9	124.2	-1.2	69.2	4.4	23.8	51.1	74.9	6.2	66.6	72.8	-0.9
	2	132.5	2.9	10.1	127.3	137.4	4.9	130.0	135.0	0.0	78.6	2.1	7.2	74.9	82.1	3.5	76.8	80.4	-0.1
	3	143.1	3.5	12.2	137.4	149.7	5.9	140.1	146.0	0.1	85.7	2.2	7.8	82.1	89.9	3.6	83.8	87.5	0.2
	4	162.7	11.5	73.7	149.7	223.4	14.3	154.0	168.3	1.4	97.7	6.9	48.7	89.9	138.6	8.6	92.5	101.0	1.4
ARV_24h	**1**	8.9	2.0	19.1	4.3	23.4	2.4	7.6	10.0	1.0	7.6	1.8	16.0	3.4	19.4	2.3	6.2	8.5	1.0
	2	9.7	2.2	17.6	3.8	21.4	2.7	8.2	10.9	1.0	8.0	1.9	14.8	3.6	18.5	2.3	6.7	9.0	1.0
	3	10.4	2.4	22.5	4.8	27.3	2.9	8.8	11.7	0.9	8.2	1.9	16.8	3.3	20.0	2.3	6.9	9.2	1.1
	4	11.7	2.7	18.4	5.5	23.9	3.5	9.7	13.2	0.7	8.8	2.1	20.1	3.8	23.9	2.6	7.3	9.9	1.2
wSD	**1**	9.9	2.6	39.0	4.6	43.5	3.2	8.2	11.3	1.7	8.1	2.1	19.5	3.6	23.1	2.5	6.7	9.2	1.1
	2	11.1	2.9	20.9	4.4	25.2	3.6	9.0	12.6	0.9	8.7	2.2	17.5	4.1	21.6	2.8	7.1	9.9	1.0
	3	12.0	3.2	26.0	5.2	31.2	4.2	9.7	13.9	1.0	9.0	2.3	17.3	4.0	21.2	2.9	7.4	10.3	1.0
	4	13.7	3.7	26.8	5.0	31.9	4.6	11.1	15.7	1.0	9.6	2.5	18.7	3.8	22.5	3.2	7.8	11.0	0.8
cv_24h	**1**	8.3	2.1	31.3	3.8	35	2.6	6.8	9.4	1.7	11.7	2.9	27.3	5	32.4	3.6	9.7	13.2	1.1
	2	8.3	2.2	17.4	3.2	20.6	2.7	6.8	9.5	0.9	11.1	2.8	22.6	5.2	27.8	3.6	9	12.6	1
	3	8.4	2.2	18.5	3.5	21.9	2.9	6.8	9.7	1.1	10.6	2.7	21.4	4.6	26	3.4	8.7	12.1	1
	4	8.4	2.2	17.6	3.2	20.9	2.8	6.9	9.6	1	9.9	2.6	21	3.2	24.3	3.4	8.1	11.4	0.9

In conclusion, the preliminary explorations indicate that as mean BP increase, BPV increases both in its mean value and in its variance, remaining normally distributed,

confirming previous work by Mancia et al[56].

3.3.3. Impact of BPV on mortality

Cox model

Cox Proportional Hazard (PH) model is used to estimate the hazard ratios (HR) for the increase in CV mortality by the various BPV indices. The hazard function for failure time T for an individual *i (i=1,2,…N)* and survival function are:

$$h(t) = h_0(t)\exp(\beta' X_i)$$

$$S(t,X) = [S_0(t)]^{\exp(\beta' X_i)}$$

$h_0(t)$ is a function of time only, which is left arbitrary but is assumed to be the same for all subjects. $\exp(\beta' X_i)$ is a quantity which depends on the individual covariates only through the regression coefficients. The covariates are assumed to be constant in time in the basic Cox model.

The Cox model is a semi-parametric model since it does not specify the form of $h_0(t)$. It does, however, specify the hazard ratio for any two individuals with covariate vector x_1 and x_2. The estimates from Cox model are obtained by maximizing the partial likelihood function, it estimates the "beta" coefficients considering the baseline hazard as a nuisance. The Cox PH model assumes that the HR is constant over time, $\hat{h}(t,X) = constant \times h_0(t,X)$. Or equally speaking, that hazard for one individual is proportional to the hazard for any other individuals, where the proportionality is constant and independent of time, as shown in the formula, $HR = \exp\sum_{i=1}^{p} \beta_i(X_{i1} - X_{i2})$, which involves no time. From the survival function we can derive this function,

$$\log[-\log S(t,x)] = \log[\beta x] + log[-logS_0(t)]$$

which indicate a constant distance from the log(-log) of the baseline survival function. Under the PH assumption, $\log[-\log S(t,x_1)]$ and $\log[-\log S(t,x_2)]$ would both exhibit a constant distance from the reference cumulative hazard $log[-logS_0(t)]$ and would therefore be parallel.

In order to apply the basic Cox model, a few assumptions should be satisfied:

1. Non-informative censoring.

2. Hazard function can be expressed by this formula:$h(t) = h_0(t)\exp(\beta' X_i)$, where the covariates are assumed to be constant in time.

3. Independent variates affect the hazard in a multiplicative way, or equivalently, affect the logarithm of the hazard in an additive way in the whole follow up period. (Proportional hazard).

4. The covariates (e.g. treatment group and gender) have independent effects on the hazard rate. In other words, there is no interaction between x_1 and x_2. In this case, the

treatment should have the same effect on males and females. However, this assumption can be relaxed by introducing an interaction term.

5. In a model with continuous covariates or covariates with multiple categories, the model assumes that the baseline log hazard is added to the same constant factor β for every increase of one unit in the value of X. If this is undesirable, alternative coding should be adopted for the variables.

6. Log-linear assumption: when a raw value of a quantitative variable is included in the model, a constant linear increase (or decrease) in the log hazard is assumed. (i.e. $\beta' X_i$is linear. But the log hazard between two comparative groups stays constant- PH assumption).

Three models are used to test the impact of BPV on CV mortality in Dublin study. Basic model is unadjusted, included only a BPV index as independent variable; the adjusted model is adjusted by age and BMI and stratified by sex, smoking and previous cardiovascular diseases. The full model contains the same stratification as in adjusted model and adjustment includes not only age and BMI, but also, in addition, the corresponding mean SBP, mean DBP and nocturnal fall, which is calculated as the ratio of night/day mean BP. Hazard ratio (HR) is calculated as one SD increase in mean and BPV for CV mortality (Table 7).

Most of the systolic and diastolic mean BP and BPV indices for day, night, 24h showed significant association with the CV mortality in the unadjusted models; with one standard deviation increase, the risk of CV mortality increased up to 82% percent. The association was weaker or absent in the adjusted model. Whereas in full model, when the mean systolic and diastolic BP of 24h, Day, Night and one of their corresponding BPV indices were included in the same model, all systolic BPV indices lost statistical significance and all mean SBP of 24h, day, night maintained their predictive effect. On the other hand, for diastolic BP and BPV, when mean BP and corresponding BPV indices were included in the same model, 24h DBP, day DBP, night BPV lost statistical significance, while night DBP, 24h wSD, 24h ARV, day ARV, day SD, 24h CoV and day CoV showed significant predictive effect for CV mortality. However the estimated increase of CV mortality risk was very small, for one standard deviation increase in these BPV indices, the risk of CV mortality increased by about 3-14%. Attention needs to be paid to the HR values estimated for diastolic BP, which are smaller than 1. This could be caused by the adjustment of the corresponding systolic BP. Indeed, in the case of increasing values of SBP, a reduction in DBP may imply a wider pulse pressure which is known to carry an increased risk of cardiovascular complications. This might partly explain the reduced HR in full model analysis when adjusting for SBP values and for BPV estimates.

Table 7. HR and 95% Confidence interval of Cox models for BPV estimates

			Basic model	Adjusted model	Full model#
Systolic	24H	Mean	1.64[1.52-1.77]‡	1.29[1.19-1.39] ‡	1.38[0.22-1.56] ‡
		ARV	1.52[1.42-1.63] ‡	1.09[1.01-1.18] ‡	1.01[0.93-1.11]
		wSD	1.47[1.37-1.58] ‡	1.11[1.02-1.20] †	1.01[0.92-1.11]
		CoV	1.27[1.18-1.37] ‡	0.98[0.90-1.07]	1.01[0.93-1.10]
	Day	Mean	1.43[1.32-1.55] ‡	1.21[1.11-1.31] ‡	1.29[1.15-1.46]‡
		ARV	1.40[1.30-1.50] ‡	1.07[0.99-1.16]	1.02[0.94-1.11]
		SD	1.46[1.37-1.57] ‡	1.11[1.03-1.21]*	1.05[0.96-1.14]
		CoV	1.33[1.24-1.43] ‡	1.03[0.95-1.12]	1.05[0.96-1.14]
	Night	Mean	1.82[1.69-1.95] ‡	1.38[1.28-1.49] ‡	1.54[1.36-1.74]‡
		ARV	1.31[1.22-1.40] ‡	1.10[1.00-1.16]	1.00[0.93-1.08]
		SD	1.23[1.14-1.32] ‡	1.03[0.96-1.12]	0.96[0.88-1.04]
		CoV	1.01[0.93-1.10]	0.91[0.84-1.00]	0.96[0.87-1.05]
Diastolic	24H	Mean	0.99[0.91-1.09]	1.14[1.05-1.25]†	0.84[0.73-0.95]*
		ARV	1.34[1.25-1.44] ‡	1.14[1.07-1.22] ‡	1.13[1.05-1.21] ‡
		wSD	1.33[1.24-1.43] ‡	1.17[1.09-1.25] ‡	1.14[1.06-1.23] ‡
		CoV	1.30[1.21-1.39] ‡	1.12[1.04-1.21]†	1.13[1.05-1.22]†
	Day	Mean	0.85[0.78-0.93]	1.08[0.99-1.18]	0.85[0.74-0.97]*
		ARV	1.27 [1.18-1.36] ‡	1.12[1.05-1.20] ‡	1.11[1.04-1.20]‡
		SD	1.33[1.23-1.43] ‡	1.15[1.08-1.24] ‡	1.13[1.05-1.22]‡
		CoV	1.34[1.26-1.44] ‡	1.13[1.05-1.22]†	1.12[1.04-1.21]†
	Night	Mean	1.22[1.13-1.33] ‡	1.22[1.12-1.32] ‡	0.82[0.72-0.94] *
		ARV	1.18[1.09-1.27] ‡	1.09[1.01-1.17] *	1.07[0.98-1.15]
		SD	1.14[1.05-1.23] ‡	1.10[1.02-1.19] *	1.08[0.99-1.17]
		CoV	1.06[0.97-1.15]	1.05[0.95-1.13]	1.09[0.99-1.19]

#: hazard ratio for one standard deviation in mean BP estimated in the full model is adjusted for the corresponding day or night SD or wSD. The HRs adjusted with corresponding ARV are similar (not shown). *p<0.05; †p<0.01 ‡:p<0.001; Basic model: unadjusted. Adjusted model: adjusted for age, BMI and stratified by sex, smoking and previous cardiovascular diseases. Full model: adjusted for age, BMI, the corresponding mean SBP, DBP and nocturnal fall, stratified by sex, smoking and previous cardiovascular diseases. All the models use standardized BPV indices (per 1 SD increase) for Hazard ratio with 95% confidence interval for CV mortality. SD: standard deviation, ARV: Average real variability, wSD, Weighted SD. Hazard ratio calculated per unit increase (Not shown) are all very similar to the result above.

3.3.4. Accelerate Failure Time Models

Using the non-parametric (Log-rank test) or semiparametric (Cox PH model) model can exempt us from specifying the hazard function completely. The utility of the proportional hazards models stems from the fact that a reduced set of assumptions is needed to provide the hazard ratios, which are easily interpreted and clinically meaningful. However, when the hazard function has a known parametric form the use

of a fully parametric model is useful to better address the goal of the analysis, in particular for prediction aims. Those models have some advantages[57], in particular, full maximum likelihood is used to estimate the parameters; the coefficients can have a more intuitive clinical interpretation; fitted values from the model can provide a direct estimates of survival time (and residuals).

Brandon et al[58] have summarized several classes of parametric models: (1) parametric proportional hazards model which takes the form of the Cox model but assumes a parametric form of the baseline hazard; (2) the additive hazards model where the predictors affect the hazard function in an additive manner instead of multiplicative; and (3) the Accelerated Failure Time (AFT) model, which is most similar to conventional linear regression. In this section, AFT model will be discussed.

A proportional hazards model estimates the hazard rate (event per time) and assumes that the effect of a covariate is to multiply the hazard by some constant. AFT model provides an alternative to the commonly used proportional hazards models, it assumes that the effect of independent variables is to multiply the survival time by some constant, usually called as acceleration factor or time ratio θ [59]. AFT models are predominantly full parametric, i.e. a probability distribution is specified for $\log(T_0)$, and the interpretation of θ in AFT models is straightforward: e.g. event of interest in the relevant life history of an individual happens θ as fast as that in the reference group. Under the AFT model, the expected survival time, median survival time of reference group are θ times as much as those of compared group. Attention needs to be paid that this does not necessarily mean that the hazard function $h(t|\theta)$ is always θ times as high.

In full generality, the accelerated failure time model can be specified as[60]:

$$h(t|\theta) = \theta h_0(\theta t)$$

$$f(t|\theta) = \theta\, f_0(\theta t)$$

$$S(t|\theta) = S_0(\theta t)$$

θ denotes the joint effect of covariates, typically $\theta = \exp(\beta_1 x_1 + .. + \beta_P x_P)$ (doesn't include the error term), $\theta > 0, t \geq 0$. The procedure Proc LifeReg in SAS fits data into models by the following equation[61]:

$$y = \beta_0 + \beta_1 x_{i1} + .. + \beta_P x_{iP} + \sigma\epsilon_i$$

where β is the regression coefficient of interest, σ is a scale parameter and ϵ is the random disturbance term, usually assumed to be an independent and identically distributed (i.i.d.) variable N (0,1) with some density function $f(\epsilon)$. All subjects share the same ϵ. These models are equivalent to AFT models when the log of the response time is the quantity being modeled, in such case, $y = \log(T_i) = \beta_0 + \beta_1 x_1 + .. + \beta_P x_P + \sigma\epsilon_i$. Taking the logarithm is in line with the fact that the survival

times are always positive. This reduces the AFT model into regression analysis (typically a linear model) where $\beta_0 + \beta_1 x_1 + .. + \beta_P x_P$ represents the fixed effects, and $\sigma\epsilon_i$ represents the noise.

Regarding the fixed effects, we can write the following formula:

$$T_i = \exp(\beta_0 + \beta_1 x_1 + .. + \beta_P x_P) \cdot \exp(\sigma\epsilon_i)$$

Thus the survival time can be seen to be multiplied by a constant effect $\exp(\beta)$ which is referred to as acceleration factor or time ratio[59], in another word, the covariate influences the survival time by a constant. Thus, if T_0 is an event time sampled from the baseline distribution corresponding to values of zero for the vector of covariates x_p, then the AFT model specifies the event time as:

$$T = \exp(\beta_0) \cdot \exp(\beta_p x_p) \cdot \exp(\sigma\epsilon_i) = T_0 \exp(\beta_p x_p)$$

$$\log(T) = \beta_p x_p + \log(T_0)$$

It shows that the moderated life time *T* is distributed such that $T\theta$ and the unmoderated life time T_0 have the same distribution, the effect of the covariates in an AFT model is to change the scale, and not the location, of a baseline distribution of failure times[61]

When holding other covariates fixed, change of the covariate *k* from *0* to *1* (from one group to another) or increase in *k* by one unit, the corresponding survival time T_1 and T_2 are:

$$\begin{aligned} T_1 &= e^{\beta_0+\beta_1 z_1+\ldots+\beta_k z_k+\ldots+\beta_p z_p} e^{\sigma\varepsilon_1} = c_1 e^{\sigma\varepsilon_1} \\ T_2 &= e^{\beta_0+\beta_1 z_1+\ldots+\beta_k (z_k+1)+\ldots+\beta_p z_p} e^{\sigma\varepsilon_2} = c_2 e^{\sigma\varepsilon_2} \end{aligned}$$

Where c_1 and c_1 are two constant related by $c_2 = c_1 * e^{\beta_k}$. The corresponding survival functions are:

$$\begin{aligned} S_1(t) &= P[T_1 \geq t] = P[c_1 e^{\sigma\varepsilon_1} \geq t] = P[e^{\sigma\varepsilon_1} \geq c_1^{-1} t], \\ S_2(t) &= P[T_2 \geq t] = P[c_2 e^{\sigma\varepsilon_2} \geq t] = P[e^{\sigma\varepsilon_2} \geq c_2^{-1} t] \end{aligned}$$

Since $\epsilon_1 = \epsilon_2$, and $c_2 = c_1 * e^{\beta_k}$, we have

$$S_2(e^{\beta_k} t) = P[e^{\sigma\varepsilon_2} \geq c_2^{-1} e^{\beta_k} t] = P[e^{\sigma\varepsilon_2} \geq c_1^{-1} e^{-\beta_k} e^{\beta_k} t] = P[e^{\sigma\varepsilon_2} \geq t] = P[e^{\sigma\varepsilon} \geq t] = S_1(t)$$

Therefore, the survival time corresponding with one unit increase are in relationship as $t_2 = e^{\beta_k} t_1$. When β_k is small, $\frac{t_2 - t_1}{t_2} = \beta_k$. Therefore, the coefficient β_k can be interpreted as the percentage increase or decrease in the average or median survival time over one unit change in covariate *k*. The exponential of β is hazard (or odds)

ratio of survival if calculated as exp(β) and of death if calculated as exp(-β).

