Final Project
(examples)

Evaluation :

- **Project**: dataset to be analyzed, presentation of the results. You can choose the statistical software you prefer (R, Python…); it could be done individually or in team (**max 3** students; **20-30** minutes)
- **Oral questions**(individual…)
- **Final mark** will be an **average** between project (team/individual) and (individual) answers

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Project guidelines

- Identification of the **scientific question** and (possibly) of the **study design** that originated the data
- Data **preprocessing**: **IDA** (initial data analysis / univariable analyses / **missing data**)
- *Model's* estimation procedures to answer the scientific question
- Report (**R markdown** or similar) explaining **analyses** and **results**.

End of the course (**Monday!**) each student/team should prepare a **5 minute** oral presentation in which:

- the selected **dataset** and **scientific question** are briefly presented
- **goals** and **roadmap** of the project should be *approximately* defined…

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Scientific question & Study design

https://pubmed.ncbi.nlm.nih.gov/7882472/

1

Clinical Trial > Circulation. 1995 Mar 15;91(6):1659-68. doi: 10.1161/01.cir.91.6.1659.

Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction. Results from an international trial of 41,021 patients. GUSTO-I Investigators

K L Lee [1], L H Woodlief, E J Topol, W D Weaver, A Betriu, J Col, M Simoons, P Aylward, F Van de Werf, R M Califf

https://pubmed.ncbi.nlm.nih.gov/36208798/

2

> Am J Kidney Dis. 2023 Mar;81(3):307-317.e1. doi: 10.1053/j.ajkd.2022.07.017. Epub 2022 Oct 5.

Cardiorenal Outcomes Among Patients With Atrial Fibrillation Treated With Oral Anticoagulants

Marco Trevisan [1], Paul Hjemdahl [2], Catherine M Clase [3], Ype de Jong [4], Marie Evans [5], Rino Bellocco [6], Edouard L Fu [7], Juan Jesus Carrero [8]

Scientific question & Study design

Acute myocardial infarction ("heart attack") is caused by the formation of a clot in one of the coronary arteries that supply blood to the heart muscle.

Mortality is substantial in the period immediately after the event, and also during the years after surviving the initial infarction. (Some patients die before reaching the hospital).

Patients seen in hospitals are reported to have an **average mortality within 30 days** around **6–15%**.

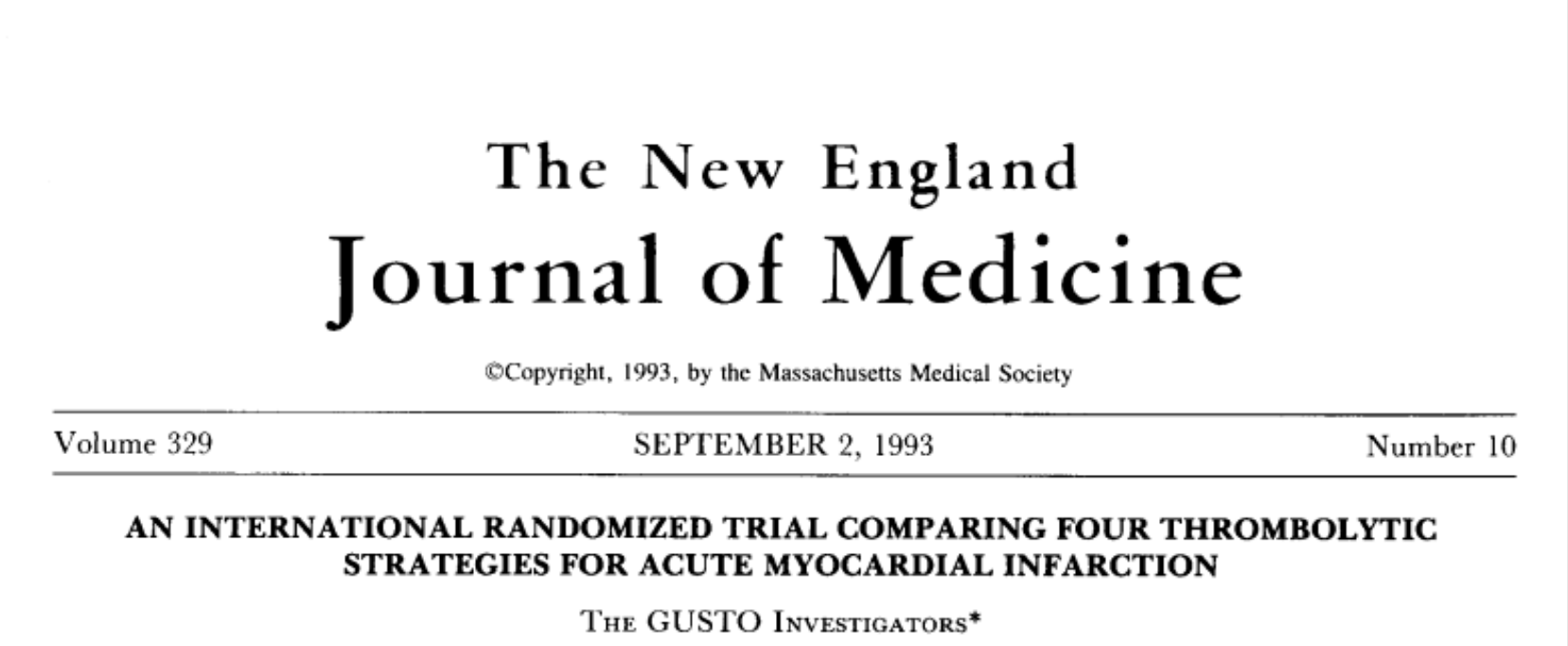
The risk of 30-day mortality strongly depends on various **prognostic factors**:

Categories	Examples
Demographics	Age, sex, weight, height, geographical site
Risk factors	Diabetes, hypertension, smoking status, hypercholesterolemia, family history of MI
Other history	Previous MI, angina, cerebrovascular disease (e.g., stroke), bypass surgery, angioplasty
Cardiac state	Location of infarction, electrocardiogram abnormalities
Presenting characteristics	Systolic and diastolic blood pressure, heart rate, left ventricular function (e.g., presence of shock, Killip class)

Scientific question & Study design

Various drugs and treatments are nowadays available for acute MI, including drugs that attack the clot ("thrombolytics") and procedures such as acute revascularization, such as percutaneous interventions ("PTCA").

GUSTO-I is one of the major **randomized controlled trials** that compared alternative treatments for acute MI.



The New England
Journal of Medicine

Volume 329 SEPTEMBER 2, 1993 Number 10

AN INTERNATIONAL RANDOMIZED TRIAL COMPARING FOUR THROMBOLYTIC STRATEGIES FOR ACUTE MYOCARDIAL INFARCTION

THE GUSTO INVESTIGATORS*

Scientific question & Study design

Table 1. Base-Line Characteristics of the Four Treatment Groups.*

Characteristic	Streptokinase and Subcutaneous Heparin (N = 9841)	Streptokinase and Intravenous Heparin (N = 10,410)	Accelerated t-PA and Intravenous Heparin (N = 10,396)	Both Thrombolytic Agents and Intravenous Heparin (N = 10,374)
Age (yr)	62 (52, 70)	62 (52, 70)	62 (52, 70)	61 (52, 70)
Female sex (%)	25	25	25	25
Diabetes (%)	15	15	15	14
Cigarette smoker (%)	43	43	43	43
Hypertension (%)	39	38	38	38
Systolic blood pressure (mm Hg)	130 (111, 144)	129 (112, 144)	130 (113, 144)	130 (112, 143)
Heart rate (beats/min)	73 (62, 85)	74 (63, 86)	73 (62, 86)	74 (62, 86)
Previous infarction (%)	16	17	17	16
Previous CABG† (%)	4	4	5	4
Time to randomization (min)	120 (90, 180)	120 (90, 180)	120 (90, 180)	120 (90, 180)
Time to treatment (min)	164 (115, 232)	165 (120, 230)	165 (120, 230)	170 (121, 237)

*Values followed by numbers in parentheses are medians, with the 25th and 75th percentiles shown inside the parentheses. There were no differences in base-line characteristics among the four groups. Time to treatment, although not strictly a base-line characteristic, did differ among the groups ($P<0.001$).

