Over keuringen. Onderzoek naar de kwaliteit van keuringen voor militair personeel van de Koninklijke Landmacht

Citation for published version (APA):

de Raad, J. (2005). *Over keuringen. Onderzoek naar de kwaliteit van keuringen voor militair personeel van de Koninklijke Landmacht*. [Doctoral Thesis, Maastricht University].
<https://doi.org/10.26481/dis.20050916jr>

Document status and date:

Published: 01/01/2005

DOI:

[10.26481/dis.20050916jr](https://doi.org/10.26481/dis.20050916jr)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

OVER KEURINGEN

Onderzoek naar de kwaliteit van keuringen voor militair
personeel van de Koninklijke Landmacht

ISBN 90 5278 479 5

Copyright © John de Raad, Maastricht 2005

Vormgeving en druk: Datawyse / Universitaire Pers Maastricht

Omslag: Invaliden te Bronbeek, 1897 door J.W. Hoijinck van Papendrecht.

OVER KEURINGEN

Onderzoek naar de kwaliteit
van keuringen voor militair personeel
van de Koninklijke Landmacht

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 16 september 2005 om 14.00 uur

door

John de Raad

Promotores

Prof. dr. F.J.N. Nijhuis

Prof. dr. J.H.B.M. Willems (Universiteit van Amsterdam)

Beoordelingscommissie

Prof. mr. dr. F. van Wijmen, voorzitter

Prof. dr. J. Th. M. van Eijk

Prof. dr. J.W. Groothoff (Rijksuniversiteit Groningen).

Dr. W.L.A.M. de Kort (Divisiedirecteur Sanquin Bloedbank Regio Zuidoost)

Dr. A. de Rijk

18. *Want geen man, in wien een
gebrek zal zijn, zal naderen
om de spijs van zijn God te offeren,
hij zij een blind man, of kreupel,
of te kort of te lang in leden;*
19. *Of een man, in wien een breuk des
voets, of een breuk der hand zal zijn;*
20. *Of die bultachtig, of dwergachtig
zal zijn, of een vel op zijn oog
zal hebben, of droge schurftheid,
of etterige schurftheid, of
die gebroken zal zijn aan zijn gemacht.*

Leviticus, vers 21, Statenvertaling 1959.

Inhoudsopgave

1. Inleiding en vraagstelling	11
1.1 Achtergrond van het onderzoek	12
1.2 Niet systematische waarnemingen	12
1.3 Systematische waarnemingen	14
1.4 Een nieuw keuringssysteem en de relatie ervan met de wet- en regelgeving	28
1.5 Theorie, hypothese en vraagstelling	39
1.6 Materiaal en methoden	51
1.7 Verdere uitwerking van het onderzoek	58
 2. De beperkingen van medische beperkingen	 65
2.1 Inleiding en vraagstelling van dit onderzoek	66
2.2 Achtergrondinformatie	66
2.3 Materiaal en methoden	68
2.4 Resultaten	71
2.5 Beschouwing	75
 3. Een onderzoek naar de reproduceerbaarheid van de basismedische eisen KL bij de pensioenkeuring	 79
3.1 Inleiding en vraagstelling van het onderzoek	80
3.2 Materiaal en methoden	81
3.3 Resultaten	83
3.4 Beschouwing	86
 4. De reproduceerbaarheid van de aanstellingskeuring op grond van de Basismedische eisen en het ABOHZIS-systeem	 91
4.1 Inleiding en vraagstelling van dit onderzoek	92
4.2 Achtergrondinformatie	93
4.3 Materiaal en methoden	94
4.4 Resultaten	96
4.5 Beschouwing	99

5. Een vergelijking tussen twee systemen voor de aanstellingskeuring bij de Koninklijke Landmacht met behulp van een gerandomiseerde, gecontroleerde studie	105
5.1 Inleiding en vraagstelling van dit onderzoek	106
5.2 Materiaal en methoden	108
5.3 Resultaten	110
5.4 Beschouwing	115
6. Analyse van het verschil in prognostische waarde tussen beide keuringssystemen	119
6.1 Inleiding en vraagstelling voor dit onderzoek	120
6.2 Materiaal en methoden	121
6.3 Resultaten	125
6.4 Beschouwing	130
7. Verschillen in inzetbaarheid tussen uitgezonden militairen	135
7.1 Inleiding en vraagstelling voor dit onderzoek	136
7.2 Materiaal en methoden	139
7.3 Resultaten	143
7.4 Beschouwing	147
8. Slotbeschouwing	151
8.1 Inleiding	152
8.2 Nabeschouwing van het verrichte onderzoek	152
8.3 Beantwoording van de vraagstelling	156
8.4 Toetsing van de hypothese	159
8.5 Aanpassing van de theorie?	160
8.6 Over keuringen: de betekenis van dit onderzoek voor de empirische spiraal	163

Bijlage: Verklaring van begrippen en afkortingen	167
Samenvatting	173
Abstract	179
Nawoord	185
Curriculum Vitae	187

10

10

10

10

10

10

10

10

1 | Inleiding en vraagstelling

1.1 Achtergrond van het onderzoek

Basis van dit onderzoek is de behoefte om de inzetbaarheid van militairen goed te kunnen voorspellen. De keuze voor het verrichten van onderzoek naar valide instrumentarium is voortgekomen uit een aantal niet-systematische waarnemingen. Deze niet-systematische waarnemingen hebben geleid tot onderzoeksvragen, met als doel om meer inzicht te krijgen in de aspecten die bijdragen tot een verantwoord instrumentarium. Uitgaande van bestaand instrumentarium is een nieuw toetsingsinstrument ontwikkeld om de inzetbaarheid van militairen mee te beoordelen. Daarmee is dit onderzoek naar de voorspelbaarheid van medische keuringen primair een valideringsonderzoek [1] geworden.

In dit proefschrift worden de aanleiding, uitvoering en interpretatie van een onderzoek naar de kwaliteit van de medische aanstellingskeuring bij de Koninklijke Landmacht (KL) beschreven. Daarbij wordt uitgegaan van een onderzoeksmodel dat vergelijkbaar is met het model van de empirische cyclus [1]. Hieronder volgt allereerst de weergave van twee niet-systematische waarnemingen, die beide aanleiding vormden voor verder onderzoek. Hierna wordt aandacht besteed aan systematische waarnemingen, die mede hebben geleid tot de ontwikkeling van een nieuw keuringssysteem, waaraan aandacht zal worden besteed. Vervolgens wordt aan de waarnemingen een theorie ontleend, waaruit met inachtneming van het nieuwe keuringssysteem een hypothese wordt gededuceerd. Dan volgt het formuleren van de onderzoeksvragen, gericht op de validering van de keuringssystematiek. Aansluitend vindt een beschrijving van materiaal en methoden plaats.

Het tweede hoofdstuk is de weergave van de primaire reactie op beide niet-systematische waarnemingen, in de vorm van een eerste eigen onderzoek. In de daaropvolgende hoofdstukken wordt separaat ingegaan op de geformuleerde onderzoeksvragen. Het afsluitende hoofdstuk begint met de nabeschuiving op de verrichte onderzoeken, waarna de onderzoeksvragen worden beantwoord, gevolgd door de toetsing van de hypothese. Uiteindelijk wordt de vraag beantwoord of de nieuw ontwikkelde keuringssystematiek een valide instrument is om de inzetbaarheid van militairen te beoordelen en wordt aandacht besteed aan de vraag op welk vervolg-onderzoek op het terrein van keuringen resteert.

1.2 Niet systematische waarnemingen

De eerste niet-systematische waarneming vond plaats tijdens het 24^e Internationale Congres voor Bedrijfsgezondheidszorg te Nice in 1993. Post-Uiterweer [2] hield tijdens dat congres een lezing over onderzoek naar de reproduceer-

baarheid van aanstellingskeuringen. Hij had 180 geanonimiseerde keuringsdossiers, waarvan de ene helft geschikt was verklaard en de andere helft ongeschikt, ter herbeoordeling aan vijf artsen van de bedrijfsgezondheidszorgdienst voor overheidspersoneel (RBB) voorgelegd. De bij dit onderzoek gevonden mate van overeenkomst viel lager uit dan wat als aanvaardbare ondergrens voor screeningstesten wordt aangehouden [1]. Omdat in Frankrijk het uitvoeren van (aanstellings-, periodieke) keuringen een prominente plaats binnen de bedrijfsgezondheidszorg inneemt, alsmede omdat de meerderheid der toehoorders uit het gastland afkomstig bleek, leidde het bekend stellen van de onderzoeksresultaten tot uitgebreide discussie met publieke inmenging. Deze waarneming leidde enerzijds tot de constatering dat er maatstaven bestonden om de kwaliteit van keuringen te beoordelen, anderzijds tot het vermoeden dat vele keurend artsen niet van deze maatstaven op de hoogte waren.

Interesse naar de vraag hoe de mate van overeenstemming van binnen de KL verrichte keuringen zou zijn, ontstond pas later. Daarvoor moest eerst een tweede niet-systematische waarneming worden gedaan.

Deze tweede waarneming vond plaats tijdens een reorganisatieproject bij de KL in 1994. De dienstplicht werd opgeschort, er zou een beroepsleger worden gevormd en de taakstelling werd uitgebreid met het uitvoeren van missies buiten de eigen landsgrenzen. De samenstelling van het functiebestand van het toenmalige beroepskader was te onevenwichtig voor de directe omvorming tot een beroepsleger. Bovendien waren bij enige duizenden beroepsmilitairen medische en/of functioneringsbeperkingen in hun personeelsdossier opgenomen. Op grond van dat laatste feit veronderstelde de KL-leiding dat deze militairen, zeker met de uit te breiden taakstelling in het vooruitzicht, als beperkt inzetbaar moesten worden beschouwd en derhalve als eersten voor ontslag in aanmerking zouden komen.

Die medische en/of functioneringsbeperkingen werden vastgesteld naar aanleiding van het "Periodiek Geneeskundig Onderzoek" (PGO). Dit PGO was een formele periodieke toetsing aan de medische dienstgeschiktheid voor beroepspersoneel. Vermindering van die geschiktheid kon leiden tot een beperking of tot ongeschikt verklaring. Op de beoordeling van militaire dienstgeschiktheid en het PGO wordt in paragraaf 1.3 nader ingegaan.

De staf van de Geneeskundige Dienst KL kreeg de opdracht om aan de hand van deze keuringsdossiers een uitspraak te doen omtrent de "algemene en brede inzetbaarheid" van militairen met beperkingen in hun personeelsdossier.

Bij toetsing van de keuringsdossiers kwamen twee knelpunten aan het licht. Ten eerste bleken de in de keuringsdossiers weergegeven beperkingen sterk risicowerend van aard te zijn; er was zelden sprake van een causaal verband tussen de gesignaleerde medische problemen en de actuele inzetbaarheid. Het gevolg

daarvan was discrepantie tussen de administratieve inzetbaarheid van vele militairen en hun effectieve inzetbaarheid. Onder de geschetste omstandigheden gingen vele militairen met succes in bezwaar tegen de opgelegde beperkingen. In de tweede plaats bleek de intra-waarnemers variatie tijdens de herbeoordeling van deze dossiers hoog te liggen.

Deze waarnemingen vormden de aanleiding om het door Post-Uiterweer verrichte onderzoek [2] binnen de KL te herhalen (Hoofdstuk 2). Het gegeven dat het gehanteerde keuringssysteem op een zo risicowerende manier kon worden gehanteerd, vormde de aanleiding om het militaire keuringssysteem nader te beschouwen.

1.3 Systematische waarnemingen

1.3.1 Beschouwing van het Militair Keuringsreglement en de lijst van ziekten en gebreken

In 1810 was Holland door Napoleon Bonaparte bij het Franse Keizerrijk ingelijfd. Tot 1810 bestonden de Napoleontische legers geheel uit vrijwillig dienend beroepspersoneel, maar in 1811 ontstond een tekort aan vrijwilligers. Daarom voerde Napoleon in 1811 de conscriptie in – een beperkte vorm van dienstplicht [3].

Voorafgaand aan het in dienst treden van deze dienstplichtigen werd de vaststelling van hun medische geschiktheid noodzakelijk geacht. Dit was de reden waarom ook in ons land het eerste (Franse) reglement op het geneeskundig onderzoek van miliciens van kracht werd. Dit reglement vormt de grondslag van het Militair Keuringsreglement (MKR), dat sedertdien bij de Koninklijke Landmacht als basis dient voor zowel de medische aanstellingskeuring als de pensioenkeuring voor militair personeel [3]. Reeds de eerste versie van het MKR bevatte naast het reglementaire gedeelte ook een lijst met ziekten en gebreken “die ongeschikt maken voor den militairen dienst”, zoals Romeyn [4] stelt.

De keurend arts onderzocht de rekrut op de aanwezigheid van die ziekten en gebreken. Indien ziekten of gebreken aanwezig bleken te zijn diende de keurend arts vervolgens te bepalen of dat lijden “op hardnekkige wijze” of “in belangrijke mate” aanwezig was. Deze beide laatste bevindingen leidden tot afkeuring. Na het vertrek van de Franse troepen uit ons land in 1815 werd dit MKR als keuringssysteem voor de krijgsmacht gehandhaafd.

Vanaf het prille begin van de dienstplichtkeuring speelde de mogelijke simulatie van ziekten en gebreken, om zo de dienstplicht te ontlopen, onder rekruten een rol. In de 19^e eeuw maakte volgens Branger [5] herkenning van simulatie onderdeel uit van de opleiding tot officier-arts. Ook in de 20^e eeuw wordt de

officier-arts door verschillende auteurs op het bestaan van simulatie geattendeerd, zoals Fouw [6], Quix, [7], Wijsman [8] en Spruyt [9].

Daarnaast ontstaat volgens Romeyn [10] al spoedig na de invoering van het MKR de tendens om actief dienende ongeschikte miliciens zo snel mogelijk af te keuren, teneinde de sterfte binnen de Landmacht te laten dalen.

In de loop der tijden werden zowel MKR als de lijst met ziekten en gebreken periodiek aangepast. Dit had onder meer te maken met het ontstaan van meer verfijnde meetmethoden (gehoor), nieuwe endemische ziekten (tuberculose), wetenschappelijke ontwikkelingen (intrede van de psychologie) en gewijzigde taakstelling (met de exploitatie van koloniën deden tropische ziekten hun intrede).

Ter verfijning van de selectiecriteria nam de KL in 1946 het Canadese PULHEEMS-selectiesysteem over, vertaald als het ABOHZIS-systeem [11]. De letters ABOHZIS stonden voor **A**lgemene fysieke toestand, **B**ovenste extremiteiten, **O**nderste extremiteiten, **H**oren, **Z**ien, **I**ntelligentie en **S**tabiliteit. Voor elk van deze deelgebieden werd een indeling in vijf classificaties gehanteerd, lopend van één tot vijf (bijvoorbeeld A₁, B₂, O₃ etc.). Op deze wijze werden de inmiddels 282 ziekten en gebreken uit het MKR over de zeven deelgebieden verspreid. Voor elke militair werd bij de keuring een ABOHZIS-classificatie opgesteld. Daarbij betekende een 1-classificatie dat men niet aan een in het MKR vermelde ziekte of gebrek leed, hetgeen in toenemende mate wel het geval was indien men een 2-classificatie of hoger kreeg. Elke 5-classificatie leidde tot ongeschiktverklaring. Hierdoor werden de begrippen "hardnekkige wijze" en "mits in belangrijke mate" nader gedetailleerd. Verminderde belastbaarheid werd aangeduid door het hanteren van één of meer graden "2", "3" of "4" in de ABOHZIS-classificatie. Deze verminderde belastbaarheid werd vervolgens door middel van een beperking verduidelijkt. Zo werd bij een bepaald perceptief verlies in het toonaudiogram de beperking "mag nog uitsluitend met dubbele gehoorbescherming en uitsluitend ten behoeve van het bijhouden van de schietvaardigheid schieten" opgelegd.

Oorspronkelijk was de combinatie van MKR met ABOHZIS-classificatie alleen bedoeld om dienstplichtigen te keuren. In 1959 werd deze classificatie ook voor het beroepspersoneel ingevoerd. Koster [12] beschrijft dat vanaf dat moment aan beroepspersoneel hogere eisen werden gesteld dan aan dienstplichtigen. Voor beroepspersoneel werd de minimale geschiktheidnorm de "2-classificatie"; voor dienstplichtigen werd de "3-classificatie" gebruikt naar mate dit (qua personele behoefte) het beste uitkwam.

Dit onderscheid is tot het einde toe blijven bestaan. De ABOHZIS-classificatie werd de maatstaf voor de algemene inzetbaarheid; op grond van die classificatie

werd beroepspersoneel in verschillende klassen ingedeeld. De klassering was bepalend voor de kansen op bevordering. Vervolgens werd deze indeling ook op actief dienend beroepspersoneel toegepast. Vanaf 1960 moest iedere beroepsmilitair om de 5 jaar op deze manier worden gekeurd. Daarmee werd het Periodiek Geneeskundig Onderzoek (PGO), zoals de KL dat tot 1996 heeft gekend, een feit. De classificatie en de bijbehorende beperkingen die uit dit PGO voortvloeiden, werden grotendeels ontleend aan het medisch dossier dat ter inzage bij die keuring werd meegegeven. Ter illustratie: de keurend arts constateerde dat de militair in kwestie jaren geleden was geopereerd aan een hernia van de rug, zocht vervolgens in het voorschrift op welke classificatie daarbij werd aanbevolen, waarna beperkingen werden opgelegd, ongeacht de status praesens van de keurling.

Het ABOHZIS-classificatie systeem functioneerde op deze wijze als een verlengstuk van de lijst van ziekten en gebreken uit het MKR. Verdere detaillering werd in 1967 bereikt met de invoering van de Ministeriële Publicatie 35-401, de "Aanwijzingen voor artsen belast met militaire keuringen". In deze regeling werd voor alle gradaties van de ziekten en gebreken een bepaalde ABOHZIS-classificatie voorgeschreven. Hierdoor was er voor de keurend arts nauwelijks meer sprake van professionele autonomie. Tegelijkertijd werd het MKR in 1967 voor de laatste maal herzien.

Op grond van het bovenstaande kunnen bij het MKR, tezamen met de lijst van ziekten en gebreken en de ABOHZIS-classificatie, de volgende kritische kanttekeningen worden geplaatst. Het keuren van dienstplichtigen betekent rekening houden met simulatie [5, 6, 7, 8, 9], hetgeen de zuivere oordeelsvorming niet ten goede kan zijn gekomen. Risicowering [10] speelt vanaf het begin van de dienstplichtkeuring een rol. Ook het vanaf 1959 gemaakte onderscheid in eisen tussen beroepspersoneel (zwaarder) en dienstplichtigen (lichter) kan worden beschouwd als een vorm van risicowering [12]. Beroepsmilitairen dienden voor het vervullen van dezelfde functie in mindere mate te lijden aan bepaalde ziekten of gebreken dan dienstplichtigen. Het systeem is diagnose-gericht en niet gebaseerd op het model belasting-belastbaarheid [13]. Op grond van aandoeningen uit het verleden worden beperkingen opgelegd. Het MKR ontwikkelde zich tot 1967 naar de samenleving. Juist sindsdien zijn de ontwikkelingen in de samenleving groter geweest dan ooit tevoren. Er is derhalve sprake van een gedateerd keuringssysteem, waarin risicowering een grote rol speelt.

Van Kessel [14] beargumenteerde al in 1984 om bovengenoemde redenen dat het ABOHZIS-systeem geen relatie met de feitelijke militaire werkzaamheden vertoont. Sindsdien is de aard van de feitelijke militaire werkzaamheden sterk

veranderd (beroepsleger, buitenland-missies), waardoor die relatie thans zeker nog ontbreekt.

Naast deze beschouwing op militaire keuringen werd vanaf de tachtiger jaren van de 20^e eeuw onderzoek verricht waardoor de kwaliteit van aanstellingskeuringen in bredere zin in ons land ter discussie kwam te staan.

1.3.2 Onderzoek naar de kwaliteit van aanstellingskeuringen

Onderzoek uit Nederland

Vanaf 1989 onderzocht De Kort, samen met wisselende andere onderzoekers zoals Lourijzen, Post-Uiterweer en Van Dijk, de omvang, kosten en kwaliteit van aanstellingskeuringen in Nederland [2, 15, 16, 17, 18, 19]. Bij deze onderzoeken werden de betrouwbaarheid, validiteit en voorspellende waarde gehanteerd als graadmeters voor de kwaliteit van de medische aanstellingskeuring als meetinstrument. Omdat de laatste twee graadmeters in de praktijk moeilijk te onderzoeken zijn voor keuringen (zie paragraaf 1.5), werd hiervoor in de plaats een literatuuronderzoek verricht [15]. Op basis van de gevonden literatuur bleken nauwelijks argumenten te vinden voor de onderbouwing van de veronderstelling dat door aanstellingskeuringen gezondheidsschade en ziekteverzuim kan worden voorkomen. Weging van voor- en nadelen bij het opzetten van het complete instrument aanstellingskeuring vindt niet plaats, evenmin als bij de verschillende afzonderlijke elementen uit de keuring. Het eindoordeel bleek tot stand te komen op basis van weging van de verschillende keuringsonderdelen, waarbij zowel de relevantie als de voorspellende waarde van de gebruikte onderzoeken ontbraken.

Voor het beoordelen van de betrouwbaarheid vond in diezelfde periode een vragenlijst-onderzoek onder 2350 bedrijven en 1750 artsen plaats. Op deze wijze wilde men inzicht krijgen in de hoeveelheid en het doel van de keuringen [16]. Uit dat vragenlijstonderzoek komt een grote pluriformiteit naar doel, werkwijze en maatstaven aan de orde bij telkens een laag percentage afgekeurden.

Daarnaast werd de mate van overeenstemming onderzocht door middel van het reproduceerbaarheidsonderzoek [2], waarvan al in paragraaf 1.2 gewag werd gemaakt. Bij dit onderzoek werden de dossiers van zowel 90 geschikt verklaarde als 90 ongeschikt verklaarde kandidaten aan een panel van vijf artsen voorgelegd. De gevonden overeenstemming was laag. Bovendien bleek een meer gedetailleerde instructie (die voor sommige functies was beschreven) slechts geringe invloed op de mate van overeenstemming te hebben.

Bovenstaande onderzoeken maakten in 1993 onderdeel uit van het proefschrift van De Kort [17]. Hij voegde hieraan enige publicaties toe. Eén ervan betreft een studie naar het aantal afwijzingen – plus de vermelde oorzaken daarvoor – bij de populatie van de toenmalige Rijksbedrijfsgezondheids- en Bedrijfsveiligheidsdienst (RBB) [18]. Op grond van de resultaten van deze studie concludeerde hij dat het percentage afkeuringen vrijwel altijd erg laag ligt en dat derhalve deze keuringen weinig toevoegen aan het wervingsproces. Voor de kleine groep die wordt afgekeurd ontbreekt bovendien een reproduceerbaar, gestandaardiseerd en valide onderzoek. Daarnaast vermeldde hij in het proefschrift een studie waarin 235 individuen met vroege uitval op het werk werden onderzocht op de relatie tussen bij de aanstellingskeuring gevonden medische bijzonderheden en de oorzaak van de uitval op het werk [16]. Met name omdat in de meerderheid van deze gevallen er helemaal geen medische bijzonderheden bij de aanstellingskeuring zijn genoteerd, concludeerde De Kort dat deze keuring slechts weinig bijdraagt aan het voorkomen van uitval op het werk. Tevens concludeerde hij dat het effect van een aanstellingskeuring op het terugdringen van arbeidsgebonden aandoeningen en vroeg optredende arbeidsongeschiktheid gering is.

In 1997 publiceerde De Kort [19] nog een artikel waarin hij het begrip aanstellingskeuring theoretisch en rekenkundig benadert. Hij definieert het preventieve effect van de aanstellingskeuring als: de benodigde inspanning – het aantal handelingen – benodigd om één uitvaller door beroepsziekte of arbeidsongeschiktheid te voorkomen. Achtereenvolgens berekent hij de prognostische positieve voorspellende waarde (PPV), het attributieve risico en het aantal af te wijzen kandidaten om één risicodrager uit te sluiten. Hieruit blijkt dat veel keuringen moeten worden verricht, altijd met een aantal fout-positieven, om een enkele risicodrager te weren.

Net als in zijn proefschrift voegt De Kort daaraan de nuancering toe dat het gebruik van testmethoden met een lage voorspellende waarde aanvaardbaar kan zijn als het te vermijden risico groot wordt geacht. Als voorbeeld wijst hij naar de keuring van luchtvaarders, waarbij als consequentie veel kandidaten worden afgewezen.

Vermeldenswaardig is ook het proefschrift van Reulings [20]. Zij onderzocht in een serie artikelen [21, 22, 23, 24, 25] in hoeverre medisch disfunctioneren bij actief dienenden te voorspellen viel en in welke mate militaire keuringen reproduceerbaar waren. Zij verschaftte eerst uitleg over het complexe keuringstraject dat voor dienstplichtigen werd gehanteerd [21]. Vervolgens verrichtte zij retrospectief cohortonderzoek onder 8714 dienstplichtigen naar het vermogen van de dienstplichtkeuring (bij 16- en 17-jarigen) als selectie-instrument [22].