As regards to the noise term, different distributional forms of ϵ imply different distributional forms of T_0, i.e. different baseline distributions of the survival time. Depending on how ϵ is distributed, the AFT model can be specified for different distribution of baseline survival time, as shown in Table 8.

Table 8. The distribution of error term and survival time in AFT models

Distribution of ϵ	Distribution of T	SAS Lifereg
Extreme values (2 par.)	Weibull	Dist=weibull (Default)
Extreme values (1 par.)	Exponential	Dist=exponential
Log-gamma	Gamma	Dist=gamma
Logistic	Log-logistic	Dist=llogistic
Normal	Log-normal	Dist=lnormal

Exponential and Weibull distribution

The simplest AFT model is the exponential model where *T* at *z* = 0 (usually referred to as the baseline) has exponential distribution with constant hazard exp(-β_0). For the estimation of $\log(T_i) = \beta_0 + \beta_1 x_{i1} + .. + \beta_P x_{iP} + \sigma\epsilon_i$, this distribution assumes that σ in set to be 1 and ϵ has a standard *extreme value* distribution, indicates that the events occur continuously and independently at a constant average rate. Automatically we have proportional hazards models.

$$h(t) = \lambda = e^{-(\beta x)}$$
$$f(t) = \lambda e^{-\lambda t}\ (\lambda > 0)$$
$$S(t) = e^{-\lambda t}$$

Therefore, if we increase the value of one covariate by one unit from x_k to x_{k+1} while holding other covariate values fixed, then the ratio of the corresponding hazards, HR, is equal to e^{β_k}, or equivalently, β_k can be interpreted as the increase in log-hazard as the value of covariate *zk* increases by one unit.

Exponential distribution can be considered as a particular case of Weibull distribution, or vice versa, Weibull can be considered as a generalized exponential distribution. The probability density function of a Weibull random variable is

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

Where k>0 is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. When k=1, Weibull distribution is the exponential distribution, or equally speaking, for the estimation of $\log(T_i) = \beta_0 + \beta_1 x_{i1} + .. + \beta_P x_{iP} + \sigma\epsilon_i$, the distribution $\sigma\epsilon$ has a standard *extreme value* distribution with the scale σ together with shape parameter k need to be estimated. In Weibull model,

$$S(t) = e^{-(\lambda t)^k}$$

$$h(t) = \lambda k(\lambda t)^{k-1}$$

$$f(t) = h(t)S(t) = \lambda k(\lambda t)^{k-1}$$

Weibull model allows the hazard form to change with time through the exponent *k-1*.

$$k\begin{cases} < 1 & h(t) \uparrow \\ > 1 & h(t) \downarrow \\ = 1 & h(t) = c \end{cases}$$

The function of the linear predictor $\beta' x$ can also be in other form instead of exponential, but exponential is preferable as it takes on positive values and no constraints on the values of β is required. The Weibull regression model can be considerd as a parametric member of the class of PH regression models, where the baseline hazard function is specified to have a form containing a function of time, $\lambda(t, x) = \lambda_0(t)g(x)$. The coefficient has the interpretation that it is the increase in log-hazard or decrease of log survival time when the value of covariate increases by one unit while other covariate values being held unchanged.

The Weibull distribution (including the exponential distribution as a special case) can be parameterized as either a PH model or an AFT model, and is the only family of distributions to have this property. The results of fitting a Weibull model can therefore be interpreted in either framework.

In Cox PH model: $HR = \frac{h(t,x_1)}{h(t,x_2)} = \frac{h_0(t)\exp(\beta' x_1)}{h_0(t)\exp(\beta' x_2)} = \exp[\beta'(x_1 - x_2)]$

$$S(t, X) = [S_0(t)]^{e^{\beta x}}$$

$$\log[-\log S(t, x)] = \log(\beta x) + log[-logS_0(t)]$$

In Weibull Model:

$$HR = \frac{h(t, x_1)}{h(t, x_2)} = \frac{\lambda k(\lambda t)^{k-1}\exp(\beta' x_1)}{\lambda k(\lambda t)^{k-1}\exp(\beta' x_2)} = \exp[\beta'(x_1 - x_2)$$

$$S(t) = e^{-(\lambda t)^k e^{\beta x}}$$

$$\log[-\log S(t, x)] = \beta x + klog\lambda + klogt$$

Under the PH assumption, $\log[-\log S(t, x_1)]$ and $\log[-\log S(t, x_2)]$ would both exhibit a constant distance from the reference cumulative hazard $log[-logS_0(t)]$ and would therefore be parallel. While if Weibull model is suitable, the plot of $\log[-\log S(t, x)]$ obtained in two subgroups should give approximately two parallel lines (linearity) when no time-dependent covariates present. The slope of the line provide a rough estimate of *k* and the difference between two intercepts corresponds to

the quantity $\beta'(x_1 - x_2)$.

Considerations from the Weibull regression model:

1. The log[-log(s(t))] is linear with log(t).

2. The shape of the hazard is monotonous, depend on *k*, and is the same for all the subgroups.

3. The effect of the covariates is to act multiplicatively [×exp(β)] on the hazard function, and the hazard ratio between two subgroups is a constant. The effect on survival function is exponential.

4. The effect of each covariate on the constant HR is multiplicative, for a unit change in *X*, HR is multiplied by a factor exp(β).

Lognormal distribution

A lognormal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Given a log-normally distributed random variable X and two parameters μ and σ, which are, respectively, the mean and standard deviation of the variable's natural logarithm, then the logarithm of X is normally distributed, and we can write X as $X = exp(\mu + \sigma z)$, with z as standard normal variable. This relationship is true regardless of the base of the logarithmic or exponential function. On a logarithmic scale, μ and σ can be called the location parameter and the scale parameter, respectively.

The log-normal model simply assumes that ϵ =N(0,1). Shape of lognormal distribution is similar to the log-logistic distribution and yields similar model results, however, it is neither a proportional hazards model nor a proportional odds model.

$$S(t) = 1 - \Phi[\log(\lambda t)/\sigma]$$

$$h(t) = f(t)/S(t)$$

$$f(t) = \left(1/\sqrt{2\pi}\,\sigma t\right) e^{-[\log(\lambda t)]^2/2\sigma^2}$$

Log-logistic distribution (proportional odds models)

The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution, it is similar in shape to the log-normal distribution but has heavier tails. The log-logistic distribution provides the most commonly used AFT model, in which the disturbance term ϵ has a standard logistic distribution. It can exhibit a non-monotonic hazard function which increases at early times and decreases at later times.

$$S(t,x) = \frac{1}{1 + (\lambda t)^k \exp(\beta x)}$$

$$f(t,x) = 1 - S(t,x) = \frac{(\lambda t)^k \exp(\beta x)}{1 + (\lambda t)^k \exp(\beta x)}$$

$$h(t) = \frac{\lambda k t^{k-1}}{1 + kt^{k}}$$

An empirical check on the suitability of the log-logistic model for the analysis of a data set in the presence of covariates is the log odds function, as odds of failure is:

$$\text{Odds} = \left[\frac{F(t,x)}{1-F(t,x)}\right] = (\lambda t)^{\eta}\exp(\beta' x)$$

$$OR = \exp[\beta'(x_1 - x_2)]$$

$$\log\left\{\frac{F(t,x)}{1-F(t,x)}\right\} = \beta' x + klog\lambda + klogt$$

which shows to be linearly related to *log(t)*. Thus if the survival time follows a log-logistic distribution, plot *log(odds)* against *log(t)* should be linear with slope *k*. In fact, log-logistic model is the only one to satisfy both the AFT and PO assumption. In log-logistic model, the hazard function is $\lambda(t) = \frac{\lambda k(\lambda t)^{k-1}\exp(\beta\prime x)}{1+\lambda k(\lambda t)^{k-1}\exp(\beta\prime x)}$, the shape of which depends on the k.

Summary of considerations from the log-logistic regression model:

1. The shape of the odds is monotonous and depends only on *k*.

2. The effect of the covariates is to act multiplicatively on the odds function, and the odds between two subgroups (odds ratio) is a constant.

3. The effect of each covariate on the constant OR is multiplicative, for a unit change in *x*, odds ratio OR is multiplied by a factor $\exp(\beta)$

Table 9. Some parametric failure time models for a homogeneous population of individuals (number of parameters in brackets)

Distribution		Form of the hazard function
Exponential (1) $\lambda > 0$	$f(t) = \lambda\exp(-\lambda t)$ $S(t) = \exp(-\lambda t)$ $\lambda(t) = \lambda$	constant
Weibull (2) $\lambda, p > 0$	$f(t) = \lambda p(\lambda t)^{p-1}\exp[-(\lambda t)^{p}]$ $S(t) = \exp[-(\lambda t)^{p}]$ $\lambda(t) = \lambda p(\lambda t)^{p-1}$	if $p < 1$, monotone decreasing; if $p > 1$, monotone increasing; if $p = 1$, constant ($\lambda(t) = \lambda$)
Gamma (2) $\lambda, k > 0$	$f(t) = \lambda(\lambda t)^{k-1}e^{-\lambda t}/\Gamma(k)$ $S(t) = 1 - \int_0^{\lambda t} u^{k-1}e^{-u}du/\Gamma(k)$ $\lambda(t) = f(t)/S(t)$	if $k < 1$, monotone decreasing with $\lim_{t\to\infty}\lambda(t) = \lambda$; if $k > 1$, monotone increasing with $\lim_{t\to\infty}\lambda(t) = \lambda$; if $k = 1$, constant ($\lambda(t) = \lambda$)
Log-normal (2) $\lambda, \sigma > 0$ $\Phi(z) = \int_{-\infty}^{z}\phi(u)du$ $\phi(z) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$	$f(t) = (1/\sqrt{2\pi\sigma t})\exp(-[\log(\lambda t)]^2/2\sigma^2)$ $S(t) = 1 - \Phi(\log(\lambda t)/\sigma)$ $\lambda(t) = f(t)/S(t)$	non-monotonic: $\lambda(0) = 0$, then increases to maximum and decreases, with $\lim_{t\to\infty}\lambda(t) = 0$
Log-logistic (2) $\lambda, p > 0$	$f(t) = \lambda p(\lambda t)^{p-1}[1 + (\lambda t)^{p}]^{-2}$ $S(t) = [1 + (\lambda t)^{p}]^{-1}$ $\lambda(t) = \lambda p(\lambda t)^{p-1}[1 + (\lambda t)^{p}]^{-1}$	if $p < 1$, monotone decreasing from ∞; if $p = 1$, monotone decreasing from $\lambda(0) = \lambda$; if $p > 1$, non-monotonic: increases from 0 to a maximum at $t = \lambda^{-1}(p-1)^{1/p}$ and decreases towards 0

Note: discrepancy in symbols: $\lambda(t)$ is the hazard and *p* is shape parameter, which are represented as *h(t)* and *k* in this thesis.

Table 9 below is the summarized by Marubuni and Valsecchi in their book <Analyzing survival data from clinical trials and observational studes>[62], which provides the

probability density function, survival function and hazard function in some AFT models. John P. Klein and Melvin L. Moeschberger have also made a summary of those functions over various distributions in their book <survival analysis techniques for censored and truncated data>, 2nd edition, Springer 2003[63]

Performance of Semi-parametric and parametric models

Depending on the shape of the baseline hazard function, there might be models for which it is equivalent to assume both PH and AFT models. In fact the Weibull model (including exponential model obviously) has the feature of being both proportional hazards and AFT, and log-logistic model being both proportional odds and AFT.

If the only basic assumption of proportional hazards is not met, parametric models are suitable alternative models to be used. A simulation study showed that whether PH assumption is met or not, the log-logistic model is the best fitted model[64]. Studies have indicated that under certain situations when the shape of the survival time is determined, the parametric models are more powerful and efficient than Cox's regression model in predicting the median survival time or the probability of a given individual surviving for at least a predefined time[65].

However, evidence from literature suggests that when the primary concern is to evaluate the effect of covariates, little is to be gained by moving from semi-parametric models, like the Cox model, to the parametric models, i.e. Cox, Weibull and log-logistic models give same view of how the prognostic factors influence survival, although among many studies have resorted to various models at the same time [59,66–69], sometime the result could be very different using Cox and AFT models[69]. The choice of the model should not be based on which gives a favorable P value, but on a proper evaluation of the distributional assumptions. It should be noted that there is no distribution that provides a perfect fit, and it is possible that more than one distribution may fit the data well[58].

A general model including all the above distribution as special cases can be used to discriminate which models are more suitable to fit the data. Such a model is the generalized F distribution discussed by Kalbfleisch and Prentice (1980)[70]. The method is computationally complicated and in practical applications might have little ability to discriminate among alternative models.

Even with the evidence that little is to be gained by moving from Cox model to the parametric models in evaluating the effect of covariates, for the purpose of efficiency and survival prediction, AFT models are applied in this report.

Application of the parametric distribution to a dataset

There are some methods based on which we choose one parametric distribution versus others as listed below.

1. The applicability of the model to a data set can be empirically checked through plots

of KM survival estimation (e.g. proc lilfetest in SAS).

- If T is exponential, then $S(t) = exp(-\lambda t)$, so $log(S(t)) = -\lambda t$, the $log(S(t))$ *vs t* plot is a straight line on time with slope= $-\lambda$ and intercept=0.

- If T is Weibull, given that $S(t) = e^{-\lambda t^k}$, so $log[-logS(t)] = log(\lambda) + klog(t)$, which is a straight line with slope *k* and intercept $log(\lambda)$ on $log(t)$. A $k=1$ indicates that the simpler exponential model could be used. In the PH Cox model, it would be enough it the curves are parallel.

- For log-logistic, the plot of $log((1-s(t))/s(t))$ against $log(t)$ should be a straight line with slope *k*.

- Qualitative and explicit forms for *f(t), S(t),* and *h(t)* from KM estimation.

2. How well a model fits the data:

    1) log-likelihood can be used to compare model fitness;

    2) Distributions with multiple parameters defining their shape may have a better fit, but if parsimony is desired, it would be better to rely on a penalized metric provided by model selection indices such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC) to choose which distribution gives the best fit with the fewest parameters among candidate distributions. These indices allow for numeric comparison which is less subjective than comparing graphs.

    3) Brier score, c-statistics and some other methods could be better choices for model discrimination and calibration. Related content will be discussed later.

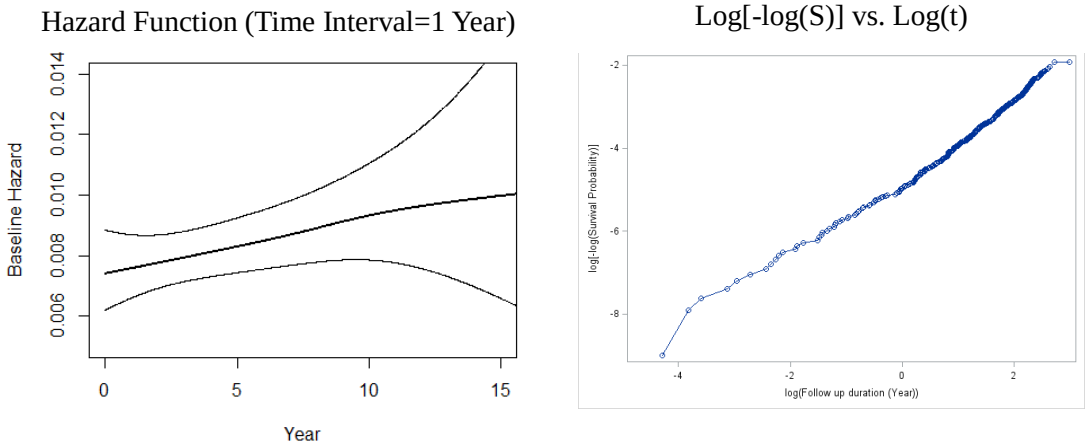
    4) If an exponential model fits the data well, then the regression coefficient estimates by Cox model or by parametric model should be approximately on the same size by absolute value, but opposite in sign but). In fact, a proportional hazards model outputs the regression coefficient estimated in log-hazard form (β_k), and a parametric model outputs the regression coefficients estimated in log-survival form. If the Weibull model is a reasonable model for the data, the regression coefficients not only have opposite signs (except possibly for the intercept) but also have different magnitude as compared with Cox model estimation(depending on whether $k > 1$ or $k < 1$).

Parametric models in Dublin data

Bshazard package is used for the non-parametric estimate of the baseline hazard function and its 95% confidence interval, with data-driven smoothing[71] (Figure 2, left). In Dublin data it showed as a constant increase with time, although limited within a very small range from 0.007-0.01. *Log[-log(S)] vs. Log(t)* plot (Figure 2, right) shows

an almost straight line, indicating that the exponential and Weibull distribution are suitable for these data. The Weibull distribution is chosen for this report for the purpose of precision.

Figure 2. the Hazard function and *Log[-log(S)] vs. Log(t)* in full dataset



Based on those, three parametric models (exponential, Weibull and log-logistic model) are used preliminarily for the estimation of HR of CV mortality per a SD increase in BPV measured by different indices. Weibull model seems to fit the Dublin data better from the log-likelihood values and the other model assessing methods including the SE of coefficient, AIC (data not shown). Furthermore, all the estimations of shape parameter in Weibull distribution are all bigger than 1, thus Weibull distribution is prefered than exponential. The results of Weibull model are in Table 10.