†CABG denotes coronary-artery bypass surgery.

The trial enrolled 41,021 patients admitted to 1081 hospitals in 15 countries. The primary **end-point** was death from any cause at 30 days of follow-up.

The hypothesis was that tPA would show a **1% absolute reduction** in 30-day mortality

Scientific question & Study design

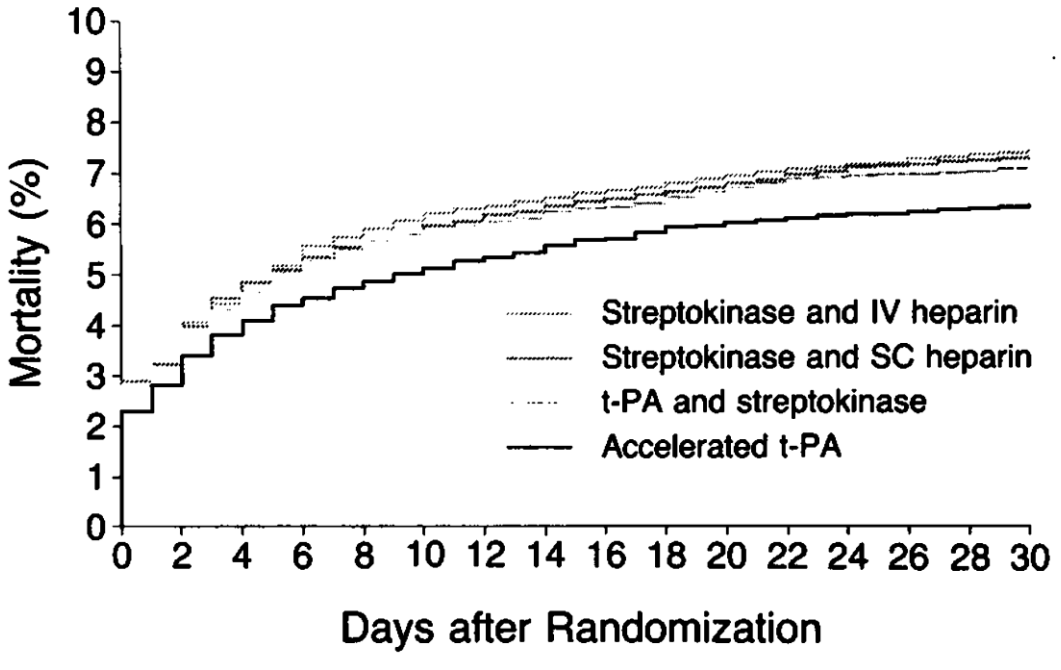
Table 2. Major Clinical Outcomes.

Outcome	Streptokinase and Subcutaneous Heparin (N = 9796)	Streptokinase and Intravenous Heparin (N = 10,377)	Accelerated t-PA and Intravenous Heparin (N = 10,344)	Both Thrombolytic Agents and Intravenous Heparin (N = 10,328)	P Value, Accelerated t-PA vs. Both Streptokinase Groups
	percent of patients				
24-hr mortality	2.8	2.9	2.3	2.8	0.005
30-day mortality	7.2	7.4	6.3	7.0	0.001
Or nonfatal stroke	7.9	8.2	7.2	7.9	0.006
Or nonfatal hemorrhagic stroke	7.4	7.6	6.6	7.4	0.004
Or nonfatal disabling stroke	7.7	7.9	6.9	7.6	0.006

These data were used also to build a **prediction model** for 30-days mortality

The trial convincingly showed a benefit of tPA treatment ($p < 0.001$) over the others.

Of note : **more expensive** thrombolytic drug (tPA) vs the cheaper drug (streptokinase).



Scientific question & Study design

Historically, atrial fibrillation (AF) patients have been treated with warfarin, a vitamin K antagonist (VKA), which effectively prevents two out of three ischemic strokes compared to placebo, but also increases the risk of bleeding, which can be a minor event or result in a fatal hemorrhage.

VKA use is limited by a narrow therapeutic index, which determines that patients have to be frequently monitored, resulting in substantial burden to them.

Therefore, Direct Oral Anticoagulants (DOACs) have been developed and showed similar or greater efficacy and safety compared to VKA in **RCTs**. The advantages of DOAC use include less drug and food interactions, more stable anticoagulant effects and reduced need for routine monitoring.

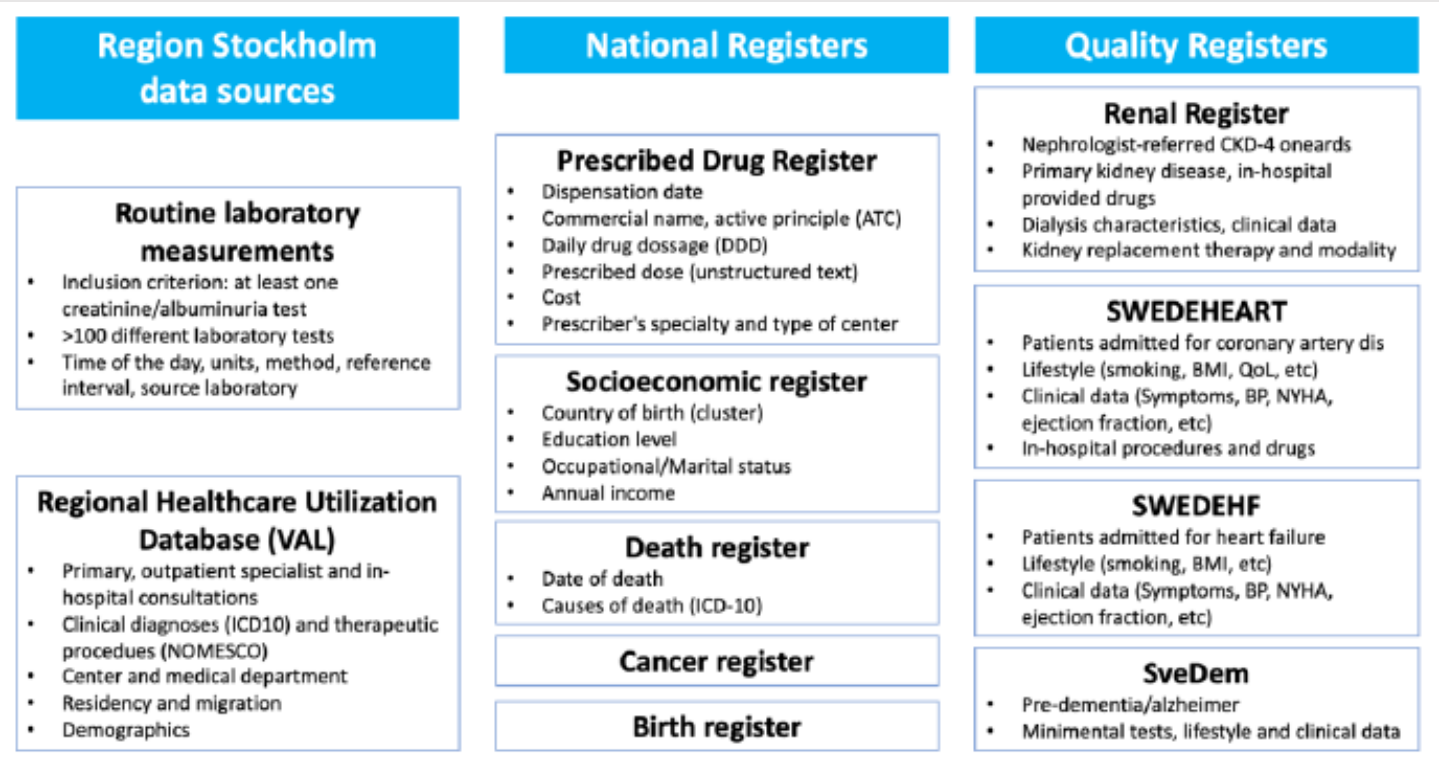
DOACs have **progressively replaced** vitamin K antagonists (VKAs) for stroke prevention in patients with nonvalvular atrial fibrillation (AF).

