Svensk översättning och reliabilitetsprövning av ”The Full Outline of UnResponsiveness” (FOUR) score

Swedish translation and evaluation of reliability of Full Outline of UnResponsiveness (FOUR) score

Annika Hickisch

Örebro universitet,
Institutionen för hälsovetenskap och medicin

Förbättringsarbete inom omvårdnadsvetenskap

Examensarbete, master 30 hp
Vårtermin 2013

SAMMANFATTNING

Bakgrund: Vid förbättringsarbete inom omvårdnad är det viktigt att använda mätinstrument med god reliabilitet för att kunna utveckla och förbättra möjligheten till att ge god, utvärderingsbar och säker vård. Att bedömning och dokumentation av patientens medvetandegrad sker på ett strukturerat sätt är stor betydelse för patientsäkerheten. The Full Outline of Unresponsiveness (FOUR) score mäter medvetandegrad mer heltäckande än de skalor som används i Sverige nu. Därför finns det behov av att översätta FOUR till svenska och pröva skalan i svensk vårdkontext. Syftet med studien var att översätta medvetandegradskalan FOUR till svenska samt att utvärdera den svenska versionens reliabilitet och att jämföra FOUR och Glasgow Coma Scale (GCS) med avseende på reliabilitet.

Metod: Den engelska versionen av FOUR översattes till svenska och testades vad gäller interbedömarreliabilitet och homogenitet på simulerade patientfall med avancerad patientsimulator. Två sjuksköterskor bedömde 30 patientfall med FOUR och GCS. Viktad kappa användes för att bestämma interbedömarreliabiliteten för skalornas delkomponenter. Intraclass Correlation Coefficient (ICC) bestämdes för totalpoäng för varje skala. Homogeniteten bestämdes med Cronbach's alfa.

Resultat: Den svenska översättningens fjärde version nådde konsensus i expertpanelen. Tillbakaöversättningen godkändes av FOUR skalans skapare. Interbedömarreliabiliteten av svenska versionen av FOUR var utmärkt. ICC på total poängen på FOUR var utmärkt, 0.986 (95 % CI: 0.970-0.993) och på GCS 0.997 (95 % CI: 0.994-0.999). Cronbach's alfa visade hög grad av homogenitet för FOUR (α= 0.917) och för GCS (α= 0.866) för första bedömaren.

Slutsats: Den svenska versionen av FOUR är ekvivalent till originalet och har vid reliabilitetsprövning med fingerade patientfall och avancerad patientsimulator visat en hög reliabilitet jämfört med GCS.

Nyckelord:
reliability, translation, Full Outline of Unresponsiveness, Hi-fidelity simulation, coma scale

INTRODUKTION

Patienten i djup koma behöver vård för att upprätthålla fri luftväg och andningsstöd vid sviktande andningsfunktion. Vid långvarig medvetslöshet krävs dessutom omfattande omvårdnad för att förebygga infektioner, ha kontroll över blodsockernivåer och kroppstemperatur samt skydda ögon och hud mot de skador som kan uppstå (Wijdicks, 2008). Att bedömning och dokumentation av patientens medvetandegrad sker på ett strukturerat sätt, har stor betydelse för patientsäkerheten, då sänkt medvetandegrad är förenat med ökad sjuklighet och dödlighet oavsett orsak. Sjuksköterskan som vårdar patienten måste ha kunskap om att korrekt och konsekvent bedöma samt kommunicera förändringar av medvetandegrad till medarbetare inom vården (Cohen, 2009). Det finns dock svårigheter i att finna användbar terminologi för beskrivning av bedömningsresultat. En studie har visat hur en och samma patient vid samma tillfälle kan beskrivas som ”somnolent”, ”svårväckt” eller ”djupt medvetslös” av tre olika observatörer (Iyer et al., 2009). Detta belyser vikten av att det finns bedömningsinstrument som är valida och reliabla för att bedöma medvetandegrad.

Medvetandegrader

Grad av medvetande omfattar i huvudsak två komponenter. En komponent är väckbarhet *(wakefulness),* som relaterar till nivån av vakenhet och som är en kvantitativ aspekt. Den är anatomiskt kopplad till tre strukturer som håller människan vaken och redo för respons. För att åstadkomma detta finns i mitthjärnan och hjärnbryggan ett område med specialiserade celler, *ascending reticular activating system* (ARAS), vilka behöver ha en fungerande interaktion med thalamus och hjärnbarken. Den andra komponenten är innehåll (*content* eller *awareness*), som är en kvalitativ aspekt av medvetandet. Denna påverkas av funktioner från hjärnbarken såsom kognitiva förmågor, exempelvis uppmärksamhet, sinnesperception, uttalat minne, språk och spatial orientering (Bordini, Luiz, Fernandes, Arruda & Teive, 2010; Wijdicks, 2008). Denna sammansättning av strukturer påverkas av skador, läkemedel, fysiologiska störningar (ex. hypoglukemi, hypoxi) eller toxin, vilket leder till försämrad funktion i hjärna och hjärnstam och därmed till olika grader av koma eller medvetandepåverkan (Wijdicks, 2008).

Skador eller förändringar i hjärnan kan yttra sig på flera olika sätt och påverka pupiller, ögonrörelser, motorisk respons och andningsmönster. Kranialnerv III (*n. oculomotorius*) som öppnar ögonlock och ger pupillsammandragning vid ljus samt kranialnerverna IV och VI (*n. trochlearis, n. abducens*) vilka deltar i horisontella och vertikala ögonrörelser, härstammar från samma område som ARAS. Detta förklarar varför ögonens öppnande, ögonrörelser samt ljusreflex kan ha störningar i samband med sänkt medvetande. Även andningsmönster kan påverkas av hjärnskada eller metabolisk påverkan på hjärnbryggan (pons). Cheyne-Stokes andning är ett mönster med rytmiskt snabbandning (hyperpné) alternerande med apnéperioder samt succesivt ökning och minskning av andetagsdjup. Vid oregelbunden andning (ataxisk) varierar andningsfrekvens, rytm och djup (Haymore, 2005). Central neurogen hyperventilation, som orsakas av skador på hjärnstammen, hypoxi eller metabiliska rubbningar, har ihållande andningsfrekvenser över 40 per minut och ger uttalad respiratorisk alkalos (Wijdicks, 2008).

Det är svårt att definiera skillnaden mellan olika stadier av sänkt medvetande och i den medicinska litteraturen finns många oprecisa termer (Wijdicks, 2008). Stupor används

främst för att identifiera ett stadium där patienten har slutna ögon som bara öppnas efter kraftig smärtstimulering samt att patienten lokaliserar smärtan genom att utföra ändamålsenliga rörelser mot det smärtande området. Wijdicks beskriver vidare koma som ett tillstånd där patienten är helt omedveten och i bästa fall kan öppna ögonen vid smärtstimulering eller har öppna ögon men, kan inte fixera eller följa med blicken. Vid smärtstimulering kan patienten i bästa fall göra undandragande rörelse eller visa motoriska reflexrörelser. Hjärnstamsreflexer kan vara intakta eller saknas i varierande grad.

Ett sällsynt tillstånd där patienten är medveten, men har samtidigt näst intill fullständig förlamning är ”locked-in syndrom”. Patienten är oförmögen att kommunicera med omvärlden förutom genom att blinka med ögonen. Frivilliga vertikala ögonrörelser samt hörsel är bevarade och vissa kan även ha kvarstående horisontala ögonrörelser. En ocklusion av arteria basilaris eller blödning i hjärnbryggan är orsaker till detta för patienten fruktansvärda tillstånd (Haymore, 2004; Wijdicks, 2008).

En noggrann bedömning av allvarlighetsgraden vid medvetslöshet baseras på följande kliniska parametrar: medvetandegrad, hjärnstamsfunktion, andningsmönster, pupillernas storlek, ljusreaktion, ögonrörelser samt motorisk respons (Kevric, Jelinek, Knott & Weiland, 2011).

Skalor för bedömning av medvetandegrad

Medvetandegradsskalor har skapats för att det finns behov av att standardisera språkanvändningen och för att underlätta den skrivna och verbala kommunikationen mellan olika hälsoprofessioner. De ger även ett konsekvent system för att följa utvecklingen av patientens medvetandegrad. Dessutom kan skalorna ge prognostisk information och därmed skapa förutsättningar för optimal behandling samt kostnadseffektiv vård (Bordini et al., 2010). Den ideala skalan för att bedöma medvetandegrad bör vara lätt att administrera och poängsätta, kunna appliceras på en större mängd patienter, ha förmåga att bedöma medvetandegrad, identifiera snabbt försämrade patienter och kunna förutsäga risk för sjuklighet och död (Kornbluth & Bhardwaj, 2011).

Glasgow Coma Scale

Glasgow Coma Scale (GCS) utarbetades år 1974 och bedömer tre aspekter av vakenhet: ögonens öppnande, verbal och motorisk respons. Från början var delkomponenterna inte poängsatta utan skalan reviderades år 1976 då de poängsattes (Ögon:1-4 p, Verbalt: 1-5 p, Motorik: 1-6 p). Delkomponenternas poänger summeras i en poängskala från 3 till 15 poäng, där 3 p är sämsta resultatet och 15 p är bästa resultatet. Internationellt har summeringen av GCS fått en universell betydelse för sjukvårdspersonal och ses som ett instrument för att bestämma allvarlighetsgraden vid skallskador, där mindre än nio poäng bedöms som svår skallskada. Den anses idag vara ”golden standard” som medvetandegradskala (Cohen, 2009; Sadaka, Patel & Lakshmanan, 2011) och används enligt Iyer et al. (2009) i flera olika bedömningssystem, som exempelvis Acute Physiology and Chronic Health Evaluation (APACHE).

Eftersom GCS inte bedömer hjärnstamsreflexer och respirationsmönster är det svårt att upptäcka komats verkliga djup med GCS enligt Kevric et al. (2011) och det går inte heller att

identifiera ”looked-in syndrom” (Kornbluth & Bhardwaj, 2011). Den verbala komponenten i GCS medför det att det svårt att korrekt bedöma intuberade, afasiska eller stämbandsskadade patienter (Bordini et al., 2010; Cohen, 2009; Kevric et al., 2011; Kornbluth & Bhardwaj, 2011). Olika sätt används för att kompensera för den verbala poängen på den intuberade patienten. Vissa verksamheter ger 1 p som verbal respons (lägsta poängen) och andra väljer att förutsäga hur patienten skulle ha svarat eller att endast ge ett ”T” utan poäng. Dessa metoder är inte universella eller validerade, vilket kan påverka data som skall förutsäga prognos från olika verksamheter (Cohen, 2009). Den verbala komponenten försvårar även bedömningen av spädbarn och barn som ännu inte börjat tala, varför en modifierad pediatrisk GCS har utvecklats (Kornbluth & Bhardwaj 2011). Det finns dock ännu ingen överenskommen ”golden standard” för pediatrisk skala (Cohen, 2009).