Table 10. Parametric Weibull model analysis per SD increase in diastolic BPV indices

Index	LLH	β	LCL(β)	UCL(β)	Exp[β[LCL, UCL]]	P	AIC
24hARV	-1965	-0.09	-0.16	-0.03	0.91[0.86-0.97]	0.0029	3952
24hwSD	-1964	-0.11	-0.17	-0.04	0.90[0.84-0.96]	0.0011	3950
24h CoV	-1965	-0.11	-0.17	-0.04	0.90[0.84-0.96]	0.0019	3951
Day ARV	-1973	-0.10	-0.16	-0.03	0.91[0.85-0.97]	0.0023	3969
Day SD	-1972	-0.12	-0.18	-0.05	0.89[0.83-0.95]	0.0004	3965
Day CoV	-1972	-0.11	-0.18	-0.05	0.89[0.84-0.95]	0.0007	3967

Note: LLH: Log-likelihood; LCL: lower 95% confidence interval; UCL: Upper 96% confidence interval. The model covariate is the same as those in Cox full model, whereas the strata are classes in Weibull model.

In Weibull models, the probability of survival or event in a given time frame can be estimated for each individual, the exp(β) is interpreted as survival time ratio or exp(-β) as HR of event per one unit change, in the case of Dublin study, one SD increase in BPV, after adjusting for the corresponding mean systolic and diastolic BP, night fall, age, gender, smoking, BMI, previous CV diseases and diabetes. For one standard deviation increase in BPV, the survival time is reduced to about 89-91%, which is in line with the results estimated by Cox PH model with HR around 1.11-1.34.

3.3.5. Find the optimal threshold values in survival data

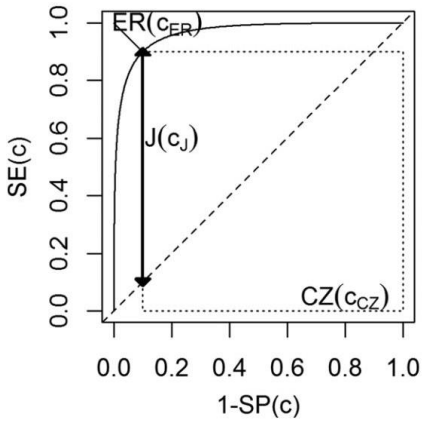
In the contest of Dublin data, a BPV index with its value v bigger than the potential cutoff value c is defined as positive or otherwise as negative if $v \leq c$. Sensitivity (SE) at c are defined as the probability of having positive BPV given that the subject died ($SE = P(V > c|M = 1)$) and specificity (SP) as the probability of having negative BPV given that the subject survived a given time ($SP = P(V \leq c|M = 0)$). A common approach to separate patients into higher and lower risk groups would be the receiver operating characteristic (ROC) curves based approaches like Youden Index, concordance probability and Euclidean Distance Formula as indicated in Figure 3.

Youden index, J= SE +SP –1, rang= [-1, 1], indicated as J(cJ) in Figure 3. Maximizing J is equal to maximizing the area under the ROC curve (AUC) for an indicator variable obtained by dichotomizing the continuous biomarker. Optimal probability cutoff is where J is maximum;

Concordance probability Cc=SE*SP, range= [0,1], indicated as CZ(cCZ) in Figure 3. The concordance probability for binary classification can be expressed as a rectangular area with width and length being the SE and SP associated with a cutoff on continuous X, such that its vertex (1-SP, SE) lies on the ROC curve. Optimal probability cutoff is where Cc/the rectangle is maximum [72];

Euclidean Distance D = Sqrt ($(1\text{-SE})^2 + (1\text{-SP})^2$), range=[0,1], also called as the point-to-(0,1), indicated as ER(cER) in Figure 3. Optimal probability cutoff is at where D is minimum.

Figure 3. Cut-point finding methods background: ROC-based functions[73]



Note: Youden index J(cJ) is the thick line segment. Concordance probability CZ(cCZ) is the area of the dotted rectangle and Euclidean distance ER(cER) is the thin line segment.

Concerning the area under the curve (AUC), its first development come from applications to diagnostic testing in radiology[74], it can be defined as the area under

ROC curve, the plot of sensitivity vs 'one minus specificity' for all possible cut-off values. This definition has been shown to be equivalent to defining AUC as the c statistics, which is probability that a given diagnostic test (or predictive model in this report) assigns a higher probability of an event to those who actually have (or develop) events[74]. The bigger the AUC, the better a biomarker performs in detecting an event.

Major problem of ROC with survival data

Dealing with time-dependent variable

In the regular setting with no time-dependent binary outcome and no censoring, the SE and SP given a certain predictor value are fixed. But it is not the case of survival data because SE and SP could vary with the change of time and censoring, and so does the ROC curve. When extending ROC curve to the case of failure time outcome, the interest could be to understand whether there will be the development of a disease/mortality or not within a given time point of clinic interest. Rather than a simple binary outcome, $Y_i=1$, a survival time can be viewed as a time-varying binary outcome by focusing on the counting process representation $N_i(t)=1(T_i \leq t)$. Under the setting of cumulative definition of case/dynamic definition of control[75], at any fixed time *t*, the entire population is classified as either case or a control on the basis of vital status at time *t*. Each individual plays the role of a control for time $t<T$, but then contributed as a case for later time, $t>T$.

The sensitivity and specificity can be defined using Bayes' theorem and KM estimation of survival function as:

$$P\{X > c|D(t) = 1\} = \frac{\{1-S(t|X > c)\}P(X>c)}{1-S(t)},\ P\{X \leq c|D(t) = 0\} = \frac{S(t|X \leq c)P(X \leq c)}{S(t)}$$

where $S(t)$ is the survival function $S(t) = P(T > t)$ and $S(t|\ X > c)$ is the conditional survival function for the subset defined by X > c, c is the potential threshold value. This formula is simple but limited by one possibility that it does not guarantee that SE or SP are monotone[75]. While the estimation using nearest neighbor estimation (NNE) of the bivariate distribution provided by Akritas [76] does not have this limitation:

$$P\{X > c|D(t) = 1\} = \frac{[1-F_x(c)]-S_m(c,t)}{1-S_m(t)},\ P\{X \leq c|D(t) = 0\} = 1 - \frac{S_m(c,t)}{1-S_m(t)}$$

where $F_x(c) = P(x \leq c)$ is the distribution function of X, $S_m(c,t)$ is an estimator of the conditional survival function characterized by a smooth parameter *m*. Furthermore, NNE estimator allows the situation when censoring process depends on the diagnostic marker.

Cumulative/dynamic definition is most appropriate when a specific time *t* is important and scientific interest lies in discriminating between subjects who die prior to a given time *t* and those that survive beyond, which is also been used in this report.

In 2005, Heagerty et al has proposed the incident definition of case/dynamic definition

of control [77]. A subject can play the role of a control for an early time, $t \leq T_i$, but then play the role of case when $t = T_i$. This dynamic status parallels the multiple contributions that a subject can make to the partial likelihood function. Here, incident sensitivity and dynamic specificity are defined by dichotomizing the risk set at time *t* into those observed to die and those observed to survive, sensitivity measures the expected fraction of subjects with a marker greater than *c* among the subpopulation of individuals who die at time *t*, while specificity measures that fraction of subjects with a marker less than or equal to *c* among those who survive beyond time t. Incidence sensitivity and dynamic specificity have some appealing characteristics relative to the alternative definitions. First of all, those definitions are based on classification of the risk set at time *t* into cases and controls, and are therefore a natural companion to hazard models. Second, the definitions easily allow extension to time-dependent covariates by introducing a time-varying marker. Finally, use of incident sensitivity and dynamic specificity allows both time-specific accuracy summaries and time-averaged summaries that directly relate to a familiar global concordance measure. In this paper[77], Heagerty has been adopted the global summary of a cutoff value as $c = p(M_j > M_k | T_j < T_k)$, which indicates the probability that a subject who died at earlier time has a larger value of the marker. With the assumption that observations M_j, T_j and M_k, T_k are independent, plus that T is continuous such that $p(T_j = T_k) = 0$, the summary $c = p(M_j > M_k | T_j < T_k)$ is a weighted average of the area under time-specific ROC curves. In many applications no a-prior time *t* of clinical interest is identified, and a global accuracy summary is desired. A substantive application that demonstrates use of cumulative/dynamic ROC curves for a Cox regression model can be found in Fan et al[78].

Etzioni et al[79] have adopted an alternative definition of time- dependent sensitivity and specificity considering incident cases/static controls, which is not discussed here as it is less appropriate to the setting of evaluating predictively of risk of disease onset over some fixed period[80].

Dealing with censored data

Time-dependent SE and SP also need to, inevitably, cope with censored data, non-parametric estimation of SE and SP can be derived following the setting of cumulative definition of cases and dynamic definition of controls. A direct estimation by weighting is originated from the consideration that subjects with observed status (cases or controls) are 'selected' from the censoring process and can be weighted to represent the subjects that are censored[81]. An inverse weighting is defined as $1/G(t)$, $G(t)$ is survival probability $S(t)=P(c>t)$ (censored time>t), which is monotonously decreasing. For a case i, $T_i < t$ and for a control j, $T_j < t$, $G(T_i) > G(T_j)$, and thus the inverse weighting $1/G(T_i) < 1/G(T_j)$. This means:

> Considering $1/G(T_j)$, the bigger is *t*, the higher probability that the censored data would otherwise turned to be cases.

Considering $1/G(T_i) < 1/G(T_j)$, the observed controls underwent greater chance of being censored during follow-up and need greater magnification than cases to represent censored data that would otherwise turned to be controls.

The probability G*(t)* can be obtained by KM or lifetable method on the whole sample subjects regardless of the test result of the grouping factor, e.g. treatment effect (Marginal weighting) or on the two subsets defined by the grouping factor (Conditional weighting).

Another direct approach would be to impute the disease status of censored subjects in terms of probability that the subject is a case or a control; give the available information on the survival time and its marker value. The survival function is estimated separately for treatment groups using KM method, and then the imputation is performed separately using the survival functions[81].

In summary, the principal concept of direct estimation is based on an inverse probability weighting scheme applied to the counts of the groups of disease-free and diseased subjects, and originates from the consideration that subjects with observed status are selected from the censoring process and can be weighted to represent the subjects that are censored. There are also indirect estimation by imputation or by Bayes theorem, which relies on writing SE and SP in terms of quantities that are estimable from the available data in the presence of censoring, and on plugging-in the estimates. Those two approaches have been proved to be equivalent[81].

In the paper by M. Rota et al[73], Euclidean Distance Formula, together with Youden J-statistics and concordance probability have been compared for their performance on cut-off value selection in the case of censored failure time outcome under direct estimation by marginal weighting setting. Euclidean Distance Formula approach is considered as having the best performance, as its estimation of the cutoff values in simulated datasets are less biased. Between Youden J-statistics and the concordance probability, X. Liu[72] suggested the latter, as the area representing Cc rectangle is always covered by the area under ROC curve (AUC), the estimation of cutoff values by concordance probability is always smaller than or equal to that by J index. When the two cutoff estimations are the same, the variance by Youden J-statistics may be much larger than that of concordance probability. Nevertheless, all those three approaches are being used in Dublin data for cut-off selection.

3.3.6. Optimal cutoff values for Dublin BPV indices

Considering previous analyses results, BPV indices that seem to have an association with CV mortalities are diastolic BPV estimated in 24h and daytime (24h/Day diastolic SD, wSD, ARV and Cov), as they were previously observed to have a significant impact on risk of CV mortality. Thus only for those BPV indices cutoff values are to be estimated. In Dublin data, the median follow-up duration was 5.69 years, considering

the clinic convenience, here in this report, the meaningful time point τ has been considered as 5 years, i.e. BPV thresholds estimated for the risk of 5-year CV mortality. The best cutoff values are also been estimated for 10-year CV mortality, which shown to be very similar to those for 5 years and the results are not discussed in this thesis. The threshold values are being estimated under the setting of accumulative definition of cases/dynamic definition of controls, as it is more suitable when scientific interest lies in discriminating between subjects who die prior to a given time *t* and those that survive beyond.

Figure 4. ROC curve of the selected BPV indices on 5-Year CV mortality

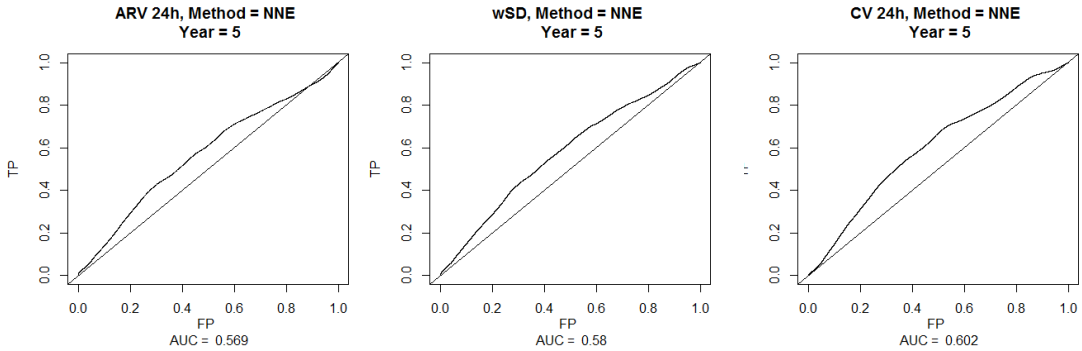


Table 11. Estimated cutoff values of diastolic BPV and ROC related statistics

BPV	Methods	Cutoff value	P(BPV>c)	SE	SP	Youden	Cc	Euclidean
24hARV	Youden	9.52	20.57%	0.32	0.81	0.13	0.26	0.70
	Euclidean	8.18	42.73%	0.53	0.59	0.12	0.31	0.62
wSD	Youden	9.66	30.93%	0.44	0.71	0.15	0.32	0.63
	Euclidean	8.97	41.60%	0.54	0.60	0.14	0.32	0.61
CoV 24h	Youden	10.98	42.19%	0.59	0.60	0.19	0.35	0.57
	Euclidean	10.98	42.19%	0.59	0.60	0.19	0.35	0.57
Day ARV	Youden	9.57	23.58%	0.35	0.78	0.13	0.27	0.69
	Euclidean	8.52	37.11%	0.47	0.64	0.11	0.30	0.64
Day SD	Youden	9.40	38.68%	0.53	0.63	0.16	0.33	0.60
	Euclidean	9.32	39.90%	0.54	0.62	0.16	0.33	0.60
CoV Day	Youden	11.64	30.91%	0.44	0.71	0.15	0.31	0.63
	Euclidean	10.81	39.58%	0.52	0.62	0.14	0.32	0.61

Note: P(BPV>c) are the percentage of patients whose BPV values were above the estimated cutoff values. Cc: Concordance probability.

ROC curve of the selected BPV indices on 5-Year CV mortality are in Figure 4. Only BPV calculated for 24h are presented as examples. The AUC for day BPV are similar, all around 0.60. Tables 11 shows the results of the estimated cutoff values in different BPV indices, their SE, SP and ROC curve related statistics. Since the cutoff values proposed by concordance probability function are all equal to those proposed by Euclidean distance, only results of Euclidean distance are shown, which are always

smaller than those proposed by Youden J statistics, except for the CoV24h. This could be explained by the above mentioned paper by X. Liu[72] that the estimation of cutoff values by concordance probability is always smaller than or equal to that by J index. In fact, in Dublin study, the Youden J statistic classifies only patients with top 20-42% BPV values as high risk group, while Euclidean distance takes a much higher percentage of patients (37-42%) as having higher risk of CV mortalities. Both Youden J statistic and Euclidean distance tends to choose higher specificity over sensitivity, although this tendency is less obvious for the Euclidean distance estimation.

Cutoff values proposed by Euclidean distance and Youden J statistics are used in Cox PH full model and Weibull full model to estimate HR and Survival between subjects whose BPV were below/above the cutoff values (Table 12). HRs are bigger than those estimated per one standard deviation increase, the 95% CI are also wider, the survival time estimated by Weibull full model is in agreement.

Table 12. The classification matrix of observed and estimated outcome in 5 years

BPV	Cutoff	Group*	Observed outcomes			Estimated outcomes^		Cox HR	Weibull Survival
			Alive	Death	Censored	Death	Alive		
24hARV	Youden	0	4088	183	4063	274.2	8059.8		
		1	1428	111	619	136.7	2021.3	1.42(1.10-1.82)	0.68(0.49-0.95)
	Euclidean	0	2826	115	3068	174.7	5834.3		
		1	2690	179	1614	236.4	4246.6	1.45(1.13-1.85)	0.69(0.49-0.95)
wSD	Youden	0	3494	146	3607	222.3	7024.7		
		1	2022	148	1075	188.7	3056.3	1.52(1.20-1.94)	0.61(0.44-0.85)
	Euclidean	0	2876	120	3131	181.8	5945.2		
		1	2640	174	1551	229.0	4136.0	1.40(1.09-1.78)	0.70(0.51-0.98)
CoV 24h	Youden &	0	6065	3081	117	171.9	5893.1		
	Euclidean	1	4427	2435	177	245.8	4181.2	1.39(1.16-1.68)	0.65(0.47-0.90)
Day ARV	Youden	0	3883	169	3966	254.2	7763.8		
		1	1633	125	716	155.1	2318.9	1.59(1.25-2.03)	0.55(0.39-0.76)
	Euclidean	0	3077	132	3389	202.6	6395.4		
		1	2439	162	1293	208.4	3685.6	1.48(1.16-1.88)	0.60(0.44-0.83)
Day SD	Youden	0	6434	3022	118	180.8	6253.2		
		1	4058	2494	176	228.4	3829.6	1.60(1.25-2.04)	0.55(0.40-0.77)
	Euclidean	0	6306	2946	115	175.8	6130.2		
		1	4186	2570	179	233.0	3953.0	1.69(1.24-2.03)	0.56(0.40-0.77)
CoV Day	Youden	0	7249	3657	140	165.2	7083.8		
		1	3243	1859	154	249.1	2993.9	1.48(1.17-1.88)	0.61(0.44-0.84)
	Euclidean	0	6339	3177	112	165.2	6173.8		
		1	4153	2339	182	249.1	3903.9	1.60(1.26-2.05)	0.55(0.40-0.77)

Note: *Group=0 contains subjects whose BPV were below the cutoff values; Group=1 contains subjects whose BPV were above the cutoff values. ^Number of estimated deaths are calculated by conditional weighting, HR are estimated using the full Cox model and survival time using full Weibull model.