DOACs cause fewer bleeding complications, but their other advantages, **particularly related to kidney outcomes**, remain inconclusive.

Scientific question & Study design

Aim: compare the risk of kidney outcomes (CKD progression or AKI) among patients with non-valvular AF initiating DOAC versus VKA.

Cohort: all adults that started oral anticoagulants (OACs) between 1st January 2011 and 31st December 2018, with a diagnosis of AF in the preceding 5 years.
New users of OAC identified as individuals with a **first prescription** of DOAC or VKA drugs with no previous dispensations of any OAC. The date of OAC initiation was defined as **index date** and start of the follow-up (T0).

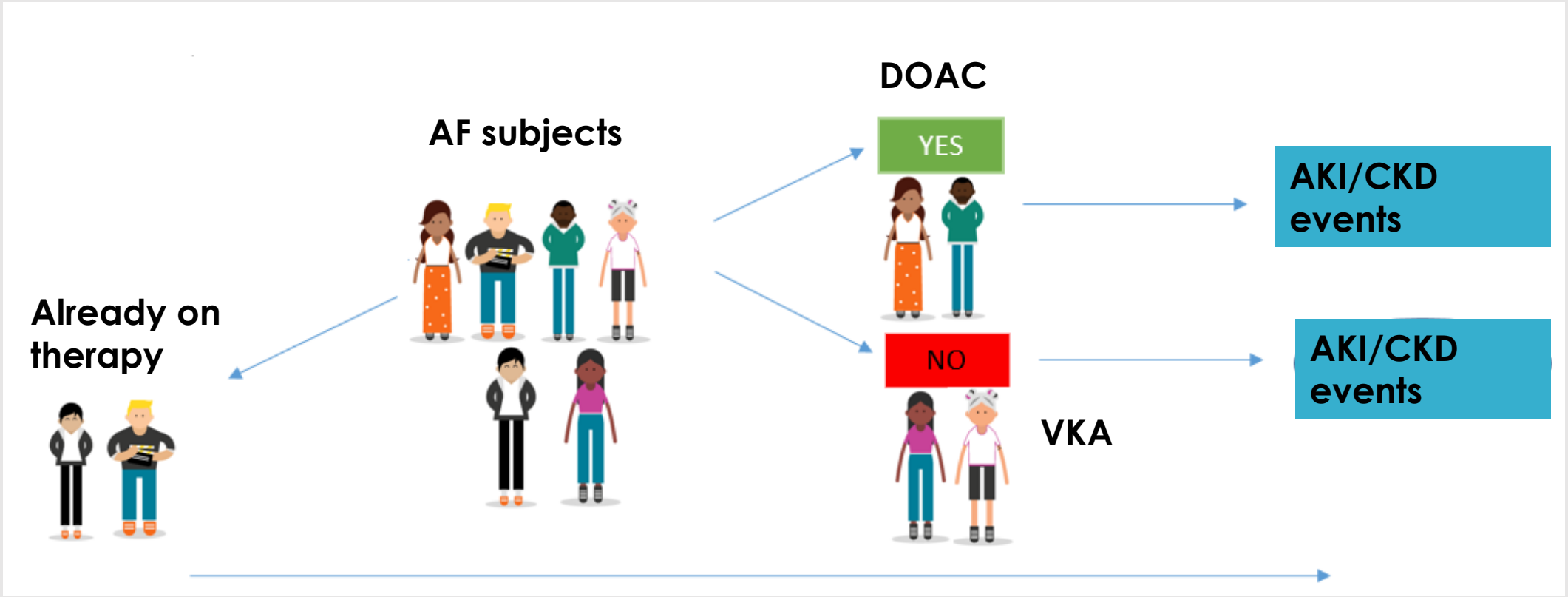


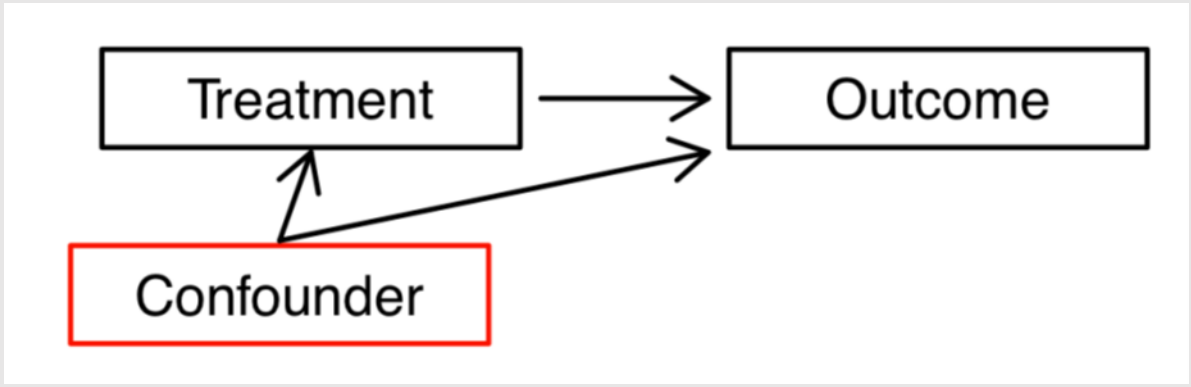
We **censored** patients at **treatment discontinuation**, defined as absence of a refill before the end of the estimated pill supply plus a lag-phase of 120 days, or switch from DOAC to VKA or vice versa.

Adminstrative censoring at 31/12/2018.

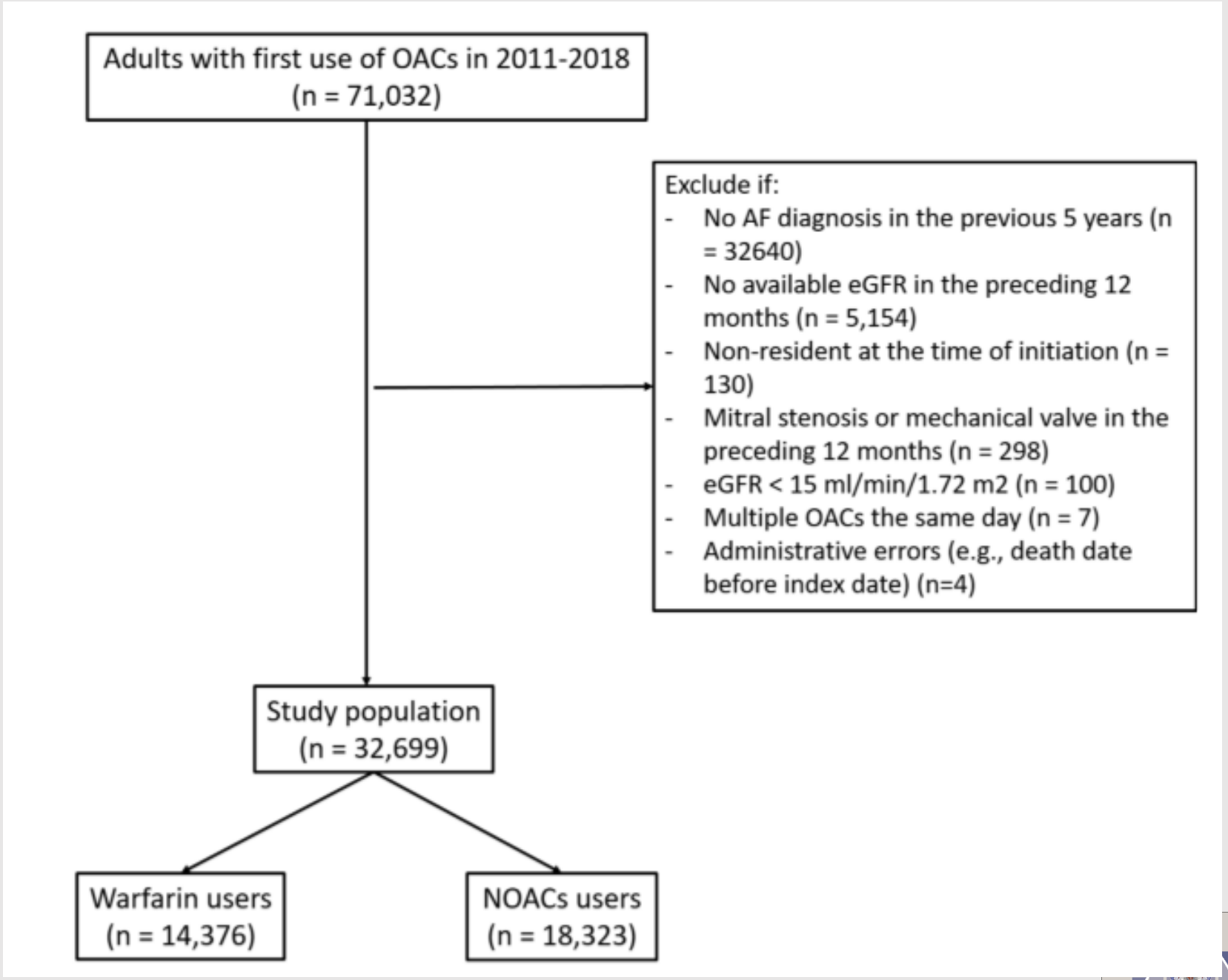
UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Scientific question & Study design