Statistisch bleek 6% na opkomst medisch te disfunctioneren, hoofdzakelijk ofwel op grond van aandoeningen aan rug en onderste extremiteiten ofwel op grond van onvoldoende psychische stabiliteit. In alle drie deze situaties bleek een hoge (minder geschikt aanduidende) ABOHZIS-classificatie tijdens de dienstplichtkeuring een goede voorspeller te zijn. In een derde onderzoek [23] werden de afkeuringsgronden en afkeuringspercentages van de destijds verschillende regionale keuringsraden voor de dienstplichtkeuring onderling vergeleken. De aanname daarbij was dat er geen regionale gezondheidsverschillen bestonden en de gevonden percentages tussen de verschillende keuringsraden vergelijkbaar zouden moeten zijn. Dit bleek niet het geval te zijn, hetgeen voor een deel zou worden verklaard door procedureverschillen tussen die keuringsraden, en voor een ander deel zou berusten op grond van verschillen tussen individuele artsen. Hierna werd de reproduceerbaarheid van dienstplichtkeuringen onderzocht met behulp van geanonimiseerde casussen door artsen van verschillende keuringsraden [24]. De gevonden overeenstemming was daarbij laag. Tenslotte werd ditzelfde onderzoek herhaald, maar nu met dossiers van beroepsmilitairen afkomstig uit het periodiek geneeskundig onderzoek [25, zie hoofdstuk 2].

Reulings wees vervolgens op een aantal nadelen van het gebruik van de ABOHZIS-classificatie, zoals het risicowerende karakter ervan en het ontbreken van eisen waartegen de op grond van die classificatie opgelegde beperkingen kunnen worden afgezet.

Omdat het overgrote gedeelte van de dienstplichtigen terecht bleek goedgekeurd, constateerde zij dat de negatieve voorspellende waarde van de dienstplichtkeuring hoog ligt.

Ook buiten artikelen over de kwaliteit van keuringen zijn in de literatuur risicofactoren gevonden die voor de kwaliteit van keuren relevant kunnen zijn. Door Van Deursen, Otten, Smulders, Houtman en Bongers is in de periode van 1997 tot 1999 onderzoek verricht naar de relatie tussen gezondheid en verzuim [26, 27, 28]. Hieruit bleek dat risicofactoren zoals de gezondheidstoestand, waarvan vooraf werd aangenomen dat zij belangrijk waren, feitelijk slechte voorspellers waren, terwijl andere minder verwachte factoren zoals de privé-situatie of verschillen op grond van leeftijd en geslacht nu juist wel relevant bleken te zijn.

Internationaal

Ook internationaal is gezocht naar informatie over de kwaliteit van (aanstelings-) keuringen. In de eerste plaats werden wereldwijd via de militaire attachés gerichte vragen gesteld omtrent de specifieke keuringssituatie bij verschillende krijgsmachten in het buitenland. Naar aanleiding hiervan werd teleurstellend

weinig materiaal ontvangen. In totaal werden slechts 3 reacties ontvangen: uit Japan, Frankrijk en India. Het betreffende materiaal gaf alleen inzicht in de gebruikte testen en apparatuur (vergelijkbaar met die van de KL), zoals voor visus en gehoor, samenhangend met een medisch onderzoek. Standaarden of normen werden, ondanks het verzoek daartoe, niet specifiek vermeld.

Daarnaast is ook in de internationale literatuur gezocht¹. Hierbij werden geen artikelen gevonden die vergelijkbaar waren met het onderzoek uit eigen land. De gevonden artikelen behandelen onderdelen van keuring of selectie en zijn hoofdzakelijk gericht op het meten of vergroten van fysieke prestaties. Veel literatuur bleek uit militaire kring afkomstig te zijn. In die literatuur is ook aandacht voor onderwerpen zoals trainingsopbouw, blessurepreventie en onderzoek naar oorzaken van vroege uitval na indiensttreding.

Ondanks de hierboven geschetste tekortkomingen zitten er toch onderwerpen tussen die relevant genoeg zijn om te vermelden. Er worden belangrijke risicofactoren voor vroegtijdige uitval beschreven. Als onderdeel daarvan wordt het ontstaan en de achtergrond van de fysieke selectie, die ook bij de KL rondom de medische aanstellingskeuring een rol speelt, geschetst. Uiteindelijk kan ook uit de internationale literatuur lering voor verder onderzoek worden getrokken.

Eén van de bij een keuring gebruikelijke instrumenten is de anamnese. Bij veel keuringen wordt een vragenlijst als (onderdeel van de) anamnese gehanteerd. Jacobs [29] onderzocht gelijktijdig de reproduceerbaarheid en validiteit ("gouden standaard": tredmolentest, vitale capaciteit, percentage lichaamsvet) van 10 in de VS gebruikelijke vragenlijsten naar lichamelijke inspanning onder 78 mannen en vrouwen van 20 tot 59 jaar. De betrouwbaarheid gedurende de eerste maand bleek in alle gevallen hoog te zijn, nadien daalde de betrouwbaarheid. De correlatie met de gouden standaard was in alle gevallen laag. Hij adviseerde voorts in zulke vragenlijsten ook navraag te doen naar inspanningen gedurende vrijetijdsbesteding en huishoudelijk werk.

Weller [30] onderzocht de test-hertest betrouwbaarheid van de Canada Fitness Survey questionnaire met de Canadian Aerobic Fitness Test als "gouden standaard" bij 127 onderzoekspersonen (64 mannen en 63 vrouwen) tussen de 15 en 80 jaar. Voor de meeste activiteitsvariabelen lag de correlatie bij mannen hoger ($r_1 = 0,28 - 0,60$) dan bij vrouwen ($r_1 = 0,38 - 0,65$). Zij concludeerde dat

1 Databases: Medline, DIRLine, HealthSTAR, HISTLINE, OLDMEDLINE, POPLINE, SDLINE, OSH-ROM, National Library of Medicine. Zoektermen: validity/predictive value, health examination/(pre-employment medical) assessment occupational (health), physical/psychological selection, attrition.

deze vragenlijst slechts matig betrouwbaar is voor het meten van de meeste lichamelijke activiteiten.

Er is tevens veel internationaal onderzoek verricht naar de effecten van fysieke training. Zowel Vogel [31] als Stacey [32] onderzochten het trainingseffect op rekruten (Britse, respectievelijk Nieuw-Zeelandse) door de VO_2 max met fiets-ergometrie te (her-) meten bij opkomst en aan het eind van de opleiding. Beiden vonden een stijging van de VO_2 max en een daling van het percentage lichaamsvet onder gelijkblijvend lichaamsgewicht en beschouwden dit als een bewijs voor de effectiviteit van het trainingsprogramma.

Voorts is veel onderzoek verricht naar de risicofactoren voor het oplopen van blessures, onder meer door Jones [33, 34], Jordaan [35], Almeida [36] en Pope [37]. De onderzoeksopzet is telkens prospectief van aard: het meten van prestaties bij rekruten direct na opkomst en nogmaals aan het einde van de militaire opleiding, tezamen met het afnemen van een vragenlijst. Alom worden door deze auteurs de volgende risicofactoren geduid: gebrekkige basisconditie, eerdere blessures, te intensieve training(-sprogramma's). Vrouwelijke rekruten zijn meer blessuregevoelig dan mannelijke. De belangrijkste aanbevelingen zijn om maatwerk trainingsprogramma's te hanteren, voldoende aandacht te schenken aan blessurepreventie, maar ook om selectie te plegen op grond van een minimaal vereiste basisconditie.

In de hierboven vermelde studies is de relatie tussen de keuringsbevindingen (aan het bewegingsapparaat) en eventuele uitvalsoorzaken niet onderzocht. Ross [38] onderzocht die relatie wel. Hij verzamelde in een retrospectieve case-control studie (bij Australische rekruten) elke anatomische afwijking aan de onderste extremiteiten uit de keuringsdossiers, niet zijnde van de huid, als variabele. Deze anatomische afwijkingen bleken risicofactoren te vormen. Bij onderzoek naar de fysieke screening op voetafwijkingen bij Australische rekruten concludeerde Rudzki [39] echter een slechte sensitiviteit en specificiteit, evenals een slechte voorspellende waarde, van deze onderzoeksmethode ten aanzien van optredende blessures.

Stevenson [40] verrichtte onderzoek naar de minimaal vereiste basisconditie voor het Canadese leger. Op grond van de gevonden resultaten adviseerde hij een ondergrens voor geschiktheid bij de fysieke test voor alle rekruten tezamen op de vijfde percentiel. Dit betekent echter in de praktijk dat vrouwelijke kandidaten aan hogere eisen moeten voldoen dan mannelijke.

Door een minimum aan conditie te eisen (conditietest) bij aanname wordt het oplopen van blessures teruggedrongen. Maar zo'n fysieke test voorspelt niet of een rekrut met succes de militaire opleiding zal afsluiten. Bij onderzoek van Chin [41] bleek geen statistisch significant verschil te zijn tussen de gemiddelde

VO₂max van de geslaagden en de drop-outs tijdens de opleiding van Amerikaanse luchtmachtrekruten, evenmin als bij een onderlinge vergelijking van het resultaat van de 2-mijlsloop.

Williams [42] evalueerde de doeltreffendheid van het Britse trainingsprogramma voor rekruten. Hij onderzocht fitness, kracht en het vermogen om met militair materieel om te gaan. Het programma blijkt vooral de fitness te verhogen, maar omgang met het militaire materieel en de kracht worden weinig beïnvloed door het programma. Het programma faalt derhalve in het opzicht om rekruten op hun militaire taak voor te bereiden.

Rayson [43] vermeldde in een serie artikelen de totstandkoming van zogenaamde taakcriterium tests (eenmalig tillen, dragen, herhaald tillen en marsen met bepakking) om tot een betere fysieke indeling van Brits militair personeel te komen. De beschreven werkwijze vertoont sterke overeenkomst met de wijze waarop de KL haar personeel indeelt.

Er bestaan nog meer risicofactoren voor het niet behalen van de militaire opleiding dan louter fysieke. In 1990 wist McCraw [44] rekruten, die op grond van aanpassings- en motivatieproblemen de militaire opleiding niet zouden afronden, met behulp van een zelf in te vullen persoonlijkheidstest te onderscheiden van de overige rekruten. Carbone [45] analyseerde 59 variabelen op het gebied van motivatie, biografie en psychologische test-uitkomsten onder 632 Amerikaanse Luchtmachtrekruten, waarvan de ene helft de opleiding afrondde en de andere helft niet. Hij concludeerde dat met behulp van een klein aantal van die variabelen tot op 70% nauwkeurig kan worden voorspeld of iemand de opleiding zal halen. Optimisme en motivatie ("Ik ga die opleiding halen", "Ik vind het leuk om nieuwe dingen te leren"), de hoeveelheid spreekuurcontacten met de behandelend militair arts en lichamelijk misbruik in de anamnese hebben in dat model de grootste bijdrage. Ook het onderzoek van Crawford [46] toont het bestaan van misbruik in de anamnese aan als risicofactor voor uitval tijdens de militaire opleiding in de VS.

Interpretatie van de literatuur voor verder onderzoek

De teneur in de onderzoeken van De Kort is dat de aanstellingskeuring ontoereikend is voor het gestelde doel. Wel wordt door hem erkend [17, 19] dat de noodzaak voor een keuring meer gerechtvaardigd zou kunnen zijn indien het te vermijden risico groot wordt geacht.

Uit de artikelen van De Kort komen een aantal argumenten naar voren die nut en noodzaak van aanstellingskeuringen ter discussie stellen. Het aantal afkeuringen is vaak laag [16, 18]. Hij signaleert voorts een ontbrekende relatie tussen medische bijzonderheden en uitval op het werk [17]. Daarnaast rekent hij het gebrek aan reproduceerbaarheid voor [19]. Hij toont aan hoeveel onnodig afge-

keurden, die theoretisch wel over de juiste kwaliteiten zouden beschikken, alsnog optreden[19].

De enige kanttekening die men bij zijn onderzoek zou kunnen plaatsen is dat hij zich in diverse artikelen [15, 16, 18] hoofdzakelijk richt op het beperkte effect van een aanstellingskeuring op het terugdringen van arbeidsgebonden aandoeningen en vroeg optredende arbeidsongeschiktheid. Klein Hesselink [47] vermeldt vele factoren (onderverdeeld in persoonsgebonden, bedrijfsgebonden en maatschappelijke oorzaken) die uitval door ziekteverzuim kunnen beïnvloeden. Deze factoren kunnen onderzoek naar het effect van de aanstellingskeuring en op uitval door ziekmelding of arbeidsongeschiktheid beïnvloeden. Bovendien behoeft het in mindere mate voldoen aan de gestelde (keurings-)eisen niet altijd tot uitval op grond van beroepsziekten of arbeidsongeschiktheid te leiden. Er zijn ook andere uitingen van het minder aan de gestelde eisen voldoen denkbaar, zoals het vertonen van een lagere productienorm, het vertonen van herhaald kortdurend verzuim of alleen maar het uiten van een lagere arbeidssatisfactie tijdens het inlooppreekuur van de bedrijfsarts.

De conclusies van het onderzoek van Reulings [20] sluiten in meerdere opzichten aan bij die van De Kort. Ook zij vond een laag percentage afgekeurden [22], en uit haar onderzoeken bleek evenzeer de reproduceerbaarheid van de betreffende keuring gering te zijn [23, 24, 25]. Ook bij haar onderzoek vallen echter enige kanttekeningen te maken. Haar onderzoek was hoofdzakelijk gericht op dienstplichtigen, waarvan sinds 1996 geen sprake meer is. Van dienstplichtigen is in paragraaf 1.3.1 een aantal factoren vermeld, die van invloed kunnen zijn op de uitslag van de keuring, zoals simulatie [5, 6, 7, 8, 9] en risicowering [10]. Deze factoren kunnen het aantal afkeuringen onnodig hebben verhoogd en de reproduceerbaarheid negatief hebben beïnvloed. Zij constateerde voorts dat een hoge ABOHZIS-classificatie – als maatstaf voor een serieus afwijkende keuringsbevinding – een goede voorspeller zou zijn voor uitval als dienstplichtige [22]. Tegelijkertijd constateerde zij echter ook dat de reproduceerbaarheid waarmee die classificaties worden gehanteerd onder de vereiste norm blijft [24, 25]. Dat laatste betekent dat de validiteit van het gehanteerde keuringssysteem in elk geval volgens Bouter en Van Dongen[1] onvoldoende moet zijn geweest.

In de internationale literatuur komen de volgende feiten aan bod. Vragenlijsten naar de mogelijkheden voor fysieke inspanning blijken alleen op korte termijn betrouwbaar [29]. Op grond van fysieke risicofactoren voor uitval tijdens de militaire opleiding [33, 34, 35, 36] (geslacht, gebrekkige basisconditie, eerdere blessures) wordt een aantal preventieve maatregelen geadviseerd, waaronder de

vaststelling van een minimaal vereiste basisconditie bij aanname [40, 43]. Naast oude of nieuwe blessures spelen motivatie, artsenbezoek en misbruik in de anamnese een rol bij uitval.

Opmerkelijk genoeg kan de voorspellende waarde van keuringsbevindingen aan het bewegingsapparaat in relatie tot uitval niet worden aangetoond in zowel een onderzoek van De Kort [17] als dat van Rudzki [39].

Ook bij de internationale literatuur moeten kanttekeningen worden geplaatst. Een eerste kanttekening is dat al het militaire onderzoek niet verder reikt dan de korte opleidingsfase. Langere termijn effecten zijn niet onderzocht. Een tweede kanttekening is dat de omstandigheden waaronder en de populatie waarop onderzoek is verricht, niet per definitie vergelijkbaar zijn met de situatie in eigen land. Ter illustratie hiervan wordt een vergelijking getrokken tussen Amerikaanse luchtmachtrekruten onderzocht door Talcott [48] en KL-militairen over 1999, dit laatste gebaseerd op interne onderzoeksgegevens, naar aard en hoeveelheid uitvallers tijdens de militaire opleiding. Hieruit blijkt dat zowel de hoeveelheid uitvallers als de oorzaken voor die uitval onderling volstrekt onvergelijkbaar zijn. In de VS valt 6,2% van de rekruten uit tegenover 25,6% bij de KL. De voornaamste oorzaken voor de uitval in de VS zijn medisch, psychisch en disciplinair, terwijl dit in de KL voornamelijk bedrijfsgebonden oorzaken, de thuisomstandigheden en heimwee zijn. Ter verklaring kan aan meerdere mogelijke oorzaken worden gedacht: een verschil in selectie of opleiding tussen beide organisaties, een verschil in motivatie en perceptie tussen de rekruten, verschillen tussen beide samenlevingen of combinaties van deze factoren. Met de interpretatie van internationale onderzoeksgegevens naar de situatie in eigen land moet in elk geval terughoudendheid worden betracht.

Op grond van het bovenstaande kunnen de volgende notities voor gebruik bij verder onderzoek worden gemaakt:

- Om het gevaar voor derden te verkleinen kan een organisatie gerechtvaardigde redenen hebben om scherpe keuringscriteria te hanteren.
- Het hanteren van een minimaal vereiste basisconditie moet in theorie een verlagend effect hebben op uitval door nieuw optredende blessures.
- Behoudens blessures in de anamnese is er geen duidelijk verband aangetoond tussen medische bijzonderheden en uitval na tewerkstelling; dat verband is wel (zij het onder buitenlandse militairen) vastgesteld voor de mate waarin men zelf een taak (bijvoorbeeld opleiding) meent aan te kunnen, eventueel misbruik als kind en de hoeveelheid artsenbezoek.
- Telkens is alleen de relatie tussen keuring en uitval onderzocht. Het in mindere mate voldoen aan de gestelde (keurings-)eisen behoeft niet altijd te leiden tot uitval op grond van beroepsziekten of arbeidsongeschiktheid. Er

zijn nog andere uitingen van het minder aan de gestelde eisen voldoen denkbaar dan alleen uitval door ziekte.

- In geen enkel Nederlands onderzoek is de beleving of motivatie van de betrokken werknemers meegenomen. In het algemeen wordt volstaan met het keuringsdossier en de verzuimgegevens of andere data. Broersen, Weel en Van Dijk [49] stellen dat de beleving van gezondheidsbelastende factoren mede bepaalt hoe het effect ervan op de werknemer zal zijn. Zich gezond of ziek voelen heeft niet alleen te maken met de objectieve gezondheidsrisico's, maar ook met de wijze waarop de werknemer die risico's zelf ervaart. De persoonlijke beleving kan mede de onderzoeksuitkomsten (ziekmelding, uitval) beïnvloeden. Het gebruik van bijvoorbeeld een periodieke vragenlijst zou op grond hiervan een toegevoegde waarde kunnen hebben bij verder onderzoek.
- Als in zo'n vragenlijst bovendien ook nog de te stellen eisen zouden worden opgenomen, zou op die manier van de werknemer zelf kunnen worden vernomen of deze zichzelf daartoe in staat acht.
- Als naast motivatie ook de mate van artsbezoek een rol bij uitval speelt, is het zinvol om bij verder onderzoek ook daaraan aandacht te besteden.

1.3.3 De Wet op de Medische Keuringen (WMK) en de praktische uitvoering ervan

Het onderzoek van De Kort, Lourijsen, Post-Uiterweer en Van Dijk [2, 15, 16, 17, 18, 19] was één van de factoren die een rol speelden bij de totstandkoming van nieuwe wetgeving op het gebied van (aanstellings-) keuringen in Nederland: de Wet op de Medische Keuringen (WMK, 1998). Centraal in deze wet staat dat er een specifiek doel moet bestaan voor de uitvoering van een aanstellingskeuring: het noodzakelijkheidscriterium (Artikel 4). Aanstellingskeuringen mogen alleen plaatsvinden als er bijzondere eisen aan de medische geschiktheid moeten worden gesteld, op basis van risico's in de betreffende functie. Bovendien moeten keuringen worden beperkt naar hun aard, inhoud en omvang tot het doel waarvoor zij worden verricht (Artikel 2 lid 1). Tevens mogen bij een keuring geen vragen worden gesteld en geen medische onderzoeken worden verricht die een onevenredige inbreuk betekenen op de persoonlijke levenssfeer van de keurling (Artikel 3 lid 1).

Op deze wetgeving werd een uitzondering gemaakt middels de Wijziging van de Wet op de Medische Keuringen (1998). Die uitzondering geldt de specifieke situatie voor militair personeel en wordt in de Memorie van Toelichting van deze wet nader verklaard. Ten aanzien van het functiegericht keuren, zoals bedoeld in Artikel 4 lid 1 van de Wet op de Medische Keuringen, wordt opgemerkt dat de militaire ambtenaar niet is aangesteld in een *functie*, maar als *militair*.

Deze militair dient alle in het belang van de taakuitoefening van de krijgsmacht opgedragen werkzaamheden en diensten te kunnen verrichten. Daarbij kan de militair overal ter wereld onder allerlei omstandigheden worden ingezet. Zowel deze ruim geformuleerde ("alle") taakstelling als de uiteenlopende mogelijke arbeidsomstandigheden ("overal ter wereld onder allerlei omstandigheden") maken dat aan de militair zware fysieke en geestelijke eisen moeten worden gesteld. Om nu de risico's bij die taakuitoefening voor de militair en voor derden zo gering mogelijk te houden worden aan elke militair (onder meer medische) eisen gesteld die betrekking hebben op diens totale operationele inzetbaarheid.

In feite stelt de wetgever hiermee dat het scala van zich mogelijk voordoende taken en arbeidsomstandigheden voor een militair zo groot is, dat bij de aanstelling met vele uiteenlopende eisen rekening moet worden gehouden.

Naast het stellen van eisen voor de uitvoering van een keuring was in de Wet op de Medische Keuringen ook de bepaling opgenomen dat door de wetgever nadere regels zouden worden gesteld inzake de te volgen procedure voor bezwaar en bezwaar, indien daarover door de sociale partners geen consensus zou worden bereikt. Het uitblijven van deze consensus leidde in 2001 vervolgens tot nadere regelgeving, vastgelegd in de Staatsbladen 597 en 598. Op grond hiervan is een onafhankelijke klachtencommissie ingesteld waar keurlingen hun bezwaar kunnen doen over de aard, inhoud en procesgang van en bij de keuring (Staatsblad 598). Daarnaast worden de in de Wet op de Medische Keuringen vermelde toetsingscriteria voor het verrichten van een aanstellingskeuring scherper omschreven (Staatsblad 597). Keuringen mogen alleen worden verricht indien aan de betreffende functie bijzondere eisen op het punt van medische geschiktheid worden gesteld en dan alleen nog indien de (te benoemen) risico's niet met gangbare maatregelen kunnen worden gereduceerd. De keuringvrager moet de te stellen eisen schriftelijk vastleggen en inzake de rechtmatigheid ervan advies vragen bij een Arbo-dienst.

Voor het verder operationaliseren van aanstellingskeuringen door Arbo-diensten is door Hulshof [50] in opdracht van de Branche Organisatie Arbo-diensten de Algemene Richtlijn Aanstellingskeuringen (ARA) ontwikkeld. De richtlijn bevat een stappenplan, alsmede een overzicht van functie-eisen met bijbehorende gezondheidscriteria en onderzoeksmethoden.

Het stappenplan bevat achtereenvolgens de volgende vijf stappen:

- Het informeren van de klant.
- Het opstellen van functie-eisen. Hierbij wordt een model verstrekt aan de hand waarvan kan worden bepaald voor welke functie-eisen een bijzonder beroep op de medische geschiktheid wordt gedaan. De ARA bevat een

beoordelingsformulier waarmee aan de hand van 18 veel voorkomende functie-eisen kan worden nagegaan of er sprake is van bijzondere eisen ten aanzien van de medische geschiktheid voor de functie. Dit formulier geeft een globale klassering in termen variërend van afwezig (klasse 1) tot zeer zwaar (klasse 5). Essentieel is dat het om risicofactoren moet gaan die de werkgever niet redelijkerwijs met preventieve of andere maatregelen kan terugdringen.

- Het maken van de keuze of er al dan niet een aanstellingskeuring moet plaatsvinden.
- Het vaststellen van de inhoud van de aanstellingskeuring.
- Het uitvoeren van de aanstellingskeuring.

Bij het opstellen van de eisen en het bepalen van het al dan niet uitvoeren van een aanstellingskeuring moeten volgens dit plan de volgende vragen worden beantwoord:

- Zijn er functie-eisen met een ongewoon hoge belasting?
 - Is die belasting niet met gangbare maatregelen te reduceren?
 - Is er een kans op gezondheidsschade?
 - Zijn de eisen te vertalen in gezondheidscriteria?
- Is medisch onderzoek mogelijk?
 - Zijn er normen?
 - Vormt de gehanteerde methode een valide indicator?
 - Staat het te verrichten onderzoek in verhouding tot het doel van de keuring (het proportionaliteitscriterium)?
 - Bestaat er geen minder belastende of ingrijpende onderzoeksmethode (subsidiariteitscriterium)?
 - Leidt de onderzoeksmethode niet tot ongerechtvaardigde differentiatie en/of discriminatie van groepen in de samenleving?
 - Is het te verrichten onderzoek relevant voor het doel van de keuring?
- Is keuren toegestaan?
 - Zijn de eisen door een bedrijfsarts vastgesteld?