In the search of the best-possible cutoff that can best separate subjects into higher and lower risk of CV mortality, the Euclidean distance and concordance probability yield almost identical estimation of cutoff values for each BPV indices, however the Youden index provides different estimation that is bigger in values.

Nevertheless, the KM survival curves between groups classified by Youden index and Euclidean distance show very similar trends, as shown by ARV 24h as an example in Figure 5 (Left). Youden index is preferred as it identifies higher cutoff values, and thus separate only subjects with very high BPV values, who carries higher risk of mortality from the rest of the group. Once the selection method is been decided, we can round the cutoff values into integer numbers as a suggestion for the possible clinical practice. KM survival curves between 24h BPV subgroups classified by Youden cutoff value are presented in Figure 5(Right). The 5-year survival was about 0.97-0.98 and 0.94-0.95, respectively for the low/high risk groups, classified by all the four diastolic BPV indices.

Figure 5. KM curves of 5-year survival of ARV thresholds estimated by Youden J statistics and Euclidean methods (Left) and of all 24h BPV index thresholds by Youden J statistics (Right)

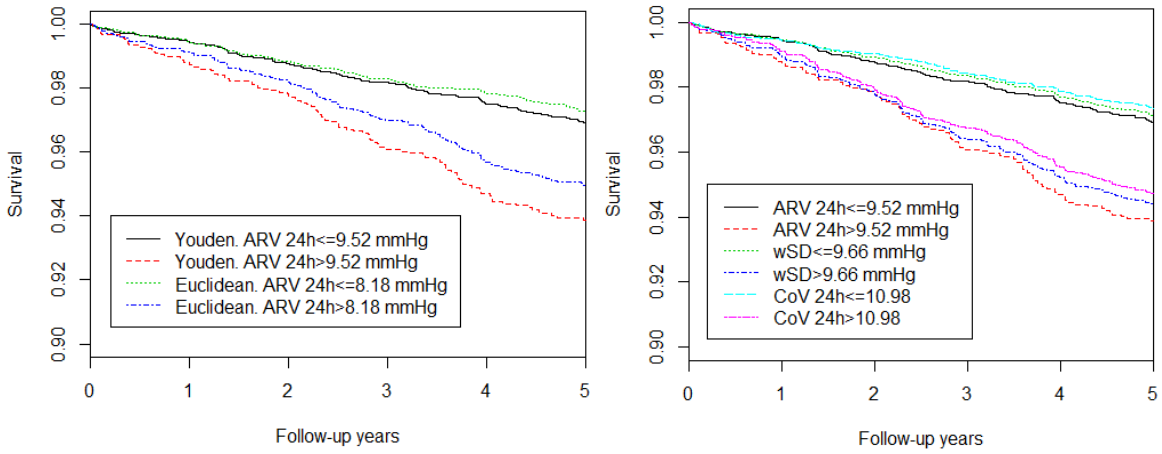
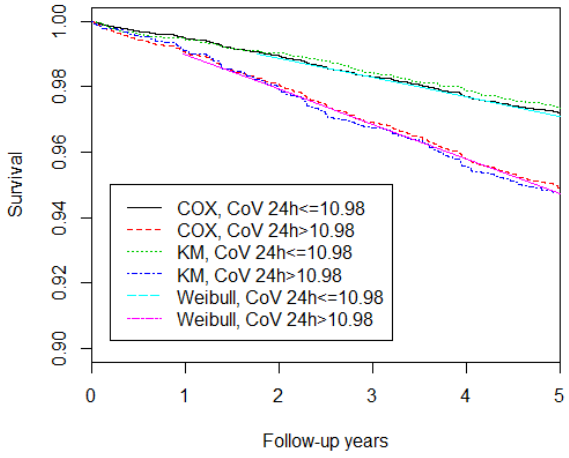


Figure 6.Survival curve by KM, Cox and Weibull in subjects by CoV cutoff groups



Youden J statistics in this report is considered as the best candidate, due to the fact that

The estimated survival curves by the PH Cox model and Weibull using CoV 24h cutoff value as single predictor together with KM curve are presented as an example in Figure 6. The survival curves are tightly clustered following the same trend, indicating a nice fit of those two models, at least when BPV considered as a single predictor.

Effect of diastolic BPV on CV mortality is weak, borderline, and tightly associated with the other covariates including mean BP levels, setting up the Euclidean cutoff values that label almost half the population as high risk may dramatize the effect of BPV on CV mortality; on the other side, Youden selection, which has better specificity, is more concentrated on subjects whose BPV was on the top of the pyramid, who might be really bearing an increased risk of CV mortality.

The above selection of BPV cutoff is based on the conditional survival probability estimated using KM methods and a single predictor, i.e., the BPV index. Considering the fact the subjects with elevated BP levels have also high risk of CV death, regardless of the BPV levels, it would be important to know if the same cutoff values could apply for patients of all BP levels. Taking advantage of the large sample size of Dublin study, the total of 10492 subjects were then divided into four subgroups according to their BP levels, including Normotensive Group: 1649 normotensives, SysHT Group: 1435 systolic-only hypertensives, DiaHT Group: 682 diastolic-only hypertensives, and Sys&Dia HT Group: 6726 hypertensives for both systolic and diastolic pressure. Cox models (full model with a term G indicting the subgroups replacing mean BP) are performed to test if the effect of these preselected BPV indices, 24h ARV, wSD, 24h CoV, Day ARV, Day SD, Day CoV, all diastolic, are similar between the subgroups. AUCs are also estimated in those 4 subgroups.

All the interaction terms of G with BPV as continuous variable or as categorical are not significant, indicating that the BPV effect is not found to be different in the subgroups. HRs estimated by full Cox model between Youden J estimated cutoff groups are not significant for most of the subgroups except for sys&dia HT group (Table 13), however, this could be caused by the greater sample size of this subgroup, as compared with the others. Furthermore, the absolute value of AUC are all small (only a bit better than the random toss of coins) and differences between AUCs from those subgroups are all small.

Table 13. Estimated AUC and HR of diastolic BPV indices in the four BP groups

		24h ARV	wSD	24h Cov	Day ARV	Day SD	Day Cov
Normotensive	AUC	0.49	0.51	0.54	0.47	0.48	0.60
n=1649	HR	0.82[0.24-2.78]	1.73[0.78-3.84]	1.90[0.90-4.01]	1.58[0.67- 3.73]	1.32[0.61-2.83]	2.39[1.17-4.87]
SysHT	AUC	0.49	0.52	0.57	0.50	0.52	0.57
n=1435	HR	1.69[0.85-3.34]	1.35[0.68-2.68]	1.28[0.67-2.46]	1.80[0.89- 3.61]	2.65[1.39-5.06]	1.46[0.79-2.72]
DiaHT	AUC	0.61	0.65	0.68	0.66	0.61	0.69
N=682	HR	0.67[0.12-3.80]	0.98[0.16-6.07]	2.91[0.52-16.4]	1.17[0.22-6.35]	1.28[0.29-5.69]	1.84[0.41-8.25]
Sys&Dia HT	AUC	0.60	0.60	0.60	0.59	0.61	0.63
N=6726	HR	1.42[1.07-1.89]	1.60[1.20-2.12]	1.45[1.09-1.93]	1.61[1.21-2.13]	1.59[1.18-2.14]	1.45[1.09-1.93]

3.3.7. Future research possibilities for BPV cutoff selection

Previous analysis indicates a strong increase of CV mortality risk per one standard deviation increase in the selected diastolic BPV indices (Table 9, of Cox/AFT), as estimated by both Cox PH model and AFT models, however, the predictive ability in BPV is weak, as seen by the AUC example in Figure 4, and both Youden and Euclidean estimation of SE and SP are low; this could be attributed, at least partially, by the study limitations, which will be discussed in the discussion section, and by the low incidence of CV mortality (≈3% in 5 years), which could be related to both high false negative proportion and high false positive proportion.

Multiple cut-offs

In future study, we could consider further not just separating subjects into two groups but into multiple groups, like that of the mean BP levels (normotensive, prehypertension, Stage 1 and Stage 2 hypertension).

The most straightforward approach is a data-based classification of BPV into tertiles or quartiles and the compared the HR. Bersabé R, Rivas T have derived a general equation to compute multiple cut-offs on a single predictor in order to classify individuals into more than two ordinal categories[82]. The equation is derived from the multinomial logistic regression model, which is an extension of the binary logistic regression model to accommodate polytomous outcome variables. From this analytical procedure, cut-off values are established at the predictor values at which an individual is as likely to be in category j as in category j+1 of an ordinal outcome variable: higher or lower than these value means the increased probability of failure or mortality. Paper applied this approach has been published[83]. However methodological studies are still required to evaluate the pro and cons of this approach and to extend it to survival data when censoring and time to event to be considered. Furthermore, this approach does not apply to the survival data which contains only binary outcome.

A possibility that one might consider, based on the previous analysis result that cutoff values estimated by Youden J statistic (c_J) and Euclidean distance (c_E) are very different, to divide subjects into three subsets by both c_J and c_E . i.e. subjects from the full dataset are divided into 3 risk groups, with lower risk group having BPV<= c_E , media risk group with c_E <BPV<= c_J and high risk group with BPV> c_J . This approach could be limited by sample size, as the high risk group composes only a small proportion of the whole sample and then has a higher hazard, this could cause small effective sample size, i.e. number at risk, at study end or even during the follow up period, and also small number of failure, particularly when the incidence is low. The limitation of sample size can also occur to the median risk group, if c_J and c_E .are close.

Prediction based on multiple risk factors

Another research possibility would be the combination of multiple predictors, for

example, in Dublin study, a comprehensive predictor can combine high/median/low BP levels and high/(median/)lower BPV levels, or even more with age and gender considered. The estimation of BPV indices within BP subgroups could also be considered as a combination, as in such case the BP level is considered, although only very roughly. The combination of different BPV indices that reflect different perspectives of the BP variability can also be considered. The concept of multiROC R package has been proposed previously[84], where a diagnostic rule is created from multiple tests, and the threshold of an individual component contained in the rule is varied to create the curve. The rule will typically consist of a Boolean expression (i.e. containing 'and', 'or') of different tests related by algebraic operators as the components of the expression. Each component of the rule is fixed at a diagnostic threshold except for one, which varies over all of its possible values and the corresponding SE and SP are plotted. ROC has been extended to the case when that score is a linear combination of several factors, usually the sum of predictors time their coefficients estimated individually from a logistic regression model[85,86]. A similar method has also been applied in survival study[87], significant risk predictors have been selected from potential risk factors using Cox PH model, and then a risk score is calculated as a sum all the significant predictors time their coefficients ($score = \beta_1 x_1 + \beta_2 x_2 + \cdots$). The risk score is then used to divide subjects into groups (3 or more). However, in the condition of combined risk factors, attention should be paid for the possible correlation among them.

3.3.8. Model fitness in survival data

The performance of predictive survival models can be assessed from two perspectives: discrimination and calibration. Discrimination quantifies the ability of the model to correctly classify subjects into one of existing categories (for instance, event and non-event). Calibration describes how closely the predicted probabilities agree numerically with the actual outcomes. Good discrimination of a model does not automatically imply good calibration or vice versa. If a choice is to be made as to which one should receive the primary focus, Harrell[88] suggested the good discrimination if preferred. This is motivated by the fact that recalibration is always possible which is not true for discrimination.

Characteristics of some performance measures are summarized by Steyerberg et al[89] and DA. Harrison et al have overviewed a variety of measures of model performance[90]. As regard to the predictive performance of survival models, there are traditional measures include:

1. Brier score[91] to indicate overall model performance (calibration)
2. R^2 statistics and Generalized R^2[26] (discrimination).
3. Overall *c* index, or concordance statistics introduced by Harrell[88] as a natural

extension of the ROC curve area to survival analysis[92–94], viewed through the Mann–Whitney statistic. (discrimination)

4. Goodness-of-fit statistics[95] (calibration).
5. Likelihood ratio test.

When the study interest is to assess the impact of a new predictor on clinic outcomes, the interest lies on how much the predictive power has been increased by adding this new predictor into a model which includes previous identified predictors. In such regard, methods listed above can also be used, while net reclassification improvement (NRI), and integrated discrimination improvement (IDI), being newly proposed, and net benefit etc. only applies to compare the predictive power of a new model or method with the old one.

AUC

The improvement in AUC for a model containing a new marker is defined simply as the difference in AUCs calculated using a model with the marker of interest and another one without. While it has been used for quantifying improvements over the last few decades, several studies have analyzed the limitations of this metric including lack of clinical relevance and difficulty in interpretation of small magnitude changes[96,97]. This limitation can be best seen in the example of high density lipoprotein (HDL) and Framingham Risk Score (FRS). When models with and without HDL were analyzed with AUC regarding effect of HDL of modifying FRS, HDL was found not to have a statistical significant effect. However, when analyzed in terms of outcomes, HDL was found to be a significant predictor of heart disease and thus should affect FRS[98]. This is also the case of BPV as a predictor for the CV mortality in this report. Also, This improvement of AUC between new and old models is often very small in magnitude; for example, Wang et al. showed that the addition of a biomarker score to a set of standard risk factors predicting CVD increases the model AUC only from 0.76 to 0.77[87]. MS Pepe et al have shown simple examples in which enormous odds ratios are required to meaningfully increase the AUC[99].

NRI

To assess and quantify the predictive ability of a new risk factor, a new model that includes the new risk factor can be constructed and then compared with the old model that contains the same covariates but this new risk factor. Pencina et al in 2008[100] has proposed net reclassification index (NRI) that attempts to quantify how well a new model reclassifies subjects as compared to an old model, typically this comparison is between an original model (e.g. hip fractures as a function age and sex) and a new model which is the original model plus one additional component (e.g. hip fractures as a function of age, sex, and weight). Consider an outcome that is a binary or ordinal and define upward movement (up) as a change of event in subject into higher category based on the new model and downward movement (down) as a change in the opposite

direction. If D denotes the event indicator, the estimators for the four probabilities comprising the NRI are:

$$\hat{P}(\text{up}|D=1) = \hat{p}_{\text{up,events}} = \frac{\#\text{ events moving up}}{\#\text{ events}}$$

$$\hat{P}(\text{down}|D=1) = \hat{p}_{\text{down,events}} = \frac{\#\text{ events moving down}}{\#\text{ events}}$$

$$\hat{P}(\text{up}|D=0) = \hat{p}_{\text{up,nonevents}} = \frac{\#\text{ nonevents moving up}}{\#\text{ nonevents}}$$

$$\hat{P}(\text{down}|D=0) = \hat{p}_{\text{down,nonevents}} = \frac{\#\text{nonevents moving down}}{\#\text{ nonevents}}$$

The NRI can be calculated as:

$$\text{NRI} = [\text{P(up|D=1)} - \text{P(down|D=1)}] - [\text{P(up|D=0)} - \text{P(down|D=0)}]$$

Calculation of NRI is category-based and easily influenced by the relationship between category cut-offs and event rate[101]. Thus a category free approach has been proposed, and been further modified and extended to survival analysis, with all probabilities estimated using the Kaplan-Meier approach[101]. Assuming that out of *n* individuals, n_U are reclassified upwards and n_D downwards:

$$NRI = \frac{P(event|up) \cdot n_U - P(event|down) \cdot n_D}{n \cdot P(event)} + \frac{(1 - P(event|down)) \cdot n_D - (1 - P(event|up)) \cdot n_U}{n \cdot (1 - P(event))} \quad (5)$$

NRI calculated in all the study subjects is equal to the sum of NRI within event group and NRI within non-event group. Unlike the previous formula, this formula does not depend on the number or even existence of risk categories as it assumes probabilities of event among those reclassified upwards or downwards that would be obtained pooling all individuals with the same reclassification. The authors have suggested two version of NRI, one with category that should be used if categories are already established in the field and one without categories that can be used universally (recommended). Furthermore, the category-free NRI has been commented by the authors that it is not affected by event incidence and thus can be compared across different studies.

Pencina et al[100] have also defined a new concept intergrated discrimination improvement (IDI), $IDI = \left(\hat{P}_{new} - \hat{P}_{old}\right) - \left(\hat{P}'_{new} - \hat{P}'_{old}\right)$where $\hat{P}$ is the mean of the new/old model based predicted probabilities of an event for those who develop events,$\hat{P}'$ is the new/old model based predicted probabilities of an event for those who don't develop events.

Chambless et al [42] have extended NRI into survival data following paper by adding the parameter *t* into the formula, and $\text{IDI(t)} = R^2_{new}(t) - R^2_{old}(t)$, where the R^2 is the proportion of variance explained by the model.

However, although the use of NRI and IDI has become increasingly popular, many articles have criticized those approaches, claiming that their use is not always safe [102,103]. NRI is considered as not adequately account for clinically important differences in shifts among risk categories, and the category-free NRI can mislead investigators by overstating the incremental value of a biomarker, even in independent validation data. When investigators want to test a null hypothesis of no prediction increment, the well-established tests for coefficients in the regression model are superior to the net reclassification index[86]. Hilden et al have shown that even the best probabilistic model can be improved on according to IDI and NRI, a model where the real marker was added was outperformed in terms of IDI and NRI by prediction models that added a random noise variable. If IDI and NRI are used to measure gain in prediction performance, then poorly calibrated models may appear advantageous, a poor prognostician can outperform a good prognostician [102], this means that spurious results may arise, and overconfident risk predictions appear advantageous.