Full Outline of Unresponsiveness (FOUR) score

År 2005 utarbetade forskare vid Mayo Clinic Saint Marys Hospital, the Full Outline of Unresponsiveness (FOUR) score (Bilaga 1). Vid användning av FOUR score sätts 0-4 poäng (0 = ingen funktion, 4 = normal funktion) på fyra funktionella delkomponenter: ögonrespons (ögonens öppnande och ögonrörelser), motorisk respons (följer uppmaning eller respons till smärtstimulering), hjärnstamsreflexer (pupill, korneal- och hostreflexer) och respiration (andningsmönster eller närvaro av egenandning under respiratorbehandling) (Wijdicks, Bamlet, Maramattom, Manno & McClelland, 2005). Om noll erhålls i alla fyra kategorierna bör undersökning utföras för att utesluta hjärndöd (Wijdicks, 2008). I och med att handgester tas med i bedömningen istället för verbalt svar, kan patienter med endotracheal intubation, afasi eller skador på stämbanden testas (Kornbluth & Bhardwaj 2011). FOUR är även översatt till franska (Weiss et al., 2009), spanska (Idrovo, et al.2010) och italienska (Marcati, et al., 2012).

Psykometriska egenskaper

Flera studier har undersökt interbedömarreliabilitet och homogenitet samt prövat validera FOUR genom att jämföra prediktion mellan FOUR med GCS som ”golden standard” med avseende på morbiditet och mortalitet. De har funnit FOUR likvärdig med GCS att förutse mortalitet och morbiditet (Cohen, 2009; Iyer et al., 2009; Marcati et al., 2011; Sadaka et al, 2011; Stead et al., 2009; Wijdicks et al., 2005). Några av studierna fann en ökning av mortaliteten för de patienter som hade lägst totalpoäng av FOUR jämfört med lägsta GCS poäng (Iyer et al., 2009; Marcati et al., 2011; Wijdicks et al., 2005).

Flertalet av studierna av FOUR har utförts på intensivvårdsavdelningar (Iyer et al., 2009; Wijdicks et al., 2005; Wolf et al., 2007) varav en på pediatrisk intensivvårdsavdelning (Cohen, 2009). FOUR har testats på akutmottagningar och även där erhållit hög interbedömarreliabilitet (Kevric et al., 2011; Stead et al., 2009).

Alla studierna jämförde interbedömarreliabilitet av FOUR med GCS och fann interbedömarreliabiliteten likvärdig eller något högre för FOUR score. I fyra studier anges graden av homogenitet. Cronbach’s alfa beräknades till mellan 0,86 - 0,95 (Iyer et al., 2009; Sadaka et al, 2011; Wijdicks et al., 2005; Wolf et al., 2007). I Tabell 1 presenteras en sammanställning av interbedömarreliabilitet av delkomponenter av FOUR och sammanvägda poäng samt homogenitet där det finns angivet i studier med den engelska versionen av FOUR.

TABELL 1. Sammanställning av FOUR, reliabilitet i tidigare studier
Interbedömarreliabilitet (viktad kappa, κ_w) och homogenitet (Cronbach's alfa)

Studier *(i kronologisk ordning)*	**Ögon** κ_w	**Motorisk** κ_w	**Hjärnstam** κ_w	**Respiration** κ_w	**Sammantagen poäng** κ_w	**Homogenitet** *Cronbach's alfa*
Cohen (2009) *N*= 60	0.98	0.86	1.00	1.00	0.95	-
Ficher et al. (2010) *N* =437	0.82	0.78	0.67	0.90	0.68	-
Iyer et al. (2009) *N* =200	0.96	0.97	0.98	1.00	0.99	0.87 bedömare 1 0.87 bedömare 2
Kevric et al. (2012) *N* =217	0.80	0.87	0.76	0.97	0.76	-
Kramer et al. (2012) *N* =906	0.88	0.91	0.89	0.94	0.92	-
Sadaka et al. (2011) *N* =51	-	-	-	-	-	0.89
Stead et al. (2009) *N* =69	0.86	0.83	0.94	0.76	0.88	
Wijdicks et al. (2005) *N* =120	0.78	0.80	0.81	0.78	0.82	0.86 bedömare 1 0.87 bedömare 2
Wolf et al., (2007) *N* =80	0.84	0.73	0.89	0.92	0.85	0.95

FOUR=Full Outline of Unresponsiveness (FOUR) score

Vid förbättringsarbete inom omvårdnad är det viktigt att använda mätinstrument med god reliabilitet och validitet för att utveckla och förbättra möjligheten till att ge god, utvärderingsbar och säker vård. För den vårdpersonal som skall göra bedömningar av patientens medvetandegrad är det viktigt att bedömningsskalan är lättförstådd och tydlig samt att resultatet mellan olika bedömare har så hög överensstämmelse som möjligt. FOUR-skalan har spridits till flera länder utanför USA med goda resultat och eftersom FOUR mäter medvetandegrad på ett mer heltäckande sätt än de skalor som används i Sverige nu så finns det behov av att översätta FOUR till svenska och pröva FOUR i svensk vårdkontext.
Vid översättning av bedömningsinstrument är det dock viktigt att undersöka om den översatta versionen är likvärdig originalet utifrån språkliga, begreppsmässiga, kulturella och mätmässiga aspekter.

SYFTE

Syftet med studien var att översätta medvetandegradskalan FOUR till svenska samt att utvärdera den svenska versionens reliabilitet och att jämföra FOUR och GCS med avseende på reliabilitet.

METOD

Denna studie har två faser, där fas ett innebär översättning av bedömningsinstrumentet FOUR till svenska. I fas två utvärderas den svenska versionens reliabilitet och jämförs med GCS.

Fas 1: Översättning av bedömningsinstrument FOUR

Målet med översättning av ett mätinstrument är att nå ekvivalens, dvs. likvärdighet, mellan originalet och den översatta versionen av instrumentet. Enligt Streiner och Norman (2008) finns olika typer av ekvivalens. Konceptuell ekvivalens omfattar begreppsmässighet och bedömer om begreppet finns i båda kulturerna och delkomponent (*items*) ekvivalens hur relevanta mätinstrumentets delkomponenter är för målgruppen. Semantisk ekvivalens omfattar den språkliga översättningen men även kulturella uttryck. Den underliggande meningen hos originalets ord bör bevaras snarare än exakta ordalydelsen (Polit & Beck, 2008). Den operationella ekvivalensen omfattar formulär, instruktioner, administrationssätt till målgruppen och slutligen finns mätmässig ekvivalens som testar om reliabilitet och validitet är lika i båda versionerna (Streiner & Norman, 2008).

En struktur för översättning av mätinstrument, *"Process of translation and adaptation of instruments"*, från World Health Organization (WHO) användes för att översätta medvetandegradskalan FOUR till svenska. Processen beskrivs nedan samt i figur (Bilaga 2).

Översättning (*forward translation*) från originalspråket (*source language, SL*) engelska till svenska (*target language, TL*) genomfördes av författaren, som är tvåspråkig med svenska som modersmål. Bedömningsskalan FOUR översattes med perspektiv på ett naturligt och accepterat språk för användargruppen dvs. sjukvårdspersonal inom främst akutsjukvård.

En tvåspråkig expertpanel med goda kunskaper i engelska och svenska bildades. Expertpanelen bestod av fyra personer, en specialistläkare inom anestesi-IVA, en professor tillika anestesisjuksköterska med erfarenhet av översättning av bedömningsinstrument, en forskare med erfarenhet av mätmetoders utveckling samt en specialistsjuksköterska inom intensivvård. Syftet med expertpanelen var att identifiera och lösa inadekvata begrepp eller uttryck i översättningen. Deltagarna skickade kommentarer och synpunkter på respektive översättning via e-post. En sammanställning av dessa synpunkter samt ett förslag till förnyad version skickades tillbaka till expertpanelen av författaren. Arbetet med översättningen upprepades flera gånger till dess en tillfredsställande version uppnåtts.

En tillbakaöversättning (*back-translation)* av expertgruppens överenskomna svenska översättning till engelska, utfördes därefter av en läkare med engelska som modersmål och som inte hade någon vetskap om FOUR skalan. Denna tillbakaöversättning (SL 2) jämfördes med originalet (SL 1) av expertpanelen men även av Dr Wijdicks som utarbetat den ursprungliga FOUR skalan.

Fas 2: Reliabilitetsprövning av svensk version av FOUR

Den svenska versionen av FOUR testades därefter för interbedömarreliabilitet och homogenitet (*internal consistency*) av två bedömare som bedömde simulerade patientfall på en avancerad patientsimulator. En avancerad patientsimulator är en verklighetstrogen docka, kopplad till en dator och en monitor. Dockan har synliga andningsrörelser och det går att lyssna på andningsljud, hjärtljud, tarmljud samt känna pulsar. Dessutom är dockan utrustad med inbyggda högtalare så när operatören talar i en mikrofon kan dockan "tala". På monitorn avläses flera valbara parametrar, såsom exempelvis SpO_2 blodtryck och EKG. Det går att utföra olika åtgärder bl.a. intubation, insättande av perifer venkateter och pleuradränage på dockan.

Bedömare

För att genomföra reliabilitetsprövningen valdes att två bedömare med likartad bakgrund skulle genomföra testet. Inklusionskriterierna för bedömare var a) sjuksköterska med specialistutbildning inom intensivvård, anestesi eller ambulanssjukvård, b) minst 5 års erfarenhet inom akutsjukvård samt c) erfarenhet av vård av patient med ventilator. Information med beskrivning av studien och inbjudan att delta som bedömare anslogs på ambulansenhet, akutmottagning, operations- och intensivvårdsavdelning vid ett akutsjukhus i Mellansverige. Då inga svar erhölls gjordes en personlig förfrågan till två av författaren kända sjuksköterskor, som i samband med förfrågan erhöll skriftlig och muntlig information om studiens syfte, tillvägagångssätt, beräknad tidsåtgång samt att utbildning skulle ske innan bedömartestet. De båda sjuksköterskorna, som tackade ja till att ingå i studien var specialistutbildade med vardera mer än 10 respektive 25 års erfarenhet av akutsjukvård och hade erfarenhet av patientvård med ventilatorer inom intensivvård respektive anestesisjukvård.

Patientsimuleringsfall

Urvalstorlek för patientfall baseras på uppskattningar av *Intraclass Correlation Coefficient* (ICC), antalet observationer per subjekt och konfidensintervallet. Om antalet bedömare är två och *Standard Error* (SE) uppskattas till 0.05 och ICC till 0.90 krävs ca 30 patientfall för att bedöma interbedömarreliabilitet (Streiner & Norman, 2008).
För att kunna bestämma Cronbach's alfa (homogeniteten) krävs enligt Streiner och Norman (2008) en urvalsstorlek på ca 30 patientfall om Cronbach's alfa förväntas bli ca α =0,90 (C.I. ± 0,15) och skalan innehåller 4 delkomponenter (*items*). Dessa antagna värden på Cronbach's alfa och konfidensintervallets bredd baserades från tidigare psykometriska studier av FOUR.

De fingerade patientfallen utarbetades utifrån exempel på olika cerebrala tillstånd och dess effekter på ögonrespons, motorik, hjärnstamsreflexer och respiration beskrivna av Wijdicks (2008). För att få en spridning över hela FOUR- skalans mätområden utarbetades de trettio patientfallen, så de representerade fyra kategorier av olika nivåer av medvetandegrad enligt AVPU-skalan. Denna förenklade skala användes ursprungligen vid utbildning av *Advanced Trauma Life Support for Doctors* (American College of Surgeons Committee on Trauma) enligt McNarry och Goldhill (2004). Patienten kan bedömas som alert och orienterad (*alert*=A) eller att patienten ger respons vid tilltal (*voice=V*) alternativt smärtstimulering (*pain=P*) eller visar ingen respons på smärtstimulering (*unresponsive=U*). De 30 fingerade patientfallen fördelades antalsmässigt som sju fall i respektive kategori A och V samt åtta fall i respektive kategori P och U. Respektive patientfall erhöll en beteckning utifrån kategori och ett löpnummer, exempelvis A3 eller P4.

Datainsamling

Studien utfördes på ett kliniskt träningscentrum (KTC) vid ett universitet i Mellansverige. Före bedömartestet erhöll bedömarna en muntligt och skriftlig instruktion om FOUR skalan och såg instruktionsfilmen *"The comatose patient*" som medföljer boken med samma namn (Wijdicks, 2008) och som visar bedömning av patienter med FOUR. Därpå gavs en skriftlig instruktion och muntlig genomgång av GCS. Det bedömningsformulär som användes vid testet visades och diskuterades samt frågor besvarades. Testledare (författare) demonstrerade bedömning med FOUR samt GCS varpå bedömarna fick träna ett fall vardera med båda skalorna och fylla i testformulären. Utbildningen varade i sammanlagt ca 50 minuter.

Bedömartestet utfördes därefter på en trådlös avancerad patientsimulator i form av en vuxendocka i helkroppsmodell, *Laerdals SimMan Essential Bleeding* ® som var placerad i en sjuksäng. Vitala parametrar såsom andningsfrekvens, kunde avläsas på en monitor. I de patientfall där mekanisk ventilation förekom, kopplades en ventilator inställd på en andningsfrekvens på 14 andetag per minut till patientsimulatorn via endotrakealtub. Patientsimulatorn manövrerades av en för simulatorn utbildad operatör och enligt ett protokoll för respons vad gäller andningsmönster, andningsfrekvens, blinkning vid testning av kornealreflex med hjälp av tops (Bild 1) eller hosta vid sugning av luftvägar vid respektive patientfall. Vid bedömning av övriga faktorer (följa med blicken, motoriskt svar på smärtstimulering och uppmaning) beskrev eller demonstrerade testledare (tillika författare) responsen för bedömarna.



Bild 1. Bedömning av kornealreflex på avancerad patientsimulator. Foto: A. Hickisch

Patientfallen bedömdes med FOUR samt GCS på patientsimulatorn av de två sjuksköterskorna var för sig. Randomisering skedde till vilken skala som användes först. I hälften av patientfallen användes FOUR skalan först och därefter användes GCS. I resterande fallen bedömdes skalorna i omvänd ordning. Bedömarparet bedömde i samma ordning varje enskilt patientfall, d.v.s. patientfallet bedömdes med samma skala först av båda bedömarna. Protokollen hade markerats med vilken av skalorna som först skulle användas. Ingen information utbyttes mellan de båda bedömarna om resultatet och de var inte samtidigt i rummet när bedömningen skedde. I bedömningsordningen varvades fall ur de fyra kategorierna (A, V, P, U).

Statistisk analys

FOUR och GCS består av kvalitativa kategoriska variabler på ordinalskalenivå och analyserades med icke-parametrisk statistik (Streiner & Norman, 2008; Polit & Beck, 2008). Informationen från bedömningsformulären överfördes till datablad (Microsoft Excel) för statistisk bearbetning. Interbedömarreliabiliteten för varje delkomponent beräknades med kvadratiskt viktad kappa, κ_w (Streiner & Norman, 2008) för FOUR och GCS. κ_w används för att ge olika vikt till bristande överensstämmelse beroende på hur stort skillnaden är (Altman, 1991). För både FOUR och GCS beräknades Intraclass Correlation Coefficient, ICC(2,1) d.v.s. *Two-way Random, Absolute agreement, single measure*, på sammanlagda poängen på respektive skala (Shrout & Fleiss, 1979). Homogenitet utvärderades med Cronbach's alfa (Bland & Altman, 1997) och beräknades för varje delkomponent och respektive bedömare hos både FOUR och GCS.

Cronbach's alfa och ICC analyserades med SPSS *(Statistical Package for the Social Sciences)* version 20. κ_w beräknades via Vasserstats hemsida (http://vassarstat.net/). Konfidensintervall sattes till 95 %. I vissa av mätningarna var dock konfidensintervallet (CI) inte möjligt att beräkna på grund av att datamatrisen innehöll för många tomma celler vid beräkning av kvadratiskt viktad kappa på Vasserstats hemsida och inget SE angavs. Beräkningar utfördes därför i ett annat statistiskt mjukvaruprogram, Stata® 11 (*StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP*). Detta program angav SE utöver κ_w för de olika delkomponenterna och därmed kunde 95 % CI beräknas manuellt för varje κ_w värde.

Forskningsetiska aspekter

De etiska riktlinjerna för omvårdnadsforskning i Norden bygger på etiska principer från FN:s deklaration om de mänskliga rättigheterna samt Helsingforsdeklarationen och skall följa de etiska principerna om, rättvisa, autonomi, att göra gott och att inte skada (Sykepleiernas samarbeid i Norden [SSN] 2003). Informationen till bedömarna var både muntlig och skriftlig. Studien omfattade inga patienter och ingen etisk ansökan har därför ej gjorts. Författaren har gjort en bedömning av att studien kan vara till nytta för vårdnadens utveckling och har eftersträvat att förhindra eventuella risker eller skador under studiens genomförande. Ett tillstånd för att översätta och testa FOUR skalan erhölls av Dr Wijdicks, som utarbetat skalan, efter förfrågan via e-post.

RESULTAT

Fas 1: Svensk version av FOUR

Expertpanelen nådde koncensus om den fjärde versionen av den svenska översättningen (TL4) av FOUR. Tillbakaöversättningen till engelska jämfördes med det engelska originalet av expertpanelen. Panelen beslöt att den svenska översättningen (TL 4) inte skulle ändras utan beslutade gemensamt att detta var finalversionen. Finalversionen presenteras i Tabell 2. Båda engelska versionerna (SL 1 och SL 2) av FOUR översändes med en förfrågan om synpunkter om likvärdigheten på tillbakaöversättningen till Dr Wijdicks, som gav sitt godkännande.

Några av begreppen och orden föranledde djupare diskussioner inom expertpanelen under översättningsprocessen. Här följer resultatet av dessa diskussioner och motivering till slutliga val av begrepp.

"*Eyelids open*" översattes med " ögonen öppna", då detta är ett vanligare begrepp som används istället för "ögonlock öppna" vid dessa bedömningar inom svensk sjukvård. Begreppet "*tracking*" (spårar) dvs. att patienten tittar på bedömarens finger och följer med ögonen fingrets rörelser, används inte på samma sätt inom svensk sjukvård utan benämns vanligen som "följer med blicken". I originalversion av FOUR används begreppet *"respons to pain"* dvs. "respons på smärta". Expertpanelen valde ordet "smärtstimulering" då det indikerar en aktiv handling som undersökaren utför på patienten och att skalan bedömer svaret på ett specifikt stimuli som skall vara reproducerbart. Vid översättning av "*cornealreflex*" beslöt expertpanelen att använda kornealreflex, som är ett vedertaget begrepp

i medicinska sammanhang, i stället för hornhinnereflex. Begreppet ”*not intubated*” fanns det flera invändningar mot inom expertpanelen. Det finns olika former av andningsstöd utan att patienten är intuberad. Expertpanelen kom slutligen fram till att behålla originalets strikta och enkla begrepp ”inte intuberad”. ”*Breathes above ventilatory rate*” förtydligades något i den svenska översättningen till ”andningsfrekvensen högre än inställd frekvens på ventilatorn”. Genomgående valdes att använda ”ventilator” i stället för ”respirator” då ventilator överensstämmer med originaltexten och är en mer korrekt och internationellt använd benämning.

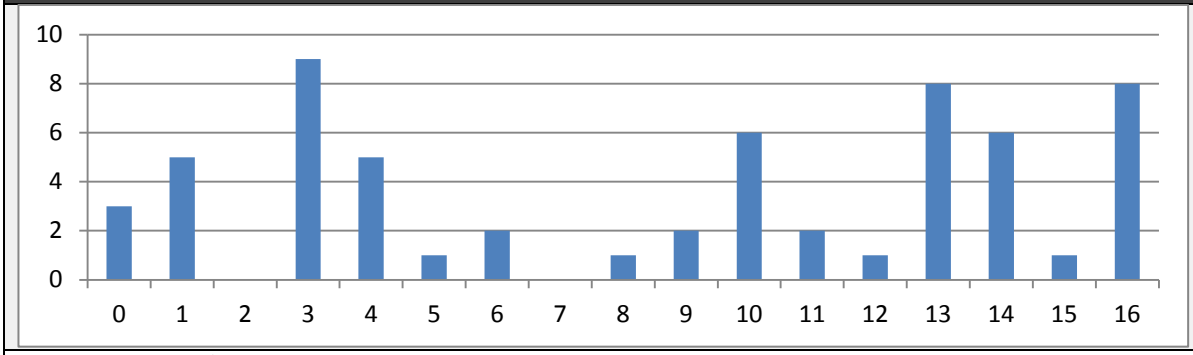
TABELL2. Svensk version av Full Outline of Unresponsiveness (FOUR) score (Wijdicks, 2005)

Komponent	Respons	Poäng
ÖGON	Ögonen öppna eller öppnas, följer med blicken eller blinkar på uppmaning	4
	Ögonen öppna men följer inte med blicken	3
	Ögonen slutna men öppnas vid tilltal med hög röst	2
(Eyes = E)	Ögonen slutna men öppnas vid smärtstimulering	1
	Ögonen förblir slutna under smärtstimulering	0
MOTORISKT SVAR	På uppmaning utför*:* tummen upp, knuten hand eller fredstecken	4
	Lokaliserar vid smärtstimulering	3
	Böjer vid smärtstimulering	2
(Motor = M)	Sträcker vid smärtstimulering	1
	Ingen respons på smärtstimuli eller generellt myokloniskt status	0
HJÄRNSTAMS-REFLEXER	Pupill- och kornealreflexer finns	4
	En pupill vidgad och ljusstel	3
	Pupill- ELLER kornealreflexer saknas	2
(Brainreflex = B)	Pupill- OCH kornealreflexer saknas	1
	Pupill-, korneal- och hostreflex saknas	0
ANDNINGS-MÖNSTER	Inte intuberad, regelbundet andningsmönster	4
	Inte intuberad, Cheyne-Stokesandning	3
	Inte intuberad, oregelbundet andningsmönster	2
(Respiration = R)	Andningsfrekvensen högre än inställd frekvens på ventilator	1
	Andningsfrekvensen samma som inställd frekvens på ventilator eller apné	0

Fas 2 Reliabiliteten hos svensk version av FOUR och jämfört med GCS

När de trettio simulerade patientfallen bedömdes med FOUR fördelade sig total poängen (0 p-16 p) av FOUR enligt sammanställningen som presenteras i Figur 1.

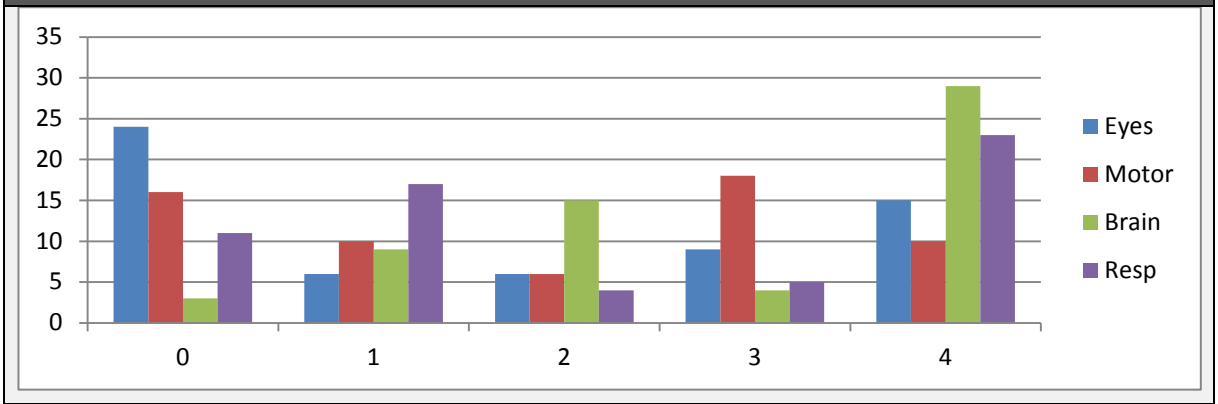
Figur 1 . Poängfördelning FOUR, total poäng för två bedömare, n=60



FOUR=Full Outline of Unresponsiveness (FOUR) score, (*0 p = sämst värde, 4 p = bästa värde, alert)*

Delkomponternas poäng fördelning (0 p – 4 p) vid bedömning av FOUR presenteras i Figur 2.

Figur 2 . Poängfördelning av respektive delkomponent i FOUR för två bedömare, n=60



FOUR=Full Outline of Unresponsiveness (FOUR) score, (*0 p = sämst värde, 4 p = bästa värde, alert)*
Eyes= ögonrespons, Motor= motorisk respons, Brain= hjärnstamsreflexer, Resp= andningsmönster

Interbedömarreliabiliteten hade mycket hög överenstämmelse för delkomponenternas poäng, vilket presenteras i Tabell 3.

TABELL 3 . Interbedömarreliabilitet - Viktad kappa, κ_w (kvadratisk)

FOUR			GCS		
n=30	**Viktad kappa, κ_w**	**95 % CI**	**n= 30**	**Viktad kappa, κ_w**	**95 % CI**
Ögon respons	0.935	0.58 – 1	Ögon respons	0.974	0.62 – 1
Motorisk respons	1	0.64 – 1	Motorisk respons	0.996	0.64 – 1
Hjärnstamsreflexer	0.952	0.60 – 1	Verbalt svar	1	0.64 – 1
Andningsmönster	0.988	0.63 – 1			

κ_w = viktad kappa, CI= konfidens intervall, n= antal patientfall
FOUR=Full Outline of Unresponsiveness (FOUR) score, GCS= Glasgow Coma Scale

Interbedömarreliabiliteten på total poängen för respektive skala bedömdes med *Intraclass Correlation Coefficient* (ICC) som var 0.986 (95 % CI: 0.97 – 0.99) för FOUR och för GCS 0.997 (95 % CI: 0.99 – 0.99).

Cronbach's alfa (α) visade en hög grad av homogenitet för den svenska versionen av FOUR skalan (α = 0.917 för bedömare A och α = 0.908 för bedömare B) och för GCS skalan (α = 0.866 för bedömare A och α = 0.872 för bedömare B).

Både interbedömarreliabilitet och homogenitet för FOUR och GCS visar värden som är likvärdiga när reliabiliteten jämförs med GCS som anses vara "golden standard".

DISKUSSION

Fas 1 Översättningen till svensk version av FOUR

Metoden som användes vid översättningen, *”Process of translation and adaptation of instruments”* (WHO, n.d.) hade en tydlig struktur som följdes med små modifieringar. Den del i processen som omfattar pretest och kognitiva intervjuer ansågs ersatt av reliabilitetsprövningen. Den slutliga dokumentationen i processen har beskrivits i denna studie som del i resultatet med presentationen av finalversion och expertpanelens arbete med rekommendationer.

Expertpanelens sammanssättning med professionella inom hälso- och sjukvården, medförde att ord eller begrepp anpassades efter språkbruk inom svensk sjukvård. Samtidigt skedde en bedömning om FOUR skalans utformning var relevant för att bedöma medvetandegrad, vilket här anses utvärdera konceptuell, komponent, operationell och semantisk ekvivalens (Streiner & Norman, 2008). Författaren sammanställde expertpanelens respons efter varje översättningsversion för att alla skulle ta del av varandras synpunkter och komma överens om bästa alternativ. Avsikten var att respektive deltagare inte skulle påverka varandra utan reflektera över ordval och därefter gemensamt lämna förslag när de såg sammanställningen efter varje version. Några deltagare i expertpanelen skickade sina synpunkter till övriga innan sammanställningen skickats ut, vilket kan ha haft viss påverkan på andra deltagare i panelen. Det var en brist i instruktionen till expertpanelen att det inte angavs att deras respektive svar endast skulle skickas till författaren för sammanställning, som de senare skulle få ta del av.

Tillbakaöversättningen som utfördes och godkändes av Dr Wijdicks är en viktig del för att få optimal semantisk ekvivalens (Maneesriwongul & Dixon, 2004).

I beskrivningar om tillvägagångssätt vid översättningar hänvisar flera författare till Brislin (1970, 1986) som anger tillbakaöversättningen som en väsentlig del i processen för att nå semantisk ekvivalens. Metodiken vid översättningar beskrivna i andra studier (Beck, Bernal & Froman, 2003; Jones, Lee, Phillips, Zhang & Jaceldo, 2001; Ægisdottir, Gerstein & Cinarbas, 2008) överensstämmer till stor del med varandra och även med processen (WHO, n.d.) beskriven i denna studie.

Metoder för översättningar av mätinstrument är främst avsedda för översättning av tvärkulturella och internationella forskningsinstrument, exempelvis självskattningsformulär för specifik population i viss kulturell kontext. Fokus ligger då på att nå tvärkulturell likvärdighet och att kunna jämföra resultat från olika kulturers mätinstrument eller självskattningsformulär. Målgruppen vid översättning av bedömningsskalor liknande FOUR är den hälso- och sjukvårdspersonal som utför bedömningarna. Därmed har professionell hälso- och sjukvårdspersonal som deltagit i expertpanelen kunnat bidra till översättningens ekvivalens med sin kunskap (Yu, Lee & Woo, 2004). Med utgångspunkt i Streiner och Normans (2008) beskrivning av olika typer av ekvivalens som bör bestämmas vid översättning av instrument, kan anses att den svenska översättningen av FOUR uppnått konceptuell, komponent, operationell och semantisk ekvivalens.

Fas 2 Reliabilitetsprövning av svenska FOUR

Homogenitet och interbedömarreliabilitet

Resultatet av reliabilitetsprövningen visar att det finns en hög korrelation mellan de olika delkomponenterna och att de mäter samma egenskap. Cronbach's alfa som påvisar homogeniteten var mycket hög för den svenska versionen av FOUR vilket överensstämmer med resultatet i tidigare studier (Iyer et al., 2009; Sadaka et al., 2011; Wijdicks et al., 2005; Wolf et al., 2007). Cronbach's alfa (α) värden är beroende av antal delkomponenter i skalan och homogeniteten kan enkelt ökas genom att fördubbla deras antal i skalan enligt Streiner och Norman (2008). De anger vidare att α inte bör överstiga 0,90 då detta kan bero på att delkomponenterna motsvarar samma fråga och är därmed onödiga, samtidigt som de nämner studier som anger att α skall överstiga 0.90 om det är kliniska instrument. Även om α översteg 0.9 finns det inga komponenter som kan utslutas eller är onödiga i FOUR-skalan. Det finns skilda åsikter i hur α ska värderas. Bland och Altman (1997) anger α 0.90 som minimum för kliniskt ändamål, men att det är önskvärt med α 0.95.

Resultatet i studien visar även att den svenska versionen av FOUR har utmärkt interbedömarreliabilitet på totalpoäng med ICC över 0.9 samt κ_w över 0.9 på de fyra delkomponenterna. Viktade kappavärden under 0.75 är den undre gränsen för att bedömas som god och under 0.60 är knappast acceptabelt (Streiner & Norman, 2008). Den höga reliabiliteten hos FOUR är i överensstämmelse med den engelska versionen. Interbedömarreliabilitet för de fyra delkomponenterna inom FOUR var utmärkt. Motorisk respons hade högst värde. Tidigare studier visar inget säkert mönster om vilka delkomponenter som får högre respektive lägre värde (Iyer et al., 2009;Ficher et al., 2010; Kevric et al., 2012; Sadaka et al, 2011; Stead et al., 2009; Wijdicks et al., 2005; Wolf et al., 2007).

Oviktad kappa statistik har en nackdel att den inte tar hänsyn till grad av bristande överensstämmelse vid flergradig skala. När kategorierna är rangordnade kan det vara en fördel att ge olika vikt till bristande överensstämmelse beroende på hur stort skillnaden är. Det finns både linjär och kvadratisk viktad kappa som beräknar på olika sätt hur stor skillnaden i överensstämmelsen är. Det system för viktning som rekommenderas är kvadratisk viktning, som baseras på kvadraten av mängden av oenighet (Streiner & Norman, 2008). Detta ligger till grund för valet att använda kvadratisk viktad kappa. Kvadratiskt viktad viktning medför att bristande överensstämmelse påverkar kappa värdet hårdare än linjär viktning. Tyvärr gick det inte att rapportera med konfidensintervall (CI) i tre av fyra delkomponenter eftersom det inte var möjligt att beräkna då kontingenstabellen innehöll för många tomma celler i flera av mätningarna. Vid linjärt viktad kappa kunde dock CI beräknas i tre av fyra delkomponenter. Förmodligen var den höga överenstämmelsen orsaken till att statistikprogrammet i Vassarstat inte kunde beräkna CI eller Standard Error (SE) på flera av delkomponenterna, då hög överensstämmelse innebär många tomma celler i en kontingenstabell. För att kunna beräkna CI användes ytterligare ett statistikprogram, Stata®11, som efter inmatning av data kunde analysera kvadratiskt viktad kappa med SE för varje delkomponent. Med hjälp av SE beräknades sedan 95 % CI ($\kappa \pm 1.96$ x SE) för kappavärdet på varje delkomponent manuellt.

Reliabiliteten av FOUR jämfört med GCS

Homogeniteten var utmärkt för både svenska FOUR och GCS och skillnaden marginell mellan skalorna. Även i tidigare studier blev resultatet för homogenitet hos de båda skalorna likartat (Iyer et al., 2009; Sadaka et al, 2011; Wijdicks et al., 2005; Wolf et al., 2007). Vad gäller ICC på totala poängen av FOUR och GCS var skillnaden liten, vilket överensstämmer

med studie av Stead et al. (2009). Den svenska versionen av FOUR har bedömts vara reliabel jämfört med en svensk version av GCS med avseende på interbedömmarreliabilitet och homogenitet.

Vid jämförelse med andra översatta versioner av FOUR, som jämfört mätegenskaper med GCS är resultaten likartade. Weiss et al. (2009) översatte FOUR till franska och fann interbedömarreliabiliteten hos den franska versionen av FOUR hög (κw 0.86) och likvärdig jämfört med GCS (κw 0.85). Idrovo et al. (2010) genomförde en studie av en spansk version av FOUR. Interbedömarreliabiliteten var utmärkt för FOUR (κw 0.93) samt för GCS (κw 0.96). Marcati et al. (2012) översatte FOUR till italienska och fann interbedömarreliabiliteten utmärkt för FOUR (κw 0.95) och för GCS (κw 0.93).

Flertalet av författaren granskade studier har vid psykometrisk utvärdering av FOUR jämfört resultatet med GCS, som ”golden standard” (Cohen, 2009; Ficher et al., 2010; Idrovo, et al.2010; Iyer et al., 2009; Kevric et al.,; Marcati, et al., 2012; Sadaka et al., 2011; Stead et al., 2009; Weiss et al., 2009; Wijdicks et al., 2005; Wolf et al., 2007). För att möjliggöra jämförelser med resultatet från denna studie med andra internationella studier valdes GCS som jämförande skala och inte Reaction Level Scale (RLS 85), som nästan uteslutande används enbart i Sverige (Kornbluth & Bhardwaj, 2011). En annan fördel med att använda GCS som jämförande skala var att förutsättningarna för de båda skalorna blev mer lika, då bedömarna inte påverkades av tidigare erfarenheter av för dem känd skala.

Fingerade patientfall

Ett skäl till den höga reliabiliteten som erhållits i studien kan vara att vi i och med de fingerade patientfallen har fått en spridning av patientfall över hela poängskalan (0 p-16 p). Det skiljer sig dock inte nämnvärt till verkliga patientstudier. Flera studier (Stead et al., 2009; Wijdicks et al., 2005; Wolf et al., 2007) har eftersträvat en jämn fördelning av olika medvetandegrader och inkluderat patientfall i nedanstående fyra kategorier enligt Ropper (1986, p. 954):

> Awake:.. patients open their eyes without prompting, looked about and conversed (if not aphasic)
>
> Drowsy: ..patients had their eyes closed, opened them when they were addressed in a conversational voice or shaken gently, responded with sensible speech, usually becoming sleepy again within seconds of the end of conversation and had natural movements of the limbs on the side without hemiparesis
>
> Stupor:.. when sleepiness was maintained for several minutes, the lids were closed, there was only momentary or no arousal in response to shaking and yelling, mumbling or single words constituted the best verbal response (often there was none), and the patient consistently pushed aside the examiner`s prodding hand with his or her good arm
>
> Comatose:.. the patients had no speech, did not open their eyelids in response to vigorous shaking of the torso, and made no attempt to push away the examiners hand

De fingerade patientfallen i föreliggande studie delades in i de fyra kategorierna A (alert), V (reagerar på verbalt tilltal), P (reagerar på smärtstimuli) och U (reagerar ej på smärtstimuli), för att patienterna skulle fördelas över hela poängskalan och för att kunna jämföra med andra studier, som använt sig av Ropper’s kategorier. Vid jämförelse av de fingerade patientfallen kategoriserade enligt AVPU och kategorierna enligt Ropper (1986), kan konstateras att ”*awake*” motsvaras av A, ”*drowsy*” av V och ”*comatose*” av U. Kategorin ”*stupor*” beskrivs

som att patienten främst reagerar på smärta men kan kortvarigt vakna upp vid höga rop eller kraftigt skakande, vilket då kan motsvara både V eller P. Detta kan innebära att vid utarbetandet av patientfallen i AVPU-kategorier med fördelning av sju fall i vardera A och V samt åtta fall i vardera P och U, så blev de fingerade patientfallen i kategorin P något mer medvetandesänkta, jämfört med studier med patientfallen i kategorier enligt Ropper (1986). Detta kan vara anledningen till att fördelningen av totalpoängen av FOUR, i denna studie, är relativt spridd över 0-16 poäng och att det finns fler fall med låga poäng. I Wijdics et al. (2005) studie med totalt 240 observationer, erhöll ca 33 % av patienterna högsta poäng (16 p) och i denna studie fick åtta av 60 observationer 16 poäng vilket motsvarar ungefär 13 %.

Avsikten var från början att det psykometriska testet skulle utföras på en intensivvårdsavdelning i Mellansverige. På grund av svårigheter att inom begränsad tidsrymd kunna inkludera tillräckligt antal patienter för att genomföra studien, beslöts att utföra testet på avancerad patientsimulator med fingerade patientfall. En brist i studien är att de fingerade patientfallen inte verifierats av specialist inom neurologi eller neurokirurgi. Avsikten var att detta skulle ske men neurologen som avsett att bekräfta de fingerade patientfallen ur medicinsk synvinkel, fick förhinder att göra detta före det psykometriska testets utförande. Men exemplen på olika cerebrala tillstånden angivet av Wijdicks (2008), som de fingerade fallen baseras på, kan anses vara likvärdigt med verifiering av neurolog. Resultaten på testet talar för detta genom den höga överensstämmelsen med tidigare studier med patienter.

Avancerad patientsimulator

Flera studier visar på större användningsområde för avancerade simuleringsdockor allt eftersom de blir mer verklighetstrogna. Vid en studie av Musacchio et al. (2010) utvärderades bedömning och åtgärder av tre neurologiska akuta tillstånd som läkare under utbildning fick genomföra på avancerad patientsimulator. Deras slutsats blev att genom användningen av avancerad patientsimulator kunde de förbättra utbildningen inom akutsjukvård och att de skulle utöka med fler scenarier och träning för att ytterligare förbättra bedömningarna.

Användningen av avancerade patientsimulatorer ökar i utbildning på universitet och på arbetsplatser inom vården men även inom forskningen. Ross, Kodate, Anderson, Thomas och Jaye (2012) beskriver i sin sammanställning över 320 studier om simulering i tidskrifter inom anestesi hur studier med simuleringsdocka har ökat markant (127 %) under den period som studerats (2001-2010). De nämner även att simulatorerna kan användas till att utvärdera instrument, särskilt under utvecklingsfas och dessa kan i ökande grad kopplas till utvärdering, tester och validering av tekniker och utrustning.

Ibland argumenteras till patientsimulatorernas nackdel, att det inte känns realistiskt med en docka. I en studie fick traumateam träna med avancerad patientsimulator och med standardiserad patient (skademarkör). Teamen fick ange om det upplevdes verklighetstroget och om de kände sig förlägna i situationen. Wisborg, Brattebø, Brinchmann-Hansen och Schrøder Hansen (2009) fann ingen skillnad mellan att träna med docka eller med mänsklig skademarkör. Nakstad och Sandberg (2011) tror på en utveckling där studier med avancerade patientsimulatorer kan användas för att utvärdera tekniker och även av etiska skäl. Men de anser även att studien måste vara bra designad för att kunna bli acceptabel ersättning för verkliga patienter. De flesta av dessa studier utvärderar kunskap eller färdigheter hos de som tränar. Ingen studie om psykometrisk utvärdering på patientsimulator har påträffats. Det finns behov av studier som jämför skillnader mellan bedömning på simulatordocka eller verkliga patienter.

Konklusion

Slutsatsen är att den svenska versionen av FOUR kan anses vara ekvivalent till det engelska originalet, att den har utmärkt reliabilitet med avseende på homogenitet och interbedömarreliabilitet samt kan anses vara likvärdiga med svensk version av GCS med avseende på reliabilitet. Till min vetskap är detta den första översättningen och reliabilitetsprövningen av FOUR i Sverige och möjligen även den första utvärderingen av ett bedömningsinstrument på avancerad patientsimulator. Ytterligare psykometrisk utvärdering av svensk version av FOUR behöver utföras på verkliga patienter. Det finns behov utifrån patientsäkerhet och vårdkvalitet att få en förbättrad bedömning av hur djup medvetandesänkningen hos patienten är och basera omvårdnadshandlingarna efter det.

Acknowledgement

Stort tack till:

- Gunnar från Laerdal för lån av avancerad patientsimulator
- Beatrice och Anna-Lena som med glädje och engagemang ställde upp som bedömare på sin fritid och kollegan Jan, som i egenskap av operatör av patientsimulatorn, fick ”patienten” att verka levande.
- Expertpanelens deltagare, Kjell, Lena, Marie och Ulrica som konkret och engagerat deltog i processen vid översättningen av FOUR.
- David som utförde tillbakaöversättningen och Sören, bibliotekarien, som inget är omöjligt för och som fixat alla böcker och artiklar.
- Anders som bistod med hjälp vid statistisk beräkning och med statistiskt program

Och slutligen ett stort tack till min handledare Marie för konstruktiv kritik och uppmuntran.

Referenser

Altman, D. G. (1991). *Pratical statistics for medical research*. London: Chapman & Hall.

Beck, C. T., Bernal, H. & Froman, R. D. (2003). Methods to document semantic equivalence of a translated scale. *Research in Nursing & Health, 26,* 64-73.

Bland, M., J. & Altman, D. G. (1997). Cronbach's alpha. *BMJ, 314*, 572.

Bordini, A. L., Luiz, T. F., Fernandes, M., Arruda, W. & Teive, H. A. (2010).Coma scales: A historical review. *Arquivos de Neuro-Psiquiatria 68*(6), 930-937.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-cultural Psychology, 1*(3), 187-216.

Brislin, R. W. (1986). The wording and translation of research instruments. In W.L. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Newbury Park, CA: Sage.

Cohen, J. (2009). Interrater reliability and predictive validity of the FOUR score coma scale in a pedriatric population. *Journal of Neuroscience Nursing, 41*(5), 261-267.

Ficher, M., Rüegg, S., Czaplinski, A., Strohmeier, M., Lehmann, A., Tschan, F. … Marsch, S. C. (2010). Inter-rater reliability of the Full Outline of Unresponsiveness score and the Glasgow coma scale in critically ill patients: A prospective observational study. *Critical Care, 14,* R64. Hämtad 2013-02-24 från Critical Care, http://ccforum.com/content/14/2/R64

Haymore, J. (2004). A neuron in a haystack: Advanced neurologic assessment. *AACN Clinical Issues: Advanced Practice in Acute and Critical Care, 15*(4), 568-581.

Idrovo, L., Fuentes, B., Medina, J., Gabaldón, L., Ruiz-Ares, G., Abenza, M.J. … Diez-Tejedor, E. (2010). Validation of the FOUR Score (Spanish Version) in acute stroke: An interobserver variability study. *European Neurolology, 63(*6), 364-369.
Doi: 10.1159/000292498.

Iyer, V., Mandrekar, J., Danielson, R., Zubkov, A., Elmer, J. & Wijdicks, E. (2009). Validity of the FOUR score coma scale in the Medical Intensive Care Unit. *Mayo Clinic Proceedings, 84*(8), 694-701.

Jones, P., S., Lee, J., W, Phillips, L., R., Zhang, X., E. & Jaceldo, K., B. (2001). An adaption of Brislin's translation model for cross-cultural research. *Nursing Research, 50* (5), 300-304.

Kevric, J., Jelinek, G., A., Knott, J. & Weiland, T. J. (2011) Validation of the Full Outline of Unresponsiveness (FOUR) scale for conscious state in the emergency department: Comparison against the Glasgow Coma Scale. *Emergency Medicine Journal, 28,* 486-490. Doi: 10.1136/emj2009.085845.

Kornbluth, J. & Bhardwaj, A. (2011) Evaluation of coma: A critical appraisal of popular scoring systems. *Neurocritical Care, 14,* 134-143. doi: 10.1007/s12028-010-9409-3

Kramer, A. A., Wijdicks, E. F., Snavely, V. L., Dunivan, J. R., Naranjo, L. L., Bible, S., … Dickess, S. M. (2012). A multicenter prospective study of interobserver agreement using the Full Outline of Unresponsiveness score coma scale in the Intensive Care Unit. *Critical Care Medicine, 40*(9), 2671-2676. doi:10.1097/CCM.0b013e318258fd88.

Maneesriwongul, W. & Dixon, J., K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing, 48* (2), 175-186.

Marcati, E., Ricci, S., Casalena, A., Toni, D., Carolei, A. & Sacco, S. (2012). Validation of the Italian version of a new coma scale: The FOUR score. *Internal and Emergency Medicine,7*(2), 145-152. doi: 10.1007/s11739-011-0583-x.

McNarry, A. F. & Goldhill, D. R. (2004). Simple bedside assessment of level of consciousness: Comparison of two simple assessment scales with the Glasgow Coma Scale. *Anaesthesia, 59*, 34–37. doi/10.1111/j.1365-2044.2004.03526.

Polit, D. F. & Beck, C. T. (2008). *Nursing Research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Lippincott Williams &Wilkins.

Musacchio, M. J., Smith, A. P., McNeal, C. A., Munoz, L., Rothenberg, D. M., Roenn, K. A. & Byrne, R. W. (2010). Neuro-critical care skills training using a human patient simulator. *Neurocritical Care, 13*, 169-175. Doi: 10.1007/s12028-101-9405-7.

Nakstad, A. R. & Sandberg, M. (2011). Airway management in simulated restricted access to a patient – Can manikin-based studies provide relevant data? *Scandinavian Journal of Trauma, Resuscitation an Emergency Medicine, 19* (36). Hämtad 2012-03-07 från http://www.sjtrem.com/content/pdf/1757-7241-19-36.pdf

Ropper, A.H. (1986). Lateral displacement of brain and level of consiousness in patients with an acute hemisheral mass. *The New England Journal of Medicine,314* (15) 953-958.

Ross, A. J., Kodate, N., Anderson, J. E., Thomas, L. & Jaye, P. (2012). Review of simulation studies in anesthesia journals, 2001-2020: Mapping and content analysis. *British Journal of Anesthesia, 109* (1), 99-109. Doi: 10.1093/bja/aes184.

Sadaka, F., Patel, D. & Lakshmanan. R. (2011). The FOUR score predicts outcome in patients after traumatic brain injury. *Neurocritical Care, 16*, 95-101.Doi 10.1007/s12028-011-9717-5

Shrout, P. E. & Fleiss, J. L.(1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*, (2), 420-428.

Stead, L. G., Wijdicks, E. M., Bhagra, A., Kashyap, R., Bellolio, M. F., Nash, D. L., …William, B. (2009). Validation of a new coma scale, the FOUR score, in the emergency department. *Neurocritical Care, 10*, 50-54. DOI: 10.1007/s12028-008-9145-0

Streiner, D. & Norman, G. (2008). *Health Measurements scales: A practical guide to their development and use* (4th ed). Oxford: Oxford University Press.

Sykepleiernas samarbeid i Norden (SSN) (2003) *Etiska riktlinjer för omvårdnadsforskning i Norden.* Hämtad 2012-11-10 från SSN http://www.sykepleien.no/ikbViewer/Content/337889/SSNs%20etiske%20retningslinjer.pdf

Weiss, N., Mutlu, G., Essardy, F., Nacabal, C., Sauve, C., Bally, C., … Bolgert, F. (2009). The French version of the FOUR score: A new coma score. [Article in French, abstract]. *Revue neurologique Société de neurologie de Paris, 165*(10),796-802. doi: 10.1016/j.neurol.2009.01.045.

Wijdicks, E., Bamlet, W., Maramattom, B., Manno, E. & McClelland, R. (2005). Validation of a New Coma Scale: The FOUR Score. *Annals of Neurology, 58* (4), 585-593. Doi:10.1002/ana.20611

Wijdicks, E. (2008). *The comatose patient.* New York: Oxford University Press.

Wisborg, T., Brattebø, G., Brinchmann-Hansen, Å. & Schrøder Hansen, K. (2009). Mannequin or standardized patient: Participant's assessment of two training modalities i trauma team simulation. *Scandinavian Journal of Trauma, Resuscitation an Emergency Medicine, 17* (59), 1-4. Doi: 10.1186/1757-7241-17-59.

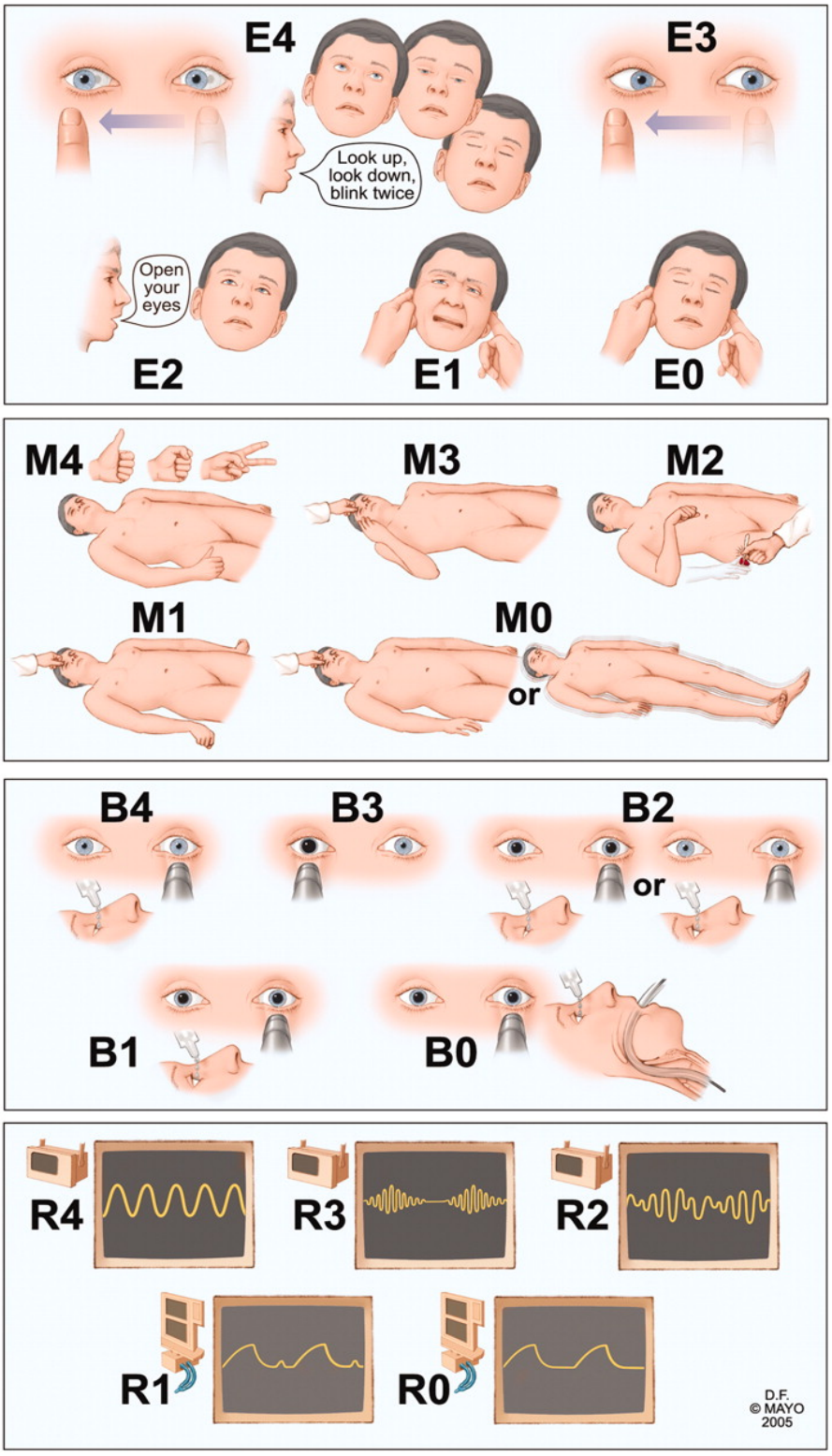
Wolf, C., Wijdicks, E., Bamlet, W., Manno, E., & McClelland, R. (2007). Further Validation of the FOUR Score Coma Scale by Intensive Care Nurses. *Mayo Clinic Proceedings, 82* (4), 435-438.

World Health Organization (WHO) (n.d.): *Process of translation and adaptation of instruments.* Hämtad 2012-10-24 från WHO http://www.who.int/substance_abuse/research_tools/translation/en/index.html

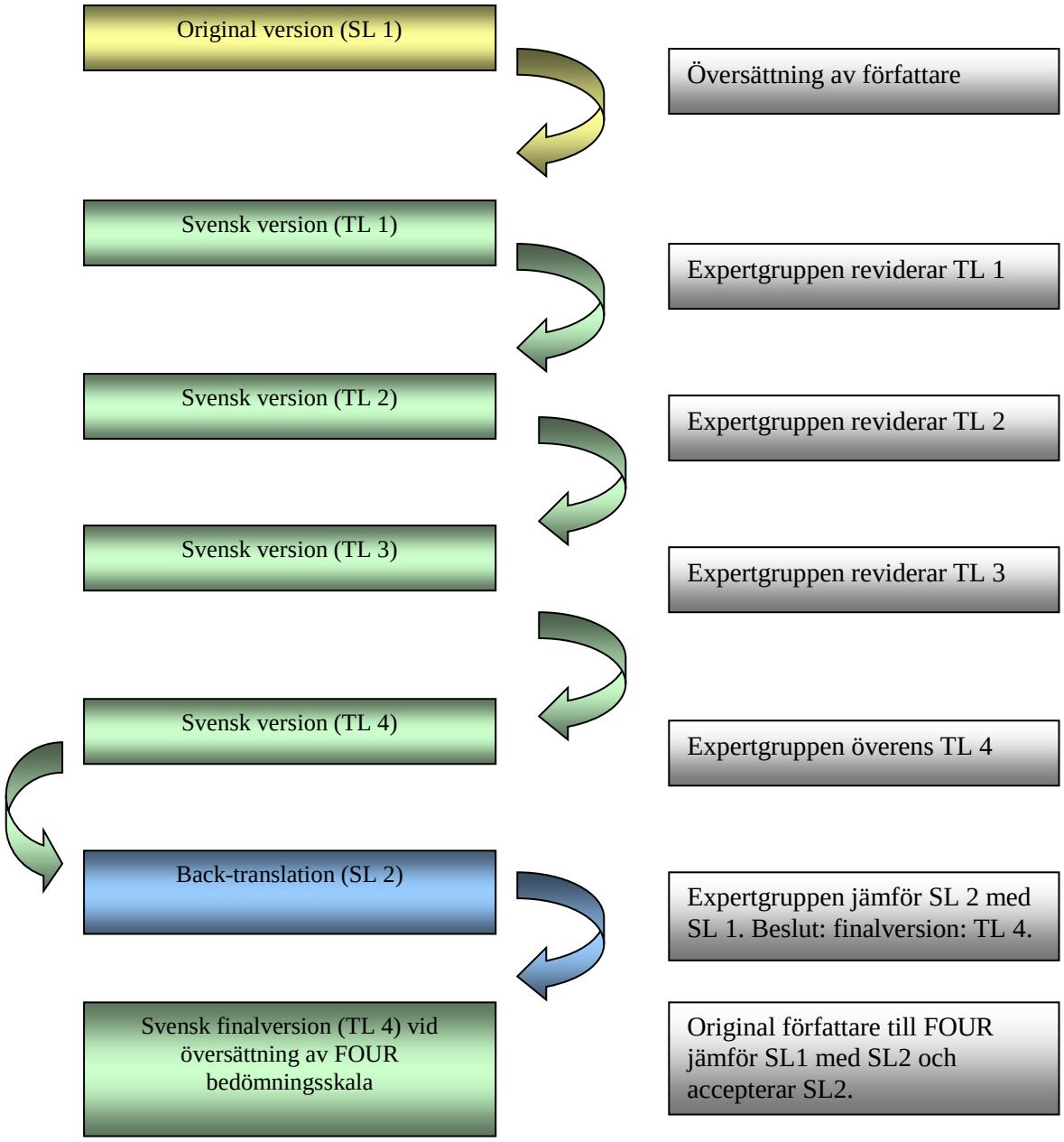
Yu, D. S., Lee, D. T. & Woo, J. (2004). Issues and challenges of instrument translation. *Western Journal of Nursing Research, 26* (3), 307-320. Doi: 10.1177/0193945903260554.

Ægisdottir, S., Gerstein, L. H. & Cinarbas, D. C. (2008).Methodological issues in cross-cultural counseling research: Equivalence, bias and translations. *The Counseling Psychologist, 36*, 188-219. doi:10.1177/0011000007305384

Bilaga 1 Figur: Full Outline of Unresponsiveness (FOUR) score



Bilaga 2 Figur over översättningsprocessen



Ämnesfördjupning

Reliabilitet, validitet och
översättning av mätinstrument inom vård

Annika Hickisch

Örebro universitet, Institutionen för hälsovetenskap och medicin
Förbättringsarbete inom omvårdnadsvetenskap

Examensarbete, master 30 hp

Vårtermin 2013

OMVÅRDNADSPROCESSEN

Omvårdnadsprocessen är en teori om hur sjuksköterskan organiserar vården för individen, familjen och i samhället (Lunney, 2009). Yura & Walsh (1973) beskrev omvårdnadsprocessen som en systematiserad, logisk, målinriktad process som skyddar mot uppkommande problem. Oavsett i vilket område eller vilken specialitet sjuksköterskan verkar, i så används omvårdnadsprocess som struktur. Omvårdnadsprocessen ansågs ursprungligen på 1960-talet omfatta fyra steg: bedömning, planering, åtgärder, utvärdering. För att kunna planera vården krävs en analys och tolkning av de data som insamlats vid bedömning. Detta bidrog till att ett femte steg tillkom, omvårdnadsdiagnos. Den första konferensen i ämnet hölls 1973 i USA. Omvårdnadsdiagnostik kräver intellektuell, interpersonell och teknisk kompetens hos sjuksköterskan. Den tekniska färdigheten att utföra en bedömning och att samla in reliabla och valida data, är grunden för att använda omvårdnadsdiagnoser. Det är viktigt att sjuksköterskan utvecklar sin förmåga att genomföra en allmän och fokuserad hälsohistoria samt fysiska undersökningar (Lunney, 2009).

BEDÖMNING, första steget i omvårdnadsprocessen

Bedömningsfasen, det första steget i omvårdnadsprocessen, omfattar vårdhistorik från patienten eller från andra uppgiftslämnare samt fysisk bedömning av patienten. Den avslutas med en omvårdnadsdiagnos alternativt en bekräftelse på att patienten har hälsa (Yura & Walsh, 1973). Sjuksköterskan baserar sin planering och prioritering av åtgärder på bedömningen.

Omvårdnadsbedömning av en individ är den systematiskta insamlingen av subjektiva data, såsom hälsohistoria (vad patienten säger) och objektiva data vid fysisk bedömning av patienten (vad som ses, hörs, luktar och känns). Två typer av bedömningar utförs. Den allmänna/generella bedömningen täcker alla aspekter för att bestämma hälsostatus för individen, familjen eller samhället. Den fokuserade bedömning utförs när t.ex. specifika symtom behöver utredas ytterligare utav det som är individens besvär, såsom andnöd eller smärta (Lunney, 2009). Det finns sex faser i bedömningssteget: datainsamling, identifiering av ledtrådar, verifiering av data, sammansättning av relaterade data, identifiering av mönster, rapportering och dokumention av data.

Den första fasen, datainsamling, kan delas in i tre typer. Ankomstbedömning, som omfattar alla aspekter av hälsostatus, sker vid initial kontakt med patienten. Fokuserad bedömning görs för att bestämma status vid ett speciellt tillstånd (ex. andnöd, blodtrycksfall, smärta, förstoppning, sänkt medvetande). Snabba prioriteringsbedömningar (*Quick Priority Assessments*, QPA) är korta fokuserade och prioriterande bedömningar (ex. A-B-C-D), som utförs för att erhålla den viktigaste informationen, som behövs tas hänsyn till, vid planering av vården. Denna form av prioriterande bedömning ökar kvaliteten och patientsäkerheten, ibland vid livshotande situationer. Maslow's behovspyramid kan användas, när behov av priortering finns. Vid intervju och fysisk bedömning av patienten beaktas de etiska, sociokulturella och andliga faktorer som kan påverka patientens hälsa vid planering av vården (Alfaro-LeFevre, 2010).

Vid genomförandet av den fysiska bedömningen används inspektion, auskultation, palpation och perkussion för att kunna identifiera objektiva (observerade) data, som komplement till de subjektiva (sagda), som patienten beskriver. Det är viktigt att den fysiska bedömningen sker på ett strukturerat och systematiskt sätt. Ett exempel på detta är ”huvud-till -tå” bedömning (*Head-to-toe psysical assessment*), där sjuksköterskan systematisk genomför fysisk bedömning på patienten uppifrån och ner, från huvud till tå (Alfaro-LeFevre, 2010).

BEDÖMNINGSINSTRUMENT INOM VÅRD OCH OMVÅRDNAD

Vid förbättringsarbete inom vårdarbete är det viktigt att utveckla mätinstrument med god reliabilitet och validitet, för att utveckla och förbättra möjligheten till att ge god, utvärderingsbar och säker vård. Mätinstrument och bedömningsskalor bör testas för reliabilitet och validitet för att vara säkra att använda.

RELIABILITET

Ett mätinstruments reliabilitet avser dess tillförlitlighet och följdriktighet, med vilket instrumentet mäter den avsedda egenskapen (Streiner & Norman, 2008). Desto mindre variation ett instrument producerar vid upprepade mätningar, ju högre är dess reliabilitet. Enligt Streiner och Norman (2008) beskrivs reliabiliteten som kvoten av variationen mellan individer och de totala variationerna i skalans poäng, med andra ord, reliabiliteten mäter förhållandet av variationen i poäng som orsakats av den sanna skillnaden emellan individerna. Reliabilitetskoefficienten uttrycks som ett tal mellan 0 och 1, där 0 indikerar ingen reliabilitet och 1 indikerar perfekt reliabilitet.

Stabilitet

Med ett instruments stabilitet menas dess reproducerbarhet av mätning utförd vid olika tillfällen. Det kan gälla mellan olika bedömare (interbedömarreliabilitet, *inter-observer reliability*), mellan olika tillfällen med samma bedömare (intrabedömarreliabilitet, *intra-observer reliability*) och mellan två observationer inom viss tidsrymd på samma patient (*test-retest*). För att kunna bedöma ett mätinstruments reliabilitet, så är information om instrumentets stabilitet ett minimum (Streiner & Norman, 2008). I en bedömning av reliabilitet mellan bedömare innebär ”överensstämmelse” (*agreement*), i vilken grad två eller fler bedömare/observatörer är överens om poängsättning med ett mätinstrument. Om överensstämmelsen är stor antas att felaktigheter vid mätning minimeras (Polit & Beck, 2008).

Kappa är en vanlig analysmetod för beräkning av reliabilitetskoefficienter vid interbedömarreliabilitet. Kappa (κ), har ett maximum av 1.0 när perfekt överensstämmelse finns och värdet noll indikerar ingen överensstämmelse utöver den slumpmässiga chansen för överensstämmelse (Altman, 1991). Fördelen med kappa jämfört med procentuell överensstämmelse är att den kompenserar för slumpmässig överensstämmelse. Landis & Koch (1977) ger följande riktlinjer hur kappa värdet på överensstämmelse skall tolkas:

< 0.20 = Dålig överensstämmelse
0.21-0.40= Knapp
0.41-0.60= Måttlig
0.61-0.80= God
0.81-1.00= Mycket bra överensstämmelse

Kappastatistik har som nackdel att den inte tar hänsyn till grad av bristande överensstämmelse vid flergradig skala. När kategorierna är rangordnade (vilket de ofta är) kan det vara en fördel att ge olika vikt till bristande överensstämmelse beroende på hur stort skillnaden är. Detta kan beräknas som ”viktad kappa” (*weighted kappa,* κ_w) enligt Altman (1991). Det finns både linjär och kvadratisk viktad kappa som beräknar på olika sätt hur stor skillnaden i överensstämmelsen är. Det system för viktning som rekommenderas är kvadratisk viktning, där viktningen baseras på kvadraten av mängden av oenighet (Streiner & Norman, 2008). Om exempelvis skillnaden mellan bedömningar är två skalsteg blir viktningen $2^2 = 4$. Enligt Streiner och Norman (2008) är viktad kappa likvärdig med Intraclass Correlation Coefficient (ICC).

Homogenitet *(internal consistency)*

En annan form av reliabilitet är homogenitet som mäter statistiskt hur konsekvent ett antal komponenter korrelerar med varandra. Om de olika delkomponenterna (*items*) korrelerar med det underliggande fenomenet så korrelerar de även med varandra. Enligt Streiner och Norman (2008) är mätningar av homogenitet den vanligaste beskrivningen på reliabilitet då det går att beräkna genom ett enda administrationstillfälle av testet.

Den vanligaste metoden för att mäta homogenitet är koefficient alfa (α), som även benämns Cronbach’s alfa. Cronbach’s alfa (α) tolkas med normal variation mellan 0- +1. Ett högre värde speglar högre/bättre homogenitet, enhetlighet. (Polit & Beck, 2008). Olika författare ger olika rekommendationer om lägst acceptabelt alfa värde. Enligt Streiner och Norman (2008) så bör homogenitet överstiga 0,8 och stabilitet bör vara över 0,5. Beroende på testets användningsområde och risker vid misstolkning, så kan högre värden krävas.

VALIDERING

Validitet avser instrumentets giltighet dvs. att det mäter det som avses att mätas och om rätt slutsatser kan dras av det som undersökts. Säkerställande av validitet kan göras på olika sätt: innehållsvalidering (*content validity*), kriterievalidering (*criterion validity)* och begreppsvalidering (*construct validity*) (Streiner och Norman, 2008). Nedan beskrivs utöver dessa tre typer även ytvaliditet.

Innehållsvalidiering

Innehållsvalidering undersöker i vilken utsträckning ett mätinstrument har passande, lämpliga delkomponenter (*items*) och adekvat omfattning för det som är avsett att mätas. För detta finns inga helt objektiva metoder, utan innehållsvaliditeten baseras på bedömning och uppfattning. Bedömningen kan göras av en panel av experter, som utvärderar och dokumenterar innehållsvaliditeten hos ett mätinstrument. Detta sker ofta genom *content validity index* (CVI), där panelen granskar individuella poster samt hela skalan utifrån relevans och passande termer (Polit & Beck, 2008).

Kriterievalidering

Ett sätt att validera ett mätinstrument är att jämföra det med ett redan befintligt bedömningsinstrument *(golden standard)* och som anses ha hög validitet (Streiner & Norman, 2008). Sambandet mellan de båda skalorna mäts som en korrelation och graden av överensstämmelse, en korrelationskoefficient, anger om den testade skalan är lika precis och tillförlitlig som den tidigare validerade skalan (Polit & Beck, 2008). Huvudfrågan för kriterievalidering är om mätinstrumentet är användbart och kan förutsäga beteenden eller tillstånd enligt Polit och Beck (2008).

Kriterievalidering indelas ofta i samtidig (*concurrent validity*) och prediktiv, förutsägande validering (*predictive validity*). Vid samtidig validering jämförs den nya skalan med ett nuvarande kriterium vid ett och samma mättillfälle. Prediktiv validering, anger hur väl mätinstrumentet kan förutsäga ett kriterium som kommer att mätas i framtiden (t.ex. död eller sjukdom) (Streiner & Norman, 2008; Polit & Beck, 2008).

Begreppsvalidering

Egenskaper som kroppsvikt kan operationaliseras och mätmetoden kan definieras. Det är svårare med begrepp som exempelvis egenskaper och känslor, då variablerna är abstrakta och kan inte observeras direkt (Streiner & Norman, 2008). Enligt Polit och Beck (2008) är begreppsvalidering en strävan mot hypotestestning, kopplat till teoretiska perspektiv vad gäller begreppet. Det finns olika metoder såsom ”*Known Groups*” där mätinstrumentet administreras till grupper, där det är känt att grupperna skiljer sig åt (ex. rökare, icke-rökare) vad gäller känd egenskap (ex. rökning). En annan metod är faktoranalys, som är en statistisk metod, vilken identifierar kluster av relaterade objekt.

Ytvaliditet (*Face validity***)**

Ytvaliditet refererar till när användaren tittar på ett mätinstrument och ser att det mäter det som det är avsett att mäta, d.v.s. hur det ”presenterar” sig för användaren (Polit & Beck, 2008; Streiner & Norman, 2008). Ytvaliditet skall dock inte ses som ett starkt bevis för mätinstrumentets validitet (Polit & Beck, 2008). Enligt Streiner och Norman (2008) har ytvaliditet fem funktioner: 1) ökar motivationen och samarbetet hos respondenter/användare, 2) kan attrahera potentiella användare, 3) minskar missnöje bland användare, 4) gör det mer möjligt att beslutsfattare och andra accepterar resultatet, 5) förbättrar relationer med ex. media och domstolar.

Att påvisa validiteten och reliabiliteten hos ett mätinstrument gäller egentligen inte skalan generellt, utan när den används till en speciell grupp av personer och under speciella sammanhang som den testats för. Om en skala skall användas till en annan grupp (t.ex. patientkategori, kulturtillhörighet) och i en annan miljö (t.ex. strokeavdelning, intensivvård) kan resultaten från ursprungliga valideringsstudien inte tillämpas utan förnyade studier är nödvändiga (Streiner & Norman, 2008; Polit & Beck, 2008).

ÖVERSÄTTNING AV MÄTINSTRUMENT

När en skala är validerad och skall översättas sker det vanligen med centrerad (eller asymmetrisk) översättning, vilket har som mål att noggrant efterlikna originalskalan. Originalspråket anses vara standarden mot vilken översättningen skall jämföras. Då ett instrument är under utveckling eller när det finns en önskan om att utveckla ett mer universellt förståeligt språk, används decentrerad (eller symmetrisk) översättning, som då ger möjlighet till modifiering av delkomponenter på original instrumentet. Både den ursprungliga och den översatta versionen anses vara lika betydelsefulla och forskaren centrerar sig inte i något av de två språken (Beck, Bernal & Froman, 2003; Polit & Beck, 2008; Ægisdottir, Gerstein & Cinarbas, 2008).

En fördel med asymmetrisk översättning som är trogen originalversionen, är att gruppen i den nya kulturen som undersöks kan jämföras med originalgruppen. Asymmetrisk översättningsprocess anses mest lämplig om studien har operativa mål och när det finns ett

”outsider” perspektiv på forskningsdeltagarna (*etic approach*) enligt Jones, Mallinson, Phillips och Kang (2006).

Studier med jämförande mål och som avser att skapa en tvärkulturell konstruktion över olika kulturer, har ett ”insider” perspektiv (*emic approach*). Då är symmetrisk översättning mest lämplig. Det finns även blandformer (*derived etic approach*) som speglar både *etic* och *emic* perspektiv (Jones et al., 2006). Enligt Brislin (1986) refererar *etic* till ett fenomen som har gemensam mening, meningens kärna, över flera kulturer. *Emic* kallas de aspekter, som är olika i kulturer, men varje *emic* är relaterad till en gemensamt delad *etic* kärna. En beskrivning på varje kulturs fenomen bör bestå av *etic* kärna samt kulturens *emic*-aspekter enligt Brislin (1986).

Målet med översättning av ett mätinstrument är att nå likvärdighet, ekvivalens, mellan originalet och den översatta versionen av skalan. Enligt Streiner och Norman (2008) finns olika typer av likvärdighet:

- Konceptuellt *(begreppsmässig, finns begreppet i båda kulturerna?*)
- Komponenterna (*item, relevanta poster för målgruppen?)*
- Semantiskt (*språklig översättning, som omfattar även kulturella uttryck)*
- Operationellt *(samma format, instruktioner, administrationssätt till målgruppen?)*
- Mätmässigt (*testar om reliabilitet och validitet är lika i båda versionerna*)

En av de första frågorna som ska ställas inför en översättning är, om det definierade begreppet i skalan speglar begreppets mening inom kulturen, som översättningen riktar sig till (konceptuell likvärdighet). Experter med kunskap i frågan inom målkulturen, bör tillfrågas tidigt i processen. Den semantiska likvärdigheten jämför, i vilken grad betydelsen av varje delkomponent är densamma hos målkulturen efter översättningen, som den var hos originalet. Bokstavliga översättningar är sällan tillfredsställande. Det är den underliggande meningen hos originalets ord som bör bevaras snarare än exakta ordalydelsen (Polit & Beck, 2008).

Översättningsprocessen har främst beskrivits av Richard W. Brislin (Ægisdottir, Gerstein & Cinarbas, 2008). Översättarna, som gärna bör vara minst två eller någon form av grupp, bör utföra översättningen till sitt eget modersmål. Det räcker inte bara att var tvåspråkig, utan de bör ha kunskap om det som ligger till grund för instrumentet (Polit & Beck, 2008). Den semantiska ekvivalensen stärks genom en tillbakaöversättning (Brislin, 1970, 1986). När instrumentet översatts från original språket (*source language*) till målspråket (*target language*), skall det översättas tillbaka till ursprungsspråket av en översättare, som inte har tagit del av original texten. Därefter sker jämförelse mellan original instrumentet och tillbakaöversättningen till originalspråket och eventuella förändringar görs i målversionen. Denna del av processen sker upprepade gånger, tills dess konsensus nåtts om finalversionen inom gruppen eller kommittéen som deltar i processen. Slutligen sker en psykometrisk utvärdering av det översatta instrumentet (Polit & Beck, 2008; Ægisdottir, Gerstein & Cinarbas, 2008).

REFERENSER

Alfaro-LeFevro, R. (2010). *Applying nursing process: A tool for critical thinking* (7th ed.). Philadelphia: Lippincott Williams & Wilkins.

Altman, D. G. (1991). *Pratical statistics for medical research*. London: Chapman & Hall.

Beck, C. H., Bernal, H. & Froman, R. D. (2003). Methods to document semantic equivalence of a translated scale. *Research in Nursing & Health, 26*, 64-73.doi: 10.1002/nur.10066

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-cultural Psychology, 1*(3), 187-216.

Brislin, R. W. (1986). The wording and translation of research instuments. In W.L. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Newbury Park, CA: Sage.

Jones, E. G., Mallinson, R. K., Phillips, L. & Kang, Y. (2006). Challenges in language, culture, and modality- Translating English measures into American Sign Language. *Nursing Research, 55* (2), 75-81.

Lunney, M. (2009). Assessment, clinical judgment, and nursing diagnoses: How to determine accurate diagnoses. In T. H. Herdman (Ed.), *Nursing diagnoses: Definitions and classification 2009-2011* (pp. 3-17). Chichester, U K: Wiley-Blackwell.

Polit, D. F. & Beck, C. T. (2008). *Nursing Research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Lippincott Williams &Wilkins.

Streiner, D. & Norman, G. (2008). *Health Measurements scales: a practical guide to their development and use* (4th ed.). Oxford: Oxford University Press.

Yura, H. & Walsh, M. B. (1973). *The nursing process: assessing, planning, implementing, evaluating* (2nd ed). New York: Appleton-Century-Crofts

Ægisdottir, S., Gerstein, L. H. & Cinarbas, D. C. (2008).Methodological issues in cross-cultural counseling research: Equivalence, bias and translations. *The Counseling Psychologist, 36*, 188-219. doi:10.1177/0011000007305384