- Death as a competing risk
- Confounders


Adults with first use of OACs in 2011-2018
(n = 71,032)
Exclude if:
- No AF diagnosis in the previous 5 years (n = 32640)
- No available eGFR in the preceding 12 months (n = 5,154)
- Non-resident at the time of initiation (n = 130)
- Mitral stenosis or mechanical valve in the preceding 12 months (n = 298)
- eGFR < 15 ml/min/1.72 m2 (n = 100)
- Multiple OACs the same day (n = 7)
- Administrative errors (e.g., death date before index date) (n=4)
Study population
(n = 32,699)
Warfarin users
(n = 14,376)
NOACs users
(n = 18,323)

Scientific question & Study design

Definition of the primary study outcomes were:

- CKD progression
- AKI

CKD progression : the composite of kidney failure or sustained 30% eGFR decline.

Kidney failure was defined as the presence of sustained eGFR <15 mL/ min/1.73 m2, initiation of maintenance dialysis, or kidney transplantation.

AKI was identified by a combination of diagnoses (ICD-10 codes) in outpatient or hospital care and transient creatinine elevations during hospitalization according to clinical criteria.

For these outcomes, follow up ended on the first date an end point was reached, date of last laboratory measurement, or December 31, 2018, (or death) whichever came first.

Initial Data Analysis

In the GUSTO-I trial, a comprehensive set of **prognostic factors** was collected at baseline.

Step	Specific issues	GUSTO-I model
General considerations		
Research question	Aim: predictors/prediction?	Both
Intended application	Clinical practice/research?	Clinical practice
Outcome	Clinically relevant?	30-day mortality
Predictors	Reliable measurement? comprehensiveness	Standard clinical workup; extensive set of candidate predictors
Study design	Retrospective/prospective? cohort; case control	RCT data: prospective cohort
Statistical model	Appropriate for research question and type of outcome?	Logistic regression
Sample size	Sufficient for aim?	>40,000 patients; 2851 events: excellent

Initial Data Analysis

The study considers many potential predictors. A comprehensive set of approximately **25** characteristics was considered, based on **subject matter knowledge** (input from expert clinicians, literature).

https://www.ahajournals.org/doi/10.1161/01.cir.91.6.1659?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed

A total of 28151 patients had died by 30 days. 39% of the deaths occurred within 24 h; more than half (55%) occurred within 48 h of randomization. This **number of events** provides an exceptional and excellent basis for prognostic modeling.

The outcome (30-day mortality) was complete for 40,830 of the 41,021 patients (99.5%).

Distributions of some candidate predictors were quite **skewed**, e.g., for Killip class (a measure of left ventricular function). Categories III or IV were present **in only 2%** of the patients; these categories represent patients in shock.

Initial Data Analysis

Table 1. Baseline Characteristics of Patients Initiating Oral Anticoagulants in Stockholm in 2011-2018, Overall and Stratified by Initial Treatment Group

		Oral Anticoagulant Started	
	Overall (N = 32,699)	DOAC (n = 18,323)	VKA (n = 14,376)
Age, y	75 [68-83]	75 [68-83]	76 [68-83]
Age category			
<75 y	15,336 (47%)	8,742 (48%)	6,594 (46%)
≥75 y	17,363 (53%)	9,581 (52%)	7,782 (54%)
Women	14,816 (45%)	8,399 (45%)	6,417 (45%)
Access to health care in the previous year			
Primary care visits	5 [2-8]	4 [2-8]	5 [2-8]
Outpatient visits	3 [1-6]	3 [1-7]	2 [1-5]
Issued ICD-10 codes	15 [8-27]	16 [8-29]	15 [8-26]
Procedures	4 [1-10]	4 [1-11]	3 [1-8]
Education			
Compulsory	8,730 (27%)	4,530 (25%)	4,200 (29%)
Secondary	12,951 (40%)	7,213 (39%)	5,738 (40%)
University	10,385 (32%)	6,256 (34%)	4,129 (29%)
Missing	633 (2%)	324 (2%)	309 (2%)
eGFR, mL/min/1.73 m^2	73 [59-85]	74 [60-85]	72 [57-85]
eGFR category			
15-29 mL/min/1.73 m^2	670 (2%)	189 (1%)	481 (3%)
30-59 mL/min/1.73 m^2	8,078 (25%)	4,300 (24%)	3,778 (26%)
≥60 mL/min/1.73 m^2	23,951 (73%)	13,834 (75%)	10,117 (71%)

Initial Data Analysis

	Overall (N = 32,699)	Oral Anticoagulant Started	
		DOAC (n = 18,323)	VKA (n = 14,376)
Medical history			
Hypertension	23,621 (72%)	13,156 (72%)	10,465 (73%)
Vascular disease	9,714 (30%)	4,896 (27%)	4,818 (33%)
Cancer	8,519 (26%)	4,994 (27%)	3,525 (24%)
CHF/LV dysfunction	8,089 (25%)	4,071 (22%)	4,018 (28%)
Heart failure	7,975 (24%)	3,999 (22%)	3,976 (28%)
Diabetes	6,906 (21%)	3,723 (20%)	3,183 (22%)
Stroke, TIA, or embolism	6,709 (20%)	3,649 (20%)	3,060 (21%)
Anemia	5,693 (17%)	3,203 (17%)	2,490 (17%)
Stroke	4,845 (15%)	2,632 (14%)	2,213 (15%)
Myocardial infarction	4,887 (15%)	2,366 (13%)	2,521 (17%)
Diabetic complications	4,473 (14%)	2,293 (12%)	2,180 (15%)
Prior bleeding	3,576 (11%)	2,133 (12%)	1,443 (10%)
COPD	3,566 (11%)	2,058 (11%)	1,508 (10%)
VTE	3,140 (10%)	1,648 (9%)	1,492 (10%)
PCI	2,641 (8%)	1,322 (7%)	1,319 (9%)
Rheumatoid arthritis	2,323 (7%)	1,307 (7%)	1,016 (7%)
Kidney disease	2,329 (7%)	1,225 (7%)	1,104 (8%)
Fracture	1,964 (6%)	1,180 (6%)	784 (5%)
DVT or knee/hip replacement	1,761 (5%)	904 (5%)	857 (6%)
Alcohol abuse	1,768 (5%)	1,129 (6%)	639 (4%)
AKI	890 (3%)	499 (3%)	391 (3%)
Liver disease	726 (2%)	428 (2%)	298 (2%)
Risk score			
CHA_2DS_2-VASc	3 [2-5]	3 [2-4]	3 [2-5]
Modified-$CHADS_2$	5 [3-7]	5 [3-7]	5 [3-7]
HAS-BLED	2 [2-3]	2 [2-3]	3 [2-3]
Concomitant medications			
β-blocker	26,174 (80%)	14,485 (79%)	11,689 (81%)
RAAS inhibitor	18,248 (56%)	10,005 (55%)	8,243 (57%)
Aspirin	14,538 (44%)	7,106 (39%)	7,432 (52%)
Statin	11,911 (36%)	6,339 (35%)	5,572 (39%)
Diuretic	11,240 (34%)	5,607 (31%)	5,633 (39%)

Missing data

Missing values in GUSTO-I occurred for various candidate predictors, but usually only in a **small** fraction.