Generalized R^2

Measures of explained variation, such as the coefficient of determination (R^2) in linear models, are helpful in assessing the explanatory power of a model. In survival analysis, these measures help quantify the ability of prognostic factors to predict a patient's time until death. In the censored data setting, the definition of such a measure is not straightforward; P.D. Allison has discussed a "generalized" R^2 statistic that is based on the likelihood-ratio statistic (LRT) for testing the global null hypothesis[104]. It is calculated as $1–exp[-(LRT/n)]$, where $LRT=(-2logL_0)–(-2logL_p)$, n is the number of sample size, $logL_0$ is the log-likelihood for the null model without covariates, $logL_1$ is the log-likelihood for the fitted model with *p* covariates, both logL value are given in Model Fit Statistics and the LRT test given in Global Tests table by PHREG procedure. Note that this generalized R^2 does not have a "proportion of variation explained by the model" interpretation. Allison states, ‘R-square does not tell you anything about how appropriate the model is for the data’ and ‘It's just a statistic between 0 and 1 that is larger when the covariates are more strongly associated with the dependent variable’.

Although the generalized R^2 is commonly recommended for the Cox model, its sensitivity to the proportion of censored values is not often mentioned. In fact, the expected value of R^2 decreases substantially as a function of the percent censored, with early censoring having a greater impact than later censoring. Simulations show that complete data R^2 values from the Cox model are very close to those from a similar linear model. However, average R^2 values can decrease by 20% or more (e.g., R^2 from 0.5 to 0.4) with heavy censoring (e.g., 50% censoring) compared to complete data[105].

Asayama et al[26] have used generalized R^2 to evaluate the additional risk explained in Cox regression by adding BPV to models already including the mean systolic level and covariates. The formula is similar to that proposed by PD Allison:

$$R^2=1-\exp\left\{\frac{-2}{n}[\ln L(X_2)-\ln L(X_1)]\right\}=1-\exp\left\{\frac{-\chi^2}{n}\right\}$$

where n is the number of participants, ln L(x_2) and ln L(x_2) are the log likelihood statistics of the full model and the basic model, respectively, and χ^2 is the likelihood ratio chi-square.

Brier score

Brier score was firstly proposed by GW. Brier in 1950[106], it measures the mean squared difference between the predicted probability assigned to the possible outcomes for subject *i* and the actual outcome. Therefore, the lower the Brier score is for a set of predictions, the better the predictions are calibrated. Note that the Brier score, in its most common formulation, takes on a value between 0 and 1, since this is the largest possible difference between a predicted probability (which must be between 0 and 1) and the actual outcome (which can take on values of only 0 and 1). In the original (1950) formulation of the Brier score, the range is double, from 0 to 2. The Brier score is appropriate for binary and categorical outcomes that can be structured as true or false, but is inappropriate for ordinal variables which can take on three or more values (this is because the Brier score assumes that all possible outcomes are equivalently "distant" from one another).

$$BS=\frac{1}{N}\sum_{t=1}^{N}(E_t-O_t)^2$$

E_t is the probability that was expected by the model while O_t is the actual observed outcome, N is the sample size.

The Brier score can be decomposed into 3 additive components: Uncertainty (the variance of the actual overall probability, or in another word, the actual probability changes), Reliability (how close the predicted probability change with the actual probability, in subset or overall), and Resolution (how close the subsets of actual probability change with the actual overall probability)[107]

Some of those methods will be applied in Dublin study to assess the model fitness and the additional effect by including also BPV indices into the model.

3.3.9. Assessing model fitness in Dublin study

The full Cox model excluding BPV is considered as the old model, to be compared with the new model, which is exactly the full Cox model, using AUC and Brier score.

Brier score calculated using Weibull model is very similar to those obtained in Cox model. Although NRI and IDI are much criticized for being prone to a false significant increase in predictive model, the values are all very small (around 0.003) with P value

bigger than 0.05. It could be considered thus that although BPV seem to have significant influence on the CV mortality, the incremental information added by introducing this BPV indices into the models is very small. On the other hand, “the fact that markers with very large effect sizes is needed to increase the AUC does not automatically imply that small increases are sufficient and likely to translate into meaningful gains in clinical performance”[108]. In fact, although many above mentioned statistics such as AUC have been accused of being insensitive to new information, the possible truth that many people are beginning to realize is that the incremental information offered by a new marker is very often disappointingly small, particularly when it is added to an already well-calibrated clinical data[102]. In any case, as both hazard ratio and AUC are purely statistical measurements, it would be difficult to decide which one is more informative when the results seem to be inconsistent. Decision analysis including costs and benefits are being proposed[109], however, how to evaluate the costs and benefits are too complicated and many remain unknown.

Table 14 Model fitness

Methods		BPV as continuous var			BPV as categorical var below/above cutoff value, for 5-year survival		
		old model	Full model	Delta	Old model	Full model	delta
AUC	CV 24h	0.8178	0.8204	0.0026	0.8227	0.8270	0.0043
	ARV 24h	0.8178	0.8197	0.0019	0.8227	0.8251	0.0024
	wSD	0.8178	0.8206	0.0028	0.8227	0.8265	0.0038
	Day CV	0.8230	0.8272	0.0042	0.8180	0.8261	0.0081
	ARV Day	0.8230	0.8270	0.0039	0.8180	0.8266	0.0086
	SD day	0.8230	0.8273	0.0043	0.8180	0.8272	0.0092
Brier score	CV 24h	0.8833	0.8838	0.0005	0.9265	0.9268	0.0003
	ARV 24h	0.8833	0.8839	0.0006	0.9265	0.9268	0.0003
	wSD	0.8833	0.8839	0.0006	0.9265	0.9268	0.0003
	Day CV	0.8833	0.8837	0.0004	0.9265	0.9269	0.0004
	ARV Day	0.8833	0.8837	0.0004	0.9265	0.9270	0.0005
	SD day	0.8833	0.8838	0.0005	0.9265	0.9270	0.0005
Likelihood Ratio test		0.001 for all except 0.00 for wSD			0.00 for all except 0.001 for cv day		

4. Discussion on Dublin Study

The main focus of this thesis is to understand the effect of BPV on the risk of CV mortality. A thorough literature search has been conducted on the proposed indices that can quantify BPV, and on previously published papers on the effect of BPV on CV outcomes. Despite of the great effort from researchers in recent decades on the mechanism, classification, measurement and evaluation of BPV, and after all, the possible effect that it might impose on the cardiovascular system, no clear or convincing answer has been established to this question. This discrepancy could be partially caused by the complexity in the BPV mechanism and the different research designs that has been discussed previously.

A detailed summary of classification and calculation of previous proposed BPV indices has been presented, among which three widely accepted and most representative ones (average real variability, coefficient of variation and (weighted) standard deviation) are chosen to quantify the BPV from Dublin study. 24h SD or CoV are not used because they include also the day-night fall as a major component, which is considered as a protective factor for CV system. 10492 subjects with sufficient ABPM measurements were included into data analysis, who had similar age, BMI and percentage of female, smoking, diabetes, previous CV diseases as compared with the rest of the subjects. The median follow-up was 5.69 years, and 502 died of CV mortality.

Semiparametric PH Cox model and parametric Weibull model have been used in this report for the estimation of hazard ratio or survival time following one standard deviation increase in the BPV indices. In a fully adjusted model where the corresponding mean SBP, DBP and other covariates are being adjusted, only diastolic BPV measured in 24h and daytime (i.e. 24h ARV, 24h wSD, 24h CoV, Day ARV, Day SD, Day CoV) are found to have significant impact on the hazard and survival time. With an increase of one standard deviation, the risk of CV mortality increased by about 11-14%, in correspondence, the expected survival time decrease by about 9-11%.

This study has confirmed that the 24h and daytime BPV is an independent CV mortality predictor, although diastolic BPV seem to be more valuable in the prediction than systolic BPV. Indeed, this is different from many other studies but it is not the only one which reported diastolic predictive power. The reason for the diastolic superiority over systolic is not clear, and not discussion will be addressed in this regard. Furthermore, not all the diastolic BPV are valuable, i.e. only the 24h and day time BPV are proven to be independent predictors. Variables that could introduce significant impact on the probability of CV mortality include age, gender (favoring females), BMI, previous CV diseases, nocturnal fall and smoking. Diabetes has no predictive effect.

BPV prognostic effect is not completely homogeneous across the population, as it is an effective predictors for subjects who were diastolic hypertensive patients or both

systolic and diastolic hypertensive patients (by AUC or the Cox PH model), however, this could also be caused by different sample sizes in HT subgroups. This result is in line with the data of Hansen[51].

A binary threshold for each effective BPV indices is also proposed for the 5-year survival, under the cumulative case and dynamic control definition and inverse weighting. The AUC is about 0.6 for all, when a single BPV is considered as the only predictor. The best threshold values estimated by Youden index are 9.52 mmHg for 24h ARV, 9.66 mmHg 24h wSD, 10.98 for 24h CoV, 9.57 mmHg for Day ARV, 9.40 mmHg for Day SD and 11.64 for Day CoV. Patients being considered by those threshold as having higher BPV (high CV risk as well) made up about 20-42% of all the subjects. Non-parametric Kaplan Meier method together with Cox and Weibull models are used to predict the survival curve. The selected thresholds are better to be rounded to integer if applied in clinical practice.

The outcome oriented selection of cutoff values based on extended ROC has its strength as it best combines sensitivity and specificity, instead of an arbitrary choice of the 75% percentile values or so on. The selection is based on BPV values only, without considering other covariates, for example, mean BP levels, however, the high BPV has its prognostic values only in diastolic hypertensive or both systolic and diastolic hypertensive patients. This could be important in clinical practice.

Based on the results of this report, we could see that the BPV and BP are closely associated, for example, BPV can be calculated directly using the mean (CoV). Although the regression coefficients of mean BP on BPV from the multiple regression (Table 4) are very small, and the HRs with BPV increase are also significant, the added predictive effect of BPV on CV mortality is very small, and it is very difficult to be separated from the effect of mean BP.

This study has the following limitations:

1. ABPM was recorded in an interval of every 30 minutes, which is considered as an under-sampling that may have led to inaccuracy in estimating BPV[110]. However, this may not lead to any systematic error and, if any, it may have led to an under-estimation of BPV impact on the CV mortality.

2. ABPM cannot be standardized by individual activities during the measurement, which is considered as an important limitation, as a healthy, active person may have had bigger BPV than who stays in bed for the whole day. This could create a heavy background noise when estimating the effect of BPV on certain outcomes.

3. Survival analysis is often under the assumption of non-informative censoring, e.g. censoring is independent of failure time. In Dublin study, we can assume the non-informative censoring, since we only have administrative censoring at the end of follow-up. However, we do not know if all the censored patients were actually all alive by the study end, as national computerized register of deaths offers a coverage

only within Ireland. We cannot exclude the possibility that subjects emigrated from Ireland after study recruitment to study end, and this piece of information was not made available to us. Thus they can be considered to be in the set at risk for a longer time than they actually were. This is a limitation of the study and can results in an underestimation of mortality and overestimation of survival.

4. Lack of information on antihypertensive treatment or any other treatment that may affect cardiovascular system during the follow-up period. Ambulatory BP readings were all obtained in the untreated condition; therefore, the study result cannot be extrapolated to prognostic importance of BPV under antihypertensive treatment.

5. Lack of information on cholesterol level or other risk factors at the baseline and during the follow-up.

6. For the estimation of the cut-off values, only the binary categories of BPV indices have been estimated, without taking into account of the other covariates which may also have impact on cardiovascular mortalities. Strictly speaking, this may not be considered as a limitation but rather a choice, just like the threshold of BP does not depend on age or other factors but only on itself. In fact, exploration of the interaction between BP level and BPV level has been performed and not evidence of estimating a cutoff based on the level of mean BP was found.

In conclusion, the aim of this report is to explore the added predictive effect of BPV on CV mortality over and above the effect of mean BP levels. To this aim, semiparametric Cox PH model and Parametric Weibull models are applied. 24h and daytime diastolic BPV are found to be predictive for CV mortality after being adjusted for other possible covariates, including the corresponding mean systolic and diastolic BP levels. The best cutoff values for the considered BPV estimates have been selected for 5-year survival. However, although an increased BPV does independently predict the risk of cardiovascular mortality, the effect is marginal and tightly associated with that of elevated mean BP and other covariates, it does not add more predictive power to the cardiovascular mortality.

Part II. Blood Pressure Control in a Randomized Clinical Trial- Comparing Conventional with Impedance Cardiography Based Strategies

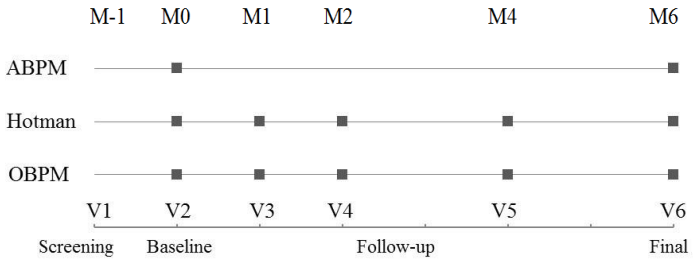
A prospective randomized clinical trial was conducted to assess the possibility that a hypertension management strategy based on hemodynamic assessment of patients through impedance cardiography might lead to a better hypertension control over 24 hours than a conventional approach only based on blood pressure measurement during clinic visits.

1. BEATUY Study

BEAUTY (BEtter control of blood pressure in hypertensive pAtients monitored Using the hoTman® sYstem) study is a multicenter prospective randomized parallel group study, testing if the HOTMAN® System, a novel impedance cardiographic device, together with a predefined algorithm of drug selection, may help the physician to better control blood pressure (BP) in patients with uncontrolled resistant hypertension. Uncontrolled resistant hypertension was defined as having office systolic (S) BP>140 mmHg and/or day SBP > 135 mmHg by Ambulatory BP monitoring (ABPM) while taking at least 2 types of drugs. Patients were randomized to Integrated Hemodynamic Management (IHM)-guided drug adjusted treatment (n=83) vs. classical clinically adjusted drug treatment (control, n=84). The BP measuring performed during all the visits are in Figure 1. ABPM was performed at baseline, which was the second visit (V2) and study end at visit 6 (V6), while office (O) BPM and Hotman measurements were performed during every visit (from V2 to V6). In IHM groups, the hemodynamic data were referred to the doctors immediately, together with 2007 ESH Guidelines, to guide antihypertension treatment while in control groups it was sealed and the drug selection followed only 2007 ESH Guidelines.

Daytime SBP by ABPM on treatment was the primary endpoint and the longitudinal office BP data are also considered with great importance for the BP reduction profile during follow-up visits. BP values measured at the time of Hotman parameters collected at baseline and on treatment were also compared as exploratory analysis, due to its intrinsic limitation for BP measuring as a hemodynamic monitor.

Figure 1. Test flow of BEAUTY study by visit



We have evaluated 315 patients with uncontrolled hypertension, among whom 148 patients who did not fulfill inclusion criteria were excluded, the remaining167 patients who had uncontrolled hypertension verified by ambulatory BP measurements were randomized to IHM-based treatment management (n=83) or to conventional treatment management (n=84) in five European Hypertension Excellence centers. At study end, 156 patients, including 79 from control group and 77 from IHM group, on whom ABPM was performed at both V2 and V6 and were included in the analysis. They all had 5 office BP measuring while only 119 patients had at least one Hotman measurements.

Patients were well randomized with demographic data similar between the IHM and

control groups. The baseline information is summarized in Table 1.

Table 1. Demographics of randomized subjects

		Total N	IHM	Control
Sex	Male	102	53 (64)	49 (58)
	Female	65	30 (36)	35 (42)
Age (Ys)			64 (11)	62 (12)
BMI	≤24.9	32	16 (19)	16 (19)
	25-29.9	66	38 (46)	28 (33)
	≥30	69	29 (35)	40 (48)
SBP (mmHg)	Day SBP		150 (12)	150 (12)
	Night SBP		130 (14)	133 (15)
	24h SBP		143 (11)	145 (12)
	Office SBP		157 (20)	156 (15)
	Home SBP		151 (16)	149 (12)

Note: data are number (percentage) for categorical variables and mean (standard deviation) for continuous variables.

2. Introduction of longitudinal data

A longitudinal study refers to an investigation where participant outcomes, possible treatments or exposures are collected at multiple follow-up times. Each longitudinal response reading can be represented as: Y_{ijp}, where i^{th} represents the subjects, $i=1, \ldots, N$; j^{th} represents the occasions, $j=1, \ldots, n$; and p^{th} variables or the predictors. The data matrix can be written as:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \cdots & X_{i1P} \\ X_{i21} & X_{i22} & \cdots & X_{i2P} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{in_i1} & Y_{in_i2} & \cdots & Y_{in_iP} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_P \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i1} \\ \vdots \\ e_{in_i} \end{pmatrix}$$

In a linear model, the probability density function for Y, denoted by $f(y)$, is the classical bell shaped curve, centered at $u_{ij} = X'_{ijp}\beta_p$ with constant variance σ_j^2. e_{ij} is normally distributed at $(0, \sigma_j^2)$, the data can be written as:

$$Y_{ij} = X'_{ijp}\beta_p + e_{ij}$$

In a longitudinal study, exposures are recorded prospectively at multiple follow-up visits in each subject before the key clinical event occurs or within a predefined follow-up period, thus the orders and amounts of exposures can be well recorded, the recall bias can be alleviated. In addition, incident events are recorded and the timing of each event can be correlated with recent changes in patient exposure and/or with previous chronic exposure. Patterns of change in outcomes and/or exposure are observed in individual levels.

