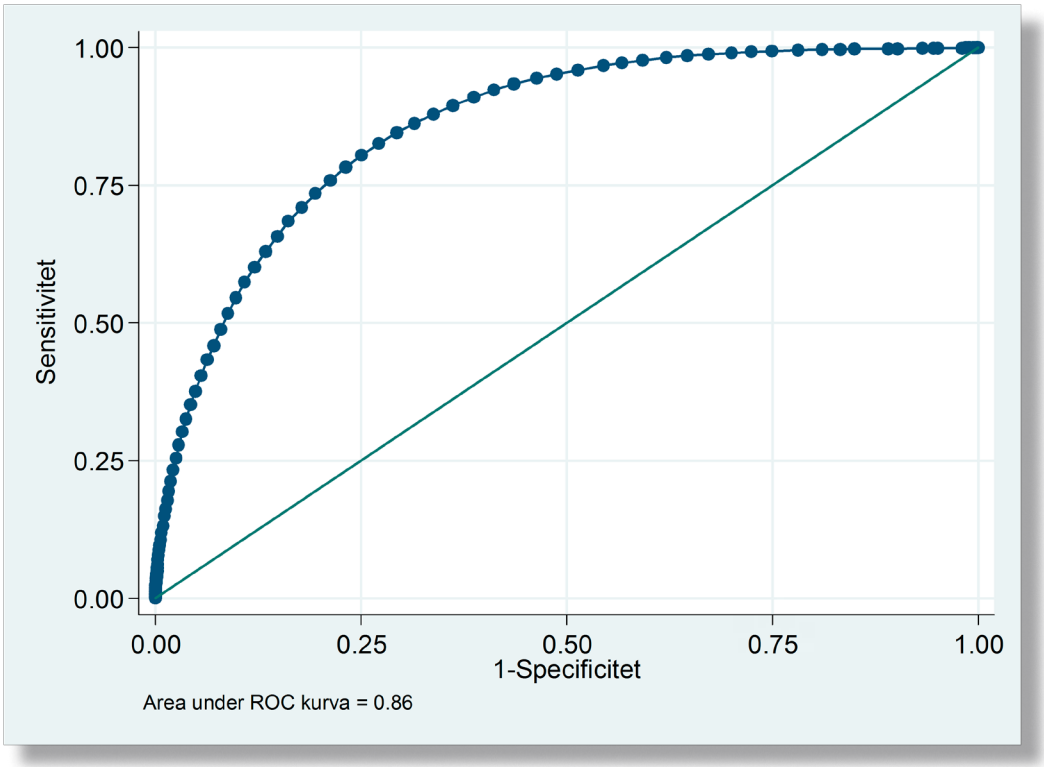
Svenska Intensivvårdsregistret
SIR

Riskjusteringsmodeller i svensk intensivvård

– en fokusrapport från Svenska Intensivvårdsregistret (version 1.00)

av Lars Engerström och Sten Walther



Innehåll

3 Bakgrund

4 Hur bedömer man riskjusteringsmodeller?
Diskriminering
Kalibrering
Andra övergripande mått

5 Riskjustering inom den allmänna intensivvården av vuxna
Simplified Acute Physiology Score 3 (SAPS3)

8 Riskjustering i thoraxintensivvård
Higgins' ICU admission score (Intensivvårds-Higgins)

10 Riskjustering vid intensivvård av barn
Pediatric Index of Mortality 2 (PIM2)

11 Generella slutsatser från arbetet med riskjusteringsmodeller i SIR

Riskjusteringsmodeller i svensk intensivvård

Bakgrund

Riskjustering av ett vårdförlopp innebär att vårdens resultat beskrivs med hänsyn tagen till riskfaktorer som kan påverka sjukdomsförloppets utfall. Ska man studera utvecklingen av vården över tid eller jämföra olika avdelningar behöver man någon form av riskjustering för att analysen ska bli meningsfull. Mortalitet utan riskjustering är av begränsat värde som mått på vårdens kvalitet eftersom patienternas överlevnad varierar med sjukdomens prognos.

Med hjälp av riskjusteringsmodeller kan en förväntad mortalitet beräknas. Denna kan användas för att beskriva hur sjuka patienterna är på en intensivvårdsavdelning. Den förväntade mortaliteten kan jämföras med den faktiskt observerade mortaliteten samt för att beräkna standardiserad mortalitets kvot (eng. standardized mortality rate, SMR). SMR beräknas som kvoten mellan observerad och förväntad (predikterad) mortalitet. Förväntad och observerad mortalitet kan också användas för att beskriva utfall med VLAD-kurvor (eng. variable life-adjusted display). SMR och VLAD kurvor för svensk intensivvård publiceras bland annat på Svenska Intensivvårdsregistrets (SIR) hemsida (www.icuregswe.org) och används av Öppna jämförelser från SKL. Ofta presenteras data rangordnande vilket lätt kan bidra till missuppfattningar om förändringars betydelse.

Ingen riskjusteringsmodell kan exakt förutsäga hur det går för den enskilde patienten. Riskjusteringsmodellens uppgift är att utifrån den information som finns tillgänglig så exakt som möjligt prediktera sannolikheten för utgången. De faktorer som ingår i modellen varierar, men vanligen ingår ålder, kronisk sjuklighet samt en beskrivning av den akuta sjukligheten med t ex inläggningsorsak och fysiologiska parametrar. De modeller som används av SIR är för vuxna patienter (≥16 år) Simplified Acute Physiology Score 3 (SAPS3). För barn och ungdomar (<16 år) används Pediatric Index of Mortality 2 (PIM2). För vuxna patienter (≥16 år) som läggs in efter att ha genomgått hjärtkirurgi med hjärt-lungmaskin används Intensivvårds-Higgins (IVA-Higgins). Alla tre modellerna beräknar patientens risk för död genom att ett antal faktorers betydelse för vårdresultatet vägs med komplexa statistiska metoder.

Det är viktigt att känna till de olika riskjusteringsmodellernas begränsningar. Bakom precisionen i siffror kan det finnas avsevärda problem i modellernas prestanda (diskriminering och kalibrering, se nedan). Dessutom är precisionen i måtten (t ex SMR och VLAD kurvor) beroende av antalet dödsfall, och när detta är lågt blir konfidensintervallen stora och därmed blir värdet av utfallsmåtten begränsat.