Missing values **were imputed** for further statistical analysis ("single imputation" approach).

Imputation was based on the *correlation* among predictors.

Although a full set of analyses was performed in patients with **complete data** for all the important predictor variables (92% of the study patients), the subset of patients with one or more missing predictor variables had a higher mortality rate than the other patients, and excluding those patients could lead to biased estimates of risk.

To circumvent this, a method for simultaneous imputation was used to estimate missing predictor variables and allow analysis of all patients.

The iterative imputation technique involved estimating a given predictor variable on the basis of multivariable regression on (possibly) transformed values of all the other predictor variables.

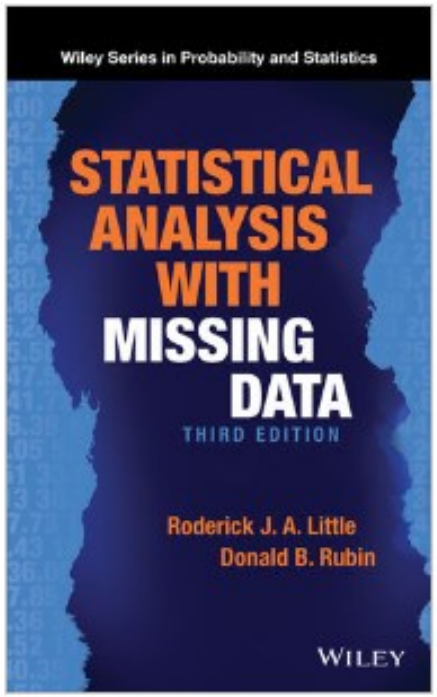
End-point data were not explicitly used in the imputation process.

Just a note here about missing data…

MCAR
Missing completely at random

the fact that data are missing is ***independent*** of the observed and *unobserved* data

no **systematic** differences between participants with missing data and those with **complete** data



MAR
Missing at random

the fact that the data are missing is **systematically** related to the observed but not the *unobserved* data

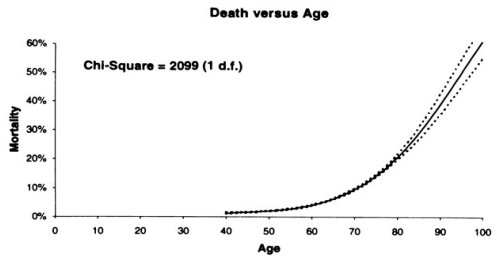
Complete case analyses may or may not result in bias. Proper **accounting** for the known factors can produce unbiased results in analysis

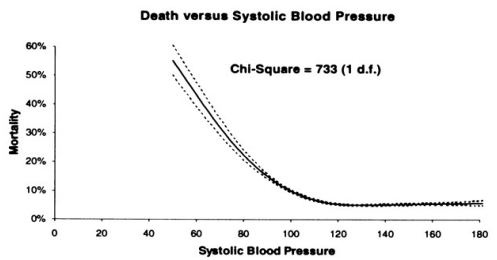
MNAR
Missing not at random

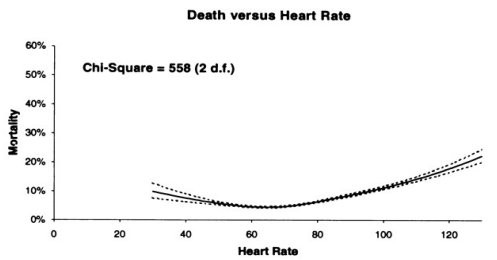
the fact that the data are missing is **systematically** related to the unobserved data…

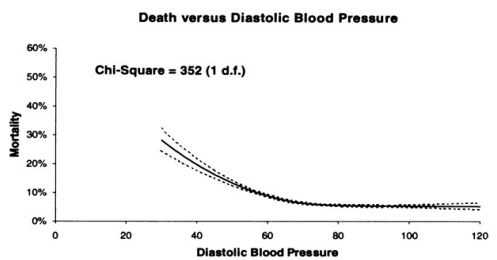
if the complete case analysis is biased this issue **cannot be** addressed…

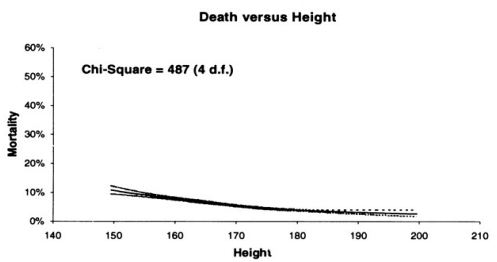
Functional forms of predictors

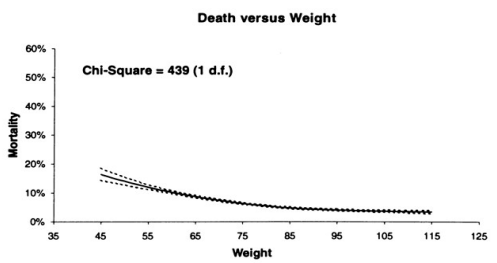


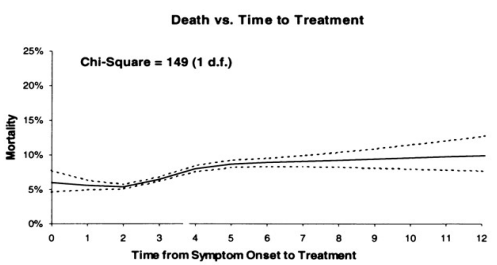












For **continuous** clinical variables, we examined the shape of the relation with 30-day mortality by use of a flexible model-fitting approach involving cubic spline functions (cubic polynomials).

Where relations were nonlinear, their shape was characterized with spline functions.

Determining how variables should be modeled was an important step in characterizing prognostic relations and identifying which variables were most strongly related to short-term mortality.

We also examined whether the prognostic relation of any important variable differed for particular levels of other important descriptors (ie, we tested for **interactions** among the prognostic clinical variables).