The analysis of longitudinal data is challenged by following characteristics:

1. Analysis of correlated data. Statistical analysis of longitudinal data require methods that can properly account for the within-subject correlation of response measurements.

2. Time-varying covariates. Although longitudinal designs offer the opportunity to associate changes in exposure with changes in the outcome, the direction of causality can be complicated by "feedback" between the outcome and the exposure. i.e., a drug influences health, on the other hand, a patient's current health status may influence the drug exposure or dosage in the future.

3. Participant loss to follow-up. Missing data are a common occurrence in a longitudinal study and can have a significant effect on the conclusions that can be drawn from the data. Understanding the reasons why data are missing can help analyzing the remaining data.

a) Missing completely at random (MCAR): when the events that lead to any particular data-item being missing are independent both of observable variables and of unobservable parameters of interest, and occur entirely at random, the analyses performed on the data are unbiased; however, data are rarely MCAR.

b) Missing at random (MAR) is when missing is related to a particular variable in the dataset, but it is not related to the value of the variable that has missing data. An example of this is accidentally omitting an answer on a questionnaire; the observed values are not necessarily a random sample of the responses. Statistical inferences about the mean response could yield valid estimates but are sensitive to any misspecification of the joint distribution of the responses.

c) Missing not at random (MNAR): data that is missing for a specific reason, the value of the variable that's missing is related to the reason it's missing (e.g. certain question on a questionnaire tend to be skipped deliberately by participants with certain characteristics). In situation of MNAR, any assumption made about the missing process are not verifiable

3. Exploring BEAUTY Data

Exploratory analysis of longitudinal data seeks to discover patterns of systematic variation across groups of patients, as well as aspects of random variation that distinguish individual patients, thus longitudinal data have two aspects that require modeling: 1). The mean response over time and 2). The covariance among repeated measures on the same individuals.

3.1. Pre/Post Analysis

Let $X_i = 0$ denotes control group; and $X_i = 1$ an exposure or intervention group, for each subject *i*, a baseline measurement is denoted as Y_{i0} and a follow-up measurement Y_{i1}. Follow-up only analysis considers only Y_{i1} when fits the model while change analysis fits the difference between $Y_{i1} - Y_{i0}$; a better approach is to use analysis of covariance (ANCOVA), setting the grouping and Y_{i0} as covariance[111]. When the selection of treatment or exposure is not randomized, ANCOVA analysis can control for "confounding due to indication", or where the baseline value Y_{i0} is associated with a greater/lesser likelihood of receiving the treatment $X_i = 1$ (Table 2).

Table 2. The different approaches of pre-post analysis in a longitudinal study

Follow-up only:	$Y_{i1} = \beta_0 + \beta_1 X_i + e_i$
Change analysis:	$(Y_{i1} - Y_{i0}) = \beta^*_0 + \beta^*_1 X_i + e^*_i$
ANCOVA:	$Y_{i1} = \beta^{**}_0 + \beta^{**}_1 X_i + \beta^{**}_2 Y_{i0} + e^{**}_i$

Table 3. Comparing the daytime SBP in control and IHM group using the 3 methods

Follow-up only: Y_{i1} =134.5-0.0075IHM				
	Mean	S.E	95%CI	
Control	134.5	1.41	[131.7, 137.3]	
IHM	134.5	1.37	[131.8, 137.2]	
Difference	-0.0075	1.96	[-3.89, 3.87]	*p*=0.75
Change analysis: $Y_{i1} - Y_{i0}$= –15.4+0.38IHM				
	Mean	S.E	95%CI	
Control	-15.4	1.63	[-18.62, -12.15]	
IHM	-15.8	1.69	[-19.12, -12.40]	
Difference in change	0.38	2.34	[-4.25, 5.01]	*P*=0.84
ANCOVA: Y= 95.36+0.26dsbpv2-0.09IHM				
	estimate	S.E	Z value/Wald test	*p*
Intercept	95.36	12.48	58.42	<0.0001
Day SBP at V2(Y_{i0})	0.26	0.08	9.95	<0.0016
IHM	-0.09	1.89	0.00	0.96

When treatment is randomized, Frison and Pocock[112] showed that $\beta_1 = \beta^*_1 = \beta^{**}_1$, each approach can provide a valid estimate of the average causal effect of treatment. However, the most precise estimate of β_1 is obtained using ANCOVA, and that final measurement analysis is more precise than the change analysis when the correlation ρ between baseline and follow-up measurements is less than 0.50. This results from $var(Y_{i1} - Y_{i0}) = 2\sigma^2(1-\rho)$, which is only less than σ^2 when $\rho > 1/2$.

As shown in Table 3, the three approaches showed similar results, as BEAUTY is a randomized study, baseline day SBP was supposed to be equal in both groups. The effect of IHM for day-time SBP at study end is almost neglectable, coefficient=−0.09 by ANCOVA, indicating very similar daytime SBP by study end in both groups.

3.2. Modelling the means

3.2.1 Analyzing response profile

The primary goal of profile analysis is to make inference about the population regression parameters, β. Response profile analysis assumes no specific time trend, instead, the times of measurement are regarded as levels of a discrete factor; it allows arbitrary patterns in the mean response over time and arbitrary patterns in the covariance of the responses. As a result, this method for longitudinal analysis has a certain robustness since the potential risks of bias due to misspecification of the models for the mean and covariance are minimal.

Table 4. Response profiles of the office SBP at baseline and follow-up visits

Effect	Group	visit	Estimate	Standard Error	DF	*t* Value	*Pr* > |*t*|
Intercept			155.09	1.98	154	78.49	<.0001
Group	IHM		3.42	2.81	154	1.22	0.2261
visit		3	-11.00	1.80	154	-6.11	<.0001
visit		4	-17.20	2.14	154	-8.04	<.0001
visit		5	-16.46	2.34	154	-7.03	<.0001
visit		6	-17.19	2.07	154	-8.32	<.0001
Group *visit	IHM	3	-1.73	2.56	154	-0.67	0.5011
Group *visit	IHM	4	0.77	3.04	154	0.25	0.7996
Group *visit	IHM	5	-2.43	3.33	154	-0.73	0.4676
Group*visit	IHM	6	-4.03	2.94	154	-1.37	0.1723

The response profile of office SBP at each follow-up visit in IHM and control groups are shown in Table 4. The effect of visit, e.g. visit 3, -11.0 is the expected average SBP decrease from baseline to visit 3 in control patients. The interaction term, e.g. IHM*visit3, is that IHM has an additional 1.73 decrease of office SBP in IHMs as compared to controls, it is an omnibus comparison of the effect of both time and

grouping. When the profiles are parallel and there is scientific interest in the main effect of time and/or group, the tests of the main effects require the model excluding the group*time interaction. In the case of BEAUTY study, the result with or without interactive term didn't change majorly.

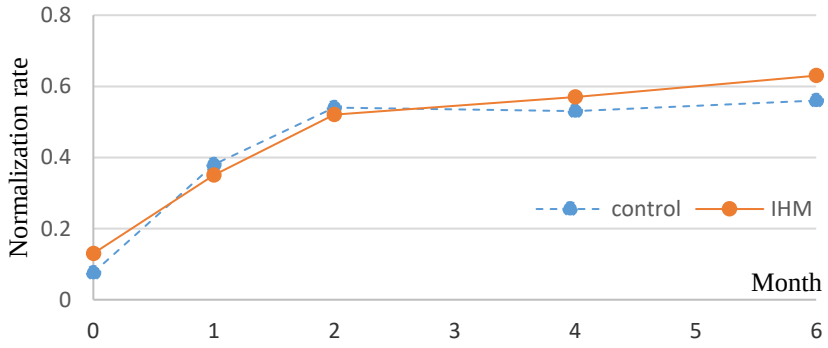
Analyzing response profile is only applicable when all individuals are measured at the same set of occasions and the number of occasions is usually small. However, even in case where the number of repeated measures is relatively small, there are two obvious drawbacks of the analysis of response profile:

1. Test of the null hypothesis of no group*time interaction is a global test and provides only a broad assessment of whether the mean response profile is the same in different groups. The rejection of null hypothesis does not indicate the specific ways in which the mean response profile differs (visits or groups) and thus requires additional analysis.

2. The times of measurement are regarded as levels of a discrete factor, by completely *ignoring the time order of the repeated measurement*, the analysis of response profiles fails to recognize that they can be considered as observations of some continuous, underlying response process over time.

From another perspective, the normalization rate of office SBP (<140 mmHg) at each follow-up visit indicated in Figure 3, can be considered as a snap of response profile. Between the IHM and control group, the figure shows no clear difference in the normalization rate.

Figure 3. Normalization rate of office SBP by visit (<140mmHg)



3.2.2 Parametric curves

When the number of occasions increases and/or when the repeated measures are irregularly timed, analyzing response profile becomes much less appealing, while the fitting of parametric or semi-parametric curves to longitudinal data can be used to describe how the mean response changes over time, which can be a relatively smooth, monotonically increasing or decreasing pattern.

The results of parametric curves in office SBP between the IHM and control groups are

shown in Table 5. It appears that mean office SBP in both groups had significant declined overtime (-3.53mmHg in controls and -4.40mmHg in IHM patients per visit), however, there is no discernible difference between IHM and control groups in the constant rate of SBP decreasing (p=0.1971).

The parametric curves provide a very parsimonious description of the mean response as an explicit function of time, i.e., the time effect is linear, as a result, there is no necessity to require that all individuals in the study have the same set of measurement times, nor even the same number of repeated measurements. It can dramatically reduce the number of model parameters. Fitting a parsimonious model for the mean response has greater power than an analysis of response profile, e.g. Akaike information criterion (AIC) =6324.2 using profile analysis while AIC=5680.9 in parameter analysis in the case of BEAUTY study. The reason for the greater power is that the tests of covariate effect focus only on a relatively narrow range of alternative hypotheses, in this case, the effect of visit and the effect of grouping; while in the response profile analysis, each visit had its effect and interactive terms with grouping. However, when each subject has a variable number of outcomes due to missing data, then mixed model regression methods is considered as a better choice than those two methods.

Table 5. Parametric linear model of the office SBP at baseline and follow-up visits

Effect	Group	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		156.55	2.35	154	66.77	<.0001
Group	IHM	5.23	3.34	154	1.57	0.1194
visit		-3.53	0.47	154	-7.51	<.0001
visit*Group	IHM	-0.87	0.67	154	-1.30	0.1971

3.3. Modelling the Covariance and correlation

Given that the variability of the difference in two observations is $Var(Y_{i2} - Y_{i1}) = Var(Y_{i2}) + Var(Y_{i1}) - 2Cov(Y_{i1}Y_{i2}) = \sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2$ and the variance of two independent observations is: $Var(Y_{i2} - Y_{i1}) = Var(Y_{i2}) + Var(Y_{i1}) = \sigma_1^2 + \sigma_2^2$, thus, providing the correlation among repeated measures in the same subject is positive, the variability of the within-individual difference is always smaller than the variability of the between individual differences.

If we further assume that the variance of the response is constant (over time in the longitudinal design, and across groups in the cross sectional design), with $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then the variance of the within-individual differences is simply $2\sigma^2(1-\rho)$, while the between-individual differences is $2\sigma^2$.

The covariance between two responses at different occasions Y_{ij} and Y_{ik} is denoted by:

$$\sigma_{jk} = E\{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\}$$

$\mu_{ij}=E(Y_{ij})$, $\mu_{ik}=E(Y_{ik})$ are the expectation or mean of Y_{ij} and Y_{ik}. It depends not only on the degree of dependence between them but also on their units of measurements.

Correlation between Y_{ij} and Y_{ik} is denoted by:

$$\rho_{jk} = \frac{E\{(Y_{ij}-\mu_{ij})(Y_{ik}-\mu_{ik})\}}{\sigma_j \sigma_k},$$

where σ_j and σ_k are the standard deviations of Y_{ij} and Y_{ik}. Correlation is a measure of linear dependence that is free of the scales of the measuremeant of Y_{ij} and Y_{ik}, by definition, it takes values between -1 and 1. The two variables can be uncorrelated without being independent since this correlation measures only linear dependence. When $\rho_{jk}>0$ it shows that between-subject variation is greater than within-subject variation. In the extreme $\rho_{jk}=1$ and $Y_{ij}=Y_{ik}$ implying no variation for repeated observations taken on the same subject.

Approaches to model the covariance among repeatedly measured responses include:

1. Unstructured covariance: it allows any arbitrary pattern of covariance among the repeated measures. With *n* repeated measures, the *n* variances at each occasion and the $n \times (n-1)/2$ pairwise covariances are estimated. It is only applicable when all individuals are measured at the same set of occasions. When the occasions are numerous, the number of covariance could be very large and the estimates are likely to be unstable.

Figure 2. Some example structures for modelling the covariance

Unstructured (type=UN)

$$\mathrm{Cov}(Y_i) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2n} \\ \sigma_{31} & \sigma_{22} & \sigma_3^2 & \dots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \dots & \sigma_n^2 \end{pmatrix}$$

compound symmetry (type=CS)

$$\mathrm{Cov}(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

Toeplitz (type=TOEP)

$$\mathrm{Cov}(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{pmatrix}$$

Auto-regression (type=AR(1))

$$\mathrm{Cov}(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

2. Covariance pattern models involves two strategies:

   A. idea borrowed from time series data, which arises from studies with a small number of individuals and a large number of repeated measures (a reversed situation with longitudinal data). This idea is, the correlations decay as the time separation

increases, quite often, the correlation among repeated measures is expressed as an explicit function of the time separation, e.g., auto-regression pattern, banded pattern, exponential pattern.

B. Many models for the variance assume that the variance does not change as a function of time, i.e. compound symmetry, as called as exchangeable model, has constant variance among all the repeated measures data.

When the number of occasions is relatively small and all individuals are assured at the same set of occasions, it may be reasonable to allow the covariance matrix to be arbitrary, with all its elements unconstrained.

4. Modelling longitudinal data

Mixed effects model has its basic premise that there is natural heterogeneity across individuals in the study population, it focuses initially on the regression relationship restricted to observations on a single individual, makes specific assumptions that the variations in observations are attributable to two kinds of variations: within-subjects variation (i.e. blood pressure fluctuation) and between-subjects variation (i.e. gender, randomization arms). The between-subject variation, determining how the average for sub-population differs across distinct values of covariates, forms fixed effects; while the within-subjects variations attributed by the natural heterogeneity are the random effect. When parameters assumptions are made regarding the within- and between-subject components of variation, maximum likelihood method is used to estimate the regression parameters β and the variance components θ (fixed /random), that maximizes the joint probability of the response variables evaluated at the observed values.

In contrast with mixed model, which is often referred to as "subject-specific models", another approach of modelling longitudinal data is "population-average models", the generalized estimating equation (GEE), which combines the generalized linear model with a marginal model. The methods rely solely on assumptions on the mean response over the population and avoids making assumptions on the distribution of the vector of responses. A distinctive feature is that the GEE separately models the mean response (the goal) and the within-subject association among the repeated responses (a nuisance characteristic of the data and must be accounted for to make correct inference about the changes in the population mean response). This separation ensures that the coefficients have interpretation that does not depend on the assumptions made on the within-subject association; specifically the coefficients describe the effects of covariates on the population mean response.

4.1 Linear Mixed model

The primary assumptions underlying a linear mixed model are: 1.the data are normally distributed (Gaussian); 2. The means (expected values) of the data are assumed to be linear in terms of a certain set of parameters.

4.1.1. Fixed and Random effect

In a longitudinal study, due to natural heterogeneity in the population, the deviation between individual observations, Yij , and the individual linear trajectory $E(Y_{ij})=\beta_{i,0} + \beta_{i,1} X_{i,j}$ is the within-subject variation:

$$E(Y_{ij}) = \beta_{i,0} + \beta_{i,1} \cdot X_{ij}$$

$Y_{ij} = \beta_{i,0} + \beta_{i,1} \cdot X_{ij} + \mathcal{E}_{ij}$,where $\mathcal{E}_{ij} \sim N(0,\sigma 2)$

The variables that don't change throughout the duration of the study, e.g. treatment, gender, cause the between-subject variation, which is indicated by variation in the intercepts $var(\beta_{i,0})$ (i.e. Grouping) and in slopes $var(\beta_{i,1})$(i.e. time):

$$\begin{pmatrix}\beta_{i,0}\\ \beta_{i,1}\end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}, \begin{pmatrix}D_{00} & D_{01}\\ D_{10} & D_{11}\end{pmatrix}\right]$$

Where the D_{00} represent the variance of intercept, D_{11} represent the variance of slope, and D_{01}, D_{10} represent the covariance between intercept and slope. $D_{01} = D_{10}$. $D_{01} = D_{10} = 0$ indicates no correlation between intercept and slope, i.e., the initial value doesn't affect how the rate of its changes with time, i.e. the slope.