De ARA is geen juridisch bindende richtlijn; men mag er beargumenteerd van afwijken. Daaraan moet echter worden toegevoegd dat in de Nota van Toelichting van de Staatsbladen 597 en 598 wordt vermeld dat de ARA een eerste poging tot ordening in de praktijk is en door de klachtencommissie kan worden gehanteerd als toetsingskader voor wat in de beroepsgroep gebruikelijk is. Dit maakt de ARA allerminst tot een vrijblijvend advies.

De tot nu toe beschreven niet-systematische zowel als de systematische waarnemingen vormden aanleiding voor de vraag of het bestaande keuringssysteem niet aan vervanging toe was.

1.4 Een nieuw keuringssysteem en de relatie ervan met wet- en regelgeving

1.4.1 De ontwikkeling van een nieuw keuringssysteem

De Wet op de Medische Keuringen staat uitvoering van aanstellingskeuringen alleen toe indien aan de medische geschiktheid bijzondere eisen moeten worden gesteld. Weliswaar wordt in de Wijziging van de Wet op de Medische Keuringen aangegeven dat bij het aanstellen van militair ambtenaren met vele mogelijke eisen rekening moet worden gehouden, maar dat laat onverlet dat er eisen moeten worden gesteld.

Bij het tot dan toe in gebruik zijnde diagnosegerichte keuringssysteem bij de KL was daarvan geen sprake. Afwijzing of ontslag vonden plaats op grond van een zekere mate van ziekte of gebrek, eisen waaraan de militair diende te voldoen waren niet geformuleerd. Er was geen sprake van een relatie tussen belasting en belastbaarheid [13]. Mede omdat deze situatie voorts in de praktijk tot rechtspositionele en juridische problemen leidde (een ontslagaanvraag omdat een militair op grond van ziekten en gebreken niet operationeel inzetbaar werd geacht, werd afgewezen door de ambtenarenrechter), gaf de KL-leiding in 1997 opdracht om functie-eisen voor militairen te laten ontwikkelen en deze op medisch gebied te laten operationaliseren [51].

Gelet op de bijzondere positie van de militair, zoals in de vorige paragraaf weergegeven, werd besloten om voor alle militairen in beginsel dezelfde algemene en brede functie-eisen te definiëren. In de loop van 1997 werden de eisen, minimaal te stellen aan KL-militairen, geformuleerd op grond van de concrete activiteiten van militairen. Die eisen strekken zich uit over de volle breedte van het functioneren als militair ambtenaar, inclusief de mogelijkheid tot wereldwijde operationele inzet. Deze eisen werden de Militaire basiseisen KL genoemd.

Om deze functie-eisen te operationaliseren is gebruik gemaakt van twee bronnen. In de eerste plaats werd gebruik gemaakt van het belastbaarheidspatroon van het Functie Informatie Systeem (FIS)². Dit FIS omvatte 28 beoorde-

2 Het belastbaarheidspatroon van het FIS is een beoordelingssysteem waarbij in de vorige eeuw voor de Nederlandse marktsector de belasting van bestaande arbeid in kaart is gebracht door middel van 28 beoordelingspunten. Bij veel van deze beoordelingspunten kan gedifferentieerd worden tussen verschillende prestatieniveaus.

lingspunten op fysiek en mentaal terrein van handelingen en gemoedstoestanden, die alle relevant bleken (omdat zij zich veelvuldig voordoen) voor het operationaliseren van de werkgeveisen. Maar hiermee werd het complete scala aan militaire werkgeveisen niet afgedekt. Zo wordt aan het kunnen schieten met een vuurwapen in het FIS geen aandacht geschonken, evenmin als de handeling graven erin voorkomt. Daarom werd daarnaast gebruik gemaakt van beoordelingspunten die specifieke militaire vaardigheden betreffen, op dezelfde wijze als het FIS ontwikkeld door sportartsen en bewegingswetenschappers van de Arbo-dienst KL. Dit leidde in december 1997 tot het opstellen van een lijst met 43 beoordelingspunten. Elk beoordelingspunt werd voorzien van een grenswaarde: de Basismedische Eis KL (BMEKL). Ter illustratie vermeldt beoordelingspunt 15 (dragen) als basiseis “het kunnen dragen van 25 kg gedurende 2 uur per werkdag”, en heeft beoordelingspunt 32 (springen) als basiseis “het van en/of over 1 m kunnen springen”.

Anders dan bij de WAO-keuring waarvoor het FIS tot december 2001 werd gebruikt, worden de BMEKL gehanteerd op een dichotome wijze: de gekeurde wordt beschouwd als zijnde wel of niet beperkt (geschikt of ongeschikt) voor het vervullen van de basale militaire functie – het gaat om de minimale geschiktheid, met inbegrip van operationele inzetbaarheid. Een overzicht van de 43 beoordelingspunten en basismedische eisen wordt in tabel 1 weergegeven.

Centraal bij het hanteren van de BMEKL staat de vraag of er bij de keuring sprake is van medische bijzonderheden. Onder een medische bijzonderheid wordt elke afwijkende medische bevinding verstaan, die in het kader van de keuring wordt geconstateerd.

Voor het onderzoek naar deze medische bijzonderheden zijn tevens keuringsprotocollen ontwikkeld, die een brug slaan tussen de keuringsgegevens en de consequenties die daaraan worden verbonden in het kader van de basismedische eisen [51]. Bij het ontwikkelen van deze keuringsprotocollen is ervoor gekozen om het onderzoek naar de meest voorkomende medische bijzonderheden bij militaire keuringen te protocolleren. Aldus werd één algemeen onderzoeksprotocol ontwikkeld, tezamen met 25 specifieke, op bepaalde lichaamsdelen, respectievelijk op specifieke aandachtsgebieden (zoals het gebruik van geneesmiddelen) en op tracti (zoals digestivus, urogenitalis) gerichte protocollen. Bij elk van deze protocollen is aangegeven wat standaard zou moeten worden gevraagd, onderzocht, gemeten en op welke wijze hieraan consequenties moeten worden verbonden.

Bij de keuring wordt standaard het algemene protocol (nummer 1) gehanteerd. In geval op het terrein van één van de specifieke keuringsprotocollen een medi-

Tabel 1. Een overzicht van de beoordelingspunten en Basismedische Eisen

Nr	Beoordelingspunt	Basismedische Eis
1	Zitten	Zitten gedurende vrijwel de gehele werkdag 2 uur aaneengesloten
2	Staan	Staan gedurende vrijwel de gehele werkdag 2 uur aaneengesloten
3	Lopen	Lopen gedurende vrijwel de gehele werkdag 1 uur aaneengesloten
4	Trappenlopen	Trappenlopen 8 keer per uur 25 treden
5	Klimmen en klauteren	Klimmen of klauteren 8 keer per uur 5 meter
6	Knielen, kruipen en hurken	Knielen, kruipen, hurken gedurende 4 uur per werkdag een half uur aaneengesloten
7	Gebogen werken	Gebogen werken gedurende vrijwel de gehele werkdag 1 kwartier aaneengesloten
8	Kortcyclisch buigen en torderen	Buigen of torderen 150 keer per uur in 90°
9	Gebruik van de nek	Gebruik van de nek is niet beperkt
10	Reiken	Reiken 500 keer per uur 80 cm
11	Bovenhands werken	Werken boven schouderhoogte 4 uur per werkdag een kwartier aaneengesloten
12	Hand- en vingergebruik	Hand- en vingergebruik is niet beperkt
13	Tillen	Tillen 150 keer per uur 20 kg
14	Duwen en trekken	Geen beperkingen ten aanzien van duwen en trekken
15	Dragen	Dragen gedurende 2 uur per werkdag 25 kg
16	Luchtverplaatsing	Blootstelling aan luchtverplaatsing is niet beperkt
17	Stof, rook, gas en damp	Blootstelling aan stof, rook, gas en damp is niet beperkt
18	Koude	Blootstelling is mogelijk langer dan 5 minuten bij temperaturen van -20°C
19	Hitte	Blootstelling is mogelijk langer dan 5 minuten bij temperaturen van + 40°C
20	Temperatuurwisselingen	Meer dan één temperatuurwisseling van 60°C binnen 5 minuten
21	Luchtvochtigheid	Blootstelling langer dan 4 uur aaneengesloten aan een relatieve luchtvochtigheid van > 90% is niet beperkt
22	Droge lucht	Blootstelling langer dan 4 uur aaneengesloten aan een relatieve luchtvochtigheid van <35% is niet beperkt
23	Huidcontact	Huidcontact met vaste en vloeibare stoffen is niet beperkt
24	Vibratiebelasting	Blootstelling aan trillingen (incl. transport) is niet beperkt
25H	Zintuigen-gehoor	Beiderzijds gehoorverlies < 30 dB t/m 3 kHz / < 65 db tussen 3-8 kHz
25Z	Zintuigen-zien	Amblyopie mc > 0.3 andere oog > 0.8; binoculair zien mc > 0.8
25K	Zintuigen-kleuronderscheiden	Ishihara > 10/16; daaronder verder onderzoek m.b.v. de Farnsworth D15 test
26	Beschermende middelen	Dragen van beschermende middelen is niet beperkt
27	Persoonlijke risico	Blootstelling aan persoonlijk risico is niet beperkt (zoals bij epilepsie wel het geval is)
28	Psychische belasting	De psychische belasting is niet beperkt (nadere specificatie subeisen van a-q) (Zie tabel 14)
29	Algemene conditie	Voldoende conditie

Nr	Beoordelingspunt	Basismedische Eis
30	Lopen met bepakking	Kan incidenteel 5 km met GVT1 (eerstelijns militaire bepakking) in één uur afleggen
31	Graven	Kan incidenteel een kubieke meter in twee uur graven
32	Springen	Kan van / over één m springen
33	Optreden te velde	Optreden te velde is op basaal niveau mogelijk
34	NBC (nucleaire, chemische en biologische protectie)	Kan incidenteel nbc-beschermende middelen gedurende één uur dragen
35	Tropen	Kan in de tropen functioneren
36	Spreekvaardigheid	Spreekvaardigheid is niet beperkt
37	Gesaneerd gebit	Dental fitness Klasse 2 (uitzendbaar zonder behandeling gedurende 6 maanden)
38	Scherfwerend vest dragen	Kan incidenteel gedurende 8 uur een scherfwerend vest dragen
39	Helm dragen	Kan incidenteel de helm gedurende 8 uur dragen
40	Hanteren vuurwapen	Kan een handvuurwapen hanteren
41	Onregelmatig functioneren	Kan incidenteel op onregelmatige tijden functioneren
42	Verdragen Klasse I (voeding)	Kan door de KL verstrekte Klasse I verdragen
43	Medicatie-afhankelijkheid	Er is geen sprake van medicatie-afhankelijkheid

sche bijzonderheid wordt verondersteld, dan vormt die veronderstelling aanleiding om vervolgens ook het betreffende specifieke protocol bij die keuring te gaan hanteren. De consequenties van de bevindingen dienen dan vervolgens conform het gestelde in de basismedische eisen te worden afgehandeld. Alle protocollen zijn opgebouwd uit dezelfde onderwerpen: anamnese, lichamelijk onderzoek, biometrische gegevens (zoals bijvoorbeeld de bloeddruk of hartfrequentie), redenen voor verwijzing (naar specialismen of voor verder onderzoek), en in welke situaties het inwinnen van inlichtingen geïndiceerd is.

Deze keuringsprotocollen hebben geen dwingend karakter, maar vormen een methodiek om geen essentiële vragen te vergeten, geen specifiek onderzoek achterwege te laten en geen stappen in het proces van oordeelsvorming over te slaan. De individuele keurend arts behoudt de vrijheid om van het in het keuringsprotocol voorgestelde oordeel af te wijken, indien dit zorgvuldig en toetsbaar gemotiveerd wordt. Dit betekent dat de arts ten opzichte van het vorige systeem meer autonomie heeft, maar tegelijkertijd over een instrument beschikt om systematisch te werken. Een overzicht van de keuringsprotocollen wordt weergegeven in tabel 2.

Voor het ontwikkelen van zowel keuringsprotocollen als basismedische eisen is een werkverband opgericht, waarin vertegenwoordigers van bedrijfsartsen, verzekeringsgeneeskundigen en sportartsen uit de krijgsmacht zitting hadden. De normstelling van de Basismedische Eisen vond plaats op grond van de erva-

ring van leden (empirie). Die ervaring is waar mogelijk aangevuld met gegevens uit de vakliteratuur. Zo doende zou gesteld kunnen worden dat het nieuwe keuringsinstrument aan de vereisten van "face validity" [1] voldoet. Alle keuringsprotocollen zijn ter beoordeling voorgelegd aan op betreffend terrein deskundig geachte klinisch specialisten. Hierdoor wordt aannemelijk dat bij het nieuwe keuringssysteem ook sprake is van "expert validity" [1].

Met behulp van de keuringsprotocollen wordt het onderzoek naar eventuele medische bijzonderheden uitgevoerd. Vervolgens moet worden vastgesteld of een gevonden medische bijzonderheid leidt tot beperkingen in het functioneren als militair. Dit vindt plaats door toetsing van de medische bijzonderheid aan de gestelde basismedische eisen. Indien dit tot de constatering van één of meer beperkingen leidt, dan is er sprake van ongeschikt verklaring. In figuur 1 wordt de te hanteren werkwijze schematisch weergegeven.

Welke verschillen bestaan er uiteindelijk tussen het nieuwe en het oude keuringssysteem? Het oude keuringssysteem ("ABOHZIS") was diagnosegericht en verplichtte de keurend arts om ziekten en gebreken te determineren, waaruit een indelingadvies volgde. Bekend was dat dit systeem geen relatie met de feitelijke militaire werkzaamheden vertoonde [14]. Er bestonden geen functie-eisen, noch werd rekening gehouden met specifieke militaire vaardigheden. In de loop van de tijd verdween de medische aandacht voor de psychische belastbaarheid. De arts beoordeelde of er sprake was ziekten of gebreken, aan de hand waarvan men werd goed – of afgekeurd, ongeacht of men aan de bepaalde functie-eisen voldeed (deze ontbraken immers).

Het nieuwe keuringssysteem ("BMEKL") is functiegericht en volgens het model belasting-belastbaarheid [13] afgeleid van de concrete eisen, te stellen aan militairen [51]. Deze eisen omvatten het complete scala aan taken met inbegrip van operationele inzet, rekening houdend met specifieke militaire vaardigheden en de mate van psychische belastbaarheid. Hierbij wordt gekeken of de gevonden medische bijzonderheden tot beperkingen leiden in het minimaal vereiste operationele functioneringsniveau. Is dit het geval dan wordt men ongeschikt bevonden. Daarbij kan nog worden aangetekend dat de eisen van het BMEKL-systeem objectiveerbaar zijn. In geval een kandidaat meent dat de uitslag niet klopt, dan kan deze in de gelegenheid worden gesteld om het gelijk aan te tonen. Ter illustratie daarvan komt het voor dat iemand bij de pensioenkeuring wordt afgekeurd (zowel het oude als het nieuwe keuringssysteem werd zowel voor de aanstellingskeuring als voor de pensioenkeuring gebruikt) op grond van het niet langer een mars met bepakking te kunnen volbrengen. In geval deze persoon de uitspraak betwist, wordt overgegaan tot het aanbieden van een meerdaags programma waarin dat vermogen in de praktijk wordt getest.

Tabel 2. De ontwikkelde keuringsprotocollen

Nr	Inhoud
1	Standaard-anamnese, lichamelijk onderzoek en biometrische gegevens
2	De beoordeling van het gebruik van geneesmiddelen
3	Het onderzoek van de voeten
4	Het onderzoek van de enkels
5	Het onderzoek van de onderbenen
6	Het onderzoek van de knieën
7	Het onderzoek van de bovenbenen en de liezen
8	Het onderzoek van de heupen
9	Het onderzoek van wervelkolom, nek en Sacro-Iliacale gewrichten
10	Het onderzoek van de schoudergordels
11	Het onderzoek van de bovenarmen
12	Het onderzoek van de ellebogen
13	Het onderzoek van de onderarmen
14	Het onderzoek van de polsen
15	Het onderzoek van de handen en vingers
16	Het neurologisch onderzoek
17	Het onderzoek van de luchtwegen
18	Het onderzoek van het hart en de bloedsomloop
19	Het onderzoek van de tractus digestivus
20	Het onderzoek van de tractus urogenitalis
21	Het dermatologisch onderzoek
22	Het onderzoek van het gehoor
23	Het onderzoek van de visus
24	De beoordeling van endocrinologische aandoeningen
25	Het psychiatrisch onderzoek
26	Het onderzoek van het bloed en de bloedvormende organen

Samengevat: kon de keurend arts vroeger na de vaststelling van een bepaalde ziekte of gebrek in een voorschrift opzoeken of daarmee de kandidaat wel of niet geschikt was, thans moet de arts met behulp van protocollen de onderzoeksbevindingen toetsen aan de functie-eisen en vervolgens afwegen of dat tot beperkingen ten aanzien van de gestelde eisen leidt of niet. In tabel 3 worden de belangrijkste verschillen op een rij gezet. Daarmee wordt een voorschot genomen op de vraag of het nieuwe keuringssysteem al dan niet voldoet aan de

Figuur 1: Beslissingsschema bij het nieuwe (BMEKL) keuringssysteem.

eisen, gesteld door de WMK. Aan het beantwoorden van deze laatste vraag wordt in de volgende paragraaf aandacht geschonken.

1.4.2 Voldoet het nieuwe systeem aan de WMK en de ARA?

Na de ingebruikneming van dit nieuwe systeem diende te worden beoordeeld of nu voldaan was aan het in de WMK gestelde. Bovendien was aansluitend op de WMK inmiddels ook Algemene Richtlijn Aanstellingskeuringen (ARA) ontwikkeld [50], terwijl een aantal bepalingen uit de WMK nog extra werden belicht door de publicatie in 2001 van de Staatsbladen 597 en 598. In Staatsblad 597 worden nogmaals alle relevante toetsingscriteria opgesomd, waaraan bij het verrichten van een aanstellingskeuring moet worden voldaan. In Staatsblad 598 wordt een onafhankelijke klachtencommissie ingesteld.

Tabel 3. De verschillen tussen het oude ("ABOHZIS") en het nieuwe ("BMEKL") keuringssysteem.

	ABOHZIS	BMEKL
Model	Ziekten en gebreken	Belasting-belastbaarheid
Functie-eisen	Nee	Ja
Relatie met taakstelling	Nee	Ja, daarvan afgeleid
Specifieke militaire vaardigheden	Nee, diagnosegericht	Ja
Psychische belastbaarheid	Niet of nauwelijks	Ja
Complete afdekking	Nee	Ja
Beoordelingskader	Afwezig	Aanwezig
Geprotocolleerd	Nee	Ja
Objectiveerbaar	Afhankelijk van oordeel arts	Ja, praktisch alle eisen
Voldoet aan WMK	Nee	Ja

Teneinde de BMEKL te toetsen aan WMK en ARA worden hieronder de vragen, zoals vermeld in het stappenplan van de ARA, achtereenvolgens beantwoord.

1. Zijn er functie-eisen met ongewoon hoge belasting?

Gelet op het in de documentatie van de basismedische eisen en de keuringsprotocollen vermelde, mag dit bevestigend worden beantwoord. Ook de werkgever geeft, zoals reeds eerder is opgemerkt, te kennen dat de samengestelde functie van de militair ambtenaar "zware fysieke en geestelijke eisen" met zich meebrengt.

Op grond van die "zware fysieke en geestelijke eisen" zijn door de werkgever – met behulp van deskundigen – in de loop van 1997 de "Militaire Basiseisen KL" opgesteld. Deze basiseisen geven een compleet overzicht van de concrete belasting die zich in het functioneren als militair ambtenaar minimaal voordoen. Deze basiseisen zijn vervolgens op medisch gebied geoperationaliseerd.

Vervolgens resteert de vraag tot welke ARA-klassering de bijzondere eisen ten aanzien van de medische geschiktheid voor de functie behoren. Dit kan variëren van afwezig (klasse 1) tot zeer zwaar (klasse 5). Gelet op het over de eisen gestelde in de Memorie van Toelichting bij de Wijziging op de Wet op de Medische Keuringen alsmede de concrete operationele taakstelling, kan worden aangenomen dat de eisen voor de militair ambtenaar een aanstellingskeuring rechtvaardigen en dat de eisen globaal op het niveau 5 van de ARA liggen. Bovendien zijn voor de militair ambtenaar nog meer eisen te formuleren dan die in de ARA als voorbeeld zijn bijgeleverd (zoals het moeten kunnen

hanteren van een vuurwapen), waarvoor basismedische eisen als bewijs kunnen gelden.

Is die belasting niet met gangbare maatregelen te reduceren?

Er is thans een Convenant Fysieke Belasting afgesloten tussen de Ministeries van Sociale Zaken en Werkgelegenheid en Defensie. In dat kader wordt onderzoek bij de krijgsmacht verricht naar mogelijkheden om op het terrein van fysieke belasting verbeteringen door te voeren. Hierbij kan worden gedacht aan tilhulpmiddelen en een lichtere standaarduitrusting. Tevens wordt onderzoek verricht naar de gebruikelijke training zoals deze bij opleiding en oefeningen voorkomt, met als doel om blessures te reduceren. Verder kan de operationele noodzaak voor met name de fysieke belasting slechts als een voldongen feit worden beschouwd: de bepakking van een infanterist weegt nu eenmaal veel, evenals het huidige scherfwerend vest. Ook naar de psychische belasting wordt veel onderzoek verricht; hieruit blijkt dat in de praktijk na elke uitzending een aantal militairen geestelijke schade op heeft opgelopen.

Is er een risico op gezondheidsschade?

De verschillende eisen (zoals diverse vormen van lichamelijke inspanning, schieten, tactisch optreden te velde) kunnen samengevat tot hetzelfde risico leiden: uitval bij de eenheid tijdens operationeel optreden, waardoor men zichzelf en/of de eenheid in gevaar brengt. Dit risico kan onder meer worden uitgesplitst naar: bedrijfsongevallen (zwaar en gevaarlijk materiaal), verkeersongevallen (verplaatsingen en tactisch optreden in oneffen terrein onder moeilijke omstandigheden), blessures (die de eenheid ophouden), het gebruik van vuurwapens, de kans op gijzeling, het plegen van gevechtshandelingen, het gebruik van beschermende middelen zoals helmen en scherfwerende vesten of het dragen van luchtmaskers plus beschermende kleding tegen chemische aanvallen.

Deze risico's worden klassiek geformuleerd als: schade voor de eigen gezondheid en veiligheid of voor die van derden. Om deze risico's te elimineren dient een militair op tekortkomingen die tot deze uitval kunnen leiden te worden beoordeeld, in relatie tot de operationele taakstelling.

Subconclusie: Op grond van het bovenstaande wordt aan het in Artikel 4 van de WMK vermelde noodzakelijkheidscriterium voldaan.

2. Zijn de eisen te vertalen in gezondheidscriteria?

Hiervoor is de procedure ingevoerd conform Figuur 1, schriftelijk weergegeven in de vorige paragraaf. De arts verzamelt medische bijzonderheden op geprotocolleerde wijze en weegt deze ten opzichte van de geoperationaliseerde basis-

medische eisen. Leidt dit tot de constatering dat er van een beperking sprake is, dan zal het ongeschiktheidsoordeel worden uitgesproken. Dit geeft de keurend arts geen vrijbrief om om het even welke medische bijzonderheid als een beperking ten aanzien van de BMEKL te beschouwen. Juist de formulering en normering van de eisen begrenst de professionele autonomie tot het gewenste terrein: alleen medische bijzonderheden die ten opzichte van de geformuleerde eisen gesignaleerde beperkingen opleveren tellen mee in de oordeelsvorming.

Subconclusie: Gelet op de documentatie van de BMEKL en de keuringsprotocollen zijn de eisen te vertalen in gezondheidsrisico's.

3. Is medisch onderzoek mogelijk?

Zijn er normen?

In de BMEKL zijn bij alle 43 beoordelingspunten duidelijke normen aangebracht. De gehanteerde methoden leveren geen risico voor de lichamelijke of geestelijke gezondheid voor de keurling op. Bij het onderzoek naar het al dan niet voldoen aan de gestelde eisen, wordt gebruik gemaakt van protocollen.

Vormt de gehanteerde methode een valide indicator?

De gehanteerde eisen zelf en de keuringsprotocollen (hoe onderzoek naar de bijzonderheden te verrichten) zijn ontwikkeld door een krijgsmachtbrede combinatie van bedrijfsartsen, verzekeringsgeneeskundigen, sportartsen, en bewegingswetenschappers. In paragraaf 1.4.1 is aannemelijk gemaakt dat voor het nieuwe keuringssysteem sprake is van zowel "face validity" [1] als van "expert validity" [1].