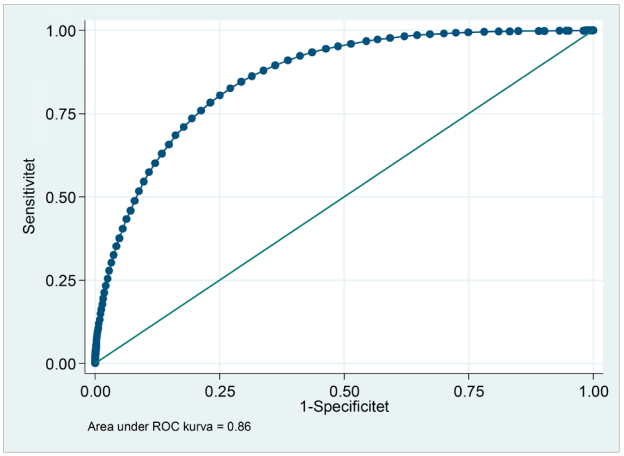
Man kan diskutera vilket resultatmått som är det mest relevanta. När dödlighet används som utfallsmått i internationella sammanhang är det oftast sjukhusmortalitet som presenteras eftersom svårigheter finns att följa upp patienterna efter att de lämnat sjukhuset. SIR, och andra, förordar ett utfallsmått som är fixerat till tiden, ofta används 30 eller 90 dagar efter inläggningen på IVA, eftersom det påverkas mindre av faktorer som är oberoende av patientens tillstånd. Det finns skäl att påpeka att andra utfallsmått än dödlighet, t ex hälsorelaterad livskvalitet och funktionsnivå kan vara viktiga komplement till informationen om patienten överlevt.

Hur bedömer man riskjusteringsmodeller?

Hur bra en riskjusteringsmodell fungerar (eng. performance) bör bedömas utifrån flera mått. Diskriminering beskriver modellens klassificeringsförmåga. Kalibrering beskriver modellens förmåga att diskriminera över hela riskintervallet. Andra mer sällan redovisade mått är Brier's score och pseudo r^2.

Diskriminering

Diskriminering beskriver i detta sammanhang modellens förmåga att skilja ut patienter som dör från de som överlever. Det vanligast måttet på modellens diskriminering är c-index, (c som i eng. concordance) som kan anta ett värde mellan 0.5 och 1, där 1 innebär perfekt klassificering och 0.5 att slumpmässig klassificering är lika bra. c-index är i princip identiskt med arean under 'receiver operating characteristic' (ROC) kurvan. ROC kurvan är en graf som illustrerar funktionen hos ett binärt klassificeringssystem (i detta fall död/levande) när diskrimineringströskeln (t ex riskpoäng) varieras. Den skapas genom att plotta fraktionen sant positiva delat med alla positiva (sensitivitet) mot fraktionen falskt positiva delat med alla negativa värden (vilket blir samma som 1-specificitet) för alla observerade tröskelinställningar (Figur 1). Ett c-index ≥0.7 brukar beskrivas som acceptabel diskriminering, ≥0.8 som utmärkt och om c-index ≥0.9 brukar diskrimineringen betraktas som utomordentlig. För att nå 1.0 måste det finnas ett tröskelvärde (beräknad risk) där sensitivitet och specificitet är 100 %, vilket är principiellt omöjligt givet det sätt som riskjusteringsmodellerna är uppbyggda.



Figur 1. ROC kurva. ROC kurvan är en graf som illustrerar funktionen hos ett binärt klassificeringssystem (i detta fall död/levande) när diskrimineringströskeln (t ex riskpoäng) varieras. Den skapas genom att plotta fraktionen sant positiva delat med alla positiva mot fraktionen falskt positiva delat med negativa värden för alla observerade tröskelinställningar (vilket blir samma som 1 – specificiteten). Arean under ROC-kurvan är arean som begränsas av ROC-kurvan.