Considering the deviation of individual observation from the population trajectory, $b_{i,0} = (\beta_{i,0} - \beta_0)$ and $b_{i,1} = (\beta_{i,1} - \beta_1)$, the model can be written as:

$$Y_{ij} = \boldsymbol{\beta_0 + \beta_1 \cdot X_{ij}} + b_{i,0} + b_{i,1} \cdot X_{ij} + \mathcal{E}_{ij}$$

A more general form is: $Y_{ij} = \beta X'_{ij} + b_i Z'_{ij} + \varepsilon_{ij}, \qquad \varepsilon \sim N(0, R)$

$\beta X'_{ij}$ is the fixed effects, β is the systematic variation, determining how the average for sub-population differs across distinct values of covariates, X_{ij}; it assumed to be the same for all the subjects and have population-averaged interpretation. b and $\mathcal{E}_{ij}$ are the random effect, representing the deviation of individual trajectory from the population average intercept and slope respectively, after the effects of covariates have been accounted for. It is a multilevel model, and generally the covariates in fixed effect matrix Z_{ij} are assumed to be a subset of all the variables in random effect matrix X_{ij}, where p is the total number of variables in X_{ij} , q is the total variables in the subset, thus $q < p$. The coefficient of covariate k for subject i is given as $(\beta_1 + b_{i,k})$ *if* $k \leq q$, and is β_k if $q < k \leq p$ (when the random effect of this variable does not exist in the subset). That is to say, in a linear mixed model some regression parameters may vary among only some individual subjects while some are common among all the subjects. This is presented in the linear trajectory figure as each subject of one particular covariate, e.g. control group, has their own intercept while may share the same slope. Such case is called a random intercept model, a special case of general linear model, assuming parallel trajectories. Another model assumes random intercept and slope, the subjects are assumed to be have their own intercept and slope, the variation includes both that between individual and common intercept, and that between individual and common slope.

4.1.2. Variance and covariance of the response

With the inclusion of random effects, the covariance among the repeated measures can be expressed as functions of time, and the between- and within- subject variability can be extinguished. For a random intercept model, the marginal variance of each response

Y_{ij}, the covariance and correlation of Y_{ij}, Y_{ik} are:

$$Var(Y_{ij}) = Var(\beta X'_{ij} + b_i + \varepsilon_{ij}) = \sigma_b^2 + \sigma^2$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_b^2 , \qquad Corr(Y_{ij}, Y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}$$

$$\mathrm{Cov}(Y_i) = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma^2 & \cdots & \sigma_b^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 + \sigma^2 \end{pmatrix}$$

For mixed model assuming random intercept and slope, the variance and covariance are also function of time:

$$Var(Y_{ij}) = g_{11} + 2t_{ij}g_{12} + t_{ij}^2 g_{22} + \sigma^2$$

$$Corr(Y_{ij}, Y_{ik}) = g_{11} + (t_{ij} + t_{ik})g_{12} + t_{ij}t_{ik}g_{22}$$

$$Cov(Y_{ij}) = Z_i G Z_i' + \sigma^2 I_{n_i}$$

Table 6 is the analysis result of office SBP from BEAUTY study assuming random intercept (Model 1) and random intercept and slope (Model 2). The interpretation is simply, taking model 1 as example, the intercept β_0 =158.65 is an estimate of the mean office SBP among the controls, who had their mean SBP declines 3.98 mmHg per visit. While at baseline, office SBP in IHM patients were 5.44 mmHg higher than controls and during the follow-up, also decreased faster, the decrease was 0.88mmHg per visit lower than that in controls, however, not significantly different. Since the focus is on the population averaged SBP, the individual random effect is not calculated here. Model assuming random intercept is slightly better than the other one, for it has a slightly smaller AIC and smaller SE.

Table 6. Different assumptions in mixed models for office SBP

$E(Y_{ij}|X_{ij}) = \beta_0 + \beta_1 \cdot Visit + \beta_2 \cdot IHM + \beta_3 \cdot Visit \cdot IHM$

	Model 1: random intercept				Model 2: random intercept and slope			
Effect	β	SE	t Value	Pr > |t|	β	SE	t Value	Pr > |t|
Intercept	158.65	2.22	71.62	<.0001	158.65	2.43	65.43	<.0001
visit	-3.98	0.46	-8.69	<.0001	-3.98	0.49	-8.19	<.0001
IHM	5.44	3.15	1.73	0.0850	5.44	3.45	1.58	0.1156
visit*IHM	-0.88	0.65	-1.34	0.1799	-0.88	0.69	-1.27	0.2062

Although it is not the case for BEAUTY data, inference using linear mixed models can

be quite sensitive to the specific random effects assumptions. When a model assumes random intercept and fix slope, and the comparison of treatment groups over time becomes statistically different, it may naively lead to rejection of the null hypothesis. This is because the model assumes that slopes do not vary among individuals but data could clearly suggest between subject variations in slopes (the observed individual trajectory showing much crossing over each over, and indicating random varying slope).

4.2 Generalized Estimating Equations (GEE)

Generalized Estimating Equation (GEE) is regression approach dealing with longitudinal data; it can be considered as the generalized linear marginal models, as it combines the generalized linear model (GLM) with a marginal model. In GEE approach, two models are specified: a regression model for the mean response and a model for the within-subject correlation. Like marginal model, the focus of the GEE is on estimating the average response over the population, i.e. population-averaged effects rather than the regression parameters that would enable prediction of the effect of changing one or more covariates on a given individual.

GEE includes two components: Random component is such that the response can be any distribution of the response that suits GLM, e.g., binomial, multinomial, normal, etc.; Systematic component is the linear predictor of any combination of continuous and discrete variables. The assumptions underlying the analyses performed by GEE are:

1. The responses are correlated or clustered within subjects, but each subject is independent of others. Thus they are marginal means $\mu_{ij} = E[Y_{ij}]$: The β keeps the same interpretation as in the independent data models;
2. Correlation is "nuisance" variation, that is, although it must be taken into account for inference on β, it is not of scientific interest, and the efficiency of estimation of correlation parameters is not an issue;
3. The homogeneity of variance does not need to be satisfied;
4. Errors are correlated among the repeated measurements;
5. Covariance structures are specified a priori, e.g. independent, unstructured; exchangeable, auto-regression order 1.

To build up a GEE model, the following marginal model needs to be assumed:

1. The conditional expectation of mean of each response, $E(Y_{ij}|X_{ij}) = \mu_{ij}$, is assumed to depend on the covariates through a know link function. $g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta$

2. The conditional variance of each Y_{ij}, given the covariates, is assumed to depend on the mean according to $Var(Y_{ij}) = \phi v(\mu_{ij})$, where $v(\mu_{ij})$ is a known "variance

function (of the mean μ_{ij})" and ϕ is a scale parameter that may be known or may need to be estimated. For balanced longitudinal designs, a separate scale parameter, ϕ_j , could be estimated at each occasion; alternately, the scale parameter could depend on the times of measurement, with $\phi(t_{ij})$ being some parametric function of t_{ij}. In practice, a limitation of many of the implementations of the GEE approach in widely available software is that they assume the scale parameter ϕ is time-invariant. The restriction on the scale parameter makes the GEE approach unappealing for analyzing longitudinal data when the response variable is continuous and the variance of the repeated measurements is not constant over the duration of the study.

3. The conditional within-subject association among the vector of repeated responses, given the covariates, is assumed to be a function of the means, μ_{ij} and an additional set of association parameters, α. For example, the components of α may represent the pairwise correlations or log odds ratio among the repeated responses.

The first two components specify the mean and variance of Y_{ij} following the standard generalized linear model formulation, and the third extends generalized linear model to longitudinal data. Next step is to extrapolate from marginal model to GEE approach.

The GEE estimator of β for marginal models arise from minimizing:

$$\sum_{i=1}^{N}(y_i - u_i\beta)'V_i^{-1}(y_i - u_i\beta)$$

A minimum of this function must solve the following generalized estimating equations:

$$\sum_{i=1}^{N} D_i'V_i^{-1}(y_i - \mu_i) = 0$$

It can be inferred that GEE depends on both the association parameters set, α, and the coefficient β. Therefore, two-stage estimation procedure is required:

1. Given current estimates of α and ϕ, V_i , the covariance matrix is estimated, and an estimate of β is obtained as the solution to the generalized estimating equation given by $\sum_{i=1}^{N} D_i'V_i^{-1}(y_i - \mu_i) = 0$.
2. Given the current estimate of β, estimates of α and ϕ are obtained based on the standardized residuals $e_{ij} = (Y_{ij} - \hat{\mu}_{ij})/\sqrt{v(\hat{\mu}_{ij})}$.

It usually iterates between step 1 and step 2 until convergence has been achieved. At the end, two models are specified: First a *regression model*, the form of which is flexible, i.e. linear model (or logistic regression model, log linear model, or any generalized linear model). Second a *correlation model* in which the within-subject correlation is specified, serving for the purpose of obtaining the weight that are applied to the vectors of each cluster in order to obtain the regression coefficient estimates; also this correlation model can provide model-based standard errors for the estimated

coefficients. Unlike mixed-effects models, GEEs belong to a class of semiparametric regression techniques because they rely on specification of only the first two components (mean and the covariance).

GEE has two important robustness properties: First, the estimated regression coefficients, $\hat{\beta}$, which are almost as precise as the maximum likelihood estimation, are broadly valid estimates that approach the correct value with increasing sample size regardless of the choice of correlation model. The correlation model is used simply to weight observations and a good correlation model choice can lead to more precise estimation of regression coefficients than a poor choice. Second, the correlation choice is used to obtain model-based SEs and these do require that the correlation model choice is correct in order to use the SEs for inference. A standard feature of GEE is the additional reporting of empirical SEs which provide valid estimates of the uncertainty in $\hat{\beta}$ even if the correlation model is not correct. Therefore, the correlation model can be any model, including one that assumes observations are independent. GEE approach can be applied equally to continuous responses as the discrete responses.

GEE estimation of office SBP assuming unstructured covariance is presented in Table 7.

Table 7. GEE analysis modeling the office SBP at V6

$$Y_{ij} = \beta_0 + \beta_1 \cdot visit + \beta_2 \cdot group + \beta_3 \cdot group \cdot visit$$

Parameter		Estimate	SE	95% Confidence Limits		Z	Pr>|Z|
Intercept		159.65	2.08	155.5745	163.7251	76.78	<.0001
Group	IHM	5.43	3.46	-1.3639	12.1930	1.57	0.1174
Visit		-4.19	0.43	-5.0252	-3.3486	-9.79	<.0001
Visit*Group	IHM	-0.88	0.69	-2.2301	0.4842	-1.26	0.2074

The interpretation of the linear GEE result is similar to that of linear mixed model, the intercept =159.65 is the estimated mean SBP value in controls at baseline, IHM at baseline were 5.43 mmHg higher SBP than controls; as SBP in controls decreased 4.19 mmHg per visit, the IHM decreased 0.88 mmHg more (5.07 mmHg in average) per visit, however the rates of decreasing cannot be considered as significantly different.

Two standard error estimates are provided with GEE: a model-based standard error that is valid if the correlation model is correctly specified; and an empirical standard errors which are valid even if the correlation model is not correct provided the data contain a large number of independent clusters. Estimation with GEE does not involve a likelihood function, rather it is based on the solution to regression equations that only use models for the mean and covariance.

4.2 Summary of mixed model and GEE

Mixed model and GEE are quite different analytic approaches which arise from

different assumptions about the joint distribution of *Yi* and the source of correlation among the repeated measures on the same individuals. The basic premise of mixed model is that there is natural heterogeneity across individuals, a subset of regression parameters are assumed to vary across individuals. Conditional on the random effects, it is assumed that the repeated measurements for any given individual are independent observations, the correlation among repeated measurements arises from their sharing of common random effect. Linear mixed model is very flexible in accommodating any degree of imbalance in longitudinal data, coupled with the ability to account for the covariance among the repeated measurements in a relatively parsimonious way. It does not require the same number of observations on each subject nor that the measurements be taken at the same set of measurement occasions. Those models are particularly well suited for inherently unbalanced longitudinal data.

In contrast, GEE makes inferences about population means, which are modelled conditionally only on the covariates and not on unobserved random effects or on previous responses. A distinctive feature is that the regression models for the mean responses and the models for the within-subject associations are specified separately. This separation ensures that the marginal model regression coefficients does not depend on the assumption made for the within-subject association. Specifically, the coefficients describe the effects of covariates on the population mean response. The summary comparison of mixed model and GEE is listed in Table 8.

Table 8. Summary comparison of linear Mixed model and GEE

	Mixed model	GEE
Basic premise	Natural individual heterogeneity	Inference on population means
Estimation	Maximum likelihood estimation	Marginal model Generalized estimation equation
Components	Between/within-subject variance Fixed effect and random effect	Regression model and covariance model are specified separately
Time varying covariates	Time is considered as a covariate, having its (linear) effect on the response	Marginal model without consideration of time-dependence; defining time-varying covariance pattern
SE estimation	SE not robust to model misspecification	Model based SE is vulnerable to wrong covariance model Empirical SE remains robust, if the sample is large
Inference	Subject-specific interpretation	Population means
Preferred study type/ example	Subject-specific effect of a treatment	Potential reduction in mortality in the population given a treatment

As both the linear mixed model and GEE count on the marginal mean response, they stay quite robust in dealing with missing data as long as the data are MCAR or MAR. However, mixed model assumes that, the population characteristics β is shared by all individuals and it explicitly distinguishes between fixed and random effect, it is not

only possible to estimate parameters for mean response changes in the population, but also can predict how individual response trajectories change over time. This is of interest when the focus of inference is on the individuals rather than the population, and distinguishes mixed model from GEE.

5. Discussion on BEAUTY Study

As this study results have been reported in full details over the use of antihypertensive drugs, the ABPM BP comparison and adverse events between the two randomized groups, this thesis is concentrated in the statistical perspectives on longitudinal data analysis. SBP values measured in office at each follow-up visit were estimated by both mixed model and GEE to understand the effect of different treatment approaches over time, the results of both approaches support the significant SBP decrease over the follow-up duration, however the decrease was found to be similar between control and IHM groups at each visit. We didn't find a better or faster BP decreasing effect promoted by Hotman hemodynamic monitoring. The normalization rate of office SBP (<140mmHg) also indicated similar results. The BPV from ABPM at baseline (Visit2) and study end (Visit6) are also calculated and compared between the two groups, no difference were found using the t-test (results are not shown).

In conclusion, those findings show that easy-to-do non-invasive monitoring of hemodynamic parameters associated with a predefined algorithm of drug selection does not contribute to improved BP control in European Hypertension Excellence Centers and induce similar reductions in 24h and daytime ambulatory BP and in office SBP, as compared to conventional clinical drug selection in patients with uncontrolled hypertension.

References

1 Mancia G, Di Rienzo M, Parati G. Ambulatory blood pressure monitoring use in hypertension research and clinical practice. *Hypertension* 1993; 21:510–524.

2 Bilo G, Giglio A, Styczkiewicz K, Caldara G, Maronati A, Kawecka-Jaszcz K, *et al.* A new method for assessing 24-h blood pressure variability after excluding the contribution of nocturnal blood pressure fall. *J Hypertens* 2007; 25:2058–2066.

3 Gavish B, Ben-Dov IZ, Kark JD, Mekler J, Bursztyn M. The association of a simple blood pressure-independent parameter derived from ambulatory blood pressure variability with short-term mortality. *Hypertens Res* 2009; 32:488–495.

4 Mena L, Pintos S, Queipo N V, Aizpúrua J a, Maestre G, Sulbarán T. A reliable index for the prognostic significance of blood pressure variability. *J Hypertens* 2005; 23:505–511.

5 Yong M, Diener H-C, Kaste M, Mau J. Characteristics of blood pressure profiles as predictors of long-term outcome after acute ischemic stroke. *Stroke* 2005; 36:2619–2625.

6 Zakopoulos N a., Tsivgoulis G, Barlas G, Papamichael C, Spengos K, Manios E, *et al.* Time rate of blood pressure variation is associated with increased common carotid artery intima-media thickness. *Hypertension* 2005; 45:505–512.

7 Palatini P, Reboldi G, Beilin LJ, Casiglia E, Eguchi K, Imai Y, *et al.* Added predictive value of night-time blood pressure variability for cardiovascular events and mortality: The ambulatory blood pressure-international study. *Hypertension* Published Online First: 2014. doi:10.1161/HYPERTENSIONAHA.114.03694

8 Kostis JB, Sedjro JE, Cabrera J, Cosgrove NM, Pantazopoulos JS, Kostis WJ, *et al.* Visit-to-Visit Blood Pressure Variability and Cardiovascular Death in the Systolic Hypertension in the Elderly Program. *J Clin Hypertens* 2014; 16:34–40.

9 Mancia G, Bombelli M, Facchetti R, Madotto F, Corrao G, Trevano FQ, *et al.* Long-term prognostic value of blood pressure variability in the general population: results of the Pressioni Arteriose Monitorate e Loro Associazioni Study. *Hypertension* 2007; 49:1265–1270.

10 Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlöf B, *et al.* Prognostic significance of visit-to-visit variability, maximum systolic blood

pressure, and episodic hypertension. *Lancet* 2010; 375:895–905.

11 Asayama K, Wei F-F, Liu Y-P, Hara A, Gu Y-M, Schutte R, *et al.* Does blood pressure variability contribute to risk stratification? Methodological issues and a review of outcome studies based on home blood pressure. *Hypertens Res* 2015; 38:97–101.

12 Howard SC, Rothwell PM. Reproducibility of measures of visit-to-visit variability in blood pressure after transient ischaemic attack or minor stroke. *Cerebrovasc Dis* 2009; 28:331–340.

13 Suchy-Dicey AM, Wallace ER, Mitchell SVE, Aguilar M, Gottesman RF, Rice K, *et al.* Blood pressure variability and the risk of all-cause mortality, incident myocardial infarction, and incident stroke in the cardiovascular health study. *Am J Hypertens* 2013; 26:1210–1217.

14 Levitan EB, Kaciroti N, Oparil S, Julius S, Muntner P. Relationships between metrics of visit-to-visit variability of blood pressure. *J Hum Hypertens* 2013; 27:589–93.

