



LUNDS UNIVERSITET

Statistiska institutionen

Metoder för prediktion av kardiovaskulär sjukdom med hjälp av njurfunktionen

Ronnie Pingel

Magisteruppsats
Nivå 91-120hp
VT 2008
Handledare: Jan Lanke

Sammanfattning

I uppsatsen undersöks i vilken utsträckning prediktionen av kardiovaskulär sjukdom bland hypertonipatienter förbättras när mått på njurfunktionen och mikroalbuminuri adderas till de klassiska riskfaktorerna.

För att undersöka prediktionsförmågan mäts diskrimineringsförmågan, kalibreringen, reklassificering och Harrells C .

Resultaten är inte entydiga. Mikroalbuminuri bör i regel inkluderas, men vilka av de övriga måtten på njurfunktion som bör väljas varierar. Vilka riskfaktorer som väljs beror därför på vilket mått av prediktionsförmåga som prioriteras.

Nyckelord: Cox-regression, diskriminering, Harrells C , kalibrering, kardiovaskulär risk, njurfunktion, prediktion, reklassificering

Abstract

This study examines if the prediction of cardiovascular disease in hypertensive patients can be improved upon when renal function and microalbuminuria are added to the classical risk factors

The predictive capability of a model is measured by discrimination, calibration, reclassification and Harrell's C .

The results are ambiguous. In most cases, microalbuminuria should be included in the model, but the results regarding the other measures of renal function are varied. Therefore, the selection of risk factors to include in the model depends on which measure of prediction one prioritizes.

Key words: calibration, cardiovascular risk, Cox-regression, discrimination, Harrell's C, prediction, reclassification, renal function

Innehåll

1	Introduktion	2
1.1	Frågeställning och syfte	2
2	Analys av överlevnadsdata	3
2.1	Överlevnadsfunktion och intensitetsfunktion	3
2.2	Cox proportionella hasardmodell	4
2.3	Skattningar med Cox proportionella hasardmodell	5
3	Metoder för att utvärdera prediktion	7
3.1	Metoder för att utvärdera prediktion av dikotoma utfall	7
3.1.1	Diskriminering - Receiver Operating Characteristic	7
3.1.2	Kalibrering	9
3.1.3	Reklassificering	10
3.2	Prediktionsutvärdering som tar hänsyn till tiden	11
4	Data	14
4.1	Variabler	14
4.2	Tidigare analys av njurfunktion och risk för kardiovaskulär sjukdom	15
5	Resultat	18
5.1	Prediktion av dikotoma utfall	18
5.1.1	Area under ROC-kurvan	19
5.1.2	Kalibrering	21
5.1.3	Reklassificering	22
5.2	Prediktionsutvärdering som tar hänsyn till tiden	23
6	Diskussion	26
A	Do-filer	29

Kapitel 1

Introduktion

1.1 Frågeställning och syfte

Hypertoni (høgt blodtryck) är en frekvent förekommande kardiovaskulär riskfaktor och blodtryckssänkande behandling är därför en primär åtgärd vad gäller att förhindra uppkomsten av kardiovaskulär sjukdom. För att fatta beslut om åtgärder bör emellertid den totala kardiovaskulära riskbilden utvärderas. Därför söker man efter nya och lätthanterliga kardiovaskulära riskmarkörer som kan medföra förbättrad riskutvärdering av hypertoni-patienter. Exempel på sådana riskmarkörer är olika mått på njurfunktion.

Det finns idag stöd för att förekomsten av *mikroalbuminuri* (MA) predikterar kardiovaskulär sjukdom bland hypertoni-patienter. Det finns även evidens för att mängden *serumkreatinin* (S-kre) samt att den *glomerulära filtrations-hastigheten* (GFR) beräknad enligt Cockcroft & Gaults formel (GFR_{CRG}) och MDRD-formeln (GFR_{MDRD}) predikterar kardiovaskulär sjukdom bland hypertoni-patienter.

Emellertid är både förekomsten av MA och minskad GFR korrelerade med klassiska kardiovaskulära riskfaktorer. Därför måste en modell skapas som tar hänsyn till både de nya riskfaktorerna – MA, S-Kre, GFR_{CRG} samt GFR_{MDRD} – och de klassiska.

Ovanstående leder fram till följande frågeställning: Hur mycket bättre blir prediktionen av kardiovaskulär sjukdom bland hypertoni-patienter om man i riskmodellen inkluderar kombinationer av eller enstaka mått på njurfunktion?

Uppsatsens syfte är att besvara denna frågeställning.

Kapitel 2

Analys av överlevnadsdata

Analys av överlevnadsdata kräver specifika statistiska verktyg. För det första är livslängder sällan normalfördelade. För det andra måste man ta hänsyn till att datamaterialet ofta är censurerat. Censurering innebär att händelser inträffar som vi inte kan observera, till exempel om individer på ett eller annat sätt försvinner ur studien innan studien är slutförd eller om man helt enkelt inte har resurser att följa alla individer tills dess att en specificerad händelse inträffar.

Notera att man med livslängd inte syftar på livslängd i kalendertid utan på tid sedan en definierad startpunkt. Vidare behöver inte livslängd innebära specifikt tid till död, utan metoden används även för att analysera tid till någon annan väldefinierad händelse. Begreppet död används således ofta som en allmän beskrivning av händelse.

2.1 Överlevnadsfunktion och intensitetsfunktion

Låt T vara en positiv slumpvariabel som bestämmer livslängden. Sannolikheten att en händelse, δ (exempelvis dödsfall), inte inträffar före tidpunkten t beskrivs med hjälp av *överlevnadsfunktionen*

$$S(t) = P(T > t) = 1 - F(t), \quad t > 0, \quad (2.1)$$

där $F(t)$ är den kumulativa fördelningsfunktionen för T .

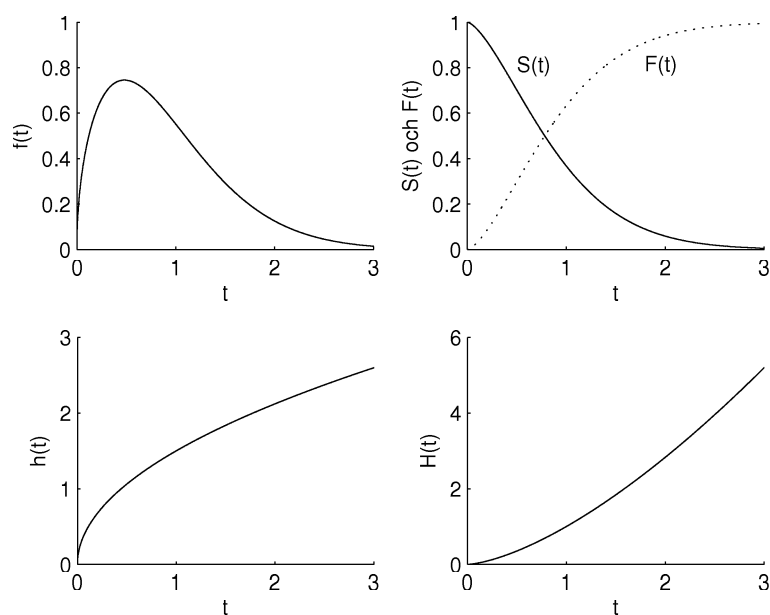
Täthetsfunktionen $f(t)$ för T erhålles från $S(t)$ genom

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}\{1 - S(t)\} = -S'(t). \quad (2.2)$$

Vidare har vi *intensitetsfunktionen* $h(t)$, vilken anger det omedelbara döds-talet vid tidsperioden t , det vill säga sannolikhetstätheten att dö före $t + \Delta t$, givet att individen överlevt till t . Relationen mellan $h(t)$ och $S(t)$ kan beskrivas med hjälp av följande resonemang.

Betrakta sannolikheten att T ligger mellan t och $t + \Delta t$ givet att T är större eller lika med t . Intensitetsfunktionen $h(t)$ är då gränsvärdet för denna sannolikhet dividerad med Δt när Δt går mot noll, det vill säga

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}. \quad (2.3)$$



Figur 2.1: Täthetsfunktion $f(t)$, överlevnadsfunktion $S(t)$ (heldragen), fördelningsfunktion $F(t)$ (streckad), intensitetsfunktion $h(t)$ och kumulativ intensitetsfunktion $H(t)$ för en Weibull-fördelad slumpvariabel med parametrarna $k = 1,5$ och $\lambda = 1$.

Utifrån formlerna för betingad sannolikhet kan detta skrivas som

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)} = \frac{f(t)}{S(t)}. \quad (2.4)$$

Vidare har vi den kumulativa intensitetsfunktionen, $H(t)$, vilken definieras av

$$H(t) = \int_0^t h(u) du \quad (2.5)$$

och som mäter den totala mängden risk som ackumulerats fram till tiden t . Det innebär att

$$H(t) = \int_0^t \frac{f(u)}{S(u)} du = - \int_0^t \frac{1}{S(u)} \left\{ \frac{d}{dt} S(u) \right\} du = - \ln\{S(t)\}. \quad (2.6)$$

Således kan vi formulera följande likheter

$$S(t) = \exp\{-H(t)\} \quad (2.7)$$

$$F(t) = 1 - \exp\{-H(t)\} \quad (2.8)$$

$$f(t) = h(t) \exp\{-H(t)\}. \quad (2.9)$$

Vi har nu kort beskrivit elementära verktyg inom överlevnadsanalys. Exempel på funktionerna åskådliggörs grafiskt i Figur 2.1.

