

Voorspelling van oorlewing in 'n chirurgiese intensiewesorgeenheid

Louis J. Roux, Margaretha Viljoen, Ester Viljoen, Pieter J. Becker, Jan P. Pretorius

Doel. Om die voorspellingswaarde van 14 bestaande voorspellingsprogramme op chirurgiese intensiewesorgpasiënte te toets en om 'n oorlewingsvoorspellingsmodel vir die chirurgiese intensiewesorgeenheid (CISE) te ontwikkel wat as instrument kan dien om kliniese besluitneming te vergemaklik.

Ontwerp. Retrospektiewe en prospektiewe versameling en rekenariserings van kliniese, fisiologiese en biochemiese veranderlikes. Student se *t*-toets, logistiese regressie, sensitiviteits- en spesifisiteitsberekening met behulp van tweerigting tabelle.

Plek. Chirurgiese Intensiewesorgeenheid, H. F. Verwoerd-hospitaal, Pretoria.

Pasiëntpopulasie. Retrospektiewe deel van studie: 188 pasiënte; prospektiewe deel: 104 pasiënte.

Bevindingsmaatstawwe. Statisties betekenisvolle verskille tussen die tellings van oorlewendes en nie-oorlewendes. Aanvaarbare sensitiviteit en spesifisiteit van ontwikkelde model.

Resultate. Al 14 bestaande voorspellingsprogramme beskik oor die vermoë om te onderskei tussen die opnamedagtellings van oorlewendes en nie-oorlewendes en het, by implikasie, dus prognostiese geldigheid. Die statistiese betekenisvolheid geassosieer met die onderskeie programme het gewissel van $P < 0.01$ tot $P < 0.0001$. 'n Reeks van opeenvolgende oorlewingsvoorspellingsmodelle, ontwikkel deur logistiese regressie-analise op die bes passende voorspellingsprogramme, het uiteindelik gelei tot die ontwikkeling van 'n model met 'n sensitiviteit van 93% en 'n spesifisiteit van 66%.

Gevolgtrekking. Verskeie bestaande voorspellingsprogramme kan bydra tot kliniese besluitneming. Voorspellingsmodelle hieruit ontwikkel moet voor implementering herhaaldelik op opeenvolgende groepe van die teikenpopulasie getoets word. Beide die programme en die modelle behoort as *bykomstige* prognostiese ondersteuningsfaktore gebruik te word eerder as absolute aanduidings van uitkoms.

S Afr Med J 1996; 86: 1417-1420.

Die behoefte om akkuraat te kan voorspel of pasiënte 'n kans op oorlewing het, bestaan reeds sedert die ontwikkeling van intensiewesorg as 'n onafhanklike dissipline. In die verlede was die kliniese beoordeling van pasiënte se siektetoestand die enigste redelike projeksie van hul kans op oorlewing. Oor die afgelope aantal jare is daar egter gepoog om deur middel van berekenbare tellings (scoring systems), gebaseer op kliniese inligting, asook biochemiese merkers, 'n meer akkurate voorspelling van oorlewing of sterfte te maak. Met die samestelling van die berekenbare tellings, of te wel voorspellingsprogramme, word verskeie faktore en kombinasies van faktore geïdentifiseer wat 'n meer realistiese en objektiewe oorsig van pasiënte se siektetoestand weerspieël.¹ Die voorspellingsprogramme moet na samestelling baie deeglik vir akkuraatheid getoets word.^{2,3}

Alle voorspellingsprogramme en indekse bevat inherente swakhede wat voorspellingsakkuraatheid direk beïnvloed.⁴ Voorbeelde hiervan sluit in (i) die ontoereikende definiëring van sekere siektetoestande soos sepsis; (ii) die onvermoë om die aanvang van orgaan-disfunksie en versaking te identifiseer; (iii) die foutiewe indeling van pasiënte in bepaalde diagnostiese kategorieë; (iv) die relatiewe variasie in die beoordeling van die graad van afwyking in die veranderlikes wat gebruik word om 'n programtelling te bereken — dit is veral van belang waar daar van meer as een waarnemer gebruik gemaak word; (v) die identifisering van vals-negatiewe en vals-positiewe waardes; (vi) die feit dat pasiënte tydens verblyf van een diagnostiese kategorie (byvoorbeeld trauma) kan verskuif na een of meer ander kategorieë (byvoorbeeld sepsis en/of multi-orgaan disfunksie); en (vii) die onvermoë om te kan voorspel wat na ontslag uit die intensiewesorgeenheid gebeur — byvoorbeeld massiewe pulmonale embolisme.⁴ Hierdie laaste feit gee daartoe aanleiding dat, in die evaluering van ontwikkelde programme, die uitkoms van pasiënte as 'oorleef' genoteer word, terwyl hulle kort na ontslag sterf. Ander faktore wat die akkuraatheid van hierdie tipe sisteem kan benadeel sluit in medikasie, intraveneuse vloeistofoediening, bloedtransfusies en kunstmatige ondersteuning van vitale funksies. Sommige voorspellingsprogramme soos die 'Acute Physiology and Chronic Health Evaluation (APACHE)' is aanvanklik ontwerp om pasiënte onafhanklik van enige terapeutiese ingrepe te beoordeel.^{4,5} In die geval van chirurgiese intensiewesorgpasiënte, is dit veral problematies aangesien programtellings kunstmatig verhoog of verlaag mag wees afhangende van die terapeutiese ingrepe in die tydperk voor opname in intensiewesorg.

Departement Fisiologie, Fakulteit Geneeskunde, Universiteit van Pretoria

Louis J. Roux, M.Sc.

Margaretha Viljoen, M.Sc., Ph.D.

Sentrum vir Epidemiologiese Navorsing in Suider-Afrika, Departement Biostatistiek, Mediese Navorsingsraad, Pretoria

Ester Viljoen, B.Sc. HONS

Pieter J. Becker, M.Sc., Ph.D.

Chirurgiese Intensiewesorgeenheid, H. F. Verwoerd-hospitaal, Pretoria

Jan P. Pretorius, M.B. CH.B., M.MED. (CHIR.)

Die meeste voorspellingsprogramme is ontwikkel met die oog op 'n spesifieke diagnostiese teikengroep. Dit wil egter voorkom asof dit ook, binne diagnostiese groeperings, populasie-spesifiek is. Verskeie outeurs het reeds aangetoon dat die sukses van voorspelling van oorlewing met behulp van programme of indekse kan verskil in verskillende pasiëntpopulasies en ook in verskillende intensiewesorgeenhede.^{6,7} Die belangrikste redes hiervoor is waarskynlik dat opnamekriteria en behandelingsprotokolle in intensiewesorgeenhede van mekaar verskil. Dit is dus wenslik om enige van die bestaande voorspellingsprogramme op die teikenpopulasie aan die ware uitkoms te toets voordat dit toegepas word.^{8,9}

Die meeste voorspellingsprogramme is vir diagnostiese homogene groepe van pasiënte geskryf. In hierdie studie is daar van 14 bestaande, hoofsaaklik diagnose-gekoppelde voorspellingsprogramme (bv. die 'Trauma Score') gebruik gemaak in 'n poging om 'n meer akkurate oorlewingsvoorspellingsmodel vir die heterogene pasiëntpopulasie van die Chirurgiese Intensiewesorgeenheid (CISE) van die H.F. Verwoerd-hospitaal te ontwikkel.

Metodes

Die 14 programme wat deel van hierdie studie gevorm het, was die 'Acute Physiology and Chronic Health Evaluation (APACHE II)', 'Acute Physiology Score (APS)', 'Simplified Acute Physiology Score (SAPS)', 'Critical Care Scoring System (CCSS)', 'Trauma Score (TS)', 'Revised Trauma Score (RTS)', 'Hanover Intensive Score (HIS)', 'Glasgow Coma Score (GCS)', 'Injury Severity Score (ISS)', 'Multiple Organ Failure Score of Goris (MOFSG)', 'System Outcome Score (SOS)', 'Organ Failure Score of Roth (OFSR)', 'Organ Failure Score of Jacobs (OFSJ)' en die 'Sepsis Severity Score (SSS)'.

Die eerste 7 genoemde programme word hoofsaaklik vanaf bekende fisiologiese veranderlikes soos sistoliese bloeddruk, temperatuur, polsspoed en respiratoriese tempo bereken. Die daaropvolgende 2 programme berus hoofsaaklik op kliniese waarnemings soos die neurologiese status van die pasiënt en die graad van besering. Die res van die programme is hoofsaaklik ontwikkel om die intensiteit van die patofisiologiese versteuring te kwantifiseer.

Die 2 uitkomstegroepe, naamlik dié wat oorleef en dié wat sterf, is met behulp van die voorspellingswaardes gebaseer op die retrospektief versamelde data ($N = 188$) uit die genoemde voorspellingsprogramme ('scoring systems') vergelyk. Die programme wat die beste geprester het is opgeneem in 'n aantal oorlewingsvoorspellingsmodelle en hierdie modelle is getoets aan die hand van die ware uitkoms van beide retrospektief- ($N = 188$) en prospektief-versamelde data ($N = 104$) van 292 pasiëntopnames oor 'n gesamentlike tydperk van 24 maande.

Vir die doel van die studie is 2 opeenvolgende pasiëntpopulasies van die CISE van die H.F. Verwoerd-hospitaal gebruik. 'n Totaal van 292 CISE pasiënte was by hierdie studie betrek. Die evaluering van die 14 prognostiese programme het eerstens berus op data wat retrospektief verkry is van 188 opeenvolgende pasiëntopnames in die

CISE oor 'n proeftydperk van 12 maande. Data is vanuit pasiëntlêers versamel en in 'n aangepaste gerekenariseerde databasis opgeneem.

Die prognostiese vermoëns van elk van die 14 programme is bereken deur die opnametellings van die oorlewendes te vergelyk met die opnametellings van die oorledenes. Betekenisvolheid is met behulp van Student se t -toets beoordeel, en met die Mann-Whitney toets geverifieer. Die gepaste t -toets, hetsy vir gelyke of vir ongelyke variansies, is gebruik na aanleiding van die uitslag van Levine se toets vir gelykheid van die variansies van die twee groepe. Die programme wat die grootste bydrae tot die voorspellingsukses gelewer het (op grond van opnametellings), is met behulp van logistiese regressie geïdentifiseer, met inagneming van multikolineariteit, en opgeneem in opeenvolgende logistiese waarskynlikheidsmodelle (P -leef' tot P -leef²) vir beide die retrospektiewe groep ($N = 188$) en die prospektiewe groep ($N = 104$). Die voorspellingsakkuraatheid van die modelle is aan die ware uitkoms van die pasiëntgroepe getoets. Sensitiwiteit en spesifisiteit is vir die logistiese regressiemodelle bereken.

Goedkeuring vir die uitvoer van die studie is van die Etiese Komitee van die H. F. Verwoerd-hospitaal verkry.

Resultate

Met betrekking tot die programtellings wat retrospektief vanuit die opnamedata van die 188 CISE pasiënte bereken is, is getoon dat vir elk van die 14 voorspellingsprogramme betekenisvol onderskei kon word tussen pasiënte wat oorleef en dié wat gesterf het (Tabel I).

Tabel I. Gemiddelde opnametellings en standaardafwykings (Std) van pasiënte wat oorleef en dié wat gesterf het*

Groep	Oorleef			Sterf			P -waarde vir t -toets
	N	\bar{x}	Std	N	\bar{x}	Std	
APACHE II	153	21.58	9.40	35	46.73	10.46	< 0.0001
APS	153	18.26	6.12	35	38.51	8.14	< 0.01
GKS	153	14.59	0.68	35	10.00	2.40	< 0.01
MOFSG	153	3.98	1.72	35	9.94	1.55	< 0.0001
SOS	153	2.40	1.20	35	6.83	1.60	< 0.01
CCSS	153	18.93	8.19	35	50.97	11.49	< 0.01
SAPS	153	11.20	4.61	35	29.3	5.02	< 0.0001
TS	103	11.39	2.08	15	4.97	1.81	< 0.0001
RTS	153	4.89	0.87	35	3.77	0.84	< 0.0001
ISS	61	18.97	8.20	11	50.97	11.49	< 0.01
OFSR	153	30.31	8.78	35	58.66	11.73	< 0.0001
HIS	153	7.65	2.33	35	17.86	2.63	< 0.0001
OFSJ	153	5.09	1.15	35	10.09	1.52	< 0.0001
SSS	153	19.57	8.75	35	50.97	11.49	< 0.01

* Betekenisvolheid (Student se t -toets) word as $P \leq 0.05$ aanvaar.

Stapsgewyse logistiese regressie-analise op die opnametellings van die 14 programme het die programme MOFSG, SAPS en HIS in die voorspellingsmodel vir

oorlewing opgeneem. Die waarskynlikheidsmodel (P -leef^f) vir hierdie programkombinasie is:

$$P(\text{leef}^f) = \frac{e^{43.251 - 1.249 \cdot \text{MOFSG} - 0.365 \cdot \text{SAPS} - 1.162 \cdot \text{HIS}}}{1 + e^{43.251 - 1.249 \cdot \text{MOFSG} - 0.365 \cdot \text{SAPS} - 1.162 \cdot \text{HIS}}} \dots\dots\dots 1)$$

By 'n afsnytpunt vir P van 0.275 waar 'n pasiënt met 'n waarde van $P \geq 0.275$ geklassifiseer word om te oorleef en om te sterf indien $P < 0.275$, het die model al 153 pasiënte wat oorleef het korrek geklassifiseer as moontlike oorlewendes, terwyl 33 van die 35 sterftes se uitkoms korrek voorspel is. Die totale korrekte voorspelde waarde vir hierdie model is 98.9%.

'n Steekproef van 141 pasiënte (75%) is hierna sistematies uit die groep van 188 getrek en in 'n stapsgewyse logistiese regressie-analise van die 14 programme se opnametellings is die programme MOFSG en HIS opgeneem, waaruit die waarskynlikheidsmodel (P -leef^f) volg. Indien $P \geq 0.625$ dui die waarskynlikheid op 'n oorlewing en indien $P < 0.625$ dui die waarskynlikheid op 'n sterfte. Getoets aan die werklike uitkoms van die oorblywende 47 pasiënte wat nie in die steekproef ingesluit is nie (38 oorlewendes en 9 sterftes), is hierdie model se sensitiviteit 100% (voorspel alle oorlewendes korrek) en die spesifisiteit 88.9% (1 van die 9 pasiënte wat gesterf het is foutiewelik voorspel om te oorleef).

$$P(\text{leef}^f) = \frac{e^{28.550 - 1.945 \cdot \text{MOFSG} - 0.883 \cdot \text{HIS}}}{1 + e^{28.550 - 1.945 \cdot \text{MOFSG} - 0.883 \cdot \text{HIS}}} \dots\dots\dots 2)$$

Die vermoë van hierdie model (P -leef^f) is getoets deur dit toe te pas op die eerste 54 pasiënte (42 het oorleef, 12 het gesterf) by wie opnamedata prospektief verkry is. By die afsnytpunt van 0.625 het hierdie model die beste gepresteer, met 'n sensitiviteit van 100% en 'n spesifisiteit van 0%. Dit verleen aan die model 'n positiewe voorspellingswaarde van 77.8% en 'n negatiewe voorspellingswaarde van minder as 1%. 'n Nuwe model is by die opnamedata van die prospektiewe groep van 54 pasiënte gepas deur weer al 14 programme se opnametellings aan logistiese regressie te onderwerp. Die gepaste model (P -leef^f) het by 'n afsnytpunt van 0.442 'n sensitiviteit van 95% en 'n spesifisiteit van 75% behaal.

$$P(\text{leef}^f) = \frac{e^{10.812 - 0.585 \cdot \text{SAPS} - 0.671 \cdot \text{HIS}}}{1 + e^{10.812 - 0.585 \cdot \text{SAPS} - 0.671 \cdot \text{HIS}}} \dots\dots\dots 3)$$

Ten einde te bepaal hoe goed hierdie model (P -leef^f) oorlewing in 'n ander prospektiewe populasie sou voorspel, is dit op 'n verdere 50 CISE pasiënte getoets. Dit het 'n sensitiviteit van 100% en 'n spesifisiteit van 44% by 'n afsnytpunt 0.442 gelewer (10 van die 18 sterfgevallen is foutiewelik voorspel om te oorleef). Die model (P -leef^f) wat by die opnamedata van al 104 prospektiewe pasiënte gepas is, is:

$$P(\text{leef}^f) = \frac{e^{7.835 - 0.416 \cdot \text{SAPS} - 0.756 \cdot \text{HIS}}}{1 + e^{7.835 - 0.416 \cdot \text{SAPS} - 0.756 \cdot \text{HIS}}} \dots\dots\dots 4)$$

Indien hierdie model teen die ware uitkoms van die groep van 104 pasiënte getoets word, lewer dit 'n sensitiviteit van 93% en 'n spesifisiteit van 66% by 'n afsnytpunt van 0.392. Van die 74 pasiënte wat die model voorspel het om te oorleef, het 67 werklik oorleef (positief voorspelde waarde 90.5%) en van die 30 pasiënte wat volgens voorspelling sou sterf, het slegs 19 gesterf (negatief voorspelde waarde 63.3%).

Bespreking

Feitlik alle individuele voorspellingsprogramme word ontwikkel om op spesifieke pasiëntpopulasies toegepas te word. Die CISE van die H.F. Verwoerd-hospitaal is egter 'n eenheid waar pasiënte uit enige van 'n aantal diagnostiese kategorieë aangetref word, byvoorbeeld trauma, algemene chirurgie, vasculêre chirurgie, multi-orgaanfunksie en/of sepsis asook brandwonde. Niteenstaande hierdie heterogene pasiëntsamesstelling het die groep pasiënte wat oorleef het en die groep wat gesterf het betekenisvol van mekaar verskil met betrekking tot die gemiddelde telling uit elk van die 14 voorspellingsprogramme ($P < 0.01$ - $P < 0.0001$). Dit is interessant dat al hierdie programme relatief goed presteer as in ag geneem word dat die meerderheid programme ontwikkel is vir spesifieke teikengroepe en eenhede. In 'n poging om die voorspellingswaarde van die program wat die beste gepresteer het verder te verbeter en aan te pas by die behoeftes van die CISE van die H.F. Verwoerd-hospitaal, is logistiese regressie gebruik om deur die gesamentlike gebruik van 'n aantal programme 'n model met 'n selfs hoër voorspellingsakkuraatheid te ontwikkel.

In beide die retrospektiewe en prospektiewe analises was die groepe die mees betekenisvol verskillend met betrekking tot die tellings afkomstig uit die MOFSG ($P < 0.0001$), SAPS ($P < 0.0001$) en HIS ($P < 0.0001$) programme. Die model (P -leef^f) het die beste resultate op grond van retrospektiewe opnamedata gelewer, met 'n sensitiviteit van 100% en 'n spesifisiteit van 88.9%. Die feit dat hierdie model nie goed gepresteer het op prospektief-versamelde data nie, mag verklaar word deur die groot verskille wat bestaan tussen data wat retrospektief uit pasiëntlêers verkry word (en dus ontbrekende en ontoereikende inligting kan bevat) en data wat prospektief in 'n intensiewesorgeenheid versamel word. (In laasgenoemde geval kan kliniese beoordeling van die pasiënt lig werp op 'n betrokke aspek wat nie retrospektief moontlik is nie.) Die waarskynlikheidsmodel (P -leef^f) wat uit die opnamedata van die 104 prospektiewe pasiëntopnames ontwikkel is, het 'n positief-voorspelde waarde van 90.5% en 'n negatief-voorspelde waarde van 63.3% gehad.

Daar bestaan etlike slaggate in die ontwikkeling van statistiese modelle en in die evaluering van individuele prognostiese programme vir gebruik in intensiewesorg. Die belangrikste hiervan is sekerlik die grootte van die databasis waarop die ontwikkeling en statistiese evaluering uiteindelik berus. As in ag geneem word dat bekende programme soos die 'APACHE II' ontwikkel is uit 5 030 intensiewesorgpasiënte, die SOS uit 5 500 en die RTS uit 2 166, kan aangeneem word dat 'n groter databasis baie daartoe sal bydra om die voorspellingsakkuraatheid van die (P -leef^f) model verder te verbeter. 'n Verdere aspek wat hier van belang is, is die akkuraatheid waarmee waarnemings gemaak word wanneer 'n programtelling bereken moet word.⁸ Dit is gewoonlik problematies wanneer meer as een waarnemer by die versameling en notering van die data betrokke is.⁹ Die waarnemer kan byvoorbeeld te staan kom voor die keuse van 'n hipotensiewe of hipertensiewe beeld, die keuse om te besluit watter orgaan die 'ergste' in versaking is, of somtyds die beoordeling van die neurologiese status van 'n gesedeerde pasiënt. Dit is dus duidelik dat die oordeel van

die waarnemer of selfs die verskillende interpretasies van 'n hele aantal waarnemers die voorspellingsvermoë van programme drasties kan beïnvloed. Dit word voorgestel dat, by die toepassing van 'n ontwikkelde model as maatstaf vir kliniese besluitneming en ondersteuning, so 'n model se voorspellingskrag van tyd tot tyd herevalueer word. Dit is veral belangrik wanneer die teikengroep, soos in hierdie geval, uit 'n heterogene populasie bestaan en waar daar 'n relatief hoë omset van personeel is.

Gevolgtrekking

Die bestaande voorspellingsprogramme toon deur hulle goeie prestasie op die heterogene groep wat ondersoek is, 'n relatiewe non-spesifisiteit wat betref die tipe van pasiënte waarop hulle toegepas word. Dit is moontlik om bestaande programme deur middel van logistiese regressie-analise in 'n voorspellingsmodel op te neem ter verbetering van die voorspellingsakkuraatheid vir 'n bepaalde teikengroep.

Hierdie werk vorm deel van Louis J. Roux se M.Sc.-verhandeling in Fisiologie aan die Fakulteit van Geneeskunde, Universiteit van Pretoria.

VERWYSINGS

1. Civetta JM, Hudston-Civetta JA, Nelson LD. Evaluation of APACHE II for cost containment and quality assurance. *Ann Surg* 1990; **212**: 266-276.
2. Smith EJ, Ward AJ, Smith D. Trauma scoring methods. *Br J Hosp Med* 1990; **44**: 8-16.
3. Gilbert J, Schoolfield J, Gaydou D, McFee A, Smith B. Modified system outcome score and outcome index — method of monitoring patient care in a special care area. *Crit Care Med* 1990; **18**: 596-602.
4. Civetta JM. 'New and improved' scoring systems. *Crit Care Med* 1990; **18**: 1487-1490.
5. Wagner DP, Knaus WA, Draper EA. Statistical validation of a severity of illness measure. *Am J Public Health* 1983; **73**: 878-884.
6. Arregui LM, Moyes DG, Lipman J, Fatti LP. Comparison of disease severity scoring systems in septic shock. *Crit Care Med* 1991; **19**: 1165-1171.
7. Chang RW. Individual outcome prediction models for intensive care units. *Lancet* 1989; **2**: 143-147.
8. Roux LJ, Viljoen M, Pretorius JP. Rekenariserings in intensiewesorg. *Nursing RSA/Verpleging* 1994; **9**: 40-42.
9. Schuster DP. Predicting outcome after ICU admission. The art and science of assessing risk. *Chest* 1992; **102**: 1861-1870.

Aanvaar 4 Junie 1996.