15 Omboni, Parati, Mancia. The trough:peak ratio and the smoothness index in the evaluation of control of 24 h blood pressure by treatment in hypertension. *Blood Press Monit* 1998; 3:201–204.

16 Omboni S, Parati G, Zanchetti A, Mancia G. Calculation of trough:peak ratio of antihypertensive treatment from ambulatory blood pressure: methodological aspects. *J Hypertens* 1995; 13:1105–12.

17 Meredith PA. How to evaluate the duration of blood pressure control: the trough:peak ratio and 24-hour monitoring. *J Cardiovasc Pharmacol* 1998; 31 Suppl 2:S17-21.

18 Parati G, Omboni S, Rizzoni D, Agabiti-Rosei E, Manciaa G. The smoothness index. *J Hypertens* 1998; 16:1685–1691.

19 Parati G, Dolan E, Ley L, Schumacher H. Impact of antihypertensive combination and monotreatments on blood pressure variability: assessment by old and new indices. Data from a large ambulatory blood pressure monitoring database. *J Hypertens* 2014; 32:1326–33.

20 Parati G, Omboni S, Rizzoni D, Agabiti-Rosei E, Mancia G. The smoothness index: a new, reproducible and clinically relevant measure of the homogeneity of the blood pressure reduction with treatment for hypertension. *J Hypertens* 1998; 16:1685–91.

21 Parati G, Ochoa JE, Lombardi C, Bilo G. Assessment and management of blood-pressure variability. *Nat Rev Cardiol* 2013; 10:143–155.

22 Stergiou GS, Kollias A, Ntineri A. Assessment of drug effects on blood

pressure variability: which method and which index? *J Hypertens* 2014; 32:1197–200.

23 Lévesque LE, Hanley J a, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* 2010; 340:b5087.

24 Pierdomenico SD, Di Nicola M, Esposito AL, Di Mascio R, Ballone E, Lapenna D, *et al.* Prognostic value of different indices of blood pressure variability in hypertensive patients. *Am J Hypertens* 2009; 22:842–847.

25 Mancia G, Facchetti R, Parati G, Zanchetti A. Visit-to-visit blood pressure variability, carotid atherosclerosis, and cardiovascular events in the European lacidipine study on atherosclerosis. *Circulation* 2012; 126:569–578.

26 Asayama K, Kikuya M, Schutte R, Thijs L, Hosaka M, Satoh M, *et al.* Home blood pressure variability as cardiovascular risk factor in the population of ohasama. *Hypertension* 2013; 61:61–69.

27 Poortvliet RKE, Ford I, Lloyd SM, Sattar N, Mooijaart SP, de Craen AJM, *et al.* Blood Pressure Variability and Cardiovascular Risk in the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER). *PLoS One* 2012; 7. doi:10.1371/journal.pone.0052438

28 Hastie CE, Jeemon P, Coleman H, McCallum L, Patel R, Dawson J, *et al.* Long-term and ultra long-term blood pressure variability during follow-up and mortality in 14 522 patients with hypertension. *Hypertension* 2013; 62:698–705.

29 Tara I. Chang, Jennifer E. Flythe, Steven M. Brunelli, Paul Muntner, Tom Greene AKC, Chertow GM. Visit-to Visit Systolic Blood Pressure Variability and Outcomes in Hemodialysis. *J Hum Hypertens* 2014; 28:18–24.

30 Gianfranco Parati, Guido Pomidossi, Fabio Albini DM and GM. Relationship of 24-Hour Blood Pressure Mean and Variability to serverity of Target-Organ Damage in Hypertension. *J Hypertens* 1987; :5:93-98.

31 Rudolph Schutte, Thijs L, Liu YP, Asayama K, Jin Y, Odili A, *et al.* Within-subject blood pressure level-not variability-predicts fatal and nonfatal outcomes in a general population. *Hypertension* 2012; 60:1138–1147.

32 Johansson JK, Niiranen TJ, Puukka PJ, Jula AM. Prognostic value of the variability in home-measured blood pressure and heart rate: The Finn-HOME study. *Hypertension* 2012; 59:212–218.

33 Rossignol P, Cridlig J, Lehert P, Kessler M, Zannad F. Visit-to-visit blood pressure variability is a strong predictor of cardiovascular events in hemodialysis: Insights from fosidiAL. *Hypertension* 2012; 60:339–346.

34 Michiaki Nagai, Satoshi Hoshide, Joji Ishikawa, Kazuyuki Shimada KK. Visit-to-visit blood pressure variations: new independent determinants for cognitive function in the elderly at high risk of cardiovascular disease. *J Hypertens* 2012; 30:1556–1563.

35 Wei FF, Li Y, Zhang L, Xu TY, Ding FH, Wang JG, *et al.* Beat-to-beat, reading-to-reading, and day-to-day blood pressure variability in relation to organ damage in untreated chinese. *Hypertension* 2014; 63:790–796.

36 Hara A, Thijs L, Asayama K, Jacobs L, Wang J, Staessen JA. Randomised Double-Blind Comparison of Placebo and Active Drugs for Effects on Risks Associated with Blood Pressure Variability in the Systolic Hypertension in Europe Trial. *PLoS One* 2014; 9. doi:10.1371/journal.pone.0103169

37 Diaz KM, Tanner RM, Falzon L, Levitan EB, Reynolds K, Shimbo D, *et al.* Visit-to-visit variability of blood pressure and cardiovascular disease and all-cause mortality: A systematic review and meta-analysis. *(Hypertension. 2014;64:965-982*

38 Johansson JK, Niiranen TJ, Puukka PJ, Jula AM. Prognostic Value of the Variability in Home-Measured Blood Pressure and Heart Rate The Finn-Home Study. 2012; :212–218.

39 Tatsuo Kawai, Mitsuru Ohishi, Norihisa Ito, Miyuki Onishi, Yasushi Takeya, Koichi Yamamoto, Kei Kamide HR. Alteration of vascular function is an important factor in the correlation between visit-to-visit blood pressure variability and cardiovascular disease. *J Hypertens* 2013; 31:1387–1395.

40 Contal C, O'Quigley J. An application of changepoint methods in studying the effect of age on survival in breast cancer. *Comput Stat Data Anal* 1999; 30:253–270.

41 Mena LJ, Maestre GE, Hansen TW, Thijs L, Liu Y, Boggia J, *et al.* How many measurements are needed to estimate blood pressure variability without loss of prognostic information? *Am J Hypertens* 2014; 27:46–55.

42 Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: Extension to survival analysis. *Stat Med* 2011; 30:22–38.

43 Diaz KM, Tanner RM, Falzon L, Levitan EB, Reynolds K, Shimbo D, *et al.* Visit-to-visit variability of blood pressure and cardiovascular disease and all-cause mortality a systematic review and meta-analysis. *Hypertension* 2014; 64:965–982.

44 L.S. M, P.M. R, J.F. P, T.G. R. Prognostic Significance of Short-Term Blood Pressure Variability in Acute Stroke: Systematic Review. *Stroke* 2015; 46:2482–2490.

45 Madden JM, O'Flynn AM, Fitzgerald AP, Kearney PM. Correlation between short-term blood pressure variability and left-ventricular mass index: a meta-analysis. *Hypertens Res* 2016; 39:171–177.

46 Stevens SL, Wood S, Koshiaris C, Law K, Glasziou P, Stevens RJ, *et al.* Blood pressure variability and cardiovascular disease: systematic review and meta-analysis. *Bmj* 2016; 354:i4098.

47 Tai C, Sun Y, Dai N, Xu D, Chen W, Wang J, *et al.* Prognostic Significance of Visit-to-Visit Systolic Blood Pressure Variability: A Meta-Analysis of 77,299 Patients. *J Clin Hypertens* 2015; 17:107–115.

48 Taylor KS, Heneghan CJ, Stevens RJ, Adams EC, Nunan D, Ward A. Heterogeneity of prognostic studies of 24-hour blood pressure variability: systematic review and meta-analysis. *PLoS One* 2015; 10:e0126375.

49 Wang J, Shi X, Ma C, Zheng H, Xiao J, Bian H, *et al.* Visit-to-visit blood pressure variability is a risk factor for all-cause mortality and cardiovascular disease: a systematic review and meta-analysis. *J Hypertens* 2017; 35:10–17.

50 Webb AJ, Rothwell PM. Blood Pressure Variability and Risk of New-Onset Atrial Fibrillation: A Systematic Review of Randomized Trials of Antihypertensive Drugs. *Stroke* 2010; :2091–2093.

51 Hansen TW, Thijs L, Li Y, Boggia J, Kikuya M, Björklund-Bodegård K, *et al.* Prognostic value of reading-to-reading blood pressure variability over 24 hours in 8938 subjects from 11 populations. *Hypertension* 2010; 55:1049–1057.

52 Asayama K, Wei F-F, Liu Y-P, Hara A, Gu Y-M, Schutte R, *et al.* Does blood pressure variability contribute to risk stratification? Methodological issues and a review of outcome studies based on home blood pressure. *Hypertens Res* 2014; 38:97–101.

53 Dolan E, Stanton A, Thijs L, Hinedi K, Atkins N, McClory S, *et al.* Superiority of ambulatory over clinic blood pressure measurement in predicting mortality: The Dublin outcome study. *Hypertension* 2005; 46:156–161.

54 Dolan E, Atkins N, McClory S, Hinedi K, Sharif S, McCormack P, *et al.* Ambulatory blood pressure measurement as a predictor of outcome in an Irish population: methodology for ascertaining mortality outcome. *Blood Press Monit* 2003; 8:143–145.

55 O'Brien E, Parati G, Stergiou G, Asmar R, Beilin L, Bilo G, *et al.* European society of hypertension position paper on ambulatory blood pressure monitoring. *J Hypertens* 2013; 31:1731–1768.

56 Mancia G, Ferrari a, Gregorini L, Parati G, Pomidossi G, Bertinieri G, *et al.* Blood pressure and heart rate variabilities in normotensive and hypertensive human beings. *Circ Res* 1983; 53:96–104.

57 Hosmer DW, Lemeshow S, May S. *Applied Survival Analysis. Regression Modeling of Time-to-Event Data*. ; 2008. doi:10.2307/1270580

58 George Brandon , Seals Samantha Seals AI. NIH Public Access. *Am Soc Nucl Cardiol* 2014; 21:686–694.

59 Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *Br J Cancer* 2003; 89:431–436.

60 Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. ; 2002. Contributor biographical information; Publisher description; Table of contents

61 Chapter 48. The LIFEREG Procedure. SAS/STAT® 9.2 User's Guide (Book Excerpt). *SAS Inst Inc 2008 SAS/STAT® 92 User's Guid Cary, NC SAS Inst Inc* 2009.

62 Marubini E, Valsecchi MG. *Analysing survival data from clinical trials and observational studies*. J. Wiley; 2004..

63 Klein JP, Moeschberger ML. *Survival analysis : techniques for censored and truncated data*. Springer; 2003.

64 Orbe J, Ferreira E, N????ez-Ant??n V. Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med* 2002; 21:3493–3510.

65 David G. Kleinbaum, Klein M. *Survival Analysis, a Self-Learning Text*. ; 2005.

66 Nardi A, Schemper M. Comparing Cox and parametric models in clinical studies. *Stat Med* 2003; 22:3597–3610.

67 Stanley C, Molyneux E, Mukaka M. Comparison of performance of exponential, Cox proportional hazards, weibull and frailty survival models for analysis of small sample size data. *J Med Stat Informatics* 2016; 4:2.

68 Gelfand LA, MacKinnon DP, DeRubeis RJ, Baraldi AN. Mediation analysis with survival outcomes: Accelerated failure time vs. proportional hazards models. *Front Psychol* 2016; 7:1–10.

69 Picciotto S, Peters A, Eisen EA. Hypothetical exposure limits for oil-based metalworking fluids and cardiovascular mortality in a cohort of autoworkers: Structural accelerated failure time models in a public health framework. *Am J Epidemiol* 2015; 181:563–570.

70 Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. Wiley; 1980.

71 Rebora P, Salim A, Reilly M. bshazard: A Flexible Tool for Nonparametric

Smoothing of the Hazard Function. *R J* 2014; 6:114–122.

72 Liu X, Jin Z. Optimal survival time-related cut-point with censored data. *Stat Med* 2015; 34:515–524.

73 Rota M, Antolini L, Valsecchi MG. Optimal cut-point definition in biomarkers: the case of censored failure time outcome. *BMC Med Res Methodol* 2015; 15:1–11.

74 Hanley AJ, McNeil JB. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 1982; 143:29–36.

75 Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; 56:337–344.

76 Akritas, M. J., 1994, Annuls of Statistics 22, 1299-1327.pdf.

77 Heagerty PJ ZY. Survival Model Predictive Accuracy and ROC Curves. *Biometrics* 2005; 61:92–105.

78 Fan VS, Au D, Heagerty P, Deyo RA, McDonell MB, Fihn SD. Validation of case-mix measures derived from self-reports of diagnoses and health. *J Clin Epidemiol* 2002; 55:371–380.

79 Etzioni R, Pepe M, Longton G, Hu C, Goodman G, Case A. New Looks at ROC Curves Incorporating the Time Dimension in Receiver Operating Characteristic Curves : (Med Decis Making 1999; 19:242-251)

80 Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med* 2006; 25:3474–3486.

81 Antolini L, Valsecchi MG. Performance of binary markers for censored failure time outcome: Nonparametric approach based on proportions. *Stat Med* 2012; 31:1113–1128.

82 Bersabé R, Rivas T. A general equation to obtain multiple cut-off scores on a test from multinomial logistic regression. Span. J. Psychol. 2010; 13:494–502.

83 Black J a., Park M, Gregson J, Falconer CL, White B, Kessel a. S, *et al.* Child obesity cut-offs as derived from parental perceptions: cross-sectional questionnaire. *Br J Gen Pract* 2015; 65:e234–e239.

84 Shultz EK. Multivariate receiver-operating characteristic curve analysis: Prostate cancer screening as an example. *Clin Chem* 1995; 41:1248–1255.

85 del Sol AI, Moons KG, Hollander M, Hofman A, Koudstaal PJ, Grobbee DE, *et al.* Is carotid intima-media thickness useful in cardiovascular disease risk assessment? The Rotterdam Study. *Stroke* 2001; 32:1532–8.

86 Marwick TH, Case CC, Siskind V, Woodhouse SP. Prediction of survival from resuscitation: a prognostic index derived from multivariate logistic model

analysis. *Resuscitation* 1991; 22:129–137.

87 Wang TJ, Gona P, Larson MG, Tofler GH, Levy D, Jacques CN-CPF, *et al.* Multiple biomarkers for the prediction of first major cardiovascular events and death: Considerable costs and limited benefits. *N Engl J Med* 2006; 9:2631–2639.

88 Harrell FE, Lee KL, Mark DB. Prognostic/Clinical Prediction Models: Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Tutorials Biostat Stat Methods Clin Stud* 1996; 15:361–387.

89 Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al.* Assessing the Performance of Prediction Models. *Epidemiology* 2010; 21:128–138.

90 Harrison D a, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006; 34:1378–1388.

91 Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical J* 2008; 50:457–479.

92 Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves 1. *Biometrics* 2005; 61:92–105.

93 Mithat Gönen. Receiver Operating Characteristic (ROC) Curves Mithat Gönen , Memorial Sloan-Kettering Cancer Center SUGI 31 Statistics and Data Analysis FN + FP. *Stat data Anal* 2001; :1–18.

94 Penciana MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Stat Med* 2004; 23:2109–2123.

95 Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16:965–980.

96 Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115:928–935.

97 Pencina MJ, D'Agostino RB, Pencina KM, Janssens ACJW, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol* 2012; 176:473–81.

98 Steyerberg EW, Van Calster B, Pencina MJ. Performance measures for prediction models and markers: evaluation of predictions and classifications. *Rev Esp Cardiol (Engl Ed)* 2011; 64:788–94.

99 Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004; 159:882–890.

100 Pencina MJ, Ralph B. D'Agostino Sr, Ralph B. D'Agostino Jr, S.Vasan R. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27(23):157–172.

101 Pencina MJ, Steyerberg EW, Sr. RBD. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. 2012; 30:11–21.

102 Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: Do not rely on integrated discrimination improvement and net reclassification index. *Stat Med* 2014; 33:3405–3414.

103 Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014; 25:114–21.

104 Allison PD, SAS Institute. *Survival analysis using SAS : a practical guide*. SAS Institute; 2010..

105 Brenda W. Gillespie. Use of generalized R-squared in Cox regression. Boston: APHA Scientific Session and Event Listing; 2006..

106 Brier GW. Verification of forecasts expersses in terms of probaility. *Mon Weather Rev* 1950; 78:1–3.

107 Murphy AH. A New Vector Partition of the Probability Score. J. Appl. Meteorol. 1973; 12:595–600.

108 Pencina MJ, D'Agostino RB, Massaro JM. Understanding increments in model performance metrics. *Lifetime Data Anal* 2013; 19:202–218.

109 Baker SG, Schuit E, Steyerberg EW, Pencina MJ, Vickers A, Moons KGM, *et al*. How to interpret a small increase in AUC with an additional risk prediction marker: Decision analysis comes through. *Stat Med* 2014; 33:3946–3959.

110 Rienzo MDI, Grassi G, Pedotti A, Mancia G. in Estimating 24-Hour Average Blood Pressure. *Hypertension* 1983; 5:264–269.

111 Vickers AJ, Altman DG. Statistics Notes: Analysing Controlled Trials With Baseline And Follow Up Measurements. *BMJ Br Med J* 2001; 323:1123–1124.

112 Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992; 11:1685–1704.