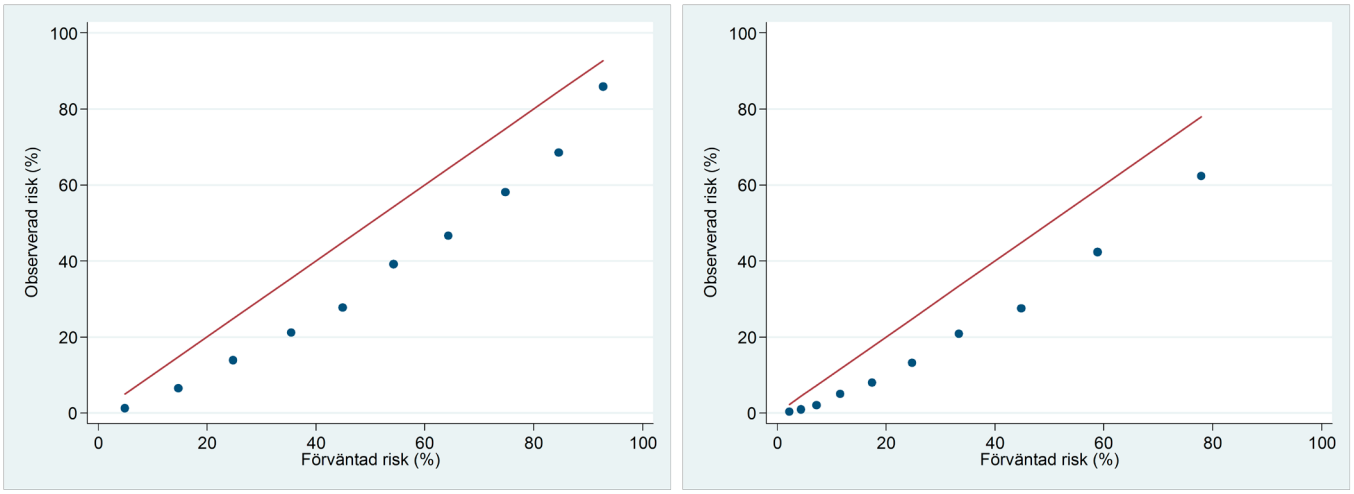
Kalibrering

Kalibrering beskriver modellens förmåga till prediktion över hela riskintervallet. Om modellen förutsäger en risk för död på 10 % för en viss riskprofil innebär detta att 1 av 10 individer med denna profil dör om modellen är väl kalibrerad. Kalibreringskurvor som visar observerad mortalitet ritad per riskintervall mot förväntad mortalitet i ett diagram ger en god bild av kalibreringen men ger föga formell information om den statistiska säkerheten i analysen (Figur 2). Ett vanligt sätt att formellt bedöma kalibreringen är att jämföra prediktion med utfall gruppvis med Hosmer-Lemeshows goodness of fit test (H-L). Två principiellt olika versioner av H-L test används; i H-L $\hat{C}$-test grupperas observationerna i lika stora grupper medan i H-L $\hat{H}$-test grupperas observationerna i lika stora riskintervall (Figur 2). För båda gäller att förväntad och observerad mortalitet beräknas varefter skillnaden mellan förväntat och observerat värde kvadreras och delas med förväntat värde. Resultatet summeras för de olika grupperna och utgör det $\chi 2$ värde varifrån ett P-värde beräknas. Om P är > 0,05 betyder det att skillnaden mellan förväntat och observerat värde inte är statistiskt säkerställd, dvs. kalibreringen är god. Tyvärr har H-L metoden låg styrka att upptäcka dålig kalibrering i små patientmaterial. Dessutom kan små betydelselösa skillnader i förväntad jämfört med observerad mortalitet leda till signifikanta utfall med H-L metoden i stora patientmaterial. En kalibreringsmetod som är mindre känslig för patientmaterialets storlek är Cox' kalibrerings regression (ska ej förväxlas med Cox' semiparametriska metod för analys av överlevnad, eng. Cox' proportional hazards model). Cox' kalibreringsregression är en regressionsanalys med förväntat utfall som enda ingående variabel och död/levande som utfall. Skärnings punkten med y-axeln (Cox' intercept) och linjens lutningskoefficient (Cox' slope) kan användas för att beskriva sambandet mellan förväntad och observerad mortalitet. I en perfekt kalibrerad modell är Cox' intercept = 0 och Cox' slope = 1. Denna analys är mindre känslig för patientmaterialets storlek, men resultatet baseras på ett genomsnitt vilket betyder att kalibreringen i vissa prediktionsgrupper kan vara dålig även om skärningspunkt och lutningskoefficient är nästan ideala.

Andra övergripande mått

Förutom ovan nämnda mått på hur väl riskmodellen fungerar finns flera andra sätt att bedöma modellens prestanda. Styrkan i kopplingen mellan beroende och oberoende variabel i en linjär regressions analys ges av korrelations koefficienten (*r*). Graden av varians, rutinmässigt uttryckt i procent, som kan förklaras av den linjära modellen kan kalkyleras som r^2. Analogt kan variansen i en binär modell kalkyleras som pseudo r^2. Detta kan göras på ett flertal sätt (t ex Nagelkerke, Efron, Cox & Snell, McFadden) där jämförbarheten och tolkningen är omdiskuterad, vilket bidragit till att detta mått sällan används för att beskriva riskmodellens funktion.

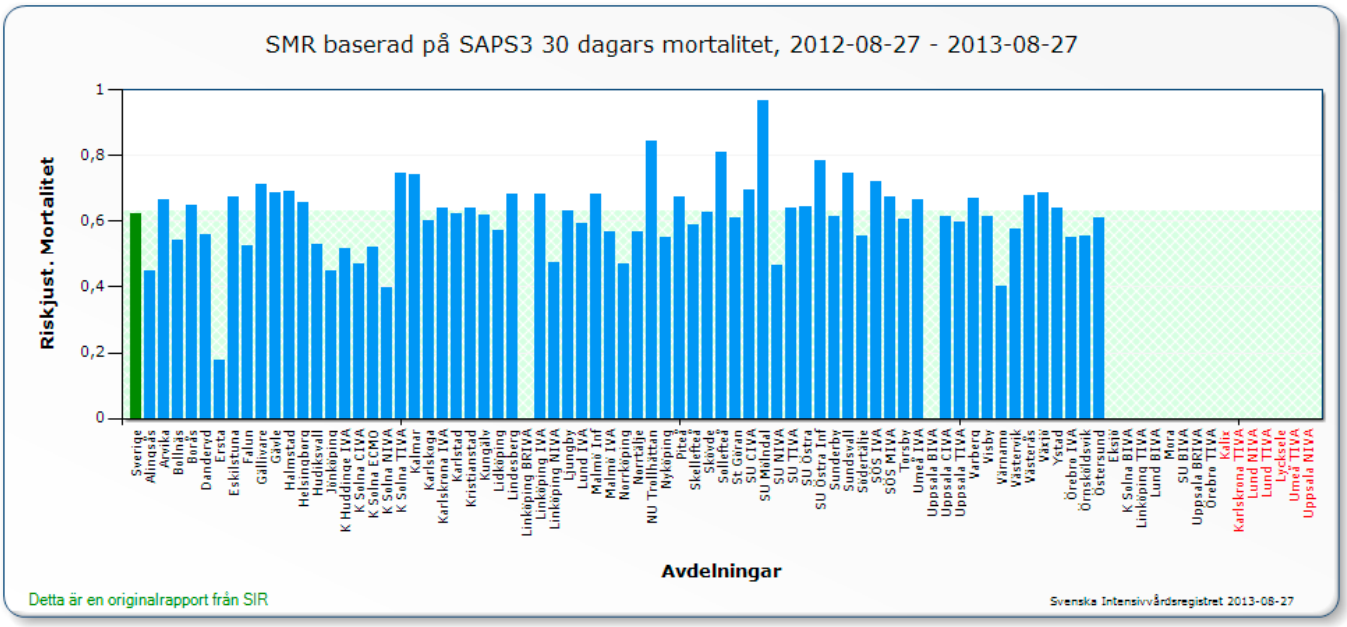
Träffsäkerhet (eng. accuracy) hos modellen kan bedömas med Brier's score. Brier är ett övergripande mått som beskriver informationsvärdet i riskskattningen och utvecklades ursprungligen inom meteorologin för att bedöma träffsäkerheten hos väderprognoser. Brier kan anta värden från 0 för en perfekt modell till 0,25 för en non-informativ modell med 50 % incidens av aktuellt utfall. Om incidensen av utfallet är lägre, t ex 18 % som för 30 dagars mortalitet, blir maximal Brier lägre (0,148). Man kan normalisera Brier enligt följande: 1- Brier dividerad med sitt maximala värde och får då en Scaled Brier som varierar mellan 0 % (non-informativ modell) och 100 % (perfekt modell).



Figur 2. Kalibrering enligt Hosmer-Lemeshow $\hat{H}$ där observationerna grupperats i 10 lika stora riskintervall (vänster bild) resp. $\hat{C}$ där observationerna grupperats i 10 ungefärligen lika stora grupper (höger bild). I båda fallen ses att kalibreringen är dålig H-L $\hat{H}$: χ^2 = 2248, df 8, P<0.001. H-L $\hat{C}$: χ^2 = 2214, df 8, P<0.001

Riskjustering inom den allmänna intensivvården av vuxna

I svensk intensivvård har APACHE (Acute Physiologic And Chronic Health Evaluation) II modellen för riskjustering använts åtminstone sedan slutet av 1980-talet då den första ansatsen att skapa en nationellt sammanhållen registrering gjordes. APACHE som utvecklades i USA på 1980-talet för att prediktera sjukhusmortalitet vid intensivvård modifierades för svenska förhållanden, t.ex. kompletterades Glasgow Come Score (GCS) med Reaction Level 85 (RLS85) för bedömning av medvetandegrad. När modellens funktion analyserades 2005 och 2011 var diskrimineringen god (c-index 0.81) men kalibreringen otillfredsställande, bl a överskattades risken i de högre åldersintervallen. APACHE II har dessutom nackdelen att de fysiologiska parametrarna registreras som sämsta värdet under de första 24 timmarna. Detta innebar att dålig vård i början av ett vårdförlopp kan ge en högre APACHE-poäng, en högre förväntad dödlighet och således en lägre riskjusterad mortalitet (ofta kallad Boyd and Grounds' effekt). SIR har sedan 2012 övergått till att endast använda SAPS3 (Simplified Acute Physiology Score) modellen vid riskjustering av allmän intensivvård av patienter ≥16 år.



Figur 3. SMR kurva från SIR:s dynamiska utdataportal 2013-08-27 (http://www.icuregswe.org/sv/Utdata/)

SAPS3

SAPS3 utvecklades i samarbete med den Europeiska intensivvårdsföreningen (ESICM) under 2002-4. Utvecklingen baserades på data från knappt 20 000 patienter som vårdades hösten 2002 på 309 intensivvårdsavdelningar i 35 länder. Alla patientkategorier inkluderades, även hjärtkirurgi som utgjorde 8.4% av den totala kohorten. Drygt 70 % av patienterna vårdades på europeiska intensivvårdsavdelningar. Patienter yngre än 16 år exkluderades. SAPS3 intagningspoäng utgör summan av tre delsummor (Faktaruta 1). Delsumma (Box) I beskriver patienten före intagning för intensivvård. Den omfattar ålder, komorbiditet, beskriver eventuell vistelse på sjukhus före intagning för intensivvård samt vissa specifika terapeutiska åtgärder. Box II beskriver omständigheter som leder till intagning för intensivvård. Dessa omfattar anledning/-ar till intagning inklusive planeringskategori, förekomst av infektion samt kirurgiskt status och, i förekommande fall, operationstyp. Box III beskriver förekomst och grad av fysiologisk störning under ett 2 timmarsintervall i anslutning till inledningen av intensivvården (1 timme före och efter ankomst).

Den statistiska modellen och dess individuella koefficienter finns offentligt tillgängliga och kan brukas helt fritt. Separata modeller för olika geografiska områden har utvecklats och finns likaledes fritt tillgängliga.

SAPS3 i SIR

SAPS3 är utvecklad för att prediktera sjukhusmortalitet men används av SIR för att prediktera utfallet 30 dagar efter att intensivvården inleddes. En annan förändring som SIR infört är att RLS85 kan användas i tillägg till GCS för bedömning av medvetandegrad.

Den ekvation som används i SIR sedan introduktionen av SAPS3 är den generella ekvationen. Den är inte kalibrerad för svenska förhållanden. Det innebär att den inte speglar aktuella förhållanden i svensk intensivvård och därför överskattar mortalitetsrisken. En konsekvens av detta är att global SMR är betydligt lägre än 1, vilket avspeglas av data som regelbundet presenteras i SIR:s årsrapporter och på hemsidans dynamiska portal (Figur 3).

Anpassning av SAPS3 till svensk intensivvård

En genomgång av SAPS3 användningen vid nästan 85 000 inläggningar i SIR på 71 IVA under perioden 2009-12 visar en god diskriminering (c-index 0,86) och god träffsäkerhet (Brier 0,12, Scaled Brier score 41 %) men dålig kalibrering med en kraftig överskattning av mortalitetsrisk (Cox' intercept -0,73, Cox' slope 1,11). Behovet av en rekalibrering (se Faktaruta 2) av SAPS3 blev därmed uppenbart, och en sådan genomfördes med utgångspunkt från SIR data. Vid arbetet användes rådata från knappt 55 000 intensivvårdstillfällen under 2011-12. När information om fysiologiska variabler (Box III) saknades fick dessa 0 poäng i enlighet med den ursprungliga SAPS3 modellen. Betydelsen av saknade variabler för modellens funktion beskrivs i detalj längre fram. Eftersom diskrimineringen var utmärkt genomfördes en andra gradens rekalibrering och den kalibrerade ekvationens utseende blev därför snarlik den ursprungliga SAPS3 ekvationen:
Logit = ln(SAPS 3 poäng +10,3352) • 7,2657 – 32,3084
(jfr Faktaruta 1).