De sensitiviteit en specificiteit van het keuringssysteem zijn niet bekend. Van de gehanteerde meetinstrumenten om medische bijzonderheden op te sporen (bijvoorbeeld voor visusbepaling) is de validiteit niet bekend, hooguit kan worden gesteld dat deze instrumenten standaard in ons land worden gehanteerd. Voor de gedetailleerde weergave van de aanstellingsprocedure en daarbij gehanteerde meetinstrumenten wordt naar de paragraaf "Achtergrondinformatie" van hoofdstuk 4 verwezen.

Is de gehanteerde methode proportioneel?

In het keuringsproces wordt eerst naar medische bijzonderheden bij de kandidaat gezocht door middel van anamnese, aanvullend onderzoek en lichamelijk onderzoek. Vervolgens wordt beoordeeld of een eventueel gevonden medische bijzonderheid tot een beperking ten aanzien van de BMEKL leidt. Die toetsing aan de BMEKL kan zonder meer als proportioneel worden beschouwd: de normen zijn afgeleid van de aan militairen te stellen eisen. Resteert de vraag of het onderzoek naar medische bijzonderheden proportioneel is. De anamnese is

enige jaren geleden reeds getoetst op proportionaliteit en waar nodig aangepast³. Behoudens die anamnese wordt een aantal instrumenten gehanteerd om aanvullend bepaalde specifieke (vooral zintuiglijke) kwaliteiten te onderzoeken, zoals het gehoor, de visus en het kleuronderscheidend vermogen. Dit geldt in feite ook voor het onderzoek naar de kwaliteit van het bewegingsapparaat door de fysiotherapeut voorafgaande aan het bezoek aan de keurend arts. Gelet op de ten tijde van het opstellen van het keuringssysteem beschikbare keuringsgegevens is destijds aangenomen dat het gebruik van het instrumentarium aan het proportionaliteitscriterium voldeed, maar het kan niet worden uitgesloten dat toekomstig onderzoek tot andere inzichten zal leiden.

Is de gehanteerde methode subsidiair?

Voor al het gebruikte instrumentarium kan worden aangenomen dat dit subsidiair is. Anamnese en lichamelijk onderzoek kunnen de toets der kritiek doorstaan; voor bij de overige onderzoeken gehanteerde methoden bestaan geen minder ingrijpende methoden, noch is het aannemelijk dat het gebruik van die methoden zelf als te ingrijpend zal worden ervaren.

Leiden de onderzoeksmethoden niet tot ongerechtvaardigde differentiatie en discriminatie van groepen in de samenleving?

Het doel van de BMEKL is om onderscheid te maken tussen een operationeel wel en niet inzetbare militair ambtenaar. De in de begeleidende documentatie van BMEKL en keuringsprotocollen vermelde gegevens geven geen aanleiding om aan te nemen dat er in deze werkwijze sprake is van ongerechtvaardigde discriminatie.

Is het te verrichten onderzoek relevant voor het verwezenlijken van het doel van de keuring (relevantiecriterium)?

Hierbij past een zelfde redenering als bij de vraag naar proportionaliteit het geval was. Gelet op de ten tijde van het opstellen van het keuringssysteem beschikbare keuringsgegevens is destijds aangenomen dat het gebruik van het instrumentarium aan het relevantiecriterium voldeed, maar het kan niet worden uitgesloten dat toekomstig onderzoek tot andere inzichten zal leiden.

Subconclusie: Voor zover redelijkerwijze te beoordelen valt, is medisch onderzoek mogelijk.

3 Naar aanleiding van vragen uit de Tweede Kamer heeft een intern werkverband bepaalde vragen uit de anamnese geschrapt.

4. Is keuren toegestaan?

Zijn de eisen door een bedrijfsarts vastgesteld?

De vertaling van de bijzondere functie-eisen is primair de taak van de bedrijfsarts. Dit is bij het opstellen van de BMEKL ook het geval geweest.

Is keuren toegestaan?

Er zijn functie-eisen met een ongewoon hoge belasting, meer zelfs dan in het voorbeeld van de ARA zijn opgenomen. Al deze eisen vallen volgens het ARA-voorbeeld in de zwaarste klasse. Deze eisen zijn te vertalen in gezondheidscriteria. Ook medisch onderzoek is mogelijk. Het keuringssysteem voldoet aan de vereisten voor face-validity en expert-validity, maar door het ontbreken van de sensitiviteit en specificiteit kan de validiteit van het keuringssysteem niet worden aangetoond; dit is overigens bij geen enkel keuringssysteem in Nederland het geval. Een voorbehoud moet voorts worden gemaakt ten aanzien van het proportionaliteitscriterium en het relevantiecriterium. Het kan niet worden uitgesloten dat het gebruik van bepaald onderzoeksinstrumentarium (visus, gehoor etc.) naar aanleiding van toekomstig onderzoek zou moeten worden bijgesteld. Het systeem is immers twee jaar voor de invoering van de ARA ontworpen. De eisen zijn door bedrijfsartsen opgesteld; de te hanteren werkwijze wordt ondersteund door protocollen.

Conclusie: Gelet op het onder de stappen 1 tot en met 4 betoogde valt het te verantwoorden dat voor een aanstelling bij de KL als militair ambtenaar een aanstellingskeuring wordt verricht, mits nog nader onderzoek wordt uitgevoerd (naar de validiteit, proportionaliteit en/of relevantie van bepaalde meetinstrumenten, evenals naar het uiteindelijke effect van het Convenant Fysieke Belasting tussen de Ministeries van SZW en Defensie).

Op de volgende pagina is met behulp van Figuur 2 het stappenplan ten aanzien van de functie voor militair ambtenaar schematisch weergegeven.

1.5 Theorie, hypothese en vraagstelling

1.5.1 Theorie

Volgens Bouter en Van Dongen [1] ontstaat een theorie door inductie van waarnemingen. Voor wat keuringen betreft zijn die waarnemingen in de voorgaande paragrafen weergegeven. De bewijslast van deze waarnemingen is overtuigend genoeg om er een theorie aan te ontleen. *Deze theorie luidt dat de kwaliteit van aanstellingskeuringen als meetinstrument te wensen over laat.*

Figuur 2: Schematisch overzicht van de volgens de ARA te beantwoorden vragen om te mogen keuren.

Het aantal afkeuringen is vaak erg laag, hetgeen de vraag oproept waarom er moet worden gekeurd [16, 18]. Ook de reproduceerbaarheid is vaak te laag [2, 19, 24, 25]. De relatie tussen medische bijzonderheden en uitval op het werk

ontbreekt [17, 39], onderzoek naar andere kwaliteitsmaatstaven dan de reproducteerbaarheid ontbreekt feitelijk [16, 17], onder meer omdat dit moeilijk te verwezenlijken valt.

1.5.2 Hypothese

Mede op grond van bovengenoemde waarnemingen is de WMK in ons land ingevoerd. Door voorwaarden te verbinden aan het mogen uitvoeren van keuringen, beoogt de WMK de kwaliteit van keuringen te verbeteren.

In paragraaf 1.3.1 is uiteengezet dat ook op het ABOHZIS-keuringssysteem de nodige kritiek te leveren viel; zo voldeed dit systeem niet aan de vereisten die door de WMK worden gesteld.

Uit de voorgaande paragraaf blijkt dat het nieuwe keuringssysteem – de BMEKL – in hoofdzaak aan zowel WMK als ARA voldoet. Hypothetisch zou dit moeten betekenen dat de kwaliteit van het BMEKL-keuringssysteem beter zou moeten zijn dan die van het ABOHZIS-systeem.

Vervolgens moeten nu de onderzoeksvragen worden geoperationaliseerd.

1.5.3 Operationalisering van de onderzoeksvragen

Inleiding

De hypothese is dat de kwaliteit van het BMEKL-keuringssysteem beter zou moeten zijn dan die van het ABOHZIS-systeem. Hoe kunnen nu vanuit deze hypothese de onderzoeksvragen worden geoperationaliseerd? Uitgangspunt is dat onderzoek dient plaats te vinden naar de kwaliteit van keuringen. In de eerste plaats zal daarbij aandacht moeten worden besteed aan de vraag welke kwaliteitscriteria voor de keuring te onderscheiden vallen. Wellicht moet daaruit een keuze worden gemaakt. Vervolgens zal een onderzoekspopulatie moeten worden gekozen. Bekeken zal moeten worden welke metingen bij die populatie moeten worden verricht om de gekozen kwaliteitscriteria te onderzoeken. Uiteindelijk zal moeten worden nagegaan onder welke omstandigheden die metingen bij de onderzoekspopulatie moeten worden verricht.

Kwaliteit

Een keuring is een meetinstrument of diagnostische test. De kwaliteit van een diagnostische test is afhankelijk van verschillende testeigenschappen. Bouter en Van Dongen [1] onderscheiden er vijf.

- De relevantie van het met behulp van de test gemeten kenmerk. Bouter en Van Dongen [1] definiëren dit als volgt: het kenmerk (bijvoorbeeld pijn) moet betekenisvol zijn, gegeven het doel dat met de toepassing van de testprocedure wordt beoogd (bijvoorbeeld vaststellen of een ziekenhuisopname is geïndiceerd).

- De praktische toepasbaarheid. De test moet zijn toegesneden op de praktijksituatie waarin hij wordt toegepast (kenmerken van de populatie: aantal, beschikbare tijd, leeftijd overwegend gezond; kenmerken van de onderzoeker: vaardigheid, beschikbare tijd). Afhankelijk van de onderzoekssituatie dienen eisen te worden gesteld aan onder andere het tijdsbeslag, de complexiteit en de aan de test verbonden risico's.
- De efficiëntie, de mate waarin de gevraagde inspanning en de gemaakte kosten zo laag mogelijk blijven.
- De reproduceerbaarheid, waarbij de onderlinge variatie in beoordelingen zo klein mogelijk dient te zijn.
- De validiteit; de test dient het diagnostische kenmerk zo nauwkeurig mogelijk te meten en derhalve een zo hoog mogelijke sensitiviteit en specificiteit te bezitten.

Eerder is echter reeds vermeld dat het ABOHZIS-systeem geen relatie meer vertoonde met de feitelijke aard van de werkzaamheden [14], terwijl het BMEKL-systeem juist was afgeleid van die feitelijke werkzaamheden [51]. Theoretisch zou daarmee de relevantie van het nieuwe systeem hoger moeten uitvallen dan het vorige ABOHZIS-systeem, maar aangetoond is dat niet.

In paragraaf 1.4.2 is uiteengezet dat bij de aanstellingskeuring van een aantal meetinstrumenten gebruik wordt gemaakt. Bij de overgang van het ene keuringsstelsel naar het andere is daarbij geen verandering gekomen. Door de keuringsprotocollen en het beoordelingskader is de beschouwing wel gewijzigd: in plaats van diagnose-gericht te oordelen, wordt in het nieuwe systeem volgens het model van belasting – belastbaarheid beoordeeld of iemand voldoet aan de te stellen eisen, die zijn ontleend aan de realiteit. Waar de praktische toepasbaarheid van het ABOHZIS-systeem door het decennialange gebruik ervan (vanaf 1946 tot 1997) empirisch is aangetoond, kan op theoretische gronden worden aangenomen dat die van het nieuwe systeem tenminste even goed en wellicht iets beter zal zijn. Ook dit is echter niet aangetoond.

Ook qua efficiëntie zullen beide systemen elkaar weinig ontlopen. Er wordt van dezelfde instrumenten, personen en infrastructuur gebruik gemaakt. Voor de ingebruikneming van de BMEKL zijn geen meerkosten gemaakt, terwijl het nieuwe keuringsproces zelf ook niet meer tijd in beslag neemt dan het vorige. Het opleiden van de betrokken artsen vond plaats gedurende twee workshops in het voorjaar van 1998, waarbij de opkomst 98% was. Na afloop gaven alle bij de aanstellingskeuring betrokken artsen in de schriftelijke evaluatie aan van mening te zijn zonder verdere begeleiding met het nieuwe systeem te kunnen werken. De speciaal daarvoor in het leven geroepen helpdesk ontving tot aan het begin van het onderzoek uit dit proefschrift (oktober 1998) geen vragen.

De reproduceerbaarheid, betrouwbaarheid of consistentie is de mate van overeenstemming tussen testuitslagen, meermalen toegepast bij dezelfde onderzoekspersonen onder gelijkblijvende condities. Dit vormt een graadmeter voor de kwaliteit van de diagnostiek.

De reproduceerbaarheid kan op verschillende manieren worden beoordeeld. Zo kan worden onderzocht of één waarnemer die de test bij dezelfde personen herhaalt, steeds tot dezelfde bevindingen komt (intra-waarnemers variatie, test-retest reliability). Hierbij moet voldoende tijd tussen de metingen zitten om herkenning te voorkomen, maar een te lange tussentijd kan werkelijke veranderingen inhouden. Ook kan worden onderzocht of twee waarnemers die de test – liefst gelijktijdig – bij dezelfde personen uitvoeren tot gelijklopende conclusies komen (inter-waarnemers variatie, inter-observer reliability).

De reproduceerbaarheid kan met behulp van verschillende maten worden onderzocht [1, 2], af te leiden van onderstaande kruistabel (tabel 4). Op deze maten zelf en de daaraan soms verbonden beperkingen [1, 52, 53] wordt in de hoofdstukken 2 tot en met 4 nader ingegaan. Hier wordt volstaan met op te merken dat volgens Bouter en van Dongen [1] de kappa één van die maten is (uitslag minimaal 0,0 tot maximaal 1,0). Volgens deze auteurs suggereren gepubliceerde onderzoeksgegevens dat een kappa tussen de 0,40 en 0,75 voor diagnostische testen in de klinische diagnostiek gebruikelijk is.

Omdat zowel het oude als het nieuwe keuringssysteem niet alleen bij de aanstellingskeuring wordt gehanteerd, maar tevens bedoeld is als instrument voor de pensioenkeuring, moet voor beide vormen van keuring de reproduceerbaarheid worden nagegaan.

Als met beoordelaar 1 en 2 twee verschillende onderzoekers worden bedoeld, kan de *inter-waarnemers variatie* worden berekend. Als met beoordelaar 1 en 2 dezelfde onderzoeker wordt bedoeld, kan de *intra-waarnemers variatie* worden berekend.

Het ABOHZIS-keuringssysteem is onderzocht op reproduceerbaarheid (hoofdstuk 2). De uitslagen van de twee daarbij gehanteerde maten voor overeenkomst waren vergelijkbaar met de uitslagen van het onderzoek van Post-Uiterweer en De Kort [2]. Het BMEKL-systeem is nog niet onderzocht op reproduceerbaarheid.

De gangbare maten om de validiteit van een test te bepalen, zijn de sensitiviteit en de specificiteit. Sensitiviteit geeft het percentage mensen met een bepaalde ziekte die door de test terecht als ziek worden geclassificeerd; sensitiviteit zegt dus iets over de gevoeligheid van de test voor het herkennen van ziektegevallen. De specificiteit geeft het percentage mensen aan dat terecht als niet ziek wordt geclassificeerd. Een goede test scoort meer dan 90% op zowel de sensitiviteit als de specificiteit [1]. Dit wordt in tabel 5 weergegeven.

Tabel 4. Kruistabel, waarin op één populatie ($a+b+c+d$) tweemaal dezelfde test wordt uitgevoerd door ofwel twee verschillende beoordelaars (1 en 2), ofwel tweemaal achtereenvolgend door dezelfde beoordelaar.

		Beoordelaar 2	
		Normale testuitslag	Afwijkende testuitslag
Beoordelaar 1	Normale testuitslag	a	b
	Afwijkende testuitslag	c	d

Tabel 5. Kruistabel waarin een ziekte of kenmerk met behulp van een test op de aan- of afwezigheid binnen een populatie ($a+b+c+d$) wordt beoordeeld, terwijl deze ziekte of kenmerk reeds eerder is vastgesteld met een andere test die over de aan- of afwezigheid daarvan absoluut uitsluitsel verschaft.

		Ziekte	
		Aanwezig	Afwezig
Test	Positief	terecht-positief a	fout-positief b
	Negatief	fout-negatief c	terecht-negatief d
		a + c	b + d
		a + b + c + d	

De sensitiviteit wordt vervolgens uitgedrukt als $Se = a/(a+c)$. De specificiteit wordt vervolgens uitgedrukt als $Sp = d/(b+d)$.

Om de validiteit te kunnen berekenen moet er zekerheid bestaan omtrent de begrippen "terecht" en "fout", er moet een test bestaan die daarover in absolute zin uitsluitsel verschaft. Zo'n test wordt de "gouden standaard" [1] genoemd. Van zowel het ABOHZIS-systeem als het BMEKL-systeem is de validiteit niet onderzocht.

In tabel 6 wordt een inschatting van de stand van zaken met betrekking tot de vijf kwaliteitsmaatstaven voor beide keuringssystemen weergegeven.

De Kort [17] en diens mede-onderzoekers beperkten zich tot de betrouwbaarheid en de validiteit als graadmeters voor de kwaliteit van de medische aanstellingskeuring als selectie-instrument. Dit doet vermoeden dat door De Kort [17] aan deze kwaliteitseisen van het instrument aanstellingskeuring meer gewicht werd toegekend dan aan andere. Tevens blijkt uit tabel 6 dat juist kennis van die specifieke kwaliteitseisen in deze situatie gebrekkig is. Op grond hiervan is het voor deze studie plausibel om juist onderzoek te verrichten naar de reproduceerbaarheid en de validiteit. Met behulp hiervan zal worden nagegaan of het

Tabel 6. Inschatting van de stand van zaken met betrekking tot de kwaliteitseisen ten aanzien van beide keuringssystemen voorafgaande aan de onderzoeksvraag.

Kwaliteitskenmerk Keuring	ABOHZIS	BMEKL	Toelichting
Relevantie	Vermoedelijk gelijkwaardig		Gemeten kenmerk moet betekenis hebben gelet op het met de test beoogde doel
Toepasbaarheid	Vermoedelijk gelijkwaardig		Toegesneden op situatie, proportioneel, subsidiair, voor alle partijen acceptabel (implementatie)
Efficiëntie	Vermoedelijk gelijkwaardig		Minimalisering van inspanningen en uitgaven
Validiteit	Beide onbekend		Operationalisatie, uitvoering, te meten concept: sensitiviteit / specificiteit, diagnostische waarde, gouden standaard
Reproduceerbaarheid	Laag	Onbekend	Intra-waarnemers en inter-waarnemers variatie

nieuwe instrument beter volstaat dan het oude. Bij een eventueel gevonden verschil zal vanzelfsprekend worden getracht om het mechanisme daarvan te achterhalen. Het primaire doel is echter om de kwaliteit van het nieuwe keuringssysteem te vergelijken met die van het oude.

Voor onderzoek naar de validiteit van één of meer keuringssystemen ontbreekt de gouden standaard, waardoor het inzicht inzake de fout-positieven en fout-negatieven ontbreekt. Er is dus niet bekend wie de “ziekte” (of in dit geval het kenmerk dienstongeschiktheid) wel of niet bezit. In dat geval is men geïnteresseerd in de voorspellende waarde van de test. De voorspellende waarde geeft de kans aan dat personen met de betreffende testuitslag de ziekte of het kenmerk bezitten of in de toekomst zullen bezitten (tabel 7).

Op grond van de testuitslag is een voorspelling mogelijk ten aanzien van de aanwezigheid, respectievelijk afwezigheid van een ziekte of het kenmerk (in dit onderzoek de dienstgeschiktheid).

Daarbij geeft een positieve voorspellende waarde (DW+) aan welk deel van de onderzochte personen met een positieve testuitslag de ziekte of het kenmerk heeft, en een negatieve voorspellende waarde (DW-) welk deel van de onderzochte personen met een negatieve testuitslag van de ziekte of het kenmerk gevrijwaard is gebleven.

Hoe groter de sensitiviteit en specificiteit zijn, des te groter is de positieve en negatieve voorspellende waarde van de test [1]. De voorspellende waarde is echter ook afhankelijk van de prevalentie van de ziekte of het kenmerk in de onderzoekspopulatie [1]. In het algemeen geldt: hoe hoger de prevalentie, des te

Tabel 7. Kruistabel waarin de voorspellende waarde binnen een populatie ($a+b+c+d$) wordt aangegeven voor het bezitten of verkrijgen van een bepaalde ziekte of kenmerk.

	Ziekte		
	Aanwezig	Afwezig	
Testuitslag positief	a	b	$DW+ = a / a + b \times 100\%$ (positieve voorspellende waarde)
Testuitslag negatief	c	d	$DW- = d / c + d \times 100\%$ (negatieve voorspellende waarde)

groter de positieve voorspellende waarde en des te kleiner de negatieve voorspellende waarde, en vice versa [1].

In de ideale onderzoeksopzet worden groeperingen kandidaat-militairen volgens één of beide systemen gekeurd en vervolgens aangenomen, opgeleid en ingezet – onafhankelijk van goed- of afkeuring (Figuur 3).

Het te werk stellen van ongeschikt verklaarde militairen was echter geen haalbare optie. Om praktische en ethische redenen geeft de KL-organisatie hiervoor geen toestemming.

Nu zou er toe kunnen worden overgegaan om de afgewezen kandidaten op dezelfde wijze te volgen gedurende dezelfde onderzoeksperiode als hun geschikt bevonden collega's. Hierbij zouden zich drie problemen voordoen. Ten eerste waren de mogelijkheden om gegevens over bij de KL te werk gestelde militairen te verzamelen aanmerkelijk uitgebreider dan die voor de externe onderzoeksgroep. Hierdoor zou selectiebias optreden. Ten tweede zouden de arbeidsomstandigheden en functie-eisen van die externe onderzoeksgroep grote individuele variatie vertonen. Door de te verwachten selectiebias zou hiervoor niet kunnen worden gecorrigeerd. Als derde zou het verkrijgen van voldoende respons problematisch kunnen blijken te zijn.

Hierdoor resteerde alleen de mogelijkheid om onderzoek te verrichten naar verschillen tussen geschikt verklaarde militairen. Dit betekent dat onderzoek naar de prognostische waarde van beide keuringssystemen zal plaatsvinden. Hierdoor zullen echter zowel sensitiviteit als specificiteit onbekend blijven.

Na het maken van deze keuze om zowel de reproduceerbaarheid als de prognostische waarde van beide keuringssystemen onderling te vergelijken moet vervolgens worden nagegaan bij welke populatie onderzoek moet plaatsvinden, en met welke metingen dat onderzoek moet worden uitgevoerd, waarna moet worden nagegaan met welke determinanten verder rekening moet worden gehouden.

Figuur 3: De ideale onderzoeksopzet voor de bepaling van de voorspellende waarde.

Alle kandidaten, geschikt zowel als ongeschikt verklaard, worden na opkomst tot het einde van de arbeidsovereenkomst gevolgd.

Keuze van de populatie

De basispopulatie wordt gevormd door gegadigden voor een aanstelling bij de KL als militair ambtenaar.

Naast grotendeels jonge gegadigden afkomstig uit de civiele sector die een overeenkomst van tijdelijke aard (in de regel 2½ jaar; "militairen BBT"⁴) krijgen, vinden aanstellingskeuringen in mindere mate ook bij reeds actief dienend militair personeel plaats dat een overeenkomst van onbepaalde tijd krijgt aangeboden ("militairen BOT"⁵). Een militair komt pas in aanmerking voor een BOT-arbeidsovereenkomst na succesvol uitdienen van een BBT-overeenkomst. Beide groepen verschillen derhalve in ervaring. Daarnaast is er een leeftijdsverschil tussen beide groepen, verloopt de militaire opleiding anders, de positionering in de hiërarchie is verschillend en de fysieke belasting kan uiteen lopen. Om die redenen wordt de populatie beperkt tot de groep der militairen BBT.

4 BBT: Beroepsmilitair Bepaalde Tijd.

5 BOT: Beroepsmilitair Onbepaalde Tijd.

Met welke metingen moet het onderzoek worden uitgevoerd?

Voor het onderzoek naar de reproduceerbaarheid van keuringsuitslagen, ongeacht of dit aanstellingskeuringen dan wel pensioenkeuringen betreft, geldt de keuringsuitslag vanzelfsprekend als de te meten effectvariabele. Voor het onderzoek naar de voorspellende waarde ligt dit ingewikkelder.

In paragraaf 1.3.2. is aandacht besteed aan onderzoek dat reeds was verricht naar de kwaliteit van aanstellingskeuringen [20, 23]. Hieruit kwam een aantal aandachtspunten naar voren, dat van nut zou zijn bij verder onderzoek, zoals vermeld in paragraaf 1.3.2. Naast een aantal bemerkingen van meer algemene strekking waren daaronder tevens bemerkingen die richting geven aan de te gebruiken effectvariabelen.

Een eerste opmerking daarover luidde dat tot op heden alleen de relatie tussen keuring en uitval was onderzocht. Het in mindere mate voldoen aan de gestelde (keurings-)eisen behoeft niet altijd tot uitval op grond van beroepsziekten of arbeidsongeschiktheid te leiden. Er zijn ook andere uitingen van het minder aan de gestelde eisen voldoen denkbaar. De vraag is vervolgens hoe dit voor militairen te operationaliseren.