Model's estimation

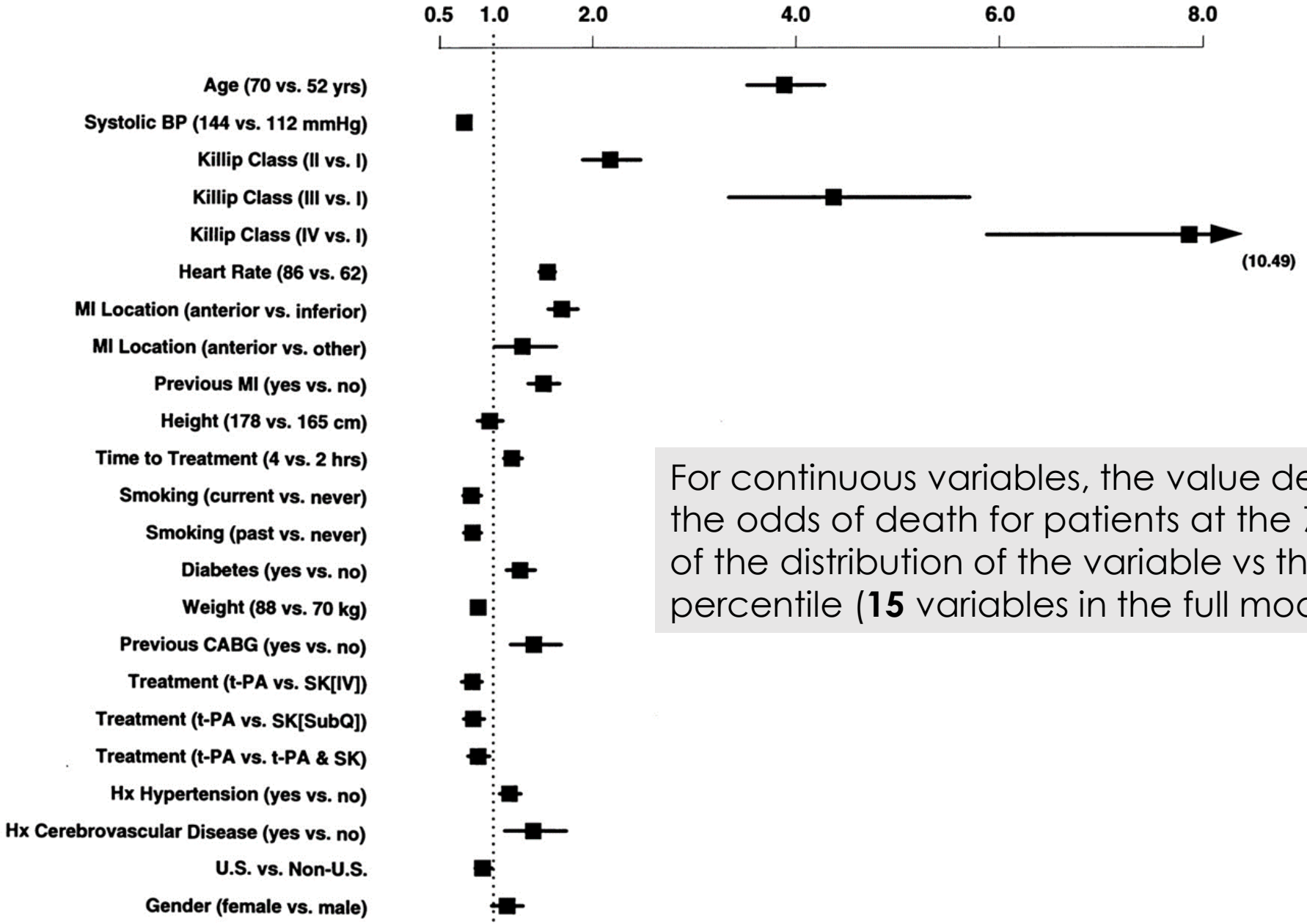
A **logistic multivariable regression model** was used to examine individual and joint relations between baseline clinical characteristics and the binary outcome of death within 30 days of randomization

Probability of death within 30 days = $1/[1 + \exp(-L)]$, where

$L = 3.81 + 0.0762$ age $- 0.0398 \min(\text{SBP}, 120) + 2.08$ [Killip class II] $+ 3.62$ [Killip class III] $+ 4.04$ [Killip class IV] $- 0.0211$ heart rate $+ 0.0394$ (heart rate $- 50)_+ - 0.0397$ height $+ 0.000184$ (height $- 154.9)_+^3 - 0.000898$ (height $- 165.1)_+^3 + 0.00159$ (height $- 172.0)_+^3 - 0.00107$ (height $- 177.3)_+^3 + 0.000194$ (height $- 185.4)_+^3 + \ldots$

Explanatory notes:

1. Brackets are interpreted as [c] = 1 if the patient falls into category c, [c] = 0 otherwise.
2. $(x)_+ = x$ if $x > 0$, $(x)_+ = 0$ otherwise.
3. For systolic blood pressure (SBP), values >120 mmHg are winsorized.



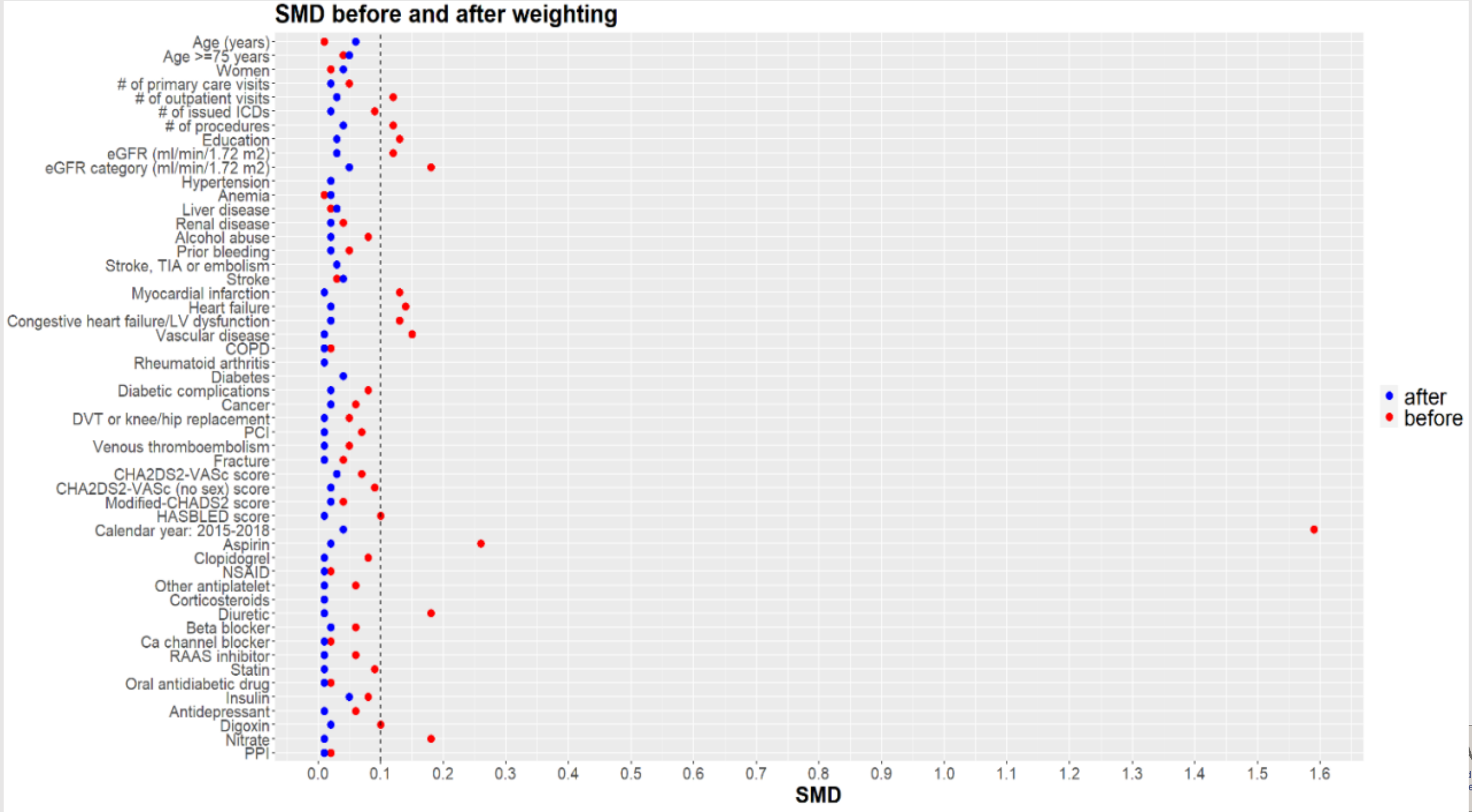
For continuous variables, the value depicted reflects the odds of death for patients at the 75th percentile of the distribution of the variable vs the 25th percentile (**15** variables in the full model).