2.2 Cox proportionella hasardmodell

Ofta önskar man undersöka förklaringsvariablers relation till framförallt $S(t)$ och $h(t)$. Ett exempel är om patienter randomiseras för att erhålla en utav två typer av behandling. Då förklaringsvariabeln är binär kan relationen

undersökas genom att grafiskt jämföra $S(t)$ och $h(t)$ för respektive behandlingsgrupp.

En alternativ metod, som även är generaliserbar till att omfatta ytterligare förklaringsvariabler, är *Cox proportionella hasardmodell*. Collett (1994) beskriver Cox proportionella hasardmodell enligt följande:

Låt $h_S(t)$ och $h_N(t)$ vara dödsintensiteten vid tiden t för patienter i två behandlingsgrupper – en grupp som får standardbehandling och en grupp som erhåller en ny behandlingsform. Till grund för modellen ligger antagandet att intensiteten vid t för patienter som får den nya behandlingen är proportionell mot intensiteten vid samma tid för gruppen som erhåller standardbehandlingen, det vill säga

$$h_N(t) = \psi h_S(t), \quad t \geq 0.$$

Konstanten ψ är den *relativa intensiteten* och om $\psi < 1$ så är dödsintensiteten vid t mindre för gruppen som erhåller den nya behandlingen relativt gruppen som får standardbehandlingen. Det innebär att den nya behandlingen i så fall har en positiv effekt. Om $\psi > 1$ gäller det motsatta.

Ovanstående beskriver hasardmodellen i dess mest elementära form. Anta nu n individer, där respektive individs intensitetsfunktion ges av $h_i(t)$, $i = 1, 2, \dots, n$. Intensitetsfunktionen för en individ som erhåller standardbehandling betecknas $h_0(t)$ och intensitetsfunktionen för en individ som får ny behandlingsform betecknas $h_0(t)\psi$. Eftersom ψ inte kan vara negativ kan vi skriva $\psi = \exp(\beta)$.

Nu kan följande intensitetsfunktion för individ i formuleras:

$$h_i(t) = \exp(\beta x_i) h_0(t), \quad (2.10)$$

där variabeln x_i beskriver individ i 's behandlingsform: $x_i = 0$ om patienten får standardbehandlingen och $x_i = 1$ för patienten som får den nya behandlingen. Modellen kan vidare utökas till att omfatta p förklaringsvariabler, x_1, x_2, \dots, x_p :

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) h_0(t). \quad (2.11)$$

Om kovariatvektorn $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ sätts till 0 får vi *basintensitetsfunktionen*, $h_0(t)$ som enbart är en funktion av tiden.

2.3 Skattningar med Cox proportionella hasardmodell

Låt oss nu se på hur β -koefficienterna estimeras. Anta n individer, bland vilka det finns r unika död tidpunkter. Anta vidare att det finns $n - r$ censurerade överlevnadstider där information saknas om när en händelse inträffar. Censurering betyder i uppsatsens kontext högercensurering, det vill säga givet att individen randomiserades vid $t = 0$ och senast observerades vid t vet vi bara att livslängden är $T > t$. Ovanstående innebär att vi saknar ties (individer som delar händelsetidpunkt).

De r ordnade död tidpunkterna betecknas $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ där $t_{(j)}$ är den j :te ordnade död tiden. Uppsättningen individer under risk vid tidpunkten $t_{(j)}$ betecknas $R(t_{(j)})$ och består av individerna som lever och är ocensurerade

vid tiden precis för $t_{(j)}$. Nu kan β -koefficienterna estimeras genom att man maximerar den partiella likelihood-funktionen

$$PL(\beta) = \prod_{j=1}^r \frac{\exp(\beta' \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' \mathbf{x}_l)}, \quad (2.12)$$

där $\mathbf{x}_{(j)}$ är kovariatvektorn för den individ som dör vid den j :te ordnade död tiden, $t_{(j)}$. Nämnaren summerar värdena av $\exp(\beta' \mathbf{x})$ för alla individer som är utsatta för en dödsrisk vid tiden $t_{(j)}$.

Funktion (2.12) benämns partiell likelihood-funktion då funktionen endast använder sig av information om händelsernas ordning och inte tar hänsyn till de faktiska tidpunkterna då en händelse inträffar.

Ett problem som i regel uppkommer är ties, något som hittills ignoreras. Man kan enkelt konstatera att det enbart är bristande precision i mätningen som gör att individer delar död tidpunkt. Även om två individer har lika många observerade dagar så dör i verkligheten en individ först och sedan en annan, vi vet bara inte i vilken ordning. Följaktligen måste funktion (2.12) modifieras. Beräkningar som inkluderar ties blir dock omfattande, därför har olika metoder för att approximera den modifierade likelihood-funktionen tagits fram. Dessa fungerar väl om andelen tied observationer vid en given död tidpunkt är relativt liten. Se Collett (1994, s. 65-66) för en beskrivning av sådana approximationer. **Stata 10** använder som default Breslows metod för ties (vilket används i den kommande analysen).

När β -koefficienterna är estimerade kan den skattade basöverlevnads-funktionen, $\hat{S}_0(t)$ erhållas, vilken är basöverlevnadssfunktionen när alla $x_i = 0$. I **Stata** erhålls $\hat{S}_0(t)$ med optionen `basesurv()`. Sedan används $\hat{S}_0(t)$ tillsammans med de skattade β -koefficienterna för att beräkna den skattade överlevnadsfunktionen för individ i :

$$\hat{S}_i(t) = [\hat{S}_0(t)]^{\exp(\hat{\beta}' \mathbf{x}_i)}. \quad (2.13)$$

Kapitel 3

Metoder för att utvärdera prediktion

Vi kommer i kapitlet behandla följande metoder som antingen utvärderar en modells förmåga att prediktera händelser eller jämför prediktionsförmågan hos två modeller:

- *Diskriminering* är ett mått på en modells förmåga att klassificera individer i två typer av kategorier och inkluderar mått som arean under ROC-kurvan (Receiver Operating Characteristic).
- *Kalibrering* mäter hur väl predikterade sannolikheter överensstämmer med observerad risk. I uppsatsen utvärderas kalibreringen med Hosmer-Lemeshows goodness-of-fit-test.
- *Reklassificering* mäter förbättrad riskkategorisering när en ny riskfaktor adderas till modellen. Genom att jämföra andelen individer som förbättrar respektive försämrar sin riskklass givet att de verkligen drabbas eller inte drabbas av en händelse erhålls en nettoreklassificeringsförbättring.
- *Rangkorrelation* används för att studera sambandet mellan två variabler som var och en innebär en rangordning, i vårt fall predikterad tid och observerad tid till händelse. Som specifik metod används Harrells C , som kan sägas vara en generalisering av rangkorrelation i fallet att man har censurerade data; noga taget är det inte C utan $2C - 1$ som kan uppfattas som en korrelation.

Ovanstående metoder kan indelas i dels metoder som analyserar dikotoma utfall, dels metoder som explicit inkluderar tidsaspekten i utvärderingen.

3.1 Metoder för att utvärdera prediktion av dikotoma utfall

3.1.1 Diskriminering - Receiver Operating Characteristic

Diskriminering mäter en modells förmåga att klassificera individer i två typer av kategorier – individer som drabbas av en händelse, δ , respektive inte drabbas av en händelse, δ^c före en viss tidpunkt, t_0 . Perfekt diskriminering råder om modellen placerar alla som råkar ut för δ i rätt kategori samt även alla δ^c i

		Sann klassificering	
		δ	δ^c
Predikterad klassificering	δ_{pred}	Sann positiv TP	Falsk positiv FP
	δ^c_{pred}	Falsk negativ FN	Sann negativ TN

Figur 3.1: Klassificeringsschema

rätt kategori. Om prediktionen är positiv, δ_{pred} , och överensstämmer med att individen de facto drabbas av δ har vi en sann positiv prediktion. Analogt innebär en positiv prediktion, δ_{pred} , där individen inte drabbas av δ , en falsk positiv prediktion.

Figur 3.1 visar schematiskt alternativen och utifrån den kan följande mått i analysen beräknas:

$$\text{Sensitivitet (Se)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.1)$$

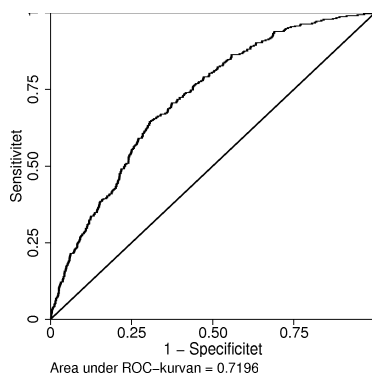
$$\text{Specificitet (Sp)} = \frac{\text{TN}}{\text{FP} + \text{TN}} = 1 - \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (3.2)$$

Ofta används en kontinuerlig variabel för att prediktera ett dikotomt utfall. Huruvida man predikterar att en individ drabbas av δ eller δ^c beror då på vilket tröskelvärde som väljs. Om man har en kontinuerlig variabel måste nämligen ett värde väljas som bestämmer gränsen för vilken kategori en individ placeras i. Väljs ett lågt tröskelvärde kommer man prediktera att många individer drabbas av δ . Detta ger hög Se men låg Sp. På motsvarande vis leder ett högt tröskelvärde till att få patienter predikteras att drabbas av δ , något som ger låg Se och hög Sp.