Voor militairen wordt inzetbaarheid als het belangrijkste product beschouwd. Dit begrip is voor deze studie geoperationaliseerd als het aantal dagen tijdens de onderzoeksperiode waarin geen sprake was van verminderde inzetbaarheid door ziekteverzuim of andere vormen van arbeidsverzuim. Door deze definitie kon het netto aantal dagen worden vastgesteld, dat deze militair theoretisch operationeel had kunnen worden ingezet. Weliswaar is in paragraaf 1.3.2 reeds gesteld dat verzuim een multifactoriële effectparameter [47] is, onder te verdelen in persoonsgebonden, bedrijfsgebonden en maatschappelijke oorzaken, maar omwille van de volgende argumenten is deze effectvariabele desalniettemin bruikbaar voor deze studie:

- het geeft de inzetbaarheid in maat en getal weer;
- voorafgaande aan de tewerkstelling vindt achtereenvolgens administratieve, psychologische en fysieke selectie plaats, waardoor diverse persoonsgebonden en maatschappelijke factoren worden uitgesloten;
- er vindt een aanstellingskeuring plaats waarbij eveneens een aantal persoonsgebonden factoren worden uitgesloten;
- men wordt binnen een enkele organisatie tewerkgesteld, zodat hooguit minimale verschillen in bedrijfsgebonden factoren resteren;
- onder de voorafgaande aan de studie onderkende mogelijke confounders (zie paragraaf 1.6.2) bevindt zich een aantal persoonsgebonden en bedrijfsgebonden factoren waarop bij gevonden verschil tussen studiegroepen kan worden gecorrigeerd.

Een tweede opmerking was dat als artsbezoek ook een rol bij uitval speelt, het zinvol is om bij verder onderzoek ook daaraan aandacht te besteden. Volgens het onderzoek van Carbone [45] zou de hoeveelheid spreekuurcontacten omgekeerd evenredig zijn met de inzetbaarheid. Binnen de KL plaatsgevonden spreekuurcontacten zouden eveneens kunnen worden geteld, maar daarnaast is er de mogelijkheid dat militairen (buiten de diensturen om) ook extern geneeskundige verzorging behoeven. Van die vorm van verzorging kan het precieze aantal contacten niet worden nagegaan, maar de erbij gedeclareerde kosten worden centraal bijgehouden. Om die reden is de medische consumptie die gedurende de onderzoeksperiode heeft plaatsgevonden als tweede effectvariabele gehanteerd.

Een derde opmerking was dat in geen enkel Nederlands onderzoek de beleving of motivatie van de betrokken werknemers is meegenomen, ondanks het feit dat de invloed van die beleving bekend was. Het gebruik van bijvoorbeeld een periodieke vragenlijst zou op grond hiervan een toegevoegde waarde kunnen hebben bij verder onderzoek.

Samenvattend blijkt uit deze paragraaf dat onderzoek zal worden verricht naar het aantal dagen dat de onderzoeksgroep operationeel inzetbaar is geweest, evenals naar de medische consumptie gedurende de onderzoeksperiode. Daarnaast zal gebruik worden gemaakt van een vragenlijst. Voor het operationaliseren van de onderzoeksvragen resteert nu nog om na te gaan onder welke omstandigheden het onderzoek moet worden verricht.

Determinanten

Welke factoren kunnen van invloed op de resultaten zijn? Zoals onderaan paragraaf 1.4.1. is uiteengezet omvat het nieuwe BMEKL-keuringssysteem het complete scala aan taken met inbegrip van operationele inzet. Een belangrijk onderscheid tussen beide keuringssystemen is in theorie dat het voormalige ABOHZIS-systeem geen rekening met deze vorm van inzet hield en het nieuwe BMEKL-systeem nu juist wel. Het complete scala aan taken met inbegrip van operationele inzet komt echter pas tot volle uiting onder operationele omstandigheden. Weliswaar wordt de militair door middel van training en opleiding ook onder niet-operationele omstandigheden op operationeel optreden voorbereid, maar behoudens die operationele omstandigheden wordt een specifieke functie vervuld. Die (vredes-)functie zelf vertoont soms sterke overeenkomsten met civiele functies. Bovendien kunnen niet alle aspecten vooraf worden getraind, zoals bijvoorbeeld het ondergaan van hoge temperaturen in een uitzendgebied. Er bestaat derhalve een onderscheid in belasting tussen het optreden onder niet-operationele omstandigheden (lichter, minder breed scala aan taken, bekende omstandigheden) en het optreden onder operati-

onele omstandigheden (zwaarder, breed scala aan taken, geheel andere omstandigheden).

Of en in hoeverre een aangenomen militair daadwerkelijk wordt uitgezonden, wordt echter niet door middel van de keuring bepaald, maar is afhankelijk van andere factoren, zoals of de betreffende eenheid aan de beurt is voor uitzendingen. Dit betekent dat bij het onderzoek rekening moet worden gehouden met het gegeven of de militair al dan niet operationeel wordt ingezet voor een buitenlandse missie.

Daarnaast kunnen ook die uitkomstvariabelen als determinanten worden beschouwd, waarvan bekend was of waarvan niet kon worden uitgesloten dat zij de inzetbaarheid en/of medische consumptie mogelijk beïnvloeden: de locatie van de initiële militaire opleiding (er bestonden verschillen in zwaarte en duur), of men een gevechtsfunctie of een gevechtsondersteunende functie vervult, het onderdeel en de functie waar men na de opleiding wordt geplaatst.

Zoals uit paragraaf 1.3.2 blijkt, kan ook aan de eigen beleving van omstandigheden en gezondheid invloed worden toegekend. Hierop wordt in paragraaf 1.6.2. nader ingegaan.

1.5.4 Vraagstelling voor dit onderzoek

Om, bij het ontbreken van een gouden standaard, de kwaliteit van beide keuringssystemen onderling te vergelijken zal onderzoek worden verricht naar zowel de reproduceerbaarheid als de prognostische waarde van beide systemen. De reproduceerbaarheid van beide keuringssystemen wordt zowel bij de aanstellingskeuring als bij de pensioenkeuring vergeleken. Daarbij wordt de keuringsuitslag als effectvariabele gehanteerd.

Door middel van de prognostische waarde wordt onderzocht of er associaties bestaan tussen (één van) beide keuringssystemen en de afhankelijke variabelen dagen inzetbaarheid en medische consumptie, met de eigen beleving van werk en gezondheid als secundaire uitkomstmaat. Dit onderzoek vindt plaats bij militairen BBT, waarbij onderscheid moet worden gemaakt tussen degenen die wel, respectievelijk niet operationeel worden ingezet.

Bovenstaande samenvatting van de operationalisering leidt tot de volgende onderzoeksvragen:

1. Voldoet de reproduceerbaarheid van het nieuwe keuringsteststelsel bij de pensioenkeuring aan de normering die gebruikelijk is voor screenings-testen⁶?

6 Volgens Bouter en Van Dongen [1] suggereren gepubliceerde onderzoeksgegevens dat een kappa tussen de 0,40 en 0,75 voor diagnostische testen in de kliniek gebruikelijk is.

2. Hoe verhoudt zich de mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring?
3. Is er een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van de geschikt verklaarde militairen BBT op het gebied van de dagen inzetbaarheid en/of de genoten medische consumptie?
4. Indien een verschil in prognostische waarde tussen beide keuringssystemen wordt gevonden, kan het mechanisme daarvan worden verklaard door analyse van de determinanten persoonskenmerken, de eigen beleving van gezondheid, inzetbaarheid en operationele inzet?
5. Bestaat er ook een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van uitsluitend de uitgezonden militairen en zo ja, kan dat worden verklaard door analyse van de determinanten persoonskenmerken, het beleven van de eigen gezondheid, de eigen beleving van gezondheid en inzetbaarheid?

1.6 Materiaal en methoden

1.6.1 Populatie

Groepsgrootte

Ten aanzien van de groepsgrootte moest rekening worden gehouden met mogelijke uitval. Bekende redenen voor uitval waren (en zijn):

- Uitval door psychologische selectie. Deze uitval bedroeg ongeveer 30% van de gegadigden.
- Uitval door afkeuring en fysieke selectie van de overgebleven kandidaten. Deze uitval was niet exact te voorspellen omdat in de gekozen onderzoeksperiode een nieuw systeem voor fysieke selectie werd ingevoerd en het Instituut Keuringen en Selectie (IKS) geen concrete informatie over afkeuringen kon verschaffen. Desgevraagd schatten de uitslaggevend seniorbedrijfsartsen van het IKS deze uitval globaal in op 20 – 30% van de na de psychologische selectie overgebleven kandidaten.
- Uitval door het veiligheidsonderzoek (militair personeel wordt voorafgaande de formele aanstelling door de Militaire Inlichtingen en Veiligheidsdienst gescreend op veiligheidsaspecten). De groepsgrootte hiervan was onbekend, maar vooraf werd als gering ingeschat.
- Uitval in de eerste opleidingsweken. Intern onderzoek toonde aan dat in de jaren 1997 (zowel als in 1998 en 1999) sprake was van een niet-medische uitval van tussen de 20 en 30% van de geschikt bevonden kandidaten.

Naar inschatting van de uitslaggevend arts werden in 1998 maandelijks tussen de 300 en 500 kandidaten medisch gekeurd. Om twee groepen van voldoende

grootte te kunnen formeren zou een periode van 6 tot 8 weken volstaan. Om onnodige variatie uit te sluiten, diende sprake te zijn van een zo kort mogelijke, maar in elk geval aaneengesloten periode.

Exclusiecriteria

Behoudens het onderzoek te beperken tot militairen met een overeenkomst van bepaalde tijd ("BBT"), diende tevens met de volgende feiten rekening te worden gehouden:

- Kandidaten bestemd voor de Nationale Reserve (Natres) worden administratief ook als gewone KL-gegadigde beschouwd, terwijl hun taakstelling niet vergelijkbaar is met die van de gewone KL-militair (geen operationele inzet buiten de landsgrenzen en in beginsel alleen in de weekeinden actief). Om deze redenen dienden deze kandidaten te worden uitgesloten.
- Er bestaan functies binnen het bestand aan KL-militairen, waarvan het risicoprofiel teveel afwijkt van het in de Wijziging van de Wet op de Medische Keuringen over militair ambtenaren geformuleerde risicoprofiel, zoals militaire muzikanten en geestelijk verzorgers. Ook bij deze gegadigden moest uitsluiting plaatsvinden.
- Bekend was dat tijdens de initiële militaire opleiding of AMO een forse uitval ontstaat – gedurende de opleidingsperiode kan de rekrut de verbintenis eenzijdig opzeggen. Wellicht zou hierin qua omvang al een onderscheid te signaleren zijn bij een vergelijking tussen twee onderzoeksgroepen (althoewel tot dan toe intern onderzoek had uitgewezen dat meer dan 95% van de bij voortijdig vertrek opgegeven redenen van niet-medische aard zijn). Het probleem is dat deze "drop-outs" vertrekken tussen de dag na opkomst en de eerste 12 weken. Hierdoor zou bij een vergelijking van diegenen die (veel langer) bij de KL blijven (en van de opleidingssituatie overgaan naar een situatie waarin zij concrete werkzaamheden moeten uitvoeren) een sterke informatie-bias optreden.
- Het moment van de aanstellingskeuring behoefde niet maatgevend te zijn voor het moment van opkomst en tewerkstelling. Ten tijde van de onderzoeksofzet was niet bekend hoeveel tijd er lag tussen het moment van keuren en het moment van daadwerkelijke opkomst. Het was de bedoeling om alle personen van de onderzoeksgroep (of groepen) gedurende dezelfde onderzoeksperiode te volgen.
- Omdat het hierboven genoemde delay ten tijde van de onderzoeksofzet onbekend was, viel evenmin uit te sluiten dat er ook nog rekruten zouden zijn die ondanks hun geschiktheid niet op zouden komen, bijvoorbeeld omdat ze in de tussenliggende periode een andere (civiele) functie hadden aanvaard.

Onderzoekperiode

Het nieuwe BMEKL-keuringssysteem zou per 1 november 1998 worden ingevoerd. Om die reden werd voordien bijscholing in het gebruik van het nieuwe systeem voor de keurend artsen georganiseerd. Hierdoor werd het mogelijk om voorafgaande aan de invoering van dit nieuwe systeem tijdelijk beide keuringssystemen tegelijkertijd te gebruiken.

Bekend was reeds dat er een verschil in beweegredenen van gegadigden bestaat tussen diegenen die zich aansluitend aan de afloop van het schooljaar (en/of in de daaropvolgende zomervakantie) aanmelden en diegenen die zich gedurende de rest van het jaar aanmelden. De laatste groep mist de natuurlijke aansluiting, hierdoor zijn de beweegredenen om zich tot de KL te wenden meer pluriform van aard dan van de eerste groep. In de zomervakantie zelf treedt mogelijk selectiebias op door vakantie van zowel de gegadigden als de keurend artsen. Hierdoor werd de ideale onderzoekperiode gedefinieerd als zijnde de maanden september en oktober 1998.

Omdat onder de exclusiecriteria is vermeld dat het delay tussen keuring en opkomst niet bekend was, werd de onderzoekperiode vastgesteld als de periode waarin de meeste kandidaten na hun militaire opleiding op functie worden geplaatst. Uiteindelijk bleek dit de periode van 1 juni 1999 tot 1 juni 2001 te zijn.

1.6.2 Procedure

Reproduceerbaarheid

Omdat het keuringssysteem voor zowel de aanstellingskeuring als de pensioenkeuring zou worden gebruikt moest reproduceerbaarheidonderzoek voor beide processen plaatsvinden.

In de ideale situatie wordt de reproduceerbaarheid beoordeeld door de kandidaten achtereenvolgens door twee verschillende artsen te laten keuren (inter-waarnemers variatie), of na verloop van tijd opnieuw door dezelfde arts (intra-waarnemers variatie).

Zelfs aan deze ideale opzet kleven nadelen, zoals rechtspositionele problemen (aannemen of afkeuren bij discongruente uitslagen?), eventuele invloed uitgeoefend door de onderzoekers (op gedrag van de keurling of op elkaars beslissing), dan wel het herkennen van de keurling bij een tweede beoordeling door dezelfde arts.

Voor de beoordeling van de reproduceerbaarheid in dit onderzoek is ervoor gekozen om dat aan de hand van de beoordeling van dossiers te laten bepalen. Alhoewel hieraan ook nadelen zijn verbonden (zoals de kans op informatie-bias), is dergelijk onderzoek al vaker verricht, hetgeen de aanvaardbaarheid ervan verhoogt.

Het bepalen van de mate van overeenkomst tussen bedrijfsartsen bij de aanstellingskeuring kon qua onderzoekspopulatie worden gecombineerd met een prospectief onderzoek, omdat dit laatste bij de aanstellingskeuring moest beginnen. Hierdoor zou in elk geval voldoende onderzoeksmateriaal ontstaan om de mate van overeenkomst te meten.

Het bepalen van de mate van overeenkomst tussen artsen bij de pensioenkeuring kon beter met behulp van een andere populatie plaatsvinden. Er was geen garantie dat een cohortstudie naar geschikt verklaarde militairen voldoende pensioenkeuringen zou opleveren om tot een aanvaardbaar minimum aan dossiers te geraken, nog afgezien van de onbekende tijdsduur die het verzamelen ervan in beslag zou nemen. Onbekend was vooraf immers wanneer militairen eventueel ongeschikt zouden worden of wanneer zij hun claim zouden indienen (rechtspositioneel mag dit op elk moment plaatsvinden, bijvoorbeeld ook eerst na afloop van de verbintenis).

Mogelijke confounders

Een confounder is een verstorende variabele die de weergave van de relatie tussen centrale determinant en onafhankelijke variabele vertekent (bijvoorbeeld roken bij onderzoek naar de relatie tussen alcoholconsumptie en het voorkomen van strottehoofdkanker) [1]. Confounding kan het best worden bestreden door correctie en randomisatie [1]. Met name op grond van de in de internationale literatuur gevonden risicofactoren, alsmede op grond van reeds door anderen verricht intern onderzoek, zijn voorafgaande aan het onderzoek de volgende mogelijke confounders onderscheiden:

- Persoonskenmerken, waarvan bekend was of werd aangenomen dat zij de inzetbaarheid en/of medische consumptie beïnvloeden: leeftijdsverschillen (op jonge leeftijd geeft een gering leeftijdsverschil al een hogere slagingskans in de militaire opleiding ten gunste van de oudere rekrut), het geslacht (vrouwen verzuimen meer dan mannen), de gezinssituatie (herkomst uit een compleet gezin is bevorderlijk voor het behalen van de opleiding), rookgewoonten (kunnen tot ziekte leiden), de schoolopleiding (hoe hoger de schoolopleiding des te groter de kans op slagen in de militaire opleiding), herkomst uit een stad dan wel van het platteland (plattelanders behalen vaker hun militaire opleiding).
- Keuringsgegevens, waarvan bekend was of werd aangenomen dat zij de inzetbaarheid en/of medische consumptie beïnvloeden zoals het wel of niet ter keuring vaststellen van medische bijzonderheden door de keurend arts (met een bijzonderheid mogelijk grotere kans op uitval), de aard van die medische bijzonderheden, de vastgestelde mate van fysieke opleidbaarheid (des te hoger, des te geringer de kans op blessures), het vetpercentage (hoe lager des te hoger de conditie), de keurend arts zelf.

- Uitkomstvariabelen, waarvan bekend was of waarvan het vermoeden bestond dat zij de inzetbaarheid en/of medische consumptie kunnen beïnvloeden: de locatie van de initiële militaire opleiding (er zijn verschillen in zwaarte en duur), of men een gevechtsfunctie of een gevechtsondersteunende functie vervult, het onderdeel en de functie waar men na de opleiding wordt geplaatst, of men al dan niet wordt uitgezonden.

Deze gegevens konden in beide onderzoeksgroepen van invloed zijn, er moest wellicht voor worden gecorrigeerd; derhalve dienden deze gegevens te worden verzameld en bijgehouden.

Randomisatie

Om het verschil te meten tussen geschikt verklaarde BBT-militairen van beide keuringssystemen moesten twee onderzoeksgroepen worden geformeerd. Randomisatie kon in theorie weliswaar op verschillende manieren plaatsvinden, maar de keuzemogelijkheden werden in de praktijk beperkt. Zo zou het dagelijks aanbieden van aselekt (bijvoorbeeld op grond van een lijst met aselekte getallen) in de twee keuringssystemen ingedeelde kandidaten aan elke keurend arts tot diverse problemen kunnen leiden: het gevaar verstrikt te raken in verschillende diagnostische classificatieschema's [1], expectation bias [1], de hoge moeilijkheidsgraad van controle door derden op de correcte uitvoering en motivatieproblemen bij de keurend arts. In de voorbereiding tot dit onderzoek leidde dit tot de volgende wijze van randomisatie: om de samenstelling van de twee studiegroepen vergelijkbaar te maken zou per week van keuringssysteem worden gewisseld. Dit leidde tot de in figuur 4 weergegeven onderzoeksopzet.

Eventuele bias kon nu op de volgende manieren worden beperkt. In de eerste plaats wist elke keurend arts op welke wijze hij / zij die week geacht werd te keuren: één periode, lang genoeg om zich op het betreffende keuringssysteem te kunnen focussen. Voorts kon de uitslaggevend seniorbedrijfsarts (die de keuringsdossiers verzamelde) relatief eenvoudig controleren of de juiste werkwijze werd gehanteerd en zo nodig corrigerend optreden. Bovendien waren alle dossiers gedateerd en kon de onderzoeker na afloop zelf ook controleren of de juiste werkwijze was gehanteerd.

In hoofdstuk 4 wordt onder de paragraaf "Achtergrondinformatie" gedetailleerd ingegaan op de aanstellingsprocedure. Hier wordt volstaan met op te merken dat de volgorde waarin kandidaten voor de aanstellingskeuring worden opgeroepen willekeurig is, waardoor het begrip randomisatie mag worden gebruikt voor de hierboven geschetste procedure.

Informatiebronnen

Voor het uitvoeren van dit onderzoek werd het raadplegen van de volgende gegevens noodzakelijk geacht: de dossiers van de aanstellingskeuring, exitinter-

views afgenomen bij dienstverlaters, centraal beheerde (over-) plaatsingsgegevens, het centraal gecontroleerde en beheerde verzuimregistratiesysteem, gegevens afkomstig van de pensioenkeuring, de interne geneeskundige verrichtingenadministratie zoals opgesteld door de militaire zorgverleners en de door de onderzoeksgroep bij de ziektekostenverzekeraar ingediende declaraties van civiele zorgverleners. Het verlenen van schriftelijke toestemming door de betreffende kandidaat voor het mogen verzamelen van bovenstaande gegevens heeft tijdens de keuring plaatsgevonden. De kandidaten waren niet op de hoogte van het onderzoek of het feit dat zij gerandomiseerd gekeurd werden. Alle tot nu toe beschreven gegevens waren afkomstig van derden en verschaften informatie over de te volgen militair zonder dat deze zelf om een mening werd gevraagd. Als aanvulling hierop is ook aandacht besteed aan de eigen beleving van de onderzoekspersonen; bovendien zouden deze gegevens ook nog als controle op de overige informatie kunnen dienen.

De gehanteerde vragenlijst (tabel 8) was gebaseerd op de "groene vragenlijst" (VAG) van het Periodiek Arbeidgezondheidkundig Onderzoek [54] (PAGO) en vrijwel op dezelfde wijze geclusterd [55, 56]. Die groene vragenlijst bestond zelf uit twee onderdelen: het onderdeel gezondheidsbeleven (in deze vragenlijst clusters 3, 4, 5 en 6) en het onderdeel werkbeleven (in deze vragenlijst clusters 8 tot en met 13). Hieraan waren voor deze vragenlijst 8 clusters toegevoegd: twee ter controle van elders verkregen gegevens (clusters 1 en 2), één voor het beoordelen van de medische consumptie (cluster 7), en vier clusters om de uit het FIS ontleende BMEKL-vragen in te plaatsen (14 tot en met 17). Het achtste nieuwe cluster – 18 – bevatte de omzetting van de specifieke militaire beoordelingspunten naar vragen. Alle vragen vanaf cluster 3 waren dichotoom. Bij de vragen uit de clusters 3 tot en met 7 werd specifiek naar de situatie van het laatste half jaar gevraagd.

Vanaf het begin van de onderzoeksperiode ontving de kandidaat deze vragenlijst halfjaarlijks, in totaal vijf maal. Positieve reacties telden voor 1 punt, negatieve voor 0 punten.

Vijf maal achtereenvolgende ontving de kandidaat gedurende de onderzoeksperiode halfjaarlijks de hierboven vermelde vragenlijst op het verblijfsadres met als doel om na te gaan of er verschillen tussen de twee studiegroepen zouden optreden in de beantwoording van de verschillende clusters.

Effectvariabelen

Voor deze studie is de inzetbaarheid geoperationaliseerd als het aantal dagen tijdens de onderzoeksperiode waarin geen sprake was van verminderde inzetbaarheid door ziekteverzuim of andere vormen van arbeidsverzuim. Omdat in

Figuur 4: De onderzoeksopzet van de gerandomiseerde gecontroleerde studie.

Na gerandomiseerd te zijn gekeurd worden de geschikt verklaarde kandidaten na opkomst gedurende de periode van paraatstelling (twee jaar) tot het einde van de arbeidsovereenkomst gevolgd.

de onderzoeksperiode een schrikkeljaar zat, bedroeg het maximale aantal dagen inzetbaarheid 731 kalenderdagen.

Inzake de medische consumptie is het volgende van belang. De krijgsmacht, respectievelijk de KL, beschikt over een eigen eerstelijns geneeskundig zorgsysteem (huisartsenzorg) en voorziet voor de belangrijkste klinisch specialismen ook in de tweede lijns gezondheidszorg. Die eerstelijns zorg geldt echter in eigen land slechts tijdens kantooruren. Tijdens operationele inzet en bij plaatsing in Duitsland wordt 24 uur per dag eerstelijns gezondheidszorg verstrekt. Buiten kantooruren kan een militair zich tot een civiele (eerste- of tweedelijns) zorgverlener wenden; hiervoor is de militair verzekerd. De verrichtingen van de eigen eerste lijn worden in een centrale database vergaard. Alle tweedelijns verrichtingen, alsmede extern geraadpleegde eerstelijns verrichtingen, worden door de ziektekostenverzekeraar vergoed na declaratie. Om een volledig overzicht te verkrijgen over de totaal gemaakte medische kosten moeten zowel de ziektekostenverzekeraar als de database worden geraadpleegd. De medische eerstelijns consumptie is berekend door alle eerstelijns consulten te rekenen naar het gemiddeld over de onderzoeksperiode geldende CTG-tarief (€ 17,92). De overige gemaakte en gedeclareerde ziektekosten zijn – inclusief in het eigen

Tabel 8. De clusterindeling van de gebruikte vragenlijst. De clusters 1 en 2 bestaan voornamelijk uit niet-dichotome controlevragen en vallen buiten verder onderzoek.

Cluster	Aantal vragen	Herkomst
Controlevragen	8	onderzoeker
Controlevragen	9	onderzoeker
Algemene gezondheid	4	PAGO-gezondheid
Bewegingsapparaat	9	PAGO-gezondheid
KNO	3	PAGO-gezondheid
Stress	4	PAGO-gezondheid
Medische consumptie	9	onderzoeker
Taakhoud	5	PAGO-werk
Werkorganisatie	5	PAGO-werk
Samenwerking	5	PAGO-werk
Toekomstperspectief	5	PAGO-werk
Duidelijkheid	3	PAGO-werk
Privé	3	PAGO-werk
FIS-bewegingsapparaat	16	onderzoeker
Fysisch-chemisch, buiten	8	onderzoeker
Zintuigen	3	onderzoeker
FIS-psychisch	14	onderzoeker
Militaire beoordelingspunten	15	onderzoeker

zorgsysteem gemaakte tweedelijns kosten – bij de ziektekostenverzekeraar opgevraagd.

1.7 Verdere uitwerking van het onderzoek

Nadere uitwerking vindt plaats in de volgende hoofdstukken. Allereerst wordt, zoals aangekondigd in paragraaf 1.2., een systematische waarneming weergegeven in hoofdstuk 2. Daarin wordt het reproduceerbaarheidsonderzoek naar de kwaliteit van het ABOHZIS-systeem bij periodieke keuringen weergegeven dat in 1996 is verricht na kennis te hebben genomen van het werk van De Kort en Post-Uiterweer [2, 13] (figuur 5).

Hierna worden alle hierboven vermelde vragen achtereenvolgens in een apart hoofdstuk (3 tot en met 7) behandeld. In figuur 6 wordt een overzicht van de daarin gehanteerde werkwijzen weergegeven, waarbij ook de verschillende hoofdstukken staan vermeld.

Figuur 5: De onderzoeksopzet van hoofdstuk 2.

24 aselect gekozen keuringsdossiers, afkomstig van één keuringsraad, worden geanonimiseerd aan twee artsen van de andere keuringsraad aangeboden met het verzoek om zowel de ABOZHIS-classificatie als de beperkingen vast te stellen, waarna het resultaat met de originele uitslag wordt vergeleken met behulp van de kappa en PD.

Figuur 6: Het overzicht van de onderzoeken uit dit proefschrift en de daarbij behorende hoofdstukken.

In de hoofdstukken 2 en 3 wordt de reproduceerbaarheid van het periodiek geneeskundig onderzoek, respectievelijk de pensioenkeuring onderzocht, analoog aan Figuur 5. Hoofdstuk 4 bevat soortgelijk onderzoek voor wat betreft de aanstellingskeuringen van dezelfde gerandomiseerde, gecontroleerde studie, van wie de twee studiegroepen in de hoofdstukken 5 en 6 gedurende hun plaatsing op functie worden gevolgd om de voorspellende waarde tussen beide keuringssystemen te kunnen vergelijken.

Figuur 7: De onderzoeksopzet van hoofdstuk 3.

In hoofdstuk 3 wordt de reproduceerbaarheid van de pensioenkeuring onderzocht, analoog aan de wijze waarop dat in Figuur 5 voor het periodiek geneeskundig onderzoek heeft plaatsgevonden.

Figuur 8: De onderzoeksopzet van hoofdstuk 4.

Met behulp van dossiers uit de gerandomiseerde, gecontroleerde studie wordt de reproduceerbaarheid van beide keuringssystemen voor wat betreft de aanstellingskeuring onderzocht. Hierbij worden zowel de nieuwe uitslagen onderling vergeleken ("Nieuw onderling"), als de nieuwe uitslagen met de oorspronkelijke uitslag ("Oud versus nieuw, Arts C versus de Artsen 1, 2 en 3).

Figuur 9: De onderzoeksopzet voor hoofdstuk 7.

In deze studie worden uitsluitend de uitgezonden militairen, gerandomiseerd gekeurd en in twee studiegroepen verdeeld, gevolgd.

In hoofdstuk 3 wordt het onderzoek naar de reproduceerbaarheid van het nieuwe keuringssysteem tijdens de pensioenkeuring weergegeven conform onderzoeksvraag 1. In dit onderzoek zijn 30 geanonimiseerde keuringsdossiers, waaronder evenveel geschikt als ongeschikt verklaarden, ter beoordeling aan twee seniorbedrijfsartsen voorgelegd, waarna de reproduceerbaarheid is berekend (figuur 7).

In hoofdstuk 4 wordt het onderzoek naar de mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring weergegeven (onderzoeksvraag 2). Hiertoe werden 227 keuringsdossiers, afkomstig van beide keuringssystemen, geanonimiseerd. Vervolgens werden deze dossiers opnieuw voorgelegd aan dezelfde drie seniorbedrijfsartsen die in het najaar van 1998 hierover al uitslag hadden gegeven, waarna de reproduceerbaarheid is berekend (figuur 8).

In hoofdstuk 5 wordt de gerandomiseerde, gecontroleerde studie voortvloeiende uit vraag 3 naar de geschikt verklaarde militairen behandeld (figuren 4 en 6). In de periode juni 1999 – juni 2001 werd een groep van 352 militairen BBT gedurende twee jaar gevolgd, nadat zij gerandomiseerd volgens het nieuwe BMEKL- dan wel het oude ABOHZIS-keuringssysteem waren beoordeeld.

In hoofdstuk 6 wordt het onderzoek, voortvloeiende uit vraag 4 beschreven. In dit hoofdstuk worden analyses gepleegd om het mechanisme, dat achter de gevonden associatie schuilgaat, aan te tonen. Daarbij wordt gebruik gemaakt van een vragenlijst die gedurende de onderzoeksperiode vijf maal aan de onderzoekspersonen is voorgelegd (figuren 4 en 6).

In hoofdstuk 7 wordt het onderzoek naar aanleiding van vraag 5 weergegeven (figuur 9). Gegeven de vraagstelling wordt in dit hoofdstuk uitsluitend aandacht besteed aan de uitgezonden militairen uit beide studiegroepen, die voor de hoofdstukken 5 en 6 zijn gebruikt. Bij dit onderzoek wordt zowel gebruik gemaakt van dezelfde effectvariabelen als in de hoofdstukken 5 en 6, als van de van de laatste drie halfjaarlijks toegezonden vragenlijsten, zijnde de periode waarin de uitzendingen plaatsvonden.

Het afsluitende hoofdstuk 8 bevat een nabeschuiving op de verrichte onderzoeken, gevolgd door het beantwoorden van de onderzoeksvragen. Hierna wordt de hypothese getoetst. Vervolgens wordt beoordeeld of en in hoeverre het onderzoek invloed heeft op de theorie, waarna wordt besloten door na te gaan welk onderzoek resteert op het terrein van (militaire) keuringen.

Geraadpleegde literatuur

1. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Bohn, Stafleu en van Loghum. Houten, 1995.
2. Kort WLAM de, Post-Uiterweer HW: Project aanstellingskeuringen dossieronderzoek bij de RBB. Deel II: panelonderzoek. Ministerie van Sociale Zaken en Werkgelegenheid, S 53-2, 1991.
3. Raad J de: De historie van het Militair Keuringsreglement en de Lijst van ziekten en gebreken. "The mummy returns?" Nederl Mil Geneesk T 2002; 55:146-151.
4. Romeyn D: Onze Militair-Geneeskundige Dienst voor honderd jaren en daaromtrent. Haarlem: De Erven F Bohn, 1913.
5. Branger JD: Het militair geneeskundig onderzoek (de visitatie) in de vorige eeuw. Nederl Mil Geneesk T 1988; 41: 167-206.
6. Fouw de: Referaat over de methode van Warnecke voor de ontmaskering van doofheids-simulanten. Nederl Mil Geneesk T 1900; 243-244.
7. Quix FH: Voordracht over "Ontmaskering van doofheidssimulanten in verband met eene nieuwe methode van bepaling der gehoorscherpthe voor toonhoor en fluisterstem". Nederl Mil Geneesk T 1903; 177-198.
8. Wijsman PW: Simulatie en aggraviatie van doofheid. Nederl Mil Geneesk T 1935; 24: 127-137.
9. Spruyt S: Over simulatie. Nederl Mil Geneesk T 1958; 11: 1-13.
10. Romeyn D: Over ongeschiktverklaring en pensionering van onze mindere militairen. Nederl Mil Geneesk T 1895; 159-173.
11. Belloni R: De bepaling van de dienstgeschiktheid. Utrecht, 1985.

12. Koster L: Keuringseisen ten aanzien van allergische aandoeningen voor de militaire dienst. *Nederl Mil Geneesk T* 1959; 7: 173-186.
13. Dijk FJH van, Dormolen M van, Kompier MAJ et al: Herwaardering model belastingbelastbaarheid. *TSG*, 1990; 68: 3-10.
14. Kessel JGFM van: Consequenties van onjuist keuren. *Ned Mil Geneesk T*; 1984; 37: 176-228.
15. Lourijen ECMP, Kort WLAM de, Dam J: De praktijk van de medische aanstellingskeuring in Nederland. Ministerie van Sociale Zaken en Werkgelegenheid, S 53-3, 1991.
16. Lourijen ECMP, Hoolboom H, Kort WLAM de: De Medische aanstellingskeuring. Een inventarisatie van de mate waarin en de wijze waarop Nederlandse bedrijven en artsen medische aanstellingskeuringen (laten) verrichten. *DGA S* 53-1, 1989.
17. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
18. Kort WLAM de, Fransman LG, Dijk FJH van: Pre-employment medical examinations in a large occupational health service. *Sacnd J Work Environ & Health* 1991; 17: 392-297.
19. Kort WLAM de, Dijk FJH van. Preventive effectiveness of pre-employment medical assessments. *Occupational Environmental Medicine* 1997; 54: 1-6.
20. Reulings PGJ: Keuren is mensenwerk. Maastricht, 1996.
21. Reulings PGJ, Horst F van der, Crebolder HFJM et al: Het keuringstraject voor dienstplichtigen van de Koninklijke Landmacht. *Nederl Mil Geneesk T* 1994; 47: 201-204.
22. Reulings PGJ, Stuurmans F, Crebolder HFJM et al: Prognostische factoren van het medisch disfunctioneren tijdens militaire dienst bij de Koninklijke Landmacht. *TSG* 1995; 73: 505-511.
23. Reulings PGJ, Crebolder HFJM, Schouten HJA et al: Medisch keuren is mensenwerk. Variatie tussen artsen bij de beoordeling van de geschiktheid voor militaire dienst. *TSG* 1996; 74: 213-216.
24. Reulings PGJ, Schouten HJA, Rethans JJ et al: Reproduceerbaarheid van de beoordeling bij dienstplichtkeuringen. *TSG* 1996; 74: 267-271.
25. Raad J de, Reulings PGJ, Schouten HJA: De beperkingen van medische beperkingen. De reproduceerbaarheid van de beoordeling van de ABOHZIS-classificatie en de medische beperkingen bij periodiek geneeskundig onderzoek bij de Koninklijke Landmacht. *Ned Mil Geneesk T*; 1996; 49: 106-111.
26. Deursen CGL van, Smulders PGW, Bongers PM: Vormt een slechte gezondheid een ziekteverzuimrisico? *TSG* 1997; 75: 157-164.
27. Otten FWJ, Smulders PGW, Houtman ILD: Werkstress en ziekteverzuim bij jong en oud. *TSG* 1998; 76: 420-428.
28. Deursen CGL van, Houtman ILD, Bongers PM: Werk, privé-situatie, riskante gewoonten en ziekteverzuim: verschillen tussen mannen en vrouwen. *TSG* 1999; 77: 105-115.
29. Jacobs DR Jr, Ainsworth BE, Hartman TJ et al: A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc* 1993; 25(1): 81-91.
30. Weller IM, Corey PN: A study of the reliability of the Canada Fitness Survey questionnaire. *Med Sci Sports Exerc* 1998; 30(10):1530-1536.
31. Vogel JA, Crowdy JP, Amor AF et al: Changes in aerobic fitness and body fat during army recruit training. *Eur J Appl Physiol* 1978; 40(1): 37-43.
32. Stacy RJ, Hungerford RL, McMahon BB: The effect of basic training on aerobic capacity and body fat in New Zealand army recruits. *N Z Med J* 1982; 95(722): 876-878.
33. Jones BH, Cowan DN, Tomlinson JP et al : Epidemiology of injuries associated with physical training among young men in the army. *Med Sci Sports Exerc* 1993; 25(2): 197-203.
34. Jones BH, Bovee MW, Harris JM 3d et al: Intrinsic risk factors for exercise-related injuries among male and female army trainees. *Am J Sports Med* 1993; 21(5): 705-710.

35. Jordaan G, Schweltnus MP: The incidence of overuse injuries in military recruits during basic military training. *Mil Med* 1994; 159(6):421-426.
36. Almeida SA, Williams KM, Shaffer RA et al: Epidemiological patterns of musculoskeletal injuries and physical training. *Med Sci Sports Exerc* 1999; 31(8): 1176-1182.
37. Pope RP, Herbert R, Kirwan JD et al: Predicting attrition in basic military training. *Mil Med* 1999; 164(10): 710-714.
38. Ross J, Woodward A: Risk factors for injury during basic military training. Is there a social element to injury pathogenesis? *J Occup Med* 1994; 36(10): 1120-1126.
39. Rudzki SJ: Injuries in Australian Army recruits. Part III: The accuracy of a pretraining orthopedic screen in predicting ultimate injury outcome. *Mil Med* 1997; 162(7):481-483.
40. Stevenson JM, Bryant JT, Andrew GM et al: Development of physical fitness standards for Canadian Armed Forces younger personnel. *Can J Sport Sci* 1992; 17(3):214-221.
41. Chin DL - Blackwood GV - Gackstetter GD: Ergometry as a predictor of basic military training success. *Mil Med* 1996; 161(2):75-77.
42. Williams AG, Rayson MP, Jones DA: Effects of basic training on material handling ability and physical fitness of British Army recruits. *Ergonomics* 1999; 49(8):1114-1124.
43. Rayson M, Holliman D, Belyavin A: Development of physical selection procedures for the British Army. Phase 2: relationship between physical performance tests and criterion tasks. *Ergonomics* 2000; 43(1):73-105.
44. McCraw RK, Bearden DL: Personality factors in failure to adapt to the military. *Mil Med* 1990; 155(3): 127-130.
45. Carbone EG, Cigrang JA, Todd SL et al: Predicting outcome of military basic training for individuals referred for psychological evaluation. *J Pers Assess* 1999; 72(2):256-265.
46. Crawford SL, Fiedler ER: Childhood physical and sexual abuse and failure to complete military basic training. *Mil Med* 1992; 157(12): 645-648.
47. Klein Hesselink DJ, Kruidenier HJ, Veerman TJ et al: Afwezigheid verklaard. Een literatuurstudie naar determinanten van ziekteverzuim en arbeidsongeschiktheid. NIA. Amsterdam, 1993.
48. Talcott GW, Haddock CK, Klesges RC et al: Prevalence and predictors of discharge in United States Air Force Basic Military Training. *Mil Med* 1999; 164(4):269-274.
49. Broersen JPJ, Weel ANH, Dijk FJH van: Periodiek Bedrijfsgeneeskundig Onderzoek; middel en maatstaf. Amsterdam, 1989.
50. Hulshof CTJ: Algemene Richtlijn Aanstellingskeuringen. Amsterdam: SKB, 1999.
51. Raad J de, Peters JHM, Groot RP de et al: Modernisering van het keuringssysteem bij de Koninklijke Landmacht; keuringsprotocollen en functioneel beoordelen. TBV 2000; 8:110-115
52. Altman DG: Practical statistics for medical research. London: Chapman and Hall, 1991.
53. Steinijans VW, Diletti E, Bömches B et al: Interobserver agreement: Cohens Kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J of Clin of Pharm And Ther.* 1997; 35; 3: 93 - 95.
54. Raad J de: Vragenlijsten bij periodiek bedrijfsgezondheidkundig onderzoek en het nut voor de Koninklijke Landmacht, een literatuurstudie. Leiden, 1993.
55. Putten DJ van, Marcelissen FGH, Winter CJ de: De vragenlijst voor PBGO 1-4. Achtergronden voor het gebruik ervan. Het wijzigen van antwoorden door bedrijfsartsen. Methoden voor gebruik van referentiegegevens. TSG 1988; 66: 115-120, 143-147.
56. Putten DJ van: De vragenlijst voor PBGO, voorstellen tot wijziging van de vragenlijst. Leiden, NIPG/TNO, 1989.

2 | De beperkingen van medische beperkingen

In belangrijke mate gebaseerd op: Raad J de, Reulings PGJ, Schouten HJA: De beperkingen van medische beperkingen. Ned Mil Geneesk T 1996; 49: 106-111.

2.1 Inleiding en vraagstelling van dit onderzoek

Naar aanleiding van de omvorming naar een beroepsleger en de uitbreiding van de taakstelling met het uitvoeren van missies buiten de eigen landsgrenzen, onderging de Koninklijke Landmacht (KL) in 1994 een reorganisatie. Er was in de KL zowel sprake van personele overtoolligheid als van personeel dat vermoedelijk niet geschikt was om de nieuwe taakstelling uit te voeren. Daarom werd een onderzoek gehouden naar de kansen op behoud van werk voor beroepsmilitairen (Project Loopbaanprognose). Als onderdeel hiervan werd beoordeeld of de beroepsmilitairen in de toekomst inzetbaar zouden zijn op medische gronden. Die beoordeling vond plaats op basis van de keuringsuitslag van het laatst gehouden periodiek geneeskundig onderzoek (PGO). Op grond van deze beoordeling deed zich een aantal vragen voor over de keuringsprocedure die bij de KL werd gevolgd bij dat PGO.

Bij het PGO werd de beroepsmilitair beoordeeld, zoals vermeld in het Militair Keuringsreglement (MKR), op aanwezigheid van ziekten en gebreken. Als consequentie werd de gekeurde militair in een bepaalde categorie op grond van de zogenoemde ABOHZIS-classificatie ingedeeld (zie onder), waarbij zo nodig (medische en functionerings-) beperkingen werden verstrekt.

Omdat aan deze handelwijze anders dan voorheen het geval was – op grond van de reorganisatie en dreigende overtoolligheid – serieuze (rechtspotionele) consequenties werden verbonden, deed zich de vraag voor of en in hoeverre de op die manier tot stand gekomen uitspraken voldoende betrouwbaar waren. Recente publicaties van Post-Uiterweer [1] en De Kort [2] inzake keuringen bij de RBB en meer algemeen in Nederland, gaven aanleiding om zich af te vragen of de betrouwbaarheid van deze keuringen bij de KL volstond.

Er was nooit onderzoek verricht naar de mate van overeenkomst tussen de beoordelingen van keurend artsen volgens de ABOHZIS-classificatie of naar het verband tussen die classificatie en de daaruit voortvloeiende te verstrekken beperkingen. Gelet op het belang dat bij de reorganisatie aan deze keuring werd toegekend, lag het voor de hand om zulk onderzoek te verrichten.

De vraagstelling voor dat onderzoek luidt: “Is de beoordeling van de ABOHZIS-classificatie en de medische beperkingen, zoals die bij het PGO van de KL worden vastgesteld, voldoende reproduceerbaar, en komt de gegeven ABOHZIS-classificatie overeen met de toegekende medische beperkingen?”

2.2 Achtergrondinformatie

Tussen 1960 en 1995 werden beroepsmilitairen van de Koninklijke Landmacht middels het PGO (doorgaans om de vijf jaar) gekeurd om te beoordelen of zij

op medische gronden nog geschikt waren om als militair te kunnen blijven functioneren. De militair werd opgeroepen om bij één van de twee keuringsraden te verschijnen. De keurend arts had bij dit onderzoek inzage in het medisch dossier van de militair in kwestie. Het PGO zelf bestond uit een gestandaardiseerde anamnese, lichamelijk onderzoek en aanvullende onderzoeksgegevens, te weten een rust-ecg, audiometrie, de bepaling van de visus, onderzoek naar het kleuronderscheidend vermogen en urineonderzoek.

Naar aanleiding van de gevonden resultaten werd elke vijf jaar voor het militair beroepspersoneel een nieuwe ABOHZIS-classificatie opgesteld. De letters ABOHZIS staan voor **A**lgemene fysieke toestand, **B**ovenste extremiteiten, **O**nderste extremiteiten, **H**oren, **Z**ien, **I**ntelligentie en **S**tabiliteit. Voor elk van deze deelgebieden werd een indeling in vijf classificaties gehanteerd, lopend van één tot vijf (bijvoorbeeld A₁, B₂, O₃ etc.). Daarbij betekende een 1-classificatie dat men niet in het minst aan een in het MKR vermelde ziekte of gebrek leed, hetgeen in toenemende mate wel in geval was indien men een 2-classificatie of hoger kreeg. De constatering van een classificatie tussen 2 en 4 leidde tot het opleggen van één of meer medische en functioneringsbeperkingen. De constatering van een 5-classificatie betekende formeel dat de betrokken militair dienstongeschikt was en moest worden afgekeurd. Voor de medische beperkingen werd van een standaardlijst gebruik gemaakt. Deze lijst was naar analogie van de ABOHZIS-classificatie eveneens verdeeld in de rubrieken A, B O, H, Z, I en S, waarbij elke specifieke beperking van een nummer was voorzien, zoals bijvoorbeeld A02 ("geen hindernisbaan"), O10 ("geen zware rugbelasting"), B10 ("geen zware nekbelasting"), S07 ("mag geen geestelijk stresserende werkzaamheden verrichten").

Zowel de medische als de daaraan gekoppelde functioneringsbeperkingen werden vervolgens aan de betreffende militair en diens commandant bekend gesteld. De commandant kon bepalen of deze militair de opgedragen werkzaamheden al dan niet kon uitvoeren. Eventueel kon de militaire dienstgeschiktheid van de militair door de commandant ter discussie worden gesteld. In het algemeen leidden de opgelegde beperkingen bij militairen niet tot verre gaande consequenties, althans tot 1994. Er bestond een gesloten personeelssysteem voor militairen, men werd geacht "voor de gehele duur van de loopbaan te zijn aangesteld".

Een groot nadeel van de gehanteerde werkwijze was dat wél een beoordeling van de belastbaarheid werd opgemaakt, maar dat er geen sprake was van concreet inzicht in de werkelijke belasting: er werden wel functioneringsbeperkingen opgelegd, maar van functie-eisen was geen sprake. Een tweede nadeel was dat de ruimte voor het opleggen van een lagere ABOHZIS-classificatie in de loop der jaren was gekrompen. Kon de arts oorspronkelijk per deelgebied in

de A, B, O, H, Z van de gradaties 1 naar 4 beperkingen opleggen, inmiddels was voor de gehele classificatie de ondergrens voor geschiktheid opgetrokken naar ten hoogste 2 of 3, waaronder zich de gradatie X bevond, welke laatste classificatie dienstongeschiktheid inhield. Hierdoor werden de mogelijkheden tot nuancering sterk ingekrompen, ofschoon die mogelijkheid tot nuancering nu juist de reden was dat dit systeem oorspronkelijk in gebruik was genomen [3]. Voorts was er kritiek op de koppeling van het keuringssysteem met de ABOHZIS-classificatie, die tot een te grote inperking van de professionele autonomie van de arts zou leiden [4]. Daarnaast bleek dat de indeling op grond van de ABOHZIS-classificatie zelf te weinig relatie met de feitelijke militaire werkzaamheden vertoonde [5].

Ten gevolge van de Defensienota uit 1991 en de Prioriteitennota uit 1993 werd de taakstelling van de KL uitgebreid, onder gelijktijdige omvorming naar een beroepsleger. De nieuwe taakstelling bracht een verzwaring van de belasting met zich mee. Daarnaast bleek er overtolligheid onder beroepsmilitairen te bestaan, waardoor niet voor elke beroepsmilitair werk kon worden gegarandeerd. Hierdoor ontstond grotere belangstelling bij de KL-leiding voor de actuele medische inzetbaarheid van de individuele militair dan voorheen. De werkgever – de Bevelhebber der Landstrijdkrachten (BLS) – wenste militairen zonder beperkingen te kunnen inzetten. Aan het te werk stellen van militairen met ernstige medische en/of functioneringsbeperkingen werd een einde gemaakt. Als gevolg daarvan werd bij het overtolligheidonderzoek dat in 1994 startte ook aandacht geschonken aan de individuele medische beperkingen.

2.3 Materiaal en methoden

2.3.1 Keuringsdossiers

Uit een bestand van 200 dossiers van recente (minder dan 6 maanden oude) PGO's, die door de arts van de ene keuringsraad waren verricht, zijn 24 dossiers aselekt gekozen. Alle 200 dossiers werden voorzien van een opeenvolgend nummer, waarna een medewerker 24 toevalsgetallen [6] uitkoos. De dossiers zijn geanonimiseerd en ontdaan van de vastgestelde ABOHZIS-classificatie en medische beperkingen. Alle overige informatie, met inbegrip van correspondentie met eventuele geconsulteerde klinisch specialisten, was overgenomen. Onder de betreffende dossiers bevonden zich 6 dossiers van militairen waarbij geen beperkingen waren vastgesteld ten aanzien van de A-, B- en O-classificatie.

Deze 24 keuringsdossiers zijn ter beoordeling aangeboden aan twee artsen van de andere keuringsraad (Figuur 10); één ervaren keurend arts en één arts met

Figuur 10: De onderzoeksopzet van hoofdstuk 2.

24 aselect gekozen keuringsdossiers, afkomstig van één keuringsraad, worden geanonimiseerd aan twee artsen van de andere keuringsraad aangeboden met het verzoek om zowel de ABOHZIS-classificatie als de beperkingen vast te stellen, waarna het resultaat met de originele uitslag wordt vergeleken met behulp van de kappa en PD.

ongeveer 6 maanden werkervaring. Aan beide artsen is gevraagd om zowel de ABOHZIS-classificatie als de medische beperkingen vast te stellen. De beoordeling van de oorspronkelijke arts (beoordelaar 1) is vervolgens vergeleken met de beoordeling van de ervaren arts (beoordelaar 2) en de minder ervaren arts (beoordelaar 3).

2.3.2 Statistische methoden

Er bestaan meerdere methoden om de mate van overeenstemming te berekenen [1, 7]. Eén van deze methoden is het berekenen van de kappa (figuur 11). De waarde van kappa varieert tussen 0 en 1 (figuur 11). Voor het beoordelen van screeningstesten wordt vaak aan een bepaalde waarde van kappa een kwalificatie gekoppeld. Bouter en van Dongen [7] vermelden dat een kappa tussen de 0,40 en 0,75 voor diagnostische testen in de klinische diagnostiek gebruikelijk is. Een wat uitgebreider indeling is die waarbij een kappa < 0,40 aangeeft dat er sprake is van een geringe overeenstemming, een kappa tussen 0,40 en 0,60 dat er sprake is van een redelijke overeenstemming, een kappa tussen 0,60 en 0,75 dat er sprake is van een goede overeenstemming en een kappa van boven de 0,75 dat er sprake is van een voortreffelijke overeenstemming. De mate van overeenkomst tussen de drie beoordelaars wordt aan de hand van de groeps-kappa

weergegeven [8]; dit geldt zowel voor de ABOHZIS-classificatie als voor de medische beperkingen.

De belangrijkste parameters van de ABOHZIS-classificatie die bij de uitslag van het PGO een rol spelen zijn de A-, B- en O-classificatie. Om die reden wordt de kappa uitsluitend voor deze drie classificaties berekend. Om de mate van overeenstemming tussen de artsen met betrekking tot de A-, B- en O-classificatie vast te stellen, worden per rubriek twee kappa-waarden berekend. De eerste kappa geeft de mate van overeenstemming op de vier afzonderlijke scores van 1, 2, X of ? weer. De tweede kappa betreft de overeenstemming op de score van 1 tegenover de combinatie 2, X of ?. De scores 2, X en ? worden daarbij als gelijkwaardig gezien en geven aan dat een ziekte of gebrek in enigerlei mate aanwezig is.

Bij de medische beperkingen wordt de overeenstemming beoordeeld tussen het wel of niet toekennen van een beperking op een bepaalde rubriek. Om dichotomie te bewerkstelligen en aldus gecorreleerde kappa's met elkaar te kunnen vergelijken is de "jackknife techniek" gebruikt, analoog aan de gepaarde t-toets [9].

De formule voor Kappa is:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e}$$

Hierbij geldt:

P_o = de waargenomen overeenstemming (observed)

P_e = de toevallige overeenstemming (chance expected)

Interpretatie:

$K < 0,40$	geringe overeenstemming
$K \ 0,40\text{--}0,60$	redelijke overeenstemming
$K \ 0,60\text{--}0,75$	goede overeenstemming
$K > 0,75$	voortreffelijke overeenstemming

Figuur 11. De betekenis van kappa.

Tenslotte wordt onderzocht of de artsen consequent omgaan met het toekennen van medische beperkingen. De overeenkomst tussen de beperkingen in termen van de A-, B- of O-classificatie en de medische beperkingen wordt voor elke arts vastgesteld met behulp van de kappa. Daarbij worden gedichotomiseerde variabelen met elkaar vergeleken. Er is overeenstemming indien de arts enerzijds een A1, B1 of O1 verstrekt terwijl hij/zij tegelijk op deze rubriek geen medische beperkingen aangeeft. Evenzeer dient bij een andere dan een 1-classificatie nu juist wel medische beperkingen te worden vermeld.

Als tweede maat om de reproduceerbaarheid vast te stellen is het percentage discordante of incongruente uitslagen gebruikt, zoals in het panelonderzoek van De Kort en Post-Uiterweer [1] is gehanteerd: $(GO + OG / GG + GO + OG + OO) \times 100\%$.

2.4 Resultaten

2.4.1 De A-, B- en O-classificatie

In tabel 9 is voor alle 24 keuringsformulieren de toegekende ABOHZIS-classificatie van de drie beoordelaars weergegeven. De oorspronkelijke beoordelaar heeft gebruik gemaakt van de mogelijkheid om een ABOHZIS-classificatie van 1, 2 respectievelijk X te verstrekken. Beoordelaars 2 en 3 gebruikten soms daarnaast “?” als score. In de dagelijkse praktijk hebben de beoordelaars 2 en 3 de mogelijkheid om dit vraagteken vervolgens voor te leggen aan het hoofd van de keuringsraad.

2.4.2 Medische beperkingen

In tabel 10 staan de medische beperkingen weergegeven, zoals de verschillende artsen deze per dossier hebben vermeld. De codes staan voor specifieke medische beperkingen, zoals onder “achtergrondinformatie” is weergegeven. Het is moeilijk aan te geven of de keurend artsen inhoudelijk overeenstemmen in de toekenning van de medische beperkingen. Inhoudelijk is de betekenis van de betreffende beperkingen niet altijd medisch van aard, regelmatig hebben beperkingen het karakter van een functioneringsbeperking. Daarnaast is er overlap tussen de rubrieken, een aantal beperkingen afkomstig uit verschillende rubrieken zijn identiek.

2.4.3 De mate van onderlinge overeenstemming bij de A-, B-, en O-classificatie en de onderlinge overeenstemming bij de medische beperkingen

In tabel 11 zijn de kappa's voor alle drie de beoordelaars tezamen weergegeven, zowel voor wat de ABOHZIS-classificatie betreft als voor de medische beperkingen.

De mate van overeenkomst is per A-, B-, of O-rubriek beoordeeld. De eerste kappa die wordt vermeld (tweede kolom; “Kappa 1, 2, X, ?”) geldt voor de afzonderlijke A-, B- of O-scores, zijnde 1, 2, X of ?. Deze eerste kappa-waarde geeft aan dat de mate van overeenstemming per rubriek verschillend is. De overeenkomst tussen de artsen bij de A-rubriek is gering; de overeenstemming bij de B- en O-rubriek is redelijk.

Tabel 9. De door de 3 artsen bij de 24 PGO's toegekende ABOHZIS-classificaties.

Formulier Nummer	Beoordelaar 1							Beoordelaar 2							Beoordelaar 3						
	A	B	O	H	Z	I	S	A	B	O	H	Z	I	S	A	B	O	H	Z	I	S
1	1	1	2	2	2	1	1	1	1	1	2	x	1	1	1	1	?	2	?	1	1
2	2	1	1	2	1	1	1	x	1	1	2	1	1	1	x	1	1	2	1	1	1
3	1	1	2	1	1	1	1	2	1	x	?	1	1	1	?	1	x	1	1	1	1
4	2	1	1	1	2	1	1	2	1	1	1	?	1	1	2	1	2	1	?	1	1
5	1	1	2	1	2	1	2	1	1	x	1	?	1	1	?	1	2	1	?	1	2
6	2	1	1	1	2	1	1	2	1	1	1	?	1	1	?	1	1	1	?	1	1
7*	1	1	2	1	2	1	1	1	1	x	1	2	1	1	?	1	2	1	2	1	1
8	2	2	2	1	2	1	2	2	2	x	?	?	1	2	?	2	x	?	2	1	2
9*	2	1	2	1	1	1	1	2	?	?	1	1	1	1	?	1	2	1	1	1	1
10	2	1	1	1	2	1	1	1	1	1	1	x	1	1	?	1	x	1	x	1	1
11	2	1	2	1	1	1	1	x	1	2	1	1	1	1	?	1	2	1	1	1	1
12*	2	1	1	1	1	1	1	2	1	1	?	?	1	1	?	2	1	?	1	1	1
13	1	1	x	1	2	1	1	?	?	x	1	2	?	?	?	1	x	1	2	1	1
14	2	1	1	1	1	1	1	x	1	1	1	1	1	1	?	1	1	1	1	1	1
15	2	1	1	2	1	1	1	x	1	1	2	1	1	1	?	1	1	?	1	1	1
16*	1	2	1	2	1	1	1	2	1	1	?	1	1	1	?	2	1	?	1	1	1
17	2	2	2	2	2	1	1	x	2	2	?	2	1	1	x	2	2	?	?	1	1
18	1	1	2	1	2	1	1	1	1	x	1	x	1	1	?	1	x	1	?	1	1
19	2	1	2	1	1	1	1	x	1	2	1	1	1	1	x	1	1	1	1	1	1
20*	2	1	1	1	1	1	1	x	1	1	2	1	1	1	2	1	1	1	1	1	1
21	2	1	1	1	2	1	1	x	1	1	1	2	1	1	x	?	?	1	?	1	1
22	2	1	x	2	1	1	1	x	1	x	x	?	x	1	?	1	2	?	x	1	1
23*	1	2	2	1	1	1	1	1	2	2	1	1	1	1	?	x	1	?	1	1	1
24	x	1	1	1	2	1	1	x	1	1	?	2	1	1	x	1	1	?	?	1	1

*: Oorspronkelijk geen medische beperkingen. Beoordelaar 1: oorspronkelijke keurend arts; beoordelaar 2: ervaren keurend arts; beoordelaar 3: niet-ervaren keurend arts.

De tweede kappa geeft de overeenstemming weer bij het vaststellen of er wel of geen ("+" / "-") ziekte of gebrek is. Daartoe zijn de 1-classificaties dichotoom tegenover alle andere classificaties (2, X, ?) geplaatst, zoals in paragraaf 2.3.2 onder "statistische methoden" reeds is aangegeven. In de volgende kolommen – de vierde en vijfde – wordt op basis van dezelfde dichotome verdeling als uit de

Tabel 10. De door de 3 artsen bij de 24 PGO's toegekende medische beperkingen.

Formulier Nummer	Beperkingen ABOHZIS		
	Beoordelaar 1	Beoordelaar 2	Beoordelaar 3
1	H: 4	H: 4	
2	A: 6	A: 6, 43, x	A: 9, 43
3	A: 20 O: 4, 18, 20	A: 20 O: 4, 19	A: 20 O: 4
4	A: 5, 6, 8, y	A: 5, 6, 34	A: 6, 9, y
5	O: 7, 11, 18, 20, 27	O: 7, 18, 19, 20, 27, 30	O: 7, 18, 27
6	A: 6, 34+ O: 30	A: 20 S: 14	
7 *			
8	A: 20 B: 12, 21 O: w S: 14	A: 20 S: 14	
9 *			
10	A: 14		
11	O: 10, 27		O: 10
12 *			
13	A: 20 O: 7, 18, 20, 27, 28, 36, z	O: 7, 27, 30	O: 7, 18, 27
14	A: 6, x	A: 6, x	
15	A: 6, 34	A: 6 H: 4	A: 10, x
16 *	H: 4		
17	B: 4, 10, 17, 21 O: 10, 18, 21, 27 H: 4 S: 11	A: 6, 9, 20B: 4, 7, 10, 14, 17 O: 10, 18, 20	A: 6 B: 7, 10
18	O: 10, 14+, 18, 20, 27, x	O: 10, 14	O: 10, 18, 27
19	A: 6, 20 O: 11, 14, 18, 20+	A: 9, 11, 24	A: 20, x
20 *		A: 6, 20, 34	
21	A: 6, 12, x S: 7	A: 5, 6, 12, 34	A: 12
22	O: 9, 12, 14, 27		
23 *			
24	A: 6, 9, 11, x	A: 9, 11, 24	A: 6, 11, x

*: oorspronkelijk geen beperkingen; +/++ gemodificeerde beperking; w/x/y niet in de lijst voorkomende beperkingen

derde kolom het percentage discordante uitslagen en de bijbehorende standaardafwijking weergegeven.

In de zesde kolom wordt de overeenstemming ("Kappa") bij de medische beperkingen beoordeeld, met behulp van de "Jackknife techniek" teneinde

Tabel 11. De mate van overeenkomst tussen de beoordelaars met betrekking tot de rubrieken A, B en O van de ABOHZIS-classificatie en de medische beperkingen, uitgedrukt in Kappa.

ABOHI- rubriek	ABOHZIS-classificatie				Medische beperkingen "Jackknife techniek"			
	Kappa 1, 2, X, ?	Kappa +/- ziekte (dichotoom)	PD%	SD%	Kappa	P	PD%	SD%
A	0,16	0,32	20,8	11,8	0,67	0,07	20,8	5,9
B	0,52	0,58	10,4	2,9	0,74	0,10	2,8	2,4
O	0,44	0,67	12,5	11,8	0,65	0,10	16,7	0

PD staat voor percentage discrepante uitslagen tussen de beoordelaars 2 en 3 versus beoordelaar 1 en sd voor de standaardafwijking daarvan. Verdere toelichting wordt in de tekst hieronder weergegeven.

dichotomie te bewerkstelligen (paragraaf 2.3.2 "statistische methoden") met in de volgende kolom de bijbehorende p-waarde (telkens $> 0,05$). Van diezelfde dichotome verdeling wordt in de laatste twee kolommen het percentage discrepante uitslagen en de bijbehorende standaardafwijking weergegeven.

2.4.4 De overeenkomst tussen de ABOHZIS-classificatie en de medische beperkingen per beoordelaar

In tabel 12 staat aangegeven in hoeverre de drie artsen bij de A-, B- en O-rubriek consequent omgaan met het toekennen van medische beperkingen. Bij elke arts is per rubriek de relatie tussen de verstrekte A-, B- of O-score en het wel of niet toekennen van een medische beperking op de betreffende rubriek vastgesteld.

De overeenkomst tussen de ABOHZIS-classificatie en de medische beperkingen bij de oorspronkelijke beoordelaar is redelijk tot goed. Bij de andere twee beoordelaars is de overeenkomst geringer, namelijk slecht tot redelijk. Alle drie artsen zijn het met elkaar eens dat bij een A1, B1 of O1 geen medische beperkingen worden toegekend. Echter, het komt vaak voor dat geen medische beperkingen worden toegekend indien de ABOHZIS-classificatie niet 1 is, maar 2, X of ?.

De percentages discrepantie (PD) tussen de beoordelaar 1 enerzijds en de beoordelaars 2 en 3 anderzijds zijn voor de relatie tussen beperkingen en A-classificatie $34,7 \pm 16,8\%$, voor de relatie tussen beperkingen en B-classificatie $13,9 \pm 4,8\%$ en voor de relatie tussen beperkingen en O-classificatie $26,4 \pm 13,4\%$.

Tabel 12. De overeenkomst tussen de toegekende ABOHZIS-classificatie en de medische beperkingen bij de drie beoordelaars.

ABOHZIS	Classificatie	Beoordelaar 1				Beoordelaar 2				Beoordelaar 3			
		Medische beperking				Medische beperking				Medische beperking			
		Nee	Ja	Kappa	SD	Nee	Ja	Kappa	SD	Nee	Ja	Kappa	SD
A	A1	7	1			5	0			1	0		
	A2, Ax, A?	5	11			6	13			13	10		
				0,43	0,32			0,47	0,32			0,10	0,18
B	B1	20	0			19	0			18	0		
	B2, Bx, B?	2	2			4	1			4	2		
				0,63	0,47			0,28	0,45			0,43	0,42
O	O1	11	0			13	0			10	0		
	O2, Ox, O?	4	9			5	6			10	4		
				0,67	0,28			0,57	0,30			0,25	0,24

sd staat voor de standaardafwijking.

2.5 Beschouwing

Hoewel de overeenkomsten tussen de artsen niet altijd als "slecht" kunnen worden gedefinieerd, sluiten de gevonden uitslagen aan bij eerdere publicaties over de tekortkomingen van keuringsuitslagen. Post-Uiterweer en De Kort [1] vonden bij hun panelonderzoek een gemiddelde kappa van $0,35 \pm 0,07$ met een gemiddelde waarde voor PD van $37 \pm 4\%$. Uit tabel 10 blijkt dat deze waarden in dit onderzoek positiever liggen ten aanzien van de kappa, maar dat de waarde van PD met name de A-classificatie die van het onderzoek van Post-Uiterweer en De Kort [1] benadert, ook ten aanzien van de toegekende beperkingen op grond van die A-classificatie.

De overeenkomst tussen de toegekende ABOHZIS-classificatie en de vastgestelde medische beperkingen uit tabel 11 is evenmin zo groot als wenselijk: vaker worden geen medische beperkingen vermeld terwijl toch bij de rubrieken A, B en O een andere score dan 1 wordt vermeld, hetgeen in principe betekent dat beperkingen moeten worden opgelegd. Weliswaar is er in alle gevallen sprake van een grote standaardafwijking bij de vermelde kappa, maar gelet op het theoretisch causale verband tussen beide fenomenen mag dit resultaat toch niet als goed worden beschouwd. Ondanks die standaardafwijkingen betekent het voor beoordelaar 3 in twee van de drie classificaties nog steeds een

niet meer dan redelijke overeenstemming. De gevonden PD-waarden van de relatie tussen medische beperkingen en A- en O-classificatie sluiten, inclusief de standaardafwijkingen, geheel aan bij de bevindingen van Post-Uiterweer en De Kort [1].

De stapel dossiers waaruit de aselechte trekking plaatsvond, kwam tot stand omdat militairen periodiek worden herkeurd. Vanaf hun aanstelling dienen zij vijfjaarlijks opnieuw te worden gekeurd. Het moment van aanstelling is dus bepalend voor het feit dat deze bewuste dossiers op de stapel lagen. Het gaat daarbij om alle militairen die vijf jaar daarvoor (of een veelvoud daarvan) zijn opgekomen. Er is echter geen aanleiding om te veronderstellen dat deze militairen een afwijkend patroon in ziekten of gebreken zouden hebben ten opzichte van andere militairen die reeds voordien waren gekeurd of nog gekeurd moesten worden. In dat opzicht zou er wel sprake kunnen zijn van selectiebias, maar die bias zou geen invloed uitoefenen op het resultaat.

Het verschil in ervaring tussen beide beoordelaars wordt in theorie opgeheven door de uitvoerige werkinstructies inzake het toekennen van de ABOHZIS-classificatie en de bijbehorende beperkingen (paragrafen 1.3.1 en 2.2).

De oorspronkelijke beoordelaar heeft de militair in persoon gezien. Beoordelaar 2 en 3 hebben dat niet; hun oordeel komt tot stand op basis van alleen schriftelijke informatie. Door het overnemen van alle informatie uit de dossiers is getracht om het beeld voor deze artsen te completeren, maar daarmee wordt de fysieke afwezigheid (non-verbale en verbale communicatie) niet bestreden. Dit probleem heeft zich echter evenzeer bij het voor deze studie als voorbeeld geldend onderzoek van Post-Uiterweer en De Kort [1] voorgedaan. Deze onderzoeksmethode wordt vaker gehanteerd om de mate van overeenstemming vast te stellen. Bovendien zijn keurend artsen bij de KL door de gehanteerde werkwijze (overname van informatie uit het meegeleverde medisch dossier) voor een gedeelte gewend om zich met schriftelijke gegevens te behelpen.

De artsen zijn het met elkaar eens over de PGO's waar oorspronkelijk geen beperkingen werden vastgesteld. Het feit dat keurend artsen onderling weinig beoordelingsverschillen vertonen indien zij een militair zonder ziekten of gebreken keuren, zal weinig verbazing wekken. De onderlinge verschillen nemen echter toe naarmate de medische problematiek groter wordt – en derhalve ook de meegeleverde correspondentie toeneemt. Dit duidt eerder op verschillen in interpretatie dan op verschillen op grond van informatie-bias.

Een ander effect dat kan zijn opgetreden, omdat de eerste keurend arts de gekeurde wel en de twee andere niet hebben gezien, is het Hawthorne-effect

ten opzichte van de eerste keurend arts. Deze zou door de onderzoekspersoon ten aanzien van de conclusies kunnen zijn gemanipuleerd. Van de andere twee artsen valt niet uit te sluiten dat zij zich sociaalwenselijk hebben opgesteld ("misschien verwacht de onderzoeker wel dat wij tot dezelfde conclusies komen"). De invloed van beide effecten valt niet te meten; aangenomen wordt dat die invloed beperkt is geweest.

De verdeling van zowel de afwijkende bevindingen (in ABHOHZIS-classificaties) als de mate ervan (1, 2 of x) en de daaraan gekoppelde beperkingen correspondeert goed met de resultaten van het loopbaanonderzoek onder 13.500 militairen in 1994 [10], waaraan onder de inleiding werd gerefereerd.

De ABOHZIS-classificatie wordt in theorie aangevuld met de "Aanwijzingen", hetgeen ertoe zou moeten leiden dat bij soortgelijke keuringsbevindingen telkens eenzelfde ABOHZIS-classificatie wordt opgemaakt. Soortgelijke "Aanwijzingen" bestaan niet voor het waarborgen dat bij dezelfde ziekten of gebreken ook telkens dezelfde beperkingen worden vastgesteld. Toch is bij de vaststelling van medische beperkingen de overeenstemming tussen de artsen vergelijkbaar met die bij de vaststelling van de ABOHZIS-classificatie. Dat sluit aan bij de conclusie van De Kort [2] dat meer gedetailleerde instructie slechts geringe invloed heeft op de mate van overeenstemming.

In tegenstelling tot de geringe praktische betekenis van de ABOHZIS-classificatie (er wordt op geen enkele wijze door de KL-organisatie van de ABOHZIS-classificaties van PGO-keuringen gebruik gemaakt) wordt door de KL wel grote betekenis toegekend aan de medische beperkingen. Zij vormen de basis voor de op te stellen functioneringsbeperkingen en kunnen aanleiding zijn om een militair geen nieuwe functie toe te wijzen, met als gevolg dat de betrokken militair sterk gereduceerde kansen heeft op behoud van zijn werk.

Men zou bezwaar tegen dit gehele hoofdstuk kunnen maken, omdat hier de reproduceerbaarheid is onderzocht van een keuringssysteem, dat momenteel niet langer in gebruik is. In de inleiding is reeds aangegeven dat het als de publicatie van een niet-systematische waarneming moet worden beschouwd. Het motief voor volledige weergave ervan in dit proefschrift is dat dit de eerste schrede op weg naar meer onderzoek was; zonder dit onderzoek zouden de overige hoofdstukken en publicaties nimmer zijn verschenen.

Tijdens dit onderzoek zijn vragen gerezen inzake de aard en inhoud van de gehanteerde lijst met medische beperkingen. Gezien de waarde die de KL eraan hecht lijkt het van het hoogste belang om juist het gebruik hiervan te protocolleren. Gelet op het teleurstellende resultaat van het protocolleren van de ABOHZIS-classificatie (die immers geen betere overeenstemming oplevert dan

de ongeprotocolleerde medische beperkingen) heeft een dergelijke protocolle-ring alleen zin indien die medische beperkingen functie(groep)-gericht zijn en kunnen worden afgezet tegenover functie-eisen.

Op grond van het bovenstaande wordt geconcludeerd dat de ABOHZIS-syste-matiek, waarvan de nadelen al langer bekend waren, zelfs met behulp van de "aanwijzingen" niet langer aan de huidige behoeften voldoet.

Geraadpleegde literatuur

1. Kort WLAM, Post-Uiterweer HW: Project aanstellingskeuringen dossieronderzoek bij de RBB. Deel II: panelonderzoek. Ministerie van Sociale Zaken en Werkgelegenheid, S 53-2, 1991.
2. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
3. Belloni R: De bepaling van de dienstgeschiktheid. Utrecht, 1985.
4. Bakker LF: Militaire keuringen en het ABOHZIS-systeem. Nederl Mil Geneesk T 1972; 25: 284-290.
5. Kessel JGFM van: Consequenties van onjuist keuren. Ned Mil Geneesk T; 1984 (37): 176-228.
6. Tafels Wiskunde Statistiek. Groningen: Wolters-Noordhoff, 1973.
7. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Houten: Bohn, Stafleu en van Loghum 1995.
8. Schouten HJA: Nominal scale agreement among observers. Psychometrika 1986; 51; 3: 453-466.
9. Schouten HJA: Klinische statistiek. Houten: Bohn Stafleu Van Loghum, 1995.
10. Raad J de, Blankestijn CDA: Loopbaanprognose bij de Koninklijke Landmacht: medische beperkingen. Carré 1995; 7/8: 10-13, 9: 14-16, 10: 20-24.

3 | Een onderzoek naar de reproduceerbaarheid van de basismedische eisen KL bij de pensioenkeuring

In belangrijke mate gebaseerd op: Raad J. de, Peters JHM., Nijhuis FJN:
De kwaliteit van een keuringssysteem. TBV 2000; 8:105-109.

3.1 Inleiding en vraagstelling van het onderzoek

In november 1988 is bij de Koninklijke Landmacht (KL) een nieuw keurings-systeem ingevoerd. Dit systeem bestaat uit 43 beoordelingspunten, waarbij aan elk beoordelingspunt een minimaal te behalen waarde is toegekend: de Basis-medische Eis KL (BMEKL). Deze Basismedische Eisen vormen de operationa-lisatie van de door de Bevelhebber der Landstrijdkrachten – de werkgever – van kracht verklaarde “basiseisen ten behoeve van militair KL-personeel” (“militaire basiseisen”). Bij het opstellen van de Basismedische Eisen is gebruik gemaakt van zowel het belastbaarheidspatroon van het Functie Informatie Systeem (FIS) als van een aantal specifieke militaire beoordelingspunten[1]. De intentie bij het operationaliseren was om volgens het principe van “belasting-belastbaarheid” [2] op medisch gebied vorm te geven aan de taakstelling van de militair ambte-naar zoals verwoord in de Wet op de Medische Keuringen. Dit keuringssysteem verving het tot dan toe in gebruik zijnde diagnose-gerelateerde zogeheten ABOHZIS-systeem.

Inmiddels had onderzoek van onder meer De Kort [3] aangetoond dat op aanstellingskeuringen in het algemeen de nodige kritiek viel te leveren. Vervol-gens had Reulings [4] de mate van overeenstemming van de diagnose-gerela-teerde ABOHZIS-keuringen binnen de KL in het bijzonder onderzocht. Die mate van overeenstemming bleek overeen te komen met de (teleurstellende) resultaten van het onderzoek van De Kort [3].

Na ingebruikneming van het nieuwe BMEKL-keuringssysteem rees dan ook de vraag of dit systeem wél tot voldoende overeenstemming zou leiden. Bij het beantwoorden van die vraag moest rekening worden gehouden met het feit dat dit keuringssysteem niet alleen wordt gebruikt voor de medische aanstellings-keuring, maar ook voor het medisch afkeuren van (post-)actief dienend perso-neel: de pensioenkeuring.

Elke KL-militair dient “algemeen en breed inzetbaar en uitzendbaar” te zijn. Daarom wordt niet alleen het ziekteverzuim geregistreerd, maar ook andere vormen van arbeidsverzuim (chronische blessures e.d.), die de operationele inzetbaarheid verminderen. Indien de operationele inzetbaarheid door een vorm van verzuim langdurig wordt aangetast, ongeacht de mate waarin de vredesfunctie wordt uitgevoerd, kan de werkgever de militair in kwestie voor een pensioenkeuring laten oproepen. Op het terrein van de sociale wetgeving is voor militairen naast de Ziektewet (ZW) en de Wet op de Arbeidsongeschikt-heid (WAO) ook het “Kaderbesluit Militaire Pensioenen” van toepassing. De daarbij behorende pensioenkeuring – beoordeling van dienstgeschiktheid en het recht op dienstverbanduitkeringen – vindt plaats bij de Verzekeringsgenees-kundige Dienst KL (VGDKL), in commissoriaal verband.

Bij dit onderzoek is de mate van overeenstemming van het nieuwe keuringssysteem bij de pensioenkeuring onderzocht. De vraagstelling voor dit onderzoek luidde: "Voldoet de reproduceerbaarheid van het nieuwe keuringssysteem bij de pensioenkeuring aan de normering die gebruikelijk is voor screeningstesten?"

3.2 Materiaal en methoden

3.2.1 Populatie en procedure

Er is voor gekozen om gebruik te maken van keuringsdossiers van militairen die een pensioenkeuring (Militair Geneeskundig Onderzoek, MGO) hebben ondergaan. Deze MGO's waren door een commissie van drie verzekeringsartsen nog beoordeeld volgens de oude systematiek van de ABOHZIS-classificatie. Om die reden kon de oorspronkelijke uitslag niet in dit onderzoek worden meegenomen.

De meerderheid van de voor een pensioenkeuring in aanmerking komende militairen wordt afgekeurd. Om inzicht te krijgen in de overeenstemming van zowel geschikt als ongeschikt verklaarden is een gelede steekproef genomen van de circa 400 dossiers van het afgelopen jaar. Zowel de geschikt verklaarden als de ongeschikt verklaarden werden apart genummerd, waarna een medewerker tweemaal vijftien toevalsgetallen uitkoos. De dossiers zijn geanonimiseerd en de conclusies zijn verwijderd (figuur 12).

Figuur 12: De onderzoeksopzet van hoofdstuk 3.

In hoofdstuk 3 wordt de reproduceerbaarheid van de pensioenkeuring onderzocht, analoog aan de wijze waarop dat in Figuur 5 voor het periodiek geneeskundig onderzoek heeft plaatsgevonden.

Deze keuringsdossiers zijn vervolgens voorgelegd aan twee seniorbedrijfsartsen van de Arbo-dienst KL met langdurige keuringservaring, met het verzoek om volgens de nieuwe keuringsmethode tot een eindoordeel te komen over de militaire dienstgeschiktheid.

De betrokken artsen hadden nog weinig praktijkervaring met het nieuwe keuringssysteem. Eén van hen had aan de totstandkoming van de basismedische eisen KL meegewerkt. Ter reductie van eventuele bias is hun kennis voorafgaande het onderzoek op hetzelfde peil gebracht.

De artsen moesten een eindoordeel vellen inzake de dienst(on-)geschiktheid. Dit oordeel komt tot stand doordat de bedrijfsarts beoordeelt of de betrokken militair al dan niet voldoet aan de Basismedische Eis van elk van de 43 beoordelingspunten. Die Basismedische Eis ligt op het prestatieniveau waardoor de geschikt verklaarde militair in staat moet worden geacht om het grootste deel van de beschikbare arbeid in ons land zonder beperkingen te kunnen verrichten. Dit betekent dat voor alle beoordelingspunten een dichotome beslissing wordt gevraagd: men is wel of niet geschikt.

3.2.2 Statistische methoden

Voor dit onderzoek worden twee maten van overeenstemming gebruikt: de kappa [5] en het percentage discordante of discongruente uitslagen (PD): $(GO + OG / GG + GO + OG + OO) \times 100\%$. De waarde van kappa ($P_o - P_e / 1 - P_e$; hierbij staat P_o voor de waargenomen overeenstemming en P_e voor de toevallige overeenstemming) varieert tussen 0 en 1. Volgens Bouter en Van Dongen [6] suggereren gepubliceerde onderzoeksgegevens dat een kappa tussen de 0,40 en 0,75 voor diagnostische testen in de kliniek gebruikelijk is. Regelmatig wordt bij de interpretatie van waarden van kappa een onderverdeling gehanteerd. Daarbij zou een kappa $< 0,40$ betekenen dat de mate van overeenstemming gering is, een kappa tussen 0,40 en 0,60 dat de mate van overeenstemming redelijk is, een kappa tussen 0,60 en 0,75 dat de mate van overeenstemming goed is en een kappa van meer dan 0,75 dat de mate van overeenstemming voortreffelijk is. De kappa is afhankelijk van een aantal factoren, zoals het aantal beoordelingscategorieën (strata): hoe meer strata des te kleiner de kappa [6]. Voorts is de kappa afhankelijk van de te verwachten prevalentie van elk van de beoordelingsuitkomsten [6, 7, 8]. Als een aandoening zeldzaam is zal de kappa laag blijven, ook bij een hoge mate van overeenstemming [7, 8]. Bij het berekenen van de kappa wordt telkens ook het 95%-betrouwbaarheidsinterval weergegeven [8].

3.3 Resultaten

De resultaten van zowel de afzonderlijke beoordelingspunten als het integrale eindoordeel worden in onderstaande tabellen weergegeven. In tabel 13 wordt de mate van overeenstemming van de beoordelingspunten van het bewegingsapparaat weergegeven. Van de 16 beoordelingspunten van tabel 13 vertonen 6 goede overeenstemming, 6 een redelijke en 4 een geringe overeenstemming. Het percentage discrepante uitslagen varieert tussen de 3,3% en 30,0%.

Tabel 13. De overeenstemming per beoordelingspunt (FIS-“bewegingsapparaat”) uitgedrukt in de kappa, het 95%-betrouwbaarheidsinterval en het percentage discrepante antwoorden PD tussen beide artsen.

Nr	Beoordelingspunt	Kappa	95% Betrouwbaarheidsinterval	Overeenstemming	PD%
1	Zitten	0,29	-0.18-0.76	Gering	13,3
2	Staan	0,76	0.50-1.02	Goed	10,0
3	Lopen	0,32	0.02-0.62	Gering	26,7
4	Trappenlopen	0,22	-0.06-0.48	Gering	30,0
5	Klimmen en klauteren	0,42	0.11-0.73	Redelijk	23,3
6	Knielen, kruipen, en hurken	0,34	0.06-0.62	Gering	23,3
7	Gebogen werken	0,47	0.10-0.84	Redelijk	16,7
8	Kortcyclisch buigen en torderen	0,41	0.05-0.77	Redelijk	20,0
9	Gebruik van de nek	0,53	0.07-0.99	Redelijk	10,0
10	Reiken	0,62	0.23-0.91	Goed	10,0
11	Bovenhands werken	0,67	0.33-1.01	Goed	10,0
12	Hand- en vingergebruik	0,78	0.36-1.20	Goed	3,3
13	Tillen	0,68	0.39-0.97	Goed	13,3
14	Duwen en trekken	0,55	0.25-0.85	Redelijk	20,0
15	Dragen	0,71	0.45-0.97	Goed	13,3
24	Vibratiebelasting	0,41	0.06-0.76	Redelijk	10,0

Tabel 14. De mate van overeenstemming per beoordelingspunt (FIS-Psychische belasting) uitgedrukt in de kappa, het 95%- betrouwbaarheidsinterval en het percentage discrepante antwoorden PD tussen beide artsen.

Nr	Beoordelingspunt	Kappa	95% Betrouwbaarheidsinterval	Overeenstemming	PD%
27	Persoonlijk risico	0,13	-0.14-0.40	Gering	36,7
28	Psychische belasting	0,60	0.28-0.92	Goed	16,7
28A	Werken onder tijdsdruk	0,51	0.14-0.88	Redelijk	16,7
28B	Dwingend werktempo	0,23	-0.18-0.64	Gering	23,3
28C	Gedwongen nietsdoen	0,28	-0.25-0.79	Gering	13,3
28D	Conflicterende functie-eisen	0,78	0.51-1.07	Goed	6,7
28E	Conflicthantering	0,45	0.25-0.85	Redelijk	16,7
28F	Verrichten monotoon werk	0,06	-0.38-0.46	Gering	13,3
28G	Contact/omgang met anderen	0,35	-0.23-0.93	Gering	10,0
28H	Verantwoordelijkheid	0,71	0.44-1.08	Goed	6,7
28I	Omgeving en lawaai	0,59	0.24-0.94	Redelijk	16,7
28J	Geen mogelijkheid tot contact	0,63	0.15-0.87	Goed	6,7
28K	Emotionele belasting	0,67	0.32-1.02	Goed	10,0
28L	Psychiatrische stoornissen	0,45	0.05-0.85	Redelijk	16,7
28M	Operationele inzet mogelijk	0,52	0.24-0.80	Redelijk	23,0
28N	Verslaving	1,0		Voortreffelijk	0,0

Tabel 14 toont de mate van overeenstemming van de beoordelingspunten van de psychische belasting. Van de 16 beoordelingspunten vertoont 1 een voortreffelijke overeenstemming; 5 vertonen goede overeenstemming, 5 een redelijke en 5 een geringe overeenstemming. De PD varieert tussen de 0,0 en 36,7%.

Tabel 15 laat de mate van overeenstemming zien van de van de "militaire" beoordelingspunten. Van de 11 beoordelingspunten vertonen 2 een voortreffelijke overeenstemming; 3 vertonen een goede overeenstemming, 3 een redelijke en 3 een geringe overeenstemming. De PD varieert tussen de 0,0 en 30,0%.

Van een aantal punten (16 tot en met 23, 25, 26, 34, 35, 36, 42) kon de kappa niet worden berekend omdat tenminste één van beide artsen uitsluitend positieve ("Voldoet aan de gestelde eis") uitslagen versterkte. Als gevolg daarvan is de waargenomen overeenstemming (P_o) minus de toevallige overeenstemming

Tabel 15. De overeenstemming per beoordelingspunt ("militaire" beoordelingspunten) uitgedrukt in de kappa, het 95%-betrouwbaarheidsinterval en het percentage discrepante antwoorden PD tussen beide artsen.

Nr	Beoordelingspunt	Kappa	95% Betrouwbaarheidsinterval	Overeenstemming	PD%
29	Algemene conditie	0,78	0.55-1.01	Goed	10,0
30	Lopen met bepakking	0,86	0.66-1.06	Voortreffelijk	6,7
31	Graven	0,55	0.55-1.01	Redelijk	10,0
32	Springen	0,36	0.04-0.68	Gering	23,3
33	Optreden te velde	0,37	0.10-0.64	Gering	30,0
37	Gesaneerd gebit	1,0		Voortreffelijk	0,0
38	Scherfwerend vest dragen	0,46	0.16-0.76	Redelijk	23,3
39	Helm dragen	0,71	0.43-1.09	Goed	6,7
40	Hanteren vuurwapen	0,78	0.51-1.07	Goed	6,7
41	Onregelmatig functioneren	0,30	0.04-0.66	Gering	33,3
43	Geneesmiddel afhankelijkheid	0,54	0.14-0.74	Redelijk	13,3

Tabel 16. Het percentage discrepante antwoorden PD tussen beide artsen voor die beoordelingspunten, waarvan de kappa niet kan worden berekend.

Nr	Beoordelingspunt	PD%
16	Luchtverplaatsing	10,0
17	Stof, rook gas en damp	0,0
18	Koude-tolerantie	13,3
19	Warmte-tolerantie	6,7
20	Temperatuurswisselingen	13,3
21	Luchtverplaatsing	10,0
22	Droge lucht	3,3
23	Huidcontact	0,0
25	Gebruik van de zintuigen	0,0
26	Beschermende middelen	3,3
34	NBC-bescherming	16,7
35	Tropische omstandigheden	0,0
36	Spreekvaardigheid	3,3
42	Verdragen dagelijkse voeding	3,3

(P_e) gelijk aan nul, zodat de waarde van kappa (zijnde $P_o - P_e / 1 - P_e$) niet kan worden berekend. Van deze punten is alleen het percentage discrepante antwoorden berekend en in tabel 16 weergegeven.

In figuur 13 wordt de mate van overeenstemming inzake het eindoordeel weergegeven. De kappa bedraagt 0,74 en de PD is 13,3%.

Eindoordeel dienstgeschiktheid: **Kappa= 0,74** (0.48–0.98)
Interpretatie overeenstemming: **goed**. PD: 13,3%

Figuur 13. Eindoordeel dienstgeschiktheid, uitgedrukt in kappa en PD.

3.4 Beschouwing

Het uiteindelijke doel van de ontwikkeling van de Basismedische Eisen was om te kunnen keuren, gericht op de aard van de werkelijke militaire activiteiten, inclusief het operationeel optreden. Dit functiegerichte keuren dient tenminste gepaard te gaan met een aanvaardbare mate van onderlinge overeenstemming. Alhoewel de validiteit van het keuringssysteem hiermee nog niet is onderzocht, is goede reproduceerbaarheid een vereiste voorwaarde voor een voldoende mate van validiteit [6].

Kijkend naar de kappa kan worden gesteld dat de mate van overeenstemming voor wat betreft het eindoordeel van de dienstgeschiktheid goed is. Voor een belangrijk deel van de separate Basismedische Eisen is de overeenstemming redelijk tot zelfs hoog te noemen. Dit geldt zowel ten aanzien van de FIS-gedeelten van de Basismedische Eisen inzake het bewegingsapparaat en de psychische belasting, als ten aanzien van de specifiek militaire Basismedische Eisen.

Ten aanzien van de percentages discordante uitslagen wordt de vergelijking getrokken met de uitkomsten van De Kort [3] bij het panelonderzoek onder RBB-artsen. Helemaal eerlijk is die vergelijking overigens niet, want bij het gememoreerde panelonderzoek waren vijf artsen betrokken en niet zoals hier slechts twee, waardoor de kans op onderlinge spreiding van de resultaten in dit onderzoek geringer is. De PD-scores uit dit onderzoek moeten dan ook eerder als indicatief worden beschouwd. In het onderzoek onder de RBB-artsen werd een gemiddelde PD-waarde van $37 \pm 4\%$ gevonden. In dit onderzoek liggen de PD-waarden voor 41 van de 43 beoordelingspunten alsmede het eindoordeel beduidend lager dan in dat van De Kort [3]. Alleen voor de beoordelingspunten, 27 (persoonlijk risico) en 42 (onregelmatig functioneren) is de score vergelijkbaar.

De beoordelingspunten kunnen globaal in drie gedeelten worden onderscheiden; zoals de tabellen aangeven: een cluster "FIS-bewegingsapparaat", een cluster psychische belastbaarheid uitgedrukt in de punten 27 en 28 en een cluster met specifiek militaire eisen. Bij 29 van de 43 Basismedische Eisen (67%) kan een uitspraak worden gedaan geven over de mate van overeenstemming met behulp van de kappa. Daaronder is bij 75% van de beoordelingspunten sprake van een redelijke of hogere mate van overeenstemming.

Bij 12 beoordelingspunten is de mate van overeenstemming gering ($\text{kappa} < 0,40$). Zoals reeds in paragraaf 3.2.2. werd gesteld, is de waarde van kappa afhankelijk van meerdere factoren, zoals de prevalentie van de bepaalde aandoening [7, 8]. Als deze bepaalde aandoening weinig voorkomt, dan lopen ook de daaraan gerelateerde beoordelingspunten het risico bij een beoordeling op overeenstemming met behulp van kappa een lage score te vertonen. Gegeven het in dit onderzoek beperkt aantal gebruikte dossiers valt niet uit te sluiten dat dit bij deze beoordelingspunten het geval is geweest. Anderzijds kan hier het gebrek aan praktijkervaring een rol hebben gespeeld, maar dit lijkt, gelet op de overige uitkomsten, weinig waarschijnlijk. Tegelijk met de lage waarde van de kappa blijkt bij deze beoordelingspunten de waarde van het percentage discrepante uitslagen niet van die van de overige beoordelingspunten te verschillen. Over het algemeen gaat een lage kappa gepaard met een hoge waarde van PD en omgekeerd. Ook deze bevinding pleit tegen een werkelijk gebrek aan overeenstemming, maar gaat meer in de richting van een lage prevalentie.

Het gebruik van MGO-keuringsdossiers van pensioenkeuringen speelt een soortgelijke rol in de afwezigheid van overeenstemming in de vorm van kappa bij de punten uit tabel 16. Pensioenkeuringen worden alleen uitgevoerd indien men in werkelijke dienst is geweest. Bepaalde aandoeningen – zoals zintuiglijke gebreken of ernstige cariës – zullen al bij de aanstellingskeuringen tot afwijzing leiden. Derhalve komen diverse aandoeningen praktisch niet voor als afkeuringgrond bij de pensioenkeuring. De PD%-scores uit tabel 15 zijn dan ook relatief laag en variëren tussen de 0,0 en 16,7%. Het feit dat de PD%-scores niet allemaal nul zijn impliceert overigens dat de artsen daarin toch soms onderlinge verschillen vertonen.

Door de aselecte gelede steekproef is in de sortering van de dossiers selectiebias voorkomen. Het gebruik van zowel dossiers van geschikt als ongeschikt verklaarden vond opzettelijk plaats, conform het onderzoek van De Kort en Post-Uiterweer [3].

Van de twee nieuwe beoordelaars had er één aan de totstandkoming van het nieuwe keuringssysteem meegeholpen, de andere niet. Zoals reeds is vermeld is

voorafgaande het onderzoek hun kennis over het nieuwe systeem op hetzelfde niveau gebracht. Het feit dat zij hierna met het nieuwe systeem konden werken sluit aan bij de ervaring die enige tijd later werd opgedaan tijdens de workshops, bedoeld om alle betrokken artsen met het nieuwe systeem te leren werken. Tijdens die bijeenkomsten – afgesloten met een toets – bleek dat alle artsen na afloop ronduit goed met de nieuwe werkwijze overweg konden. Zou bias door een verschil in kennis tussen beide artsen een rol hebben gespeeld, dan zou men bovendien eerder een lagere mate van overeenstemming hebben verwacht. Daarvan is op grond van de gevonden overeenkomst inzake het eindoordeel tussen beide artsen in dit onderzoek geen sprake.

Het beoordelen van schriftelijke dossiers om de mate van overeenstemming tussen artsen te beoordelen is een hiervoor regelmatig gebruikte methode. In het gebruik van schriftelijke dossiers schuilt mogelijk informatiebias: men keurt immers niet fysiek en mist wellicht (non-verbale en verbale communicatieve) signalen. Het voordeel van deze methode is dat de betrokken artsen per definitie uitgaan van dezelfde gegevens. Bovendien vindt de feitelijke pensioenkeuring in commissoriaal verband plaats, waarbij drie artsen betrokken zijn. Dit betekent niet dat de militair in kwestie door drie artsen wordt onderzocht, maar dat het door één arts opgestelde keuringsdossier door twee andere artsen wordt doorgenomen en (al dan niet) wordt onderschreven. Het is voor artsen bij de KL derhalve niet ongebruikelijk om een beoordeling te verrichten met behulp van alleen het dossier. Op grond hiervan wordt aangenomen dat informatiebias de onderzoeksresultaten niet onaanvaardbaar heeft beïnvloed.

Door het gebruik van dossiers in plaats van onderzoekspersonen worden bovendien het Hawthorne-effect [6], het placebo-effect [6] en het Pygmalion-effect [6] waarschijnlijk juist vermeden.

De samenstelling van de dossiers was niet representatief voor de gebruikelijke samenstelling van dergelijke dossiers. In de grote meerderheid van de gevallen leidt een pensioenkeuring ook effectief tot afkeuring en niet – zoals hier het geval was – maar in 50% van de gevallen. De uitkomsten kunnen dus niet één op één van toepassing worden verklaard op de concrete situatie bij de pensioenkeuring. Deze procedure is analoog aan die welke De Kort [3] heeft gebruikt.

Wat betreft een aantal individuele beoordelingspunten heerste bij de beoordeling van de dossiers een hoge mate van overeenstemming over de kwalificatie “betrokkene voldoet aan de Basismedische Eis”. Echter, de artsen zijn het nooit met elkaar eens wanneer iemand niet voldoet aan de Basismedische Eis. In die gevallen geldt dat $P_o = P_e$ en volgt er geen waarde uit de berekening van Kappa. Het ontbreken van deze eensgezindheid, tezamen met een hoge mate van eensgezindheid wanneer er geen sprake is van een beperking, doet evenwel een goede overeenstemming vermoeden.

De beslissing die de keurend arts in dit kader moest nemen was dichotoom: men is wel of niet geschikt. Schouten [9] stelde dat de waarde van Kappa hoger uitvalt bij een dichotome keuze-mogelijkheid dan wanneer er meerdere keuze-mogelijkheden zijn. Dit zou tot een lagere mate van overeenstemming kunnen leiden als er in de toekomst functie-eisen worden ontwikkeld, die hoger liggen dan het niveau van de Basismedische Eisen. Gelet op de te signaleren verschillen in taakstelling tussen de verschillende soorten KL-militairen (zoals bijvoorbeeld het Korps Commando Troepen of het Explosieven Opruimings Commando KL), zal het ontwikkelen van meerdere soorten functie-eisen uiteindelijk toch moeten plaatsvinden.

Samengevat kan worden geconcludeerd dat uit dit onderzoek blijkt dat de inter-waarnemer-variatie van het nieuwe keuringssysteem ruimschoots voldoet aan de normering die gebruikelijk is voor screeningstesten. De kappa van het eindoordeel ligt op 0,74, zodat daarin zelfs sprake is van goede overeenstemming.

Daarbij kunnen de volgende aanbevelingen worden gedaan. In de eerste plaats kan in een later stadium nog nader onderzoek worden verricht naar de beoordelingspunten die thans niet met behulp van de kappa te beoordelen waren, respectievelijk een te lage kappa hadden. Dit onderzoek zal vermoedelijk beter bij de aanstellingskeuring kunnen worden verricht.

Voorts zijn de Basismedische Eisen bedoeld als zo goed mogelijke afspiegeling van de concrete militaire (operationele) taakstelling. Indien die taakstelling wordt gewijzigd – bijvoorbeeld als het huidige scherfwerend vest van 15 kg wordt vervangen door een veel lichter exemplaar of verplaatsing uitsluitend nog gemotoriseerd zou plaatsvinden – dan dienen die eisen te worden herbezien in het oog van de vernieuwde taakstelling.

Geraadpleegde literatuur

1. Raad J de, Peters JHM, Groot RP de et al: Van diagnose-