Model's estimation

1. **Propensity score model**: Probability to receive DOAC versus VKA treatment as a function of the baseline covariates , estimation of **weights** (multivariable logistic regression model)

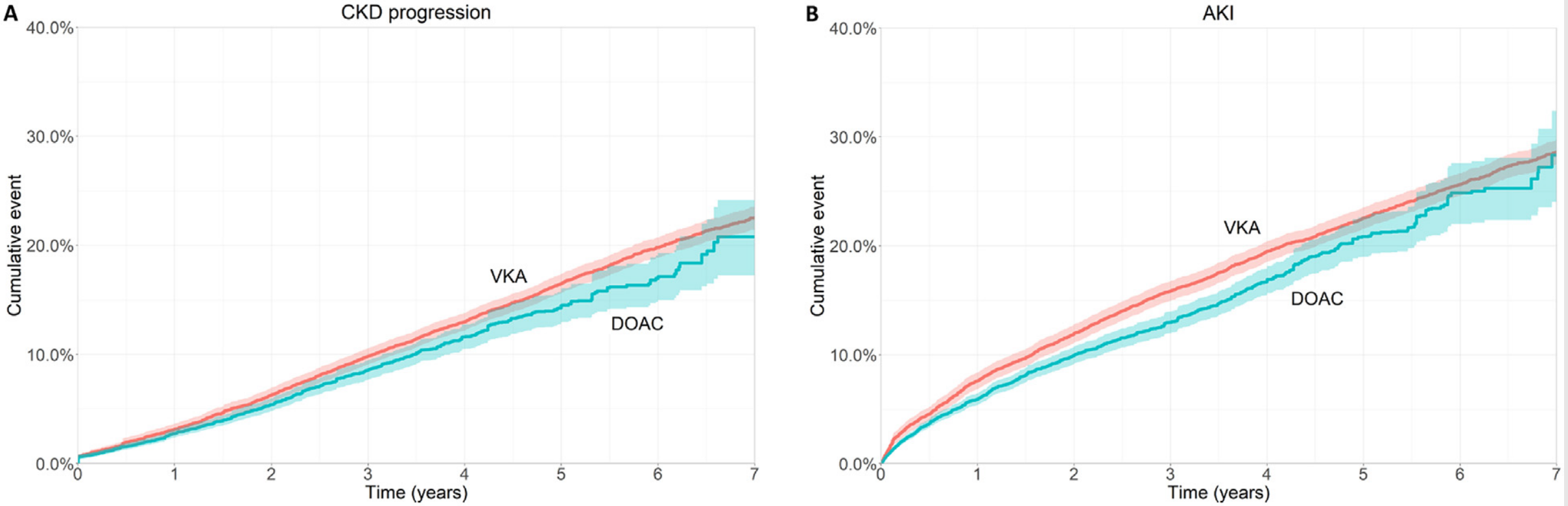
Demographics: age; sex; calendar year; education
Clinical: number of primary health care visits; number of outpatient specialist visits; number of diagnoses issued; number of procedure codes; eGFR; hypertension; anemia; liver disease; kidney disease; alcohol abuse;
Prior events/diagnosis: prior bleeding; stroke/transient ischemic stroke/embolism; stroke; myocardial infarction; heart failure; congestive heart failure; vascular disease; chronic obstructive pulmonary disease; rheumatoid arthritis; diabetes; diabetic complications; cancer; deep vein thrombosis; knee/hip surgery; percutaneous coronary intervention; venous thromboembolism; fracture; risk scores;
Other therapies: aspirin, clopidogrel, nonsteroidal anti-inflammatory drugs, other antiplatelet, corticosteroids, diuretics, β-blockers, calcium channel blockers, renin-angiotensin-aldosterone-system inhibitors, statin, insulin, other antidiabetic medications, antidepressants, digoxin, nitrate, and proton-pump inhibitors

Block 4.5


SMD before and after weighting
Age (years)
Age >=75 years
Women
# of primary care visits
# of outpatient visits
# of issued ICDs
# of procedures
Education
eGFR (ml/min/1.72 m2)
eGFR category (ml/min/1.72 m2)
Hypertension
Anemia
Liver disease
Renal disease
Alcohol abuse
Prior bleeding
Stroke, TIA or embolism
Stroke
Myocardial infarction
Heart failure
Congestive heart failure/LV dysfunction
Vascular disease
COPD
Rheumatoid arthritis
Diabetes
Diabetic complications
Cancer
DVT or knee/hip replacement
PCI
Venous thromboembolism
Fracture
CHA2DS2-VASc score
CHA2DS2-VASc (no sex) score
Modified-CHADS2 score
HASBLED score
Calendar year: 2015-2018
Aspirin
Clopidogrel
NSAID
Other antiplatelet
Corticosteroids
Diuretic
Beta blocker
Ca channel blocker
RAAS inhibitor
Statin
Oral antidiabetic drug
Insulin
Antidepressant
Digoxin
Nitrate
PPI
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
SMD
after
before

Model's estimation

2. **Outcome analysis** on the weighted cohort: cumulative incidence curves



The adjusted HR for CKD progression for DOAC users was 0.87 (95% CI, 0.78- 0.98).

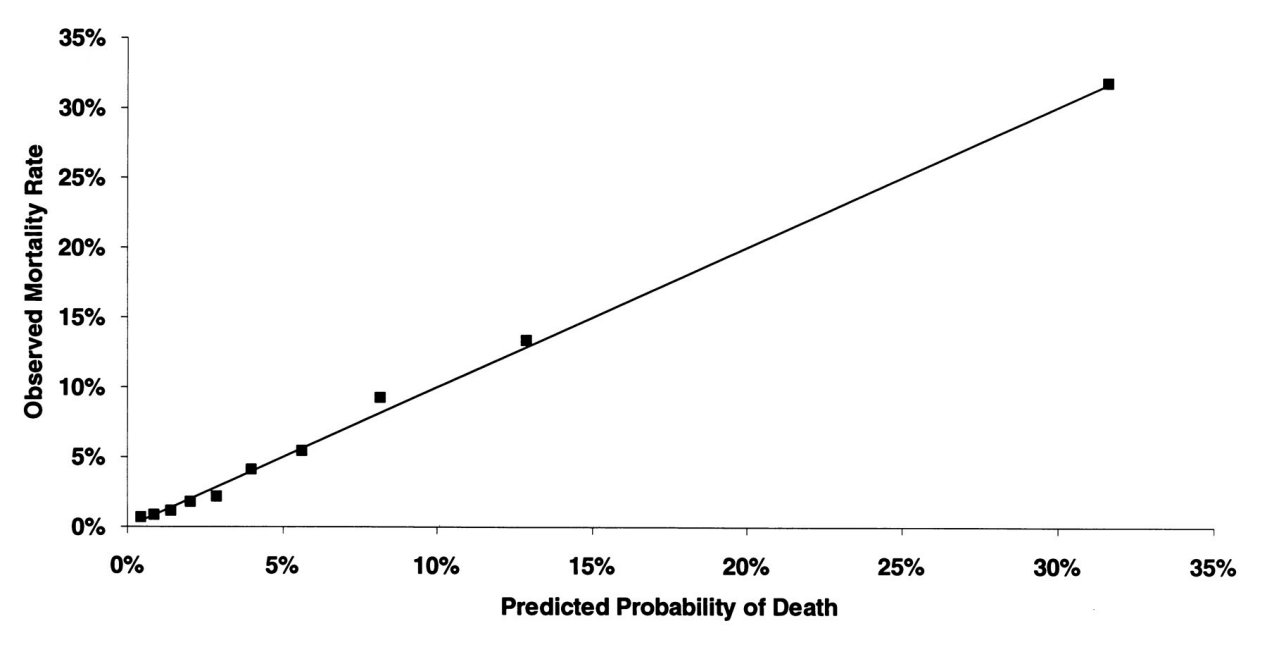
Compared with VKA use, DOAC use was associated with a lower AKI risk, with an **adjusted HR** of 0.88 (95% CI, 0.80-0.97).

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Model's performance

Discrimination and **calibration** were studied to indicate model performance.

The full study population was used in the model development process, and the predictive performance of the model was **internally validated** through cross validation and bootstrapping. First, 10-fold cross validation was performed: the model was fitted on a randomly selected subset of 90% of the study patients, and the resulting fit was tested on the remaining 10%. This process was repeated 10 times to estimate the extent to which the predictive accuracy of the model (based on the entire sample) was overoptimistic.