Genom att för olika tröskelvärden plotta andelen sant positiva prediktioner bland individer som drabbas av en händelse mot andelen falskt positiva bland individer som inte drabbas av en händelse – det vill säga Se mot $1 - \text{Sp}$ – kan modellens förmåga att diskriminera summeras i en ROC-kurva (se Figur 3.2 på nästa sida). En god modell som diskriminerar väl genererar en kurva som går nära det övre vänstra hörnet och man kan således välja tröskelvärdet så att man får både hög sensitivitet och hög specificitet. Arealen under en sådan kurva är nära 1, där 1 betyder perfekt diskriminering. Om kurvan sammanfaller med diagonallinjen, och arean under kurvan således är 0,5, diskriminerar modellen inte bättre än slumpen vid slantsingling.

Arealen under kurvan (AUC) kan visas vara sannolikheten att modellen rangordnar en slumpmässigt vald individ som drabbas av δ , högre än en slumpmässigt vald individ som inte drabbas av δ .

I Figur 3.2 på nästa sida är $\text{AUC} = 0,7196$. Anta att vi slumpmässigt drar ett par av individer från den undersökta populationen, så att paret innehåller en individ som drabbas av en händelse och en individ som inte drabbas av en händelse. Det är då 72 procent chans att modellen rangordnar individen som drabbas av en händelse högre än individen som inte drabbas av en händelse.



Figur 3.2: Exempel på en ROC-kurva från en logistisk regression baserad på datamaterialet presenterat i kapitel 4.

Om modellens prediktioner överstämmer med observationerna betecknas paret *konkordant* eller samstämmigt.

3.1.2 Kalibrering

Kalibreringen mäter hur väl de predikterade sannolikheterna att drabbas av en händelse överensstämmer med de faktiska sannolikheterna. För kalibrering används Hosmer-Lemeshows goodness-of-fit-test som grupperar individerna baserat på de predikterade sannolikheterna (Hosmer & Lemeshow 1989, s. 140-145).

Låt $\hat{F}_i(t_0)$ vara den predikterade sannolikheten att individ i senast tidpunkten t_0 drabbas av δ . Sortera individerna efter \hat{F}_i .

I huvudsak finns två strategier för klassindelning av individerna – antingen i riskklasser vilka är percentiler av antalet individer när de är sorterade enligt \hat{F}_i eller efter fasta värden av den predikterade sannolikheten. Klassindelning baserat på percentiler är att föredra om \hat{F}_i är små, därför används denna metod i analysen.

Vi väljer g riskklasser. Det förväntade antalet δ i respektive riskklass betecknas E_k , där $k = 1, \dots, g$. Då är

$$E_k = \sum_{i \in \text{riskklass } k} \hat{F}_i$$

Genom att jämföra E_k med det observerade antalet δ inom varje riskklass, O_k , kan vi utvärdera hur väl modellen överensstämmer med verkligheten. Följande testfunktion används:

$$X_{HL}^2 = \sum_{k=1}^g \frac{(O_k - E_k)^2}{n_k \frac{E_k}{n_k} (1 - \frac{E_k}{n_k})}, \quad (3.3)$$

där n_k är antalet individer i respektive riskklass.

Till skillnad från det vanliga χ^2 -testet är antalet frihetsgrader vid Hosmer-Lemeshows goodness-of-fit-test alltid $g - 2$, oberoende av hur många parametrar som estimeras. I vårt fall används $g = 10$, som är den vanligaste indelningen.

3.1.3 Reklassificering

Eftersom mått på diskriminering och kalibrering är svåra att förmedla i termer av klinisk relevans presenterar Pencina et al. (2008) en reklassificeringsmetod för att utvärdera förbättringen i prediktion.

Anta att alla individer kan klassificeras i ordinala riskkategorier att drabbas av en händelse, δ (exempelvis 0 – 6, 6 – 20, > 20 procents risk att inom 10 år drabbas av kardiovaskulär sjukdom).

Indela individerna i två kateogier – invidider som råkar ut för δ och individer som inte råkar ut för δ . Steg upp (\uparrow) till en högre riskkategori bland de individer som drabbas av δ indikerar förbättrad klassificering, medan steg nedåt (\downarrow) i samma grupp individer resulterar i en sämre klassificering. Motsvarande analys görs för individer som inte drabbas av en händelse, men då med motsatt tolkning av klassificeringsresultatet.

Anta en situation där två modeller delar alla riskfaktorer förutom en ny adderad riskfaktor i den ena modellen. Kategorisera individerna enligt riskkategorierna och korstabulera klassificeringarna. Tilldela sedan individer värdet 1 om de rör sig uppåt i riskkategorierna, värdet -1 för steg nedåt och 0 om ingen förändring sker. Den totala förbättringen av reklassificering kan då uttryckas som

$$\text{NRI} = [\text{P}(\uparrow | \delta) - \text{P}(\downarrow | \delta)] - [\text{P}(\uparrow | \delta^c) - \text{P}(\downarrow | \delta^c)] \quad (3.4)$$

där δ indikerar händelsen.

För att estimeras nettoreklassificeringsindexet (NRI) med hjälp av stickprovsdata måste de ingående komponenterna i (3.4) estimeras:

$$\begin{aligned} \hat{P}(\uparrow | \delta) &= \hat{p}_{\uparrow, \delta} = \frac{\# \delta \uparrow}{\# \delta} \\ \hat{P}(\downarrow | \delta) &= \hat{p}_{\downarrow, \delta} = \frac{\# \delta \downarrow}{\# \delta} \\ \hat{P}(\uparrow | \delta^c) &= \hat{p}_{\uparrow, \delta^c} = \frac{\# \delta^c \uparrow}{\# \delta^c} \\ \hat{P}(\downarrow | \delta^c) &= \hat{p}_{\downarrow, \delta^c} = \frac{\# \delta^c \downarrow}{\# \delta^c} \end{aligned}$$

NRI kan nu estimeras med

$$\widehat{\text{NRI}} = (\hat{p}_{\uparrow, \delta} - \hat{p}_{\downarrow, \delta}) - (\hat{p}_{\uparrow, \delta^c} - \hat{p}_{\downarrow, \delta^c}) \quad (3.5)$$

Klassificeringsförbättringen är andelen individer som rör sig uppåt minus andelen individer som rör sig nedåt bland individer som drabbas av händelsen, minus proportionen individer som rör sig uppåt plus proportionen som rör sig nedåt bland individer som inte drabbas av händelsen.

För att veta om reklassificeringen är signifikant kan estimatorn testas mot nollhypotesen $\text{NRI} = 0$ med

$$z = \frac{\widehat{\text{NRI}}}{\sqrt{\frac{\hat{p}_{\uparrow, \delta} + \hat{p}_{\downarrow, \delta}}{\# \delta} + \frac{\hat{p}_{\uparrow, \delta^c} + \hat{p}_{\downarrow, \delta^c}}{\# \delta^c}}} \quad (3.6)$$

Även de enskilda komponenterna i estimatorn kan testas:

$$z_\delta = \frac{\hat{p}_{\uparrow,\delta} - \hat{p}_{\downarrow,\delta}}{\sqrt{\frac{\hat{p}_{\uparrow,\delta} + \hat{p}_{\downarrow,\delta}}{\# \delta}}} \quad (3.7)$$

$$z_{\delta^c} = \frac{\hat{p}_{\downarrow,\delta^c} - \hat{p}_{\uparrow,\delta^c}}{\sqrt{\frac{\hat{p}_{\downarrow,\delta^c} + \hat{p}_{\uparrow,\delta^c}}{\# \delta^c}}}. \quad (3.8)$$

3.2 Prediktionsutvärdering som tar hänsyn till tiden

De i avsnitt 3.1 beskrivna metoderna använder bara en del av den information som finns i datamaterialet: det enda som utnyttjas beträffande patient i är huruvida händelsen, δ , inträffar eller ej inträffar före en viss tidpunkt, t_0 . Det innebär exempelvis att individer som kommer sent in i en studie inte kan utnyttjas. Vi ska nu visa en metod som utnyttjar datamaterialet på ett annat, mera effektivt sätt.

Utgångspunkten för den nya metoden är att man med hjälp av Cox-regressionen får en prediktion av i vilken ordning individerna kommer att dö. Cox-regression ger oss nämligen den relativa risken, eller hazardkvoten, för individ i och ju lägre denna är, dess längre kan individ i förväntas leva. Det är då naturligt att jämföra denna ordning med den verkliga observerade ordningen. Ett problem är dock att vissa individer levererar censurerade dödstillfällen och den verkliga ordningen är alltså inte fullständigt känd.

Låt oss först beskriva hur man kan mäta överensstämmelsen mellan två rangordningar. Vi har ett datamaterial (x_i, x_j) , $1 \leq i < j \leq n$, och vill få ett mått på hur väl den rangordning som induceras av x -variablerna stämmer överens med den som y -värdena ger. Ett sätt att göra detta är att beräkna Kendalls rangkorrelation τ (Everitt 1992).

Idén är att man betraktar alla de $N = n(n-1)/2$ paren

$$(x_i, y_i) \text{ och } (x_j, y_j), \quad 1 \leq i < j \leq n$$

av par och ser efter hur många av dem som är konkordanta och hur många som är diskordanta. Därvid sägs (x_i, y_i) och (x_j, y_j) vara konkordanta om det av de två paren som har störst x -värde också har störst y -värde, det vill säga om

$$(x_i - x_j)(y_i - y_j) > 0.$$

I motsatt fall, det vill säga om produkten är negativ, sägs de två paren vara diskordanta.

Låt P vara antalet konkordanta par och Q antalet diskordanta par. I det enklaste fallet är alla x_i olika, likaså alla y_j . Då är $P + Q = N$ och Kendalls rangkorrelation definieras då:

$$\tau = \frac{P - Q}{N} \quad (3.9)$$

som mäter sambanden mellan två rangordningar. Uppenbarligen blir $-1 \leq \tau \leq 1$, där både -1 och 1 är tänkbara värden. $\tau = 1$ om x -värdena och y -värdena ger precis samma rangordning, medan $\tau = -1$ om rangordningarna är precis de omvända.

Om det föreligger ties bland x -värdena och/eller y -värdena blir saken mer komplicerad eftersom ekvation (3.9) inte tar hänsyn till ties. τ kan då modifieras till

$$\tau_b = \frac{P - Q}{\sqrt{[(P + Q + X_0)(P + Q + Y_0)]}} \quad (3.10)$$

som tar hänsyn till ties. X_0 representerar antalet par av observationer som är tied endast på första variablen, det vill säga där $\#(x_i = x_j)$, och Y_0 antalet observationer av par som är tied på bara den andra variabeln, $\#(y_i = y_j)$.

Tolkningen av τ_b går dock inte att uttrycka i termer av sannolikhet. Ytterligare ett alternativ för att mäta graden av samband mellan rangordningar ges därför av Goodman och Kruskals γ . Givet att ett par inte är tied på någon av variablerna är $P/(P + Q)$ sannolikheten för konkordans och $Q/(P + Q)$ sannolikheten för diskordans. γ är då skillnaden mellan dessa sannolikheter.

$$\gamma = \frac{P - Q}{P + Q}, \quad (3.11)$$

där $-1 \leq \gamma \leq 1$.

Om ett datamaterial inte innehåller några ties är det tydligt att $\tau = \tau_b = \gamma$.

Nu har vi komplikationen att död tidpunkterna bara inducerar en partiell ordning. Harrel et al. (1982) och Harrell et al. (1996) har föreslagit följande mått för att komma till rätta med detta: Anta att vi följer en kohort av n individer och att individerna kan beskrivas med följande uppsättning variabler:

$$(t_i, d_i, hr_i) \quad i = 1, \dots, n,$$

där t_i är den observerade livslängden hos individ i och hr_i är individ i 's hasardkvot (det vill säga kvoten mellan två intensitetsfunktioner). $d_i = 1$ om individ i dör och $d_i = 0$ annars.

Från n individer kan $n(n - 1)/2$ par, (i, j) , väljas. För att ett par ska vara *användbart* för analysen måste minst en individ i paret utveckla en händelse. Om hasardkvoten är lägre för individen som levde längre betecknas prediktionerna för det paret vara konkordanta med utfallen. (i, j) är således konkordant och användbart för analysen om

$$\begin{aligned} t_i < t_j \ \& \ d_i = 1 \ \& \ hr_i > hr_j \ \text{eller} \\ t_i > t_j \ \& \ d_j = 1 \ \& \ hr_i < hr_j \ \text{eller} \\ t_i = t_j \ \& \ d_i = 1 \ \& \ d_j = 0 \ \& \ hr_i > hr_j \ \text{eller} \\ t_i = t_j \ \& \ d_i = 0 \ \& \ d_j = 1 \ \& \ hr_i < hr_j \end{aligned}$$

Analogt är (i, j) diskordant och användbart för analysen om

$$\begin{aligned} t_i < t_j \ \& \ d_i = 1 \ \& \ hr_i < hr_j \ \text{eller} \\ t_i > t_j \ \& \ d_j = 1 \ \& \ hr_i > hr_j \ \text{eller} \\ t_i = t_j \ \& \ d_i = 1 \ \& \ d_j = 0 \ \& \ hr_i < hr_j \ \text{eller} \\ t_i = t_j \ \& \ d_i = 0 \ \& \ d_j = 1 \ \& \ hr_i > hr_j \end{aligned}$$

Resterande par är inte användbara för analysen och (i, j) är *oanvändbart* om:

- (i, j) delar exakt samma observerad död tidpunkt.
- en individ i paret dog och den andra individen censurerades, men den observerade tiden i studien hos den censurerade individen är mindre än tiden till händelse hos individen som dog.
- båda individerna i paret är censurerade.

Om

$$hr_i = hr_j$$

är paret fortfarande användbart och i en sådan situation adderas $\frac{1}{2}$ istället för 1 till antalet konkordanta par.

Harrells C är proportionen konkordanta par av alla användbara par och definieras (enligt **Stata**) av:

$$C = \frac{E + \frac{1}{2}T}{U}, \quad (3.12)$$

där E är antalet konkordanta par, T är antalet ties bland användbara par, det vill säga par där $hr_i = hr_j$. Vidare är U antalet användbara par och $U = E + D + T$ där D är antalet diskordanta par.

Resultatet blir ett mått som estimerar sannolikheten för konkordans mellan predikterade och observerade utfall. 1 innebär perfekt förmåga att prediktera en rangordning medan 0,5 betyder att modellen inte har någon prediktiv förmåga. Värt att nämna är att i fallet med fullständig ordning är C identisk med AUC.

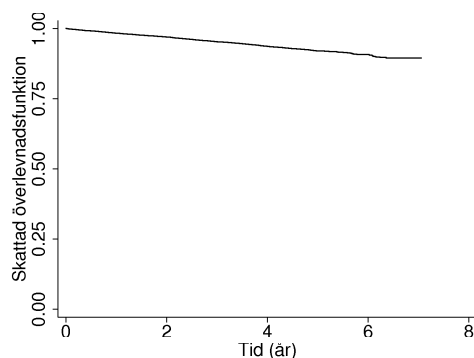
Kapitel 4

Data

Datamaterialet kommer från *The Nordic Diltiazem Study* (NORDIL). I studien randomiserades 10 881 svenska och norska hypertoni-patienter till endera av två typer av blodtryckssänkande behandling – diltiazembaserad behandling ($n = 5410$) eller diuretika och/eller β -blockerande behandling ($n = 5471$). Hypertoni-patient definierades som patient med diastoliskt blodtryck ≥ 100 mmHg vid minst två tillfällen. Studien inleddes 1992 och den genomsnittliga uppföljningstiden var 4,5 år.

4.1 Variabler

Studiens primära ändpunkter (PE) var dödlig och icke-dödlig hjärtinfarkt, annan död i hjärt- eller kärlsjukdom samt dödlig och icke-dödlig stroke. Endast tiden till patienternas första PE analyseras och totalt drabbades 802 patienter av primära ändpunkter. Figur 4.1 visar patienternas skattade överlevnadsfunktion, i det här fallet sannolikheten att inte drabbas av PE inom en viss tid.



Figur 4.1: Den skattade överlevnadsfunktionen $\hat{S}(t)$ bland de undersökta hypertoni-patienterna.

Njurfunktionen mättes med följande fyra mått:

- **Mikroalbuminuri (MA)**. Om nivån ≥ 20 $\mu\text{mol/l}$ kodas detta som 1. I övriga fall är koden 0.
- **Serumkreatinin (S-kre)** ($\mu\text{mol/l}$). I analysen används variabeln $S_KRE_transf = S_kre/10$.

- **Glomerulär filtrationshastighet enligt Cockcroft & Gaults formel** (GFR_{CRG}). Formeln är

$$GFR_{CRG} = \frac{(140 - \text{Ålder}) \times \text{Kroppsvikt(kg)}}{\theta \times \text{Serumkreatinin } (\mu\text{mol/l})}$$

där $\theta = 0,8$ om man och $\theta = 0,85$ om kvinna. Då ett högt värde minskar risken för PE används det transformerade måttet $GFR_{CRG_transf} = GFR_{CRG}/ - 10$.

- **Glomerulär filtrationshastighet enligt MDRD-formeln** (GFR_{MDRD}). MDRD står för Modification of Diet in Renal Disease och estimerar den glomerulära filtrationshastigheten enligt följande formel

$$GFR_{MDRD} = 32788 \times \text{Serumkreatinin } (\mu\text{mol/l})^{-1,154} \times \text{Ålder}^{-0,203} \times \theta,$$

där $\theta = 1$ om man och $\theta = 0,742$ om kvinna. Ett högt värde minskar risken för PE. Därför används det transformerade måttet $GFR_{MDRD_transf} = GFR_{MDRD}/ - 10$.

Notera att MA endast undersöktes i den svenska kohorten, vilket innebär att analysen av MA omfattar färre observationer ($n = 4929$). Tabell 4.2 beskriver de aktuella variablerna i studien.

4.2 Tidigare analys av njurfunktion och risk för kardiovaskulär sjukdom

Färbom et al. (2008) visar med hjälp av data från NORDIL att det finns ett signifikant samband mellan halten av serumkreatinin och risken att drabbas av PE ($p < 0,001$). Justeras modellen med hänsyn till andra kardiovaskulära riskfaktorer ser vi att givet samma riskfaktorer löper en patient 1,05 gånger större risk att drabbas av PE för 10 $\mu\text{mol/l}$ ökning av serumkreatinin ($p < 0,001$).

Även GFR_{CRG} visar sig ha ett signifikant samband med PE ($p < 0,001$). Justerat för övriga riskfaktorer innebär detta 1,06 gånger större risk att råka ut för PE ($p = 0,001$) för 10 enheters minskning av GFR_{CRG} .

Tabell 4.1: Relativ risk för PE

Mått på njurfunktion	Hasardkvot	95 % KI	<i>p</i> -värde
S_KRE.transf	1,10	1,08-1,12	< 0,001
S_KRE.transf*	1,05	1,03-1,08	< 0,001
GFR_{CRG_transf}	1,13	1,10-1,17	< 0,001
$GFR_{CRG_transf}^*$	1,06	1,02-1,10	0,001
GFR_{MDRD_transf}	1,05	1,00-1,10	0,062
$GFR_{MDRD_transf}^*$	1,07	1,02-1,13	0,007
MA	1,77	1,41-2,23	< 0,001
MA†	1,35	1,06-1,71	0,014

*Justerad för ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol

†Justerad för ålder, kön, systoliskt blodtryck, rökning, diabetes och tidigare CVD

Källa: Färbom et al. (2008)

GFR_{MDRD} är emellertid inte signifikant ($p = 0,062$) om endast den ojusterade modellen (som bara inkluderar GFR_{MDRD} och PE) betraktas. Inkluderas andra riskfaktorer blir dock sambandet signifikant ($p = 0,007$) och en patient med 10 enheter mindre GFR_{MDRD} utsätts då för 1,07 gånger större risk att drabbas av PE.

Vad gäller MA så uppvisar patientgruppen med en mikroalbuminurinivå $\geq 20 \mu\text{mol/l}$ 1,77 gånger större risk att drabbas av PE ($p < 0,001$). Beaktas även övriga riskfaktorer löper patienter med MA 1,35 gånger större risk att drabbas av PE jämfört med dem som inte har MA ($p = 0,014$).

Färbom et al. visar även att det finns en signifikant interaktion mellan MA och GFR_{CRG} samt MA och GFR_{MDRD} avseende framtida PE. Emellertid finns ingen signifikant interaktion mellan serumkreatinin och MA vad gäller risken för PE.

De olika måtten på njurfunktion är uppenbarligen signifikanta prediktorer för kardiovaskulär sjukdom. Det som återstår att undersöka är hur mycket bättre prediktionen för kardiovaskulär sjukdom blir om ovanstående riskfaktorer inkorporeras i analysen.

Tabell 4.2: Beskrivning av datamaterialet

Variabelbeskrivning	Totalt	MA-studie	Variabelnamn	Variabeltransformation
	$n = 10881$	$n = 4929$		
	Medelvärde (SD)	Medelvärde (SD)		
Dödlig hjärtinfarkt				
Icke-dödlig hjärtinfarkt				
Död i akut cerebrovaskulär sjukdom				fail (PE)
Icke-dödlig akut cerebrovaskulär sjukdom				
Annan död i hjärt- eller kärlsjukdom				
Ålder (år)	59,9 (6,5)	60,2 (6,7)	AGE	
Kön (% män)	48,6	49,8	SEX	
Systoliskt blodtryck (mmHg)	172,6 (18,7)	172,8 (18,9)	SBP	
Diastoliskt blodtryck (mmHg)	104,1 (6,8)	103,1 (7,1)	DPB	
Serumkolesterol (mmol/l)	6,4 (1,2)	6,3 (1,2)	S_KOL	
Blodglukos (mmol/l)	5,3 (1,5)	5,3 (1,6)	FBGLU	
Serumkreatinin ($\mu\text{mol/l}$)	86,7 (17,4)	86,0 (17,5)	S_KRE	S_KRE_transf = S_kre/10
GFR _{CRG} (ml/min/1,73m ²)	92,9 (24,3)	93,9 (25,0)	CRG	GFR _{CRG_transf} = GFR _{CRG} / - 10
GFR _{MDRD} (ml/min/1,73m ²)	74,1 (15,1)	74,9 (15,0)	MDRD	GFR _{MDRD_transf} = GFR _{MDRD} / - 10
Rökning (%)	22,4	20,8	SMOKE_POOLED	
Tidigare kardiovaskulär sjukdom (%)	3,5	4,0	prev_CVD	
Diabetes (%)	6,7	9,3	DIABET	
Mikroalbuminuri (%)	n.a.	22,6	U_MIC	U_MIC $\geq 20 \mu\text{mol/l} \Rightarrow \text{MA}=1$, annars 0
Body Mass Index (kg/m ²)	27,8 (4,3)	28,1 (4,4)		
Diltiazem (%)	49,7	50,1	Card	Diltiazem=1 $\Rightarrow \text{Card} = 1$, annars 0

Källa: The Nordic Diltiazem Study

Kapitel 5

Resultat

I detta kapitel presenteras resultat baserade på de i teoridelen redovisade metoderna som utvärderar modellens förmåga att prediktera händelser:

1. Diskriminering: AUC
2. Kalibrering: X^2_{HT}
3. Reklassificering: \widehat{NRI}
4. Rangkorrelation: Harrells C

Eftersom MA endast undersöks på 4929 patienter indelas analysen av patienterna i:

- beräkningar baserade på alla patienter i studien
- beräkningar baserade på de patienter där MA undersökts.

För metod 1, 2 och 3 jämförs genomgående två modeller: (1) En grundmodell som innehåller alla de klassiska riskfaktorerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol. (2) En grundmodell + en eller flera adderade riskfaktorer. Vad gäller metod 3, \widehat{NRI} , är det i sig ett mått på en jämförelse.

Exakt samma patienter ingår i varje jämförelse, vilket är ett måste om inte analysen ska leda till svårtolkade resultat eftersom patientantalet annars varierar beroende på vilka variabler som inkluderas. Således varierar även siffrorna i grundmodellerna i respektive jämförelse.

Resultaten av 1, 2 och 4 kan analyseras på två sätt; dels kan den absolut bästa modellen i termer av absoluta tal föras fram, dels kan man peka på den modell som ger den största relativa förbättringen i prediktion. Bägge ansatserna kommer att användas.

Alla beräkningar är gjorda i `Stata/SE 10.0 for Windows`.

5.1 Prediktion av dikotoma utfall

För analysen av dikotoma utfall används risken att inom fyra år drabbas av PE. Därmed analyseras endast de patienter som inledningsvis har en teoretisk möjlighet att delta minst fyra år i studien (det vill säga patienter som randomiserades senast den 31 oktober 1995); vidare tas enbart hänsyn till första

Tabell 5.1: Antal observationer för olika årsrisker

	Antal PE	Antal patienter	Antal patienter där MA mäts
3-årsrisk	493	10537	4811
4-årsrisk	459	7478	3012
5-årsrisk	314	4186	817

PE som inträffar inom fyra år. Detta innebär i praktiken att tiden efter första PE samt tiden efter fyra år är censurerad.

Beslutet att använda fyraårig risk är en avvägning mellan att inkludera tillräckligt många patienter i analysen, samtidigt som patienternas risk att drabbas av PE bör vara tillräckligt stor. Tabell 5.1 ger en jämförelse av olika patientantal. Det är tydligt att skulle en femårig risk väljas blir antalet patienter litet, särskilt i patientgruppen där förekomsten av MA mäts.

5.1.1 Area under ROC-kurvan

Vi inleder med att betrakta resultaten vad gäller AUC. Eftersom **Stata** inte har något kommando för att beräkna AUC efter en Cox-regression används de predikterade hasardkvoterna som oberoende variabel i en logistisk regression, där PE är den beroende variabeln. Därefter kan **Stata** enkelt beräkna AUC.

För AUC baserad på hela populationen ser vi att bäst modell i termer av både i termer av absoluta tal och relativ förbättring är modell 2, det vill säga när GFR_{CRG} adderas till grundmodellen (Tabell 5.2). Med grundmodell + GFR_{CRG} blir $AUC = 0,7179$, vilket är en förbättring med 0,48 % jämfört med grundmodellen. Allmänt ger ingen modell någon större förbättring i AUC.

Tabell 5.2: AUC baserad på hela populationen

	Inkluderade kovariater	AUC	(2)-(1)	(2)/(1)
1	(1) Grundmodell	0,7134	0,0023	1,0031
	(2) Grundmodell + S-kre	0,7156		
2	(1) Grundmodell	0,7145	0,0033	1,0048
	(2) Grundmodell + GFR_{CRG}	0,7179		
3	(1) Grundmodell	0,7134	0,0018	1,0025
	(2) Grundmodell + GFR_{MDRD}	0,7152		

Not 1: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolesterol.

Not 2: Eftersom angivna värden på AUC är avrundade och (2)-(1) beräknas direkt i **Stata** kan det föreligga viss diskrepans.

Ser vi enbart till gruppen patienter där MA undersöks och inte tar hänsyn till MA som riskfaktor är istället modellen med S-kre bäst ($AUC = 0,7411$, 0,37 % förbättring). Precis som tidigare ger detta ingen större förbättring i prediktion. Noterbart är att även om AUC generellt sätt är större i gruppen patienter MA undersöks, vilket motiverar de separata analyserna av patientgrupperna, så är de relativa förbättringarna i samma storleksordning.

När MA inkluderas som riskfaktor ger grundmodell + GFR_{MDRD} + MA + $GFR_{MDRD} \times MA$ bäst prediktionsförmåga ($AUC = 0,7480$). Denna modell ger

även störst procentuell förbättring, 1,3 %. Emellertid ger alla modeller som tar med S-kre, GFR_{CRG} eller GFR_{MDRD} i kombination med MA liknande resultat och när MA inkluderas förbättras AUC med ungefär en procentenhet. Det visar sig att endast grundmodell + MA ger en relativ förbättring på 0,92 %.

Vi kan konstatera att vad gäller AUC är MA överlägsen S-kre, GFR_{CRG} och GFR_{MDRD} . Bäst prediktionsförmåga erhålls emellertid när MA kombineras med någon av de övriga riskfaktorerna; vilken som väljs har ingen större betydelse.

Tabell 5.3: AUC baserad på patienter där MA undersöks

	Inkluderade kovariater	AUC	(2)-(1)	(2)/(1)
4	(1) Grundmodell	0,7384	0,0027	1,0037
	(2) Grundmodell + S-kre	0,7411		
5	(1) Grundmodell	0,7384	0,0089	1,0121
	(2) Grundmodell + S-kre + MA	0,7473		
6	(1) Grundmodell	0,7384	0,0090	1,0122
	(2) Grundmodell + S-kre + MA + S-kre×MA	0,7474		
7	(1) Grundmodell	0,7370	0,0017	1,0023
	(2) Grundmodell + GFR_{CRG}	0,7387		
8	(1) Grundmodell	0,7370	0,0085	1,0114
	(2) Grundmodell + GFR_{CRG} + MA	0,7454		
9	(1) Grundmodell	0,7370	0,0092	1,0125
	(2) Grundmodell + GFR_{CRG} + MA + GFR_{CRG} ×MA	0,7462		
10	(1) Grundmodell	0,7384	0,0022	1,0030
	(2) Grundmodell + GFR_{MDRD}	0,7406		
11	(1) Grundmodell	0,7384	0,0081	1,0110
	(2) Grundmodell + GFR_{MDRD} + MA	0,7465		
12	(1) Grundmodell	0,7384	0,0096	1,0130
	(2) Grundmodell + GFR_{MDRD} + MA + GFR_{MDRD} ×MA	0,7480		
13	(1) Grundmodell	0,7387	0,0068	1,0092
	(2) Grundmodell + MA	0,7455		

Not 1: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol.

Not 2: Eftersom angivna värden på AUC är avrundade och (2)-(1) beräknas direkt i *Stata* kan det föreligga viss diskrepans.

5.1.2 Kalibrering

Vid analys av hela populationen ges bäst kalibrering när S-kre adderas till grundmodellen (Tabell 5.4). X_{HL}^2 minskar med 25,3 %. Vad gäller de övriga två riskfaktorena ger GFR_{MDRD} en närmast oförändrad kalibreringsförmåga, medan GFR_{CRG} försämrar X_{HL}^2 med 52,3 %. Notera att låga värden på X_{HL}^2 innebär en god kalibrering.

Tabell 5.4: X_{HL}^2 baserad på hela populationen

	Inkluderade kovariater	X_{HL}^2	(2)-(1)	(2)/(1)
1	(1) Grundmodell	9,253	-2,340	0,7471
	(2) Grundmodell + S-kre	6,913		
2	(1) Grundmodell	6,627	3,463	1,5227
	(2) Grundmodell + GFR_{CRG}	10,091		
3	(1) Grundmodell	9,253	-0,351	0,9622
	(2) Grundmodell + GFR_{MDRD}	8,903		

Not 1: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolesterol.

Not 2: Eftersom angivna värden på X_{HL}^2 är avrundade och (2)-(1) beräknas direkt i *Stata* kan det föreligga viss diskrepans.

Om MA fungerade väl beträffande förbättring i AUC, är MA en sämre prediktor när man mäter med X_{HL}^2 . Grundmodell + MA ger endast 0,26 % minskning av X_{HL}^2 (Tabell 5.5). Bland patienter som undersöks för MA blir dessutom förbättringarna av X_{HL}^2 sämre för alla modeller när MA inkluderas. Störst förbättring av X_{HL}^2 erhålls av grundmodell + GFR_{MDRD} , men även grundmodell + S-kre ger ungefär samma förbättring (49,2 % respektive 47,2 % minskning).

För att vara säker på att erhålla en god kalibrering bör man således använda S-kre som prediktor (eventuellt GFR_{MDRD}) och utesluta MA ur modellen.

Tabell 5.5: X_{HL}^2 baserad på patienter där MA undersöks

	Inkluderade kovariater	X_{HL}^2	(2)-(1)	(2)/(1)
4	(1) Grundmodell	21,701	-10,237	0,5282
	(2) Grundmodell + S-kre	11,463		
5	(1) Grundmodell	21,701	-6,065	0,7205
	(2) Grundmodell + S-kre + MA	15,636		
6	(1) Grundmodell	21,701	-5,515	0,7458
	(2) Grundmodell + S-kre + MA + S-kre×MA	16,185		
7	(1) Grundmodell	20,853	-5,147	0,7532
	(2) Grundmodell + GFR _{CRG}	15,706		
8	(1) Grundmodell	20,853	-6,486	0,6890
	(2) Grundmodell + GFR _{CRG} + MA	14,368		
9	(1) Grundmodell	20,853	-1,217	0,9416
	(2) Grundmodell + GFR _{CRG} + MA + GFR _{CRG} ×MA	19,636		
10	(1) Grundmodell	21,701	-10,666	0,5085
	(2) Grundmodell + GFR _{MDRD}	11,035		
11	(1) Grundmodell	21,701	-9,351	0,5691
	(2) Grundmodell + GFR _{MDRD} + MA	12,350		
12	(1) Grundmodell	21,701	-8,686	0,5997
	(2) Grundmodell + GFR _{MDRD} + MA + GFR _{MDRD} ×MA	13,015		
13	(1) Grundmodell	24,162	-0,063	0,9974
	(2) Grundmodell + MA	24,099		

Not 1: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol.

Not 2: Eftersom angivna värden på X_{HL}^2 är avrundade och (2)-(1) beräknas direkt i *Stata* kan det föreligga viss diskrepans.

5.1.3 Reklassificering

Vid beräkningen av \widehat{NRI} väljs riskkategorierna $< 6\%$, $6 - 20\%$ och $> 20\%$ risk att inom fyra år drabbas av PE. Riskkategorierna är samma som används för normala friska patienter vid tioårig risk, men då hypertoni-patienter löper dubbelt så hög risk att drabbas av PE antar vi att dessa riskkategorier fungerar väl trots att tiden i princip är halverad.

Vi kan se i Tabell 5.6 att ingen av S-kre, GFR_{CRG} eller GFR_{MDRD} ger ett signifikant \widehat{NRI} när de adderas till grundmodellen.

Tabell 5.6: Reklassificering baserad på hela populationen

Inkluderade kovariater	\widehat{NRI}	z	z_δ	z_{δ^c}
1 Grundmodell + S-kre	0,004	0,369	0,229	0,659
2 Grundmodell + GFR _{CRG}	0,015	1,084	0,845	1,099
3 Grundmodell + GFR _{MDRD}	0,021	1,755	1,512	1,218

Not: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol

Motsvarande resultat gäller för patientgruppen där MA undersöks och även grundmodell + MA ger inget signifikant \widehat{NRI} (Tabell 5.7).

För att \widehat{NRI} ska bli signifikant måste riskfaktorn MA tas med i modellen och kombineras med någon av de övriga riskfaktorerna. Samtliga modeller förutom modell 9 (Grundmodell + GFR_{CRG} + MA + $GFR_{CRG} \times MA$) ger då signifikanta \widehat{NRI} , men även modell 9 har z -värde i närheten av de övriga modellerna. Störst \widehat{NRI} ges av Grundmodell + GFR_{MDRD} + MA ($\widehat{NRI} = 0,087$).

Betraktar vi de enskilda komponenterna i \widehat{NRI} blir resultaten något mer svårtolkade, men samma slutsats kvarstår – för att erhålla signifikanta reklassificeringar måste MA kombineras med S-kre, GFR_{CRG} och GFR_{MDRD} .

Tabell 5.7: Reklassificering baserad på patienter där MA undersöks

Inkluderade kovariater	\widehat{NRI}	z	z_δ	z_{δ^c}
4 Grundmodell + S-kre	0,010	0,519	0,535	0,000
5 Grundmodell + S-kre + MA	0,069	2,150	1,715	2,401
6 Grundmodell + S-kre + MA + S-kre \times MA	0,070	2,172	1,715	2,524
7 Grundmodell + GFR_{CRG}	-0,032	-1,312	-1,414	0,382
8 Grundmodell + GFR_{CRG} + MA	0,073	2,233	2,263	0,059
9 Grundmodell + GFR_{CRG} + MA + $GFR_{CRG} \times MA$	0,061	1,888	1,567	1,883
10 Grundmodell + GFR_{MDRD}	0,013	0,603	0,775	1,218
11 Grundmodell + GFR_{MDRD} + MA	0,087	2,780	-0,632	1,775
12 Grundmodell + GFR_{MDRD} + MA + $GFR_{MDRD} \times MA$	0,080	2,633	2,191	2,455
13 Grundmodell + MA	0,017	0,683	0,447	1,176

Not: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol

5.2 Prediktionsutvärdering som tar hänsyn till tiden

Vi inkluderar nu alla patienter i analysen oavsett tid i studien och utnyttjar därmed ytterligare information i form av tid till PE.

Vad gäller hela populationen är förändringarna i Harrells C små när S-kre, GFR_{CRG} eller GFR_{MDRD} adderas till grundmodellen. Störst C erhålls av grundmodell + GFR_{CRG} ($C = 0,7046$) vilket innebär en förbättring av C med 0,23 % (Tabell 5.8).

Tabell 5.8: Harrells C baserad på hela populationen

	Inkluderade kovariater	Harrells C	(2)-(1)	(2)/(1)
1	(1) Grundmodell	0,7011	0,0013	1,0019
	(2) Grundmodell + S-kre	0,7024		
2	(1) Grundmodell	0,7030	0,0016	1,0023
	(2) Grundmodell + GFR_{CRG}	0,7046		
3	(1) Grundmodell	0,7011	0,0003	1,0004
	(2) Grundmodell + GFR_{MDRD}	0,7014		

Not: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol

I gruppen patienter där MA undersöks ger S-kre, GFR_{CRG} eller GFR_{MDRD} endast marginella förändringar i C ; förändringarna är till och med negativa för modell 4 och 10.

Prediktorn som ger störst förändring i C jämfört med övriga riskfaktorer är MA. Studeras resultaten i Tabell 5.9 ser vi att endast grundmodell + S-kre + MA + S-kre \times MA leder till en större förbättring i C än grundmodell + MA. Grundmodell + S-kre + MA + S-kre \times MA förbättrar Harrells C med 0,47 procent, medan enbart MA ger en förbättring på 0,35 procent. Vidare är grundmodell + S-kre + MA + S-kre \times MA den modell som ger störst C ($C = 0,7247$).

Överlag rör det sig alltså om marginella förändringar, förändringar som är mindre än vid analysen av bestämda årsrisker.

Tabell 5.9: Harrells C baserad på patienter där MA undersöks

	Inkluderade kovariater	Harrells C	(2)-(1)	(2)/(1)
4	(1) Grundmodell	0,7213	-0,0005	0,9993
	(2) Grundmodell + S-kre	0,7208		
5	(1) Grundmodell	0,7213	0,0018	1,0025
	(2) Grundmodell + S-kre + MA	0,7231		
6	(1) Grundmodell	0,7213	0,0034	1,0047
	(2) Grundmodell + S-kre + MA + S-kre \times MA	0,7247		
7	(1) Grundmodell	0,7226	0,0000	1,0000
	(2) Grundmodell + GFR _{CRG}	0,7226		
8	(1) Grundmodell	0,7226	0,0017	1,0024
	(2) Grundmodell + GFR _{CRG} + MA	0,7243		
9	(1) Grundmodell	0,7226	0,0015	1,0021
	(2) Grundmodell + GFR _{CRG} + MA + GFR _{CRG} \times MA	0,7241		
10	(1) Grundmodell	0,7213	-0,0012	0,9983
	(2) Grundmodell + GFR _{MDRD}	0,7201		
11	(1) Grundmodell	0,7213	0,0013	1,0018
	(2) Grundmodell + GFR _{MDRD} + MA	0,7226		
12	(1) Grundmodell	0,7213	0,0019	1,0026
	(2) Grundmodell + GFR _{MDRD} + MA + GFR _{MDRD} \times MA	0,7232		
13	(1) Grundmodell	0,7213	0,0025	1,0035
	(2) Grundmodell + MA	0,7238		

Not: Grundmodellen inkluderar variablerna ålder, kön, systoliskt blodtryck, rökning, diabetes, tidigare CVD, behandling och kolestereol

Kapitel 6

Diskussion

Genom att jämföra olika metoder som antingen utvärderar en modells förmåga att prediktera händelser eller jämför prediktionsförmågan hos två modeller kan följande konstateras:

- Vad gäller arean under ROC-kurvan fungerar mikroalbuminuri ensamt som prediktor tillsammans med de klassiska riskfaktorerna, men bör kombineras med något av de övriga måtten på njurfunktion – serumkreatinin och den glomerulära filtrationshastigheten beräknad enligt Cockcroft & Gaults formel och MDRD-formeln – för att ge ytterligare förbättrad prediktionsförmåga.
- Mikroalbuminuri har endast en marginell effekt på kalibreringen och dessutom fungerar de övriga måtten på njurfunktionen bättre ensamma än i kombination med mikroalbuminuri.
- När reklassificeringsindexet beräknas får vi det intressanta resultatet att enbart en kombinerad modell med mikroalbuminuri plus något av de övriga måtten på njurfunktion leder till en signifikant förbättrad klassificering.
- För att i högre grad kunna utnyttja informationen i materialet beräknades även Harrells C . Det visade sig att mikroalbuminuri är det mått på njurfunktion som ensamt förbättrar C mest, och bara om mikroalbuminuri kombineras med serumkreatinin erhålls en ännu bättre modell för prediktion.

Resultaten är således inte entydiga. Mikroalbuminuri bör i regel inkluderas, men vilket av de övriga måtten på njurfunktion som bör väljas varierar. Vilken riskfaktor som väljs beror på vilket mått av prediktionsförmåga som prioriteras, något som naturligtvis beror på syftet med studien.

Även om vi kan konstatera att mikroalbuminuri är en viktig prediktor jämfört med de övriga måtten på njurfunktion är förbättringarna i prediktion generellt sett ganska blygsamma. Intressantast ur ett kliniskt perspektiv är kanske förbättringen i klassificering, vilket därför skulle kunna analyseras vidare.

Noterbart är att resultaten inte enbart varierar efter metod, utan också beror på vilken patientgrupp som används. En fråga man kan ställa sig är hur de olika måtten på prediktion påverkas av storleken på patientgrupperna. Att

kalibreringen påverkas är tydligt, men hur är fallet med övriga mått? Eller är det helt enkelt så att de svenska och norska kohorterna skiljer sig åt på något avgörande sätt?

Avslutningsvis vill jag tacka Peter Almgren och Olle Melander vid Clinical Research Centre, UMAS, för möjligheten att genomföra denna studie.

Litteraturförteckning

- [1] COLLETT D, (1994). *Modelling Survival Data in Medical Research*. Chapman & Hall, UK.
- [2] FÄRBOM P, WAHLSTRAND B, ALMGREN P, SKRTIC S, LANKE J, WEISS L, KJELDSSEN S, HEDNER T, MELANDER O, (2008). Interaction between renal function and microalbuminuria for cardiovascular risk in hypertension. The NORDIL study. *Hypertension*. I tryck.
- [3] HARRELL JR F, CALIFF R M, PRYOR D B, LEE K, ROSATI R A (1982). Evaluating the yield of statistical tests. *The Journal of the American Medical Association*, 287, 2543-2546.
- [4] HARRELL JR F, LEE K, MARK D (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy and Measuring and Reducing Errors. *Statistics in Medicine*, 15, 361-387.
- [5] HOSMER D, LEMESHOW S, (1989). *Applied Logistic Regression*. Wiley, USA.
- [6] PENCINA M J, D'AGOSTINO R B, D'AGOSTINO JR R B, RAMACHANDRAN S V, (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, 27, 157-172.

Bilaga A

Do-filer

Beräkning av 4-årsrisk

```
sort INDATE
drop in 7479/10871
sort _t
replace _t=4 in 606/7478
replace fail = 99 in 606/7478
replace PE = 0 in 606/7478
replace _d = 0 in 606/7478
```

Exempel på Do-fil för dikotoma utfall

```
log using "C:\Documents and Settings\Administrator
> \Desktop\NORDILDATA\skre1.smcl"
use "C:\Documents and Settings\Administrator\
> Desktop\NORDILDATA\NORDILJL4år.dta"
```

```
#delimiter ;
```

```
stcox AGE60 SEX SBP140 DIABET S_KOL SMOKE_POOLED
> prev_CVD Card S_KRE_transf, basesurv(S0);
predict HR;
summarize(S0);
scalar S03 = r(min);
gen S = S03^HR;
gen F = 1-S;
sort F;
drop if F==.;
summarize F;
scalar FN = r(N);
egen OBS = seq();
```

```
stcox AGE60 SEX SBP140 DIABET S_KOL SMOKE_POOLED
> prev_CVD Card, basesurv(s0);
predict hr;
summarize(s0);
scalar s03 = r(min);
gen s = s03^hr;
gen f = 1-s;
sort f;
drop if f==.;
summarize f;
scalar fn = r(N);
egen obs = seq();
```



```

logistic PE HR;
lroc, nograph;
scalar C1 = r(area);

logistic PE hr;
lroc, nograph;
scalar C2 = r(area);

#delimit cr

forval k=1/9 {
  scalar D'k'=round((FN/10)*'k'+0.5)
}

#delimit ;

summarize F if OBS<=D1;
scalar N1 = r(N);
summarize F if OBS<=D1;
scalar E1 = r(sum);
summarize F if OBS<=D2&OBS>D1;
scalar N2 = r(N);
summarize F if OBS<=D2&OBS>D1;
scalar E2 = r(sum);
summarize F if OBS<=D3&OBS>D2;
scalar N3 = r(N);
summarize F if OBS<=D3&OBS>D2;
scalar E3 = r(sum);
summarize F if OBS<=D4&OBS>D3;
scalar N4 = r(N);
summarize F if OBS<=D4&OBS>D3;
scalar E4 = r(sum);
summarize F if OBS<=D5&OBS>D4;
scalar N5 = r(N);
summarize F if OBS<=D5&OBS>D4;
scalar E5 = r(sum);
summarize F if OBS<=D6&OBS>D5;
scalar N6 = r(N);
summarize F if OBS<=D6&OBS>D5;
scalar E6 = r(sum);
summarize F if OBS<=D7&OBS>D6;
scalar N7 = r(N);
summarize F if OBS<=D7&OBS>D6;
scalar E7 = r(sum);
summarize F if OBS<=D8&OBS>D7;
scalar N8 = r(N);
summarize F if OBS<=D8&OBS>D7;
scalar E8 = r(sum);
summarize F if OBS<=D9&OBS>D8;
scalar N9 = r(N);
summarize F if OBS<=D9&OBS>D8;
scalar E9 = r(sum);
summarize F if OBS>D9;
scalar N10 = r(N);
summarize F if OBS>D9;
scalar E10 = r(sum);

summarize PE if OBS<=D1;
scalar O1 = r(sum);
summarize PE if OBS<=D2&OBS>D1;
scalar O2 = r(sum);

```

```

summarize PE if OBS<=D3&OBS>D2;
scalar O3 = r(sum);
summarize PE if OBS<=D4&OBS>D3;
scalar O4 = r(sum);
summarize PE if OBS<=D5&OBS>D4;
scalar O5 = r(sum);
summarize PE if OBS<=D6&OBS>D5;
scalar O6 = r(sum);
summarize PE if OBS<=D7&OBS>D6;
scalar O7 = r(sum);
summarize PE if OBS<=D8&OBS>D7;
scalar O8 = r(sum);
summarize PE if OBS<=D9&OBS>D8;
scalar O9 = r(sum);
summarize PE if OBS>D9;
scalar O10 = r(sum);

scalar X2 = ((O1-E1)^2)/(E1*(1-E1/N1))+
> ((O2-E2)^2)/(E2*(1-E2/N2))+((O3-E3)^2)/(E3*(1-E3/N3))+
> ((O4-E4)^2)/(E4*(1-E4/N4))+((O5-E5)^2)/(E5*(1-E5/N5))+
> ((O6-E6)^2)/(E6*(1-E6/N6))+((O7-E7)^2)/(E7*(1-E7/N7))+
> ((O8-E8)^2)/(E8*(1-E8/N8))+((O9-E9)^2)/(E9*(1-E9/N9))+
> ((O10-E10)^2)/(E10*(1-E10/N10));

scalar d1 = round((fn/100)*10+0.5);
scalar d2 = round((fn/100)*20+0.5);
scalar d3 = round((fn/100)*30+0.5);
scalar d4 = round((fn/100)*40+0.5);
scalar d5 = round((fn/100)*50+0.5);
scalar d6 = round((fn/100)*60+0.5);
scalar d7 = round((fn/100)*70+0.5);
scalar d8 = round((fn/100)*80+0.5);
scalar d9 = round((fn/100)*90+0.5);

summarize f if obs<=d1;
scalar n1 = r(N);
summarize f if obs<=d1;
scalar e1 = r(sum);
summarize f if obs<=d2&obs>d1;
scalar n2 = r(N);
summarize f if obs<=d2&obs>d1;
scalar e2 = r(sum);
summarize f if obs<=d3&obs>d2;
scalar n3 = r(N);
summarize f if obs<=d3&obs>d2;
scalar e3 = r(sum);
summarize f if obs<=d4&obs>d3;
scalar n4 = r(N);
summarize f if obs<=d4&obs>d3;
scalar e4 = r(sum);
summarize f if obs<=d5&obs>d4;
scalar n5 = r(N);
summarize f if obs<=d5&obs>d4;
scalar e5 = r(sum);
summarize f if obs<=d6&obs>d5;
scalar n6 = r(N);
summarize f if obs<=d6&obs>d5;
scalar e6 = r(sum);
summarize f if obs<=d7&obs>d6;
scalar n7 = r(N);
summarize f if obs<=d7&obs>d6;
scalar e7 = r(sum);

```

```

summarize f if obs<=d8&obs>d7;
scalar n8 = r(N);
summarize f if obs<=d8&obs>d7;
scalar e8 = r(sum);
summarize f if obs<=d9&obs>d8;
scalar n9 = r(N);
summarize f if obs<=d9&obs>d8;
scalar e9 = r(sum);
summarize f if obs>d9;
scalar n10 = r(N);
summarize f if obs>d9;
scalar e10 = r(sum);

summarize PE if obs<=d1;
scalar o1 = r(sum);
summarize PE if obs<=d2&obs>d1;
scalar o2 = r(sum);
summarize PE if obs<=d3&obs>d2;
scalar o3 = r(sum);
summarize PE if obs<=d4&obs>d3;
scalar o4 = r(sum);
summarize PE if obs<=d5&obs>d4;
scalar o5 = r(sum);
summarize PE if obs<=d6&obs>d5;
scalar o6 = r(sum);
summarize PE if obs<=d7&obs>d6;
scalar o7 = r(sum);
summarize PE if obs<=d8&obs>d7;
scalar o8 = r(sum);
summarize PE if obs<=d9&obs>d8;
scalar o9 = r(sum);
summarize PE if obs>d9;
scalar o10 = r(sum);

scalar x2 = ((o1-e1)^2)/(e1*(1-e1/n1))+
> ((o2-e2)^2)/(e2*(1-e2/n2))+((o3-e3)^2)/(e3*(1-e3/n3))+
> ((o4-e4)^2)/(e4*(1-e4/n4))+((o5-e5)^2)/(e5*(1-e5/n5))+
> ((o6-e6)^2)/(e6*(1-e6/n6))+((o7-e7)^2)/(e7*(1-e7/n7))+
> ((o8-e8)^2)/(e8*(1-e8/n8))+((o9-e9)^2)/(e9*(1-e9/n9))+
> ((o10-e10)^2)/(e10*(1-e10/n10));

gen F1 = F;
gen f1 = f;
recode F1 (min/0.06=1) (0.06000000001/0.2=2) (0.20000000001/max=3);
recode f1 (min/0.06=1) (0.06000000001/0.2=2) (0.20000000001/max=3);
generate Delta = F1-f1;
recode Delta (-2=-1)(-1=-1)(0=0)(1=1)(2=1);
count if PE == 1;
scalar delta= r(N);
count if PE == 0;
scalar deltac = r(N);
count if PE == 1 & Delta == 1;
scalar updelta = r(N);
count if PE == 1 & Delta == -1;
scalar downdelta = r(N);
count if PE == 0 & Delta == 1;
scalar updeltac = r(N);
count if PE == 0 & Delta == -1;
scalar downdeltac = r(N);
scalar pupdelta = updelta/delta;
scalar pdowndelta = downdelta/delta;
scalar pupdeltac = updeltac/deltac;

```

```

scalar pdowndeltac = downdeltac/deltac;
scalar NRI = (updelta/delta-downdelta/delta)-
> (updeltac/deltac-downdeltac/deltac);
scalar z = NRI/((pupdelta+pdowndelta)/delta+
> (pupdeltac+pdowndeltac)/deltac)^(1/2);
scalar zup = (pupdelta-pdowndelta)/
> (((pupdelta+pdowndelta)/delta)^(1/2));
scalar zdown = (pdowndeltac-pupdeltac)/
> (((pdowndeltac+pupdeltac)/deltac)^(1/2));

scalar diffC = C1-C2;
scalar diffX = X2-x2;
display C1;
display C2;
display X2;
display x2;
display NRI;
display z;
display diffC;
display diffX;
display zup;
display zdown;

log close;

```

Exempel på Do-fil för beräkning av Harrells C

```

log using "C:\Documents and Settings\Administrator
> \Desktop\NORDILDATA\Cskre.smcl"
use "C:\Documents and Settings\Administrator
> \Desktop\NORDILDATA\NORDILJL.dta"

stcox AGE60 SEX SBP140 DIABET S_KOL SMOKE_POOLED prev_CVD Card S_KRE_transf, basesurv(S0)
drop if S0==.
estat con

stcox AGE60 SEX SBP140 DIABET S_KOL SMOKE_POOLED prev_CVD Card, basesurv(s0)
estat con

log close

```