Faktaruta 1
Om SAPS3 ekvationen och dess komponenter

SAPS3 innehåller tre delsummor:

BOX 1	Ålder, komorbiditet, plats före intagning på IVA, tid på sjukhus och vasoaktiv terapi före intensivvård.
BOX 2	Planerad /akut inläggninge, intagningsorsak, opererad eller ej/akut opererad, operationstyp, nosokomial infektion eller djup luftvägsinfektion.
BOX 3	Glascow Coma Scale/RLS85, bilirubin, kroppstemp, kreatinin, hjärtfrekvens, leukocyter, trombocyter, systoliskt blodtryck, oxygenering. (+/- 1h från ankomst till IVA)

För poänger och gränsvärden se:
http://www.icuregswe.org/Documents/ Guidelines/ Riskjustering_inom_svensk_intensivvard.pdf

Sambandet med utfallet (levande/död) vid utskrivningen från sjukhus beräknas ur ekvationen
Logit = ln(SAPS3 poäng + 20.5958) • 7.3068 - 32.6659

Sannolikhet för död (mortalitetsrisk) =
$e^{Logit} / (1+ e^{Logit})$

Faktaruta 2
Om rekalibrering

En rekalibrering kan i princip utföras på två olika sätt

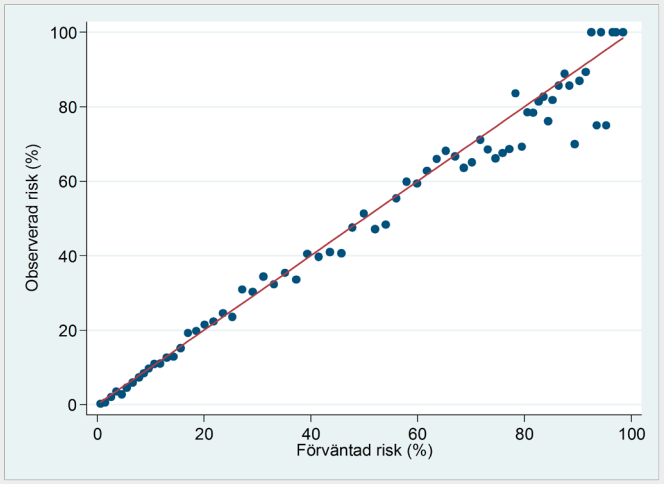
1:a graden	Rekalibrering av första graden innebär att tyngden av de ingående variablerna, liksom deras brytpunkter revideras och riskmodellens poäng beräknas om från grunden.
2:a graden	Rekalibrering av andra graden betyder att tyngderna och brytpunkterna för de ingående variablerna bevaras men poängsummans relation till det observerade utfallet korrigeras. En andra gradens rekalibrering ger i allmänhet en stabil modell då bara två koefficienter beräknas på nytt och är därför ett bra val om diskrimineringen är god

Träffsäkerheten förbättrades något och kalibreringen blev som väntat utmärkt (Figur 4). En konsekvens av den rekalibrerade modellen är att global SMR blir 1,0. Förändringen av SMR per intensivvårdsavdelning illustreras i Figur 5.

Betydelsen av saknade fysiologiska variabler i SAPS3

En farhåga vid övergången från APACHE II till SAPS3 modellen var att det smalare provtagningsfönstret (mellan 1 timme före och 1 timme efter inläggningstidpunkten på IVA jämfört med sämsta värdet de 24 första timmarna på IVA) skulle innebära ökat bortfall av fysiologiska data och att detta skulle försämra riskmodellens funktion. Därför genomfördes en analys av drygt 30 000 intensivvårdstillfällen under perioden 2009-10 för att utvärdera förekomsten av och effekten av saknade fysiologiska data. Data som beskriver patientens egenskaper innan intagning och omständigheter för intagning (Box I och II) var fullständiga, medan knappt 50 % av alla vårdtillfällen hade inkompletta fysiologiska data (Box III). De variabler som oftast saknades var oxygenering, bilirubin och pH. Saknade variabler var vanligare hos överlevande än hos patienter som inte överlevde och, i överensstämmelse med detta minskade antalet saknade variabler med ökad risk. Diskriminering och kalibrering utvärderades i hela datasetet och i delmängder där ett varierande antal variabler saknades (Tabell 1). Vi fann paradoxalt nog att träffsäkerhet och diskriminering inte försämrades med ett ökat antal saknade variabler. Dock spelade det roll vilka variabler som saknades; avsaknad av oxygenering spelade mindre roll än avsaknad av information om patientens medvetandegrad. Kalibreringen var genomgående dålig.

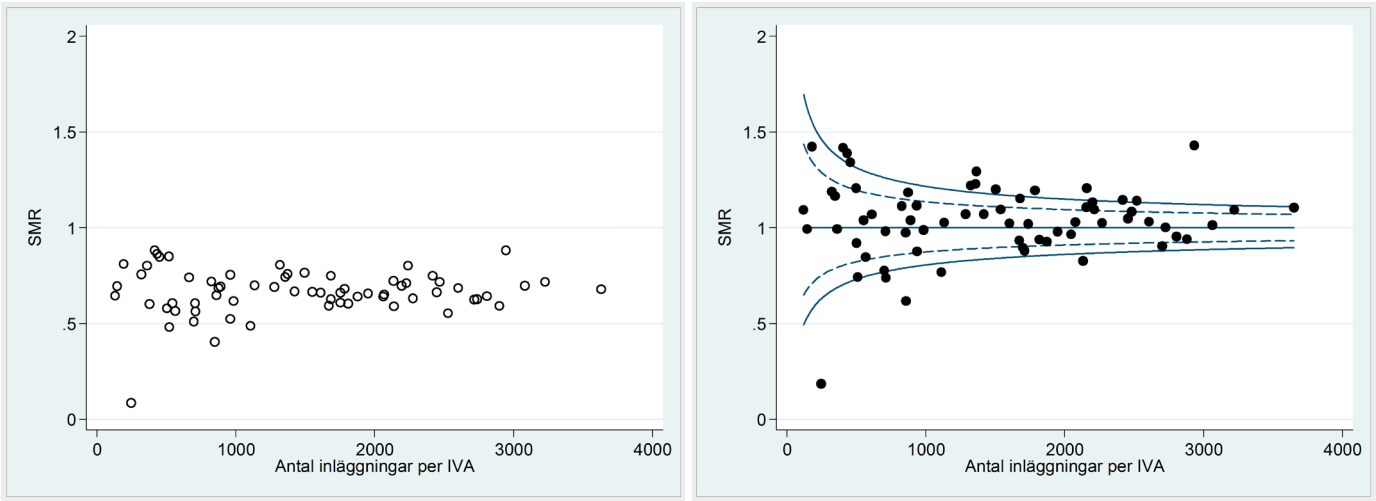
Diskriminering och kalibrering utvärderades även efter att bortfall av en variabel simulerats. För dettasena reändamålanvändeviSAPS3 registreringar utan några saknade variabler och använde en slumpalgoritm för att ersätta en variabel med ett saknat värde (0 poäng).



Figur 4. SAPS3 efter rekalibrering. Förhållandet mellan observerad och förväntad mortalitet har plottats för varje tröskelvärde. Antalet observationer i den högsta decilen av förväntad risk (90-100 %) var få (87 st). c-index förblev oförändrad (0,86) efter rekalibreringen medan Brier förbättrades från 0,12 till 0,10. Scaled Brier score 28 %. Cox' intercept och slope var 0 resp 1 för den rekalibrerade modellen.

Vi fann att ett simulerat bortfall genomgående ledde till sämre diskriminering och lägre andel korrekt klassificerade patienter (Tabell 2).

En slutsats av denna analys är att SAPS 3 presterar bra i daglig praxis med begränsad mängd saknade fysiologiska data. En annan slutsats är att bortfallet av fysiologiska data inte är slumpmässigt, det verkar finnas en koppling mellan en förutsedd låg risk för död och saknade fysiologiska data.



Figur 5. SMR före (vänster) resp. efter (höger) en andra gradens rekalibrering. Linjerna i det högra diagrammet visar global SMR (heldragen linje) samt två signalnivåers avvikelse (≈2 SD resp 3 SD).

Tabell 1. Den ursprungliga SAPS3 modellens funktion med avsaknade fysiologiska värden.

Saknade variabler	Antal obs.	Brier score (95% CI)	c-index (95% CI)	Cox' intercept	Cox' slope
Ingen saknas	16992	0,14 (0,14-0,15)	0,83 (0,83-0,84)	-0,87	1,06
1 saknas	4862	0,11 (0,10-0,11)	0,87 (0,86-0,88)	-0,81	1,21
2 saknas	3768	0,07 (0,0 -0,08)	0,88 (0,87-0,90)	-0,53	1,30
3 saknas	1891	0,08 (0,07-0,09)	0,88 (0,86-0,90)	-0,31	1,21
4 saknas	1484	0,11 (0,10-0,12)	0,86 (0,84-0,88)	-0,32	1,19
5 saknas	710	0,09 (0,08-0,10)	0,87 (0,83-0,90)	-0,25	1,22
Bilirubin	1971	0,13 (0,12-0,14)	0,86 (0,84-0,88)	-0,86	1,23
Oxygenering	1414	0,07 (0,06-0,07)	0,90 (0,88-0,93)	-0,85	1,34
Temperatur	514	0,13 (0,11-0,14)	0,86 (0,82-0,90)	-0,69	1,12
pH	347	0,08 (0,06-0,10)	0,89 (0,84-0,95)	-0,75	1,23
GCS / RLS85	282	0,15 (0,13-0,17)	0,80 (0,74-0,86)	-0,80	0,90
Trombocyter	135	0,14 (0,11-0,17)	0,85 (0,79-0,92)	-0,51	1,24

Tabell 2. Den ursprungliga SAPS3 modellens funktion med primärt resp. simulerat saknade fysiologiska värden.

Saknad variabel	Dataset	Brier score (95% CI)	c-index (95% CI)	Cox' intercept	Cox' slope
Ingen saknas	Primärt	0,14 (0,14-0,15)	0,83 (0,83-0,84)	-0,87	1,06
Bilirubin	Primärt	0,13 (0,12-0,14)	0,86 (0,84-0,88)	-0,86	1,23
	Simulerad	0,15 (0,13-0,15)	0,83(0,81-0,85)	-0,78	1,04
Oxygenering	Primärt	0,07 (0,06-0,07)	0,90 (0,88-0,93)	-0,85	1,34
	Simulerad	0,14 (0,13- 0,15)	0,81 (0,78-0,83)	-0,36	1,01
Temperatur	Primärt	0,13 (0,11-0,14)	0,86 (0,82-0,90)	-0,69	1,12
	Simulerad	0,13 (0,12-0,15)	0,85 (0,82-0,89)	-0,79	1,09
pH	Primärt	0,08 (0,06-0,10)	0,89 (0,84-0,95)	-0,75	1,23
	Simulerad	0,13 (0,12-0,15)	0,85 (0,80-0,89)	-0,97	1,24
GCS / RLS85	Primärt	0,15 (0,13-0,17)	0,80 (0,74-0,86)	-0,80	0,90
	Simulerad	0,15 (0,12- 0,17)	0,80 (0,75-0,85)	-0,84	0,77
Trombocyter	Primärt	0,14 (0,11-0,17)	0,85 (0,79-0,92)	-0,51	1,24
	Simulerad	0,15 (0,12-0,17)	0,85 (0,78-0,91)	-0,61	1,09

Riskjustering i thoraxintensivvård

En avsevärd andel hjärtkirurgiska patienter ingick i den kohort ur vilken SAPS3 utvecklades men SAPS3 har ändå visat sig fungera dåligt i denna patientkategori. SIR använder därför IVA-Higgins för riskjustering av vårdens förlopp efter hjärtkirurgi. Med hjärtkirurgi menas i detta sammanhang hjärtoperationer med hjärt-lungmaskin, kranskärlskirurgi på slående hjärta, perkutana klaffprocedurer, centralt inlagd ECMO samt öppen kirurgi på thorakalaorta.

Intensivvårds-Higgins

Intensivvårds (IVA) - Higgins utvecklades på en kohort av 2 440 patienter som genomgick kranskärlsoperation, med eller utan samtidig klaff eller carotisoperation, vid ett hjärtkirurgiskt center i USA på 1990-talet. Modellen validerades på 2 125 patienter vid samma instution. Sjukhusmortalitet som användes som utfallsvariabel var låg, endast 78 patienter dog.

IVA-Higgins i SIR

Sellgren och medarbetare gjorde 2002 en anpassning till svenska förhållanden av IVA-Higgins (Tabell 3) och denna spreds från Göteborg till andra svenska thoraintensivvårdsavdelningar (THIVA). Eftersom diskrimineringen var dålig (c-index 0.74 för en samlad svensk THIVA kohort år 2009) gjordes en första gradens rekalibrering av Freter och medarbetare där brytpunkter och poängberäkningen modifierades (Tabell 3).

Denna modell utvecklades ur 8 210 vårdtillfällen på 4 THIVA avdelningar under åren 2007-2010. Jämfört med original modellen finns tre viktiga förändringar. Data som saknas förmodas vara normala, utfallet (död eller levande) 30 dagar efter inläggning på intensivvårdsavdelning används istället för utfallet under sjukhusvistelsen samt andra typer av hjärtkirurgi än kranskärlskirurgi ingår i modellen. Funktionen hos den rekalibrerade IVA-Higgins modellen är betydligt förbättrad jämfört med den som användes tidigare. Diskrimineringen är utmärkt och kalibreringen god (Faktaruta 3). Emellertid krävs sannolikt en fortsatt anpassning av IVA-Higgins, eftersom modellen beräknats med utgångspunkt från ett litet antal utfall (196 döda, vilket trots allt är avsevärt fler än original Higgins). Risken med ett begränsat antal utfall är att koefficienterna i modellen blir alltför 'optimistiska', dvs. överanpassade till den aktuella databasen. Det leder till en sänkt Cox´ slope när man kontrollerar kalibreringen på senare data. Detta kan justeras antingen genom en krympning (eng. shrinkage) av koefficienterna från utvecklingsdatabasen eller genom att helt enkelt göra en övergripande första gradens rekalibrering med senare data.

Tabell 3. Skillnader i ingående variabler och gränsvärden mellan den ursprungliga och svenska IVA-Higgins modellerna.

Variabler	Higgins original	Göteborg 2002	SIR 2010
Kroppsyta	< 1,72 m²	< 1,72 m²	< 1,71 m²
Antal tidigare hjärtoperationer	Ingår	Ingår	Ingår ej
Tidigare perifer kärlkirurgi	Ingår	Ingår	Ingår ej
Ålder	≥ 70 år	≥ 70 år	> 71 år
Kreatinin preop	≥ 1.9 mg/dl	> 167 µmol/l	> 165 µmol/l
Albumin preop	< 3.5 mg/dl	< 35 mg/l	< 35 g/L
ECC tid	≥ 160 min	≥ 160 min	> 153 min
IABP eller VAD efter ECC	Ingår	Ingår	Ingår
A-a O2 gradient	≥ 250 mmHg	≥ 33 kPa	≥ 37,5 kPa
Hjärtfrekvens	≥ 100/min	≥ 100/min	≥ 80/min
Hjärtindex	< 2.1 l/min m2	Ingår ej	Ingår ej
Total O2-extraktion baserad på SvO2 / ScvO2	Ingår ej	>0.5 / ingår ej	> 0,46 / >0.51
CVP	≥ 17 mmHg	≥ 17 mmHg	> 11 mmHg
Arteriellt HCO_3	< 21 mmol/l	Ingår ej	Ingår ej
Basöverskott	Ingår ej	< - 4 mmol/l	< -5 mmol/l

Faktaruta 3

Om den rekalibrerade IVA-Higgins modellen

Sannolikhet för död (mortalitetsrisk) = $e^{Logit} / (1+ e^{Logit})$

Logit = -5,63059 + Kroppsyta * 0,68818
+ Ålder * 0,95102
+ Kreatinin preop * 1,57537
+ Albumin preop * 0,65316
+ ECC tid * 0,84451
+ IABP eller VAD * 1,7926
+ A-a O2 gradient * 0,63809
+ Hjärtfrekvens *0,59227
+ Total O2-extraktion * 0,94249
+ CVP * 0,79236
+ Basöverskott * 1,02394

Varje variabel blir 1 om värdet ligger över resp. under gränsvärdet (jfr tabell 3), annars blir variabeln 0. Modellen har utvecklats från en additiv modell till en logistisk modell.

Prestanda för rekalibrerad IVA-Higgins	
c-index	0,87
Brier score	0,021
Scaled Brier score	11 %
Cox' intercept	0,020
Cox' slope	1,00

Riskjustering vid intensivvård av barn

De mest använda modellerna för riskjustering av vårdförlopp på IVA är utvecklade på patientkohorter vars nedre åldergräns sällan omfattar yngre patienter. Även om SAPS3 inkluderade patienter från 16 år och uppåt var endast 2 % av patienterna 16-20 år. Den kohort ur vilken IVA-Higgins utvecklades saknade helt yngre patienter. Det finns huvudsakligen två riskjusteringsmodeller för barn, Pediatric Index of Mortality (PIM) och Pediatric Risk of Mortality (PRISM). PIM2 används spritt över hela världen, dock fortfarande inte lika mycket som PRISM II och III. PIM2 är enklare, med färre parametrar och har fördelen att värdera barnen när de anländer till intensivvården och inte efter att ha varit vårdade under 12-24 timmar. SIR använder PIM2 för vårdtillfällen inom såväl den specialiserade barnintensivvården som för den intensivvård av barn som bedrivs på andra intensivvårdsavdelningar.

PIM2

PIM2 är utvecklad i Australien på knappt 21 000 intensivvårdstillfällen med en IVA-mortalitet på 5,3 %. Modellen är byggd på information som insamlas vid tiden för barnets omhändertagande av intensivvården och utfallsvariabel är levande eller död vid utskrivning från IVA (Faktaruta 4). Patienter som överfördes mellan flera IVA ingick inte analysen.

PIM2 i SIR

PIM2 används i SIR vid all intensivvård av barn <16 år. Under perioden 2008-12 finns PIM2 registrerat för knappt 13 000 vårdtillfällen. IVA mortaliteten för dessa var ca 2 %. Ungefär två tredjedelar av dessa vårdtillfällen var på någon av landets 4 specialintensivvårdsavdelningar för barn, resterande var spridda på landets övriga intensivvårdsavdelningar. Diskrimineringen var utmärkt och träffsäkerhet god men kalibreringen mindre bra med Cox' intercept -1,17 och slope 0,93. Därför genomfördes en andra gradens rekalibrering med oförändrad relativ viktning av de ingående variablerna men med en justering av den övergripande koefficienten. Resultatet av denna visar som förväntat en bevarad diskriminering och träffsäkerhet och en förbättrad kalibrering (Figur 6, Faktaruta 4).

Faktaruta 4
Om PIM2

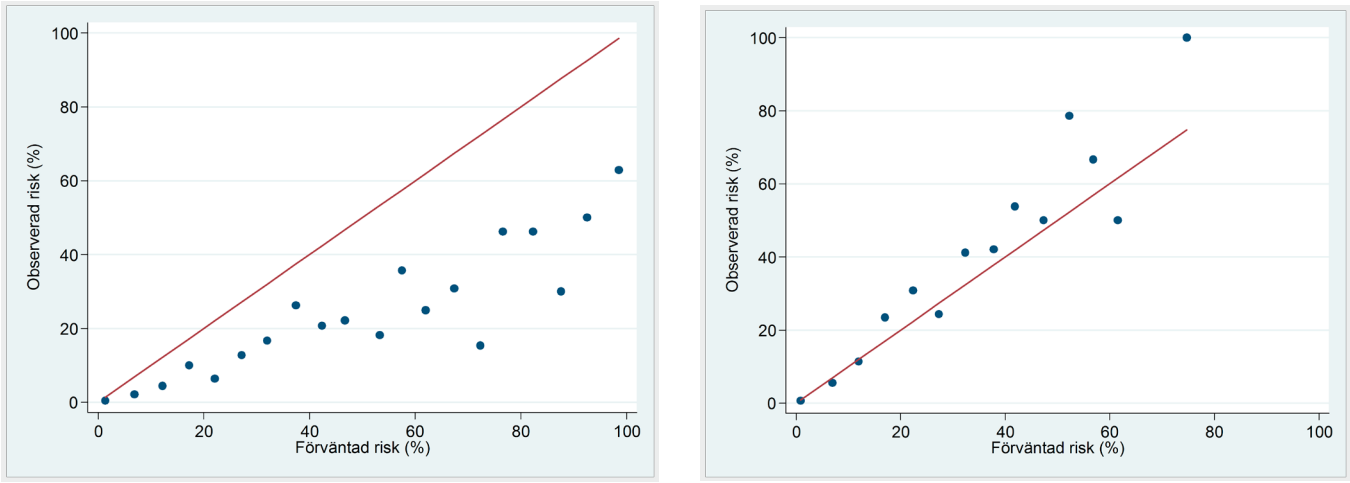
Den ursprungliga modellen:

Sannolikhet för död (mortalitetsrisk) = $e^{Logit} / (1+ e^{Logit})$

Logit = {0,01395* [absolut (systoliskt blodtryck – 120)]}
+ (3,0791*pupillreaktion 0 eller 1)
+ [0,2888*(FiO2 i %/PaO2 i kPa x 7,5)]
+ {0,1040*[absolut (+/- BE)]}
+ (1,3352* mekanisk ventilation 0 eller 1)
– (0,9282*elektiv inläggning 0 eller 1)
– (1,0244* postoperativ vård 0 eller 1)
+ (0,7507* hjärtlungmaskin 0 eller 1)
+ (1,6829* högrisk diagnos 0 eller 1)
– (1,5770 lågrisk diagnos 0 eller 1)
– 4,8841

Prestanda för rekalibrerad PIM2

c-index	0,88
Brier score	0,017
Scaled Brier score	54 %
Cox' intercept	0,00
Cox' slope	1,00



Figur 6. PIM2 före (vänster) resp efter (höger) en andra gradens rekalibrering. En enkel rekalibrering där man byter ut koefficienten -4,8841 mot -5,63353 och multiplicerar övriga koefficienter med 0,96079 får man för 2008-2012 en SMR på 1, Cox´s intercept 0 och slope 1. Trots detta leder det lilla antalet observationer bland de högre riskgrupperna till en avvikande form på kalibreringsbilden i de högre riskintervallen.

Generella slutsatser från arbetet med riskjusteringsmodeller i SIR

De tre riskjusteringsmodeller som används av SIR uppvisar alla en utmärkt diskriminering, men för att öka användbarheten har alla modellerna rekalibrerats med utgångspunkt från SIR data. IVA-Higgins har genomgått en rekalibrering med förändring av brytpunkter och vikter av ingående variabler (första gradens rekalibrering), vilket förbättrade diskrimineringen. SAPS3 och PIM2 har rekalibrerats med korrektion av poängsummans relation till utfallet (andra gradens rekalibrering). Alla tre riskjusteringsmodellerna uppvisar nu goda egenskaper, dock är det troligt att IVA-Higgins behöver genomgå en andra gradens rekalibrering inom en nära framtid. För PIM2 vore det önskvärt att inkludera patienter som transfereras mellan intensivvårdsavdelningar i modellen. En sådan utveckling kan komma att medföra ett behov av en mer omfattande rekalibrering av PIM2 modellen.

En positiv effekt av att arbeta med internt välkalibrerade modeller är möjligheten att jämföra sin avdelnings resultat med en för riket standardiserad norm (SMR). Viktigt är dock, som vi påpekat tidigare, att tänka på att konfidensintervallet kring SMR är ganska stort, eftersom det uppskattas utifrån antalet döda inom 30 dagar från intensivvårdstillfällets start. Ett problem är att en övergång till rekalibrerade modeller betyder att man inte automatiskt kan jämföra sina resultat med äldre resultat, inte heller med internationella resultat. Principiellt kan detta ändå göras genom att de äldre modellernas koefficienter bevaras och vid behov kan användas parallellt. Regelbunden revision av riskjusteringsmodellerna, förslagsvis vart 5:e år, är en trolig utveckling för att modellerna ska vara aktuella och ha bibehållet goda prestanda.

Litteraturhänvisning

Riktlinjer på SIR:s hemsida:
(http://www.icuregswe.org/Documents/Guidelines/ Riskjustering_inom_svensk_intensivvard.pdf)

Steyerberg et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 2010; 21: 128–138.

SAPS3 projektets hemsida (http://www.saps3.org/news)

Higgins et al. ICU Admission Score for Predicting Morbidity and Mortality Risk After Coronary Artery Bypass Grafting. Annals of Thoracic Surgery.1997; 64: 1050–8

Slater et al. PIM 2: a revised version of the Paediatric Index of Mortality. Intensive Care Med 2003; 29: 278-85.