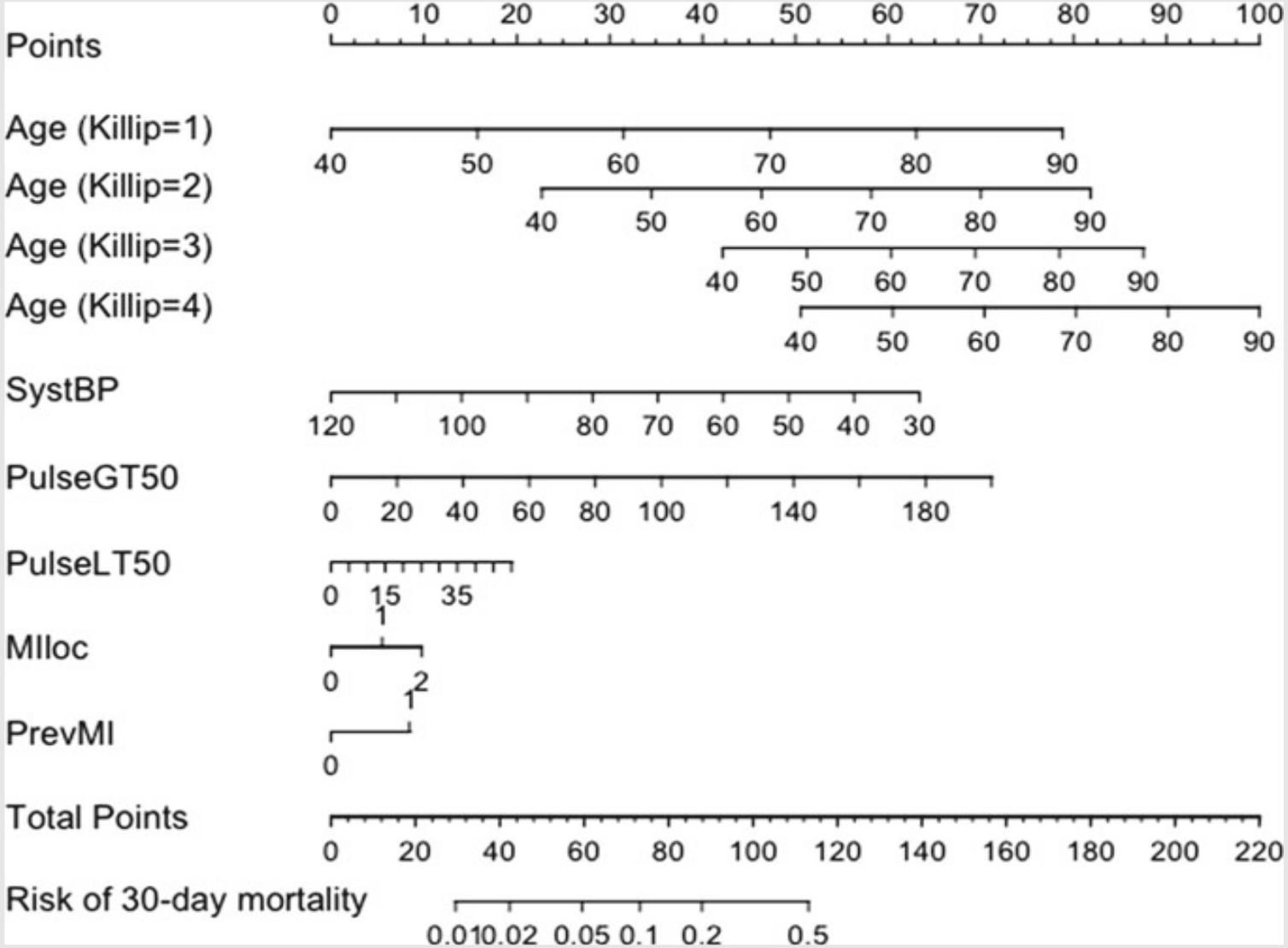


The correction to the Receiver Operating Characteristic area determined by cross validation was only 0.002 (reducing the AUC value from 0.836 to 0.834)

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Model's presentation for clinical audience

Seven (out of 15) predictors were used to derive a simpler summary score:



Age and Killip class were included as main effects and with **interaction** terms.

At younger ages, Killip class makes a substantial difference.
Equivalently, age matters among those with Killip class I, but less among those with higher Killip classes.

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

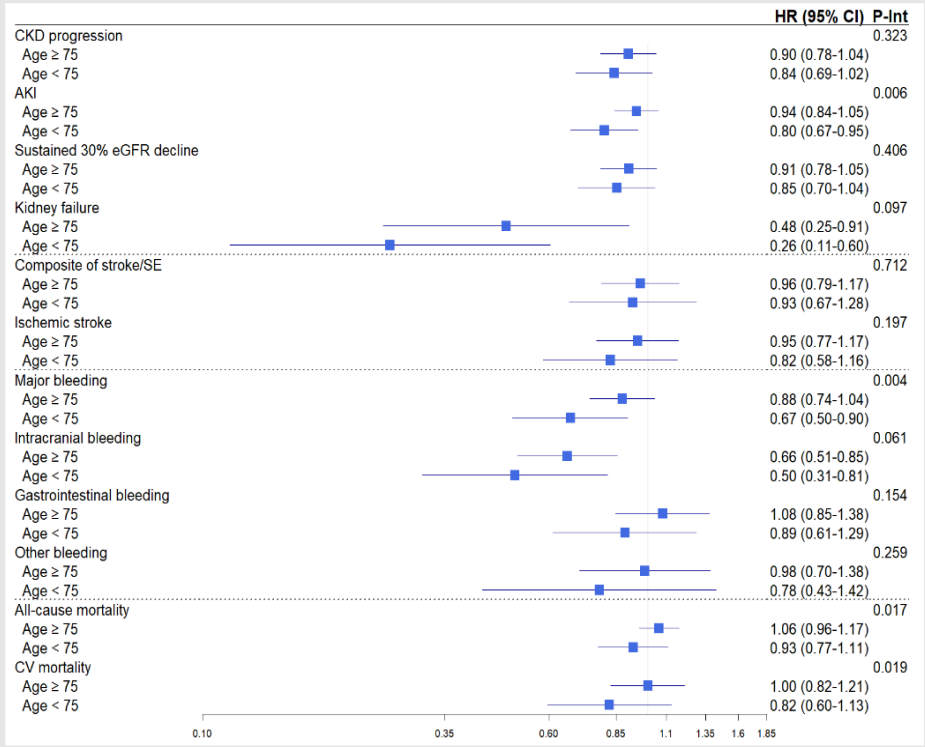
Subgroup and Sensitivity Analyses

After **accounting for the propensity of discontinuing/switching**, DOAC use was still associated with a lower risk of CKD progression (HR, 0.77 [95% CI, 0.64-0.92]) and of AKI (HR, 0.79 [95% CI, 0.71-0.89]) compared with VKA.

Stratified analyses for age/sex/GFR categories on the weighted cohorts.

Table S8. Number of events, incidence rates and adjusted hazard ratios for the association between DOAC vs VKA initiation and outcomes accounting for treatment switch and discontinuation.

	VKA: No of Events (IR/ 1000 person-years)*	DOAC: No of Events (IR/1000 person-years)*	Adjusted HR DOAC vs. VKA (95% CI)**
Kidney outcomes			
CKD progression	1335 (40.4)	1066 (32.0)	0.77 (0.64-0.92)
Sustained 30% eGFR decline	1305 (39.5)	1064 (32.0)	0.79 (0.66-0.95)
Kidney Failure	120 (3.5)	29 (0.8)	0.30 (0.15-0.59)
AKI	1868 (57.0)	1563 (46.7)	0.79 (0.71-0.89)



https://www.ajkd.org/article/S0272-6386(22)00922-2/fulltext

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Keep home messages

1

To be useful, a **prediction model** should include **all clinically relevant prognostic indicators** and should be derived from a sample of a **target population** that represents the types of patients seen in clinical practice so that reliable estimates of **true risk predictions** can be assessed.

A useful model should appropriately weight **clinically relevant predictors** and be then **externally validated** in a population with a broad spectrum of patients and hospital settings.

Pay attention that a sufficient **sample size** is available !!

2

To be useful, a **causal model** should include **all clinically relevant confounders** and should be derived from a sample of a **target population** that represents the types of patients seen in clinical practice so that reliable estimates of **true effect of the treatment/exposure** can be assessed.

A useful model should appropriately balance for the **confounders** and **sensitivity analyses** w.r.t. the causal assumptions should be done.

Pay attention that a sufficient **sample size** is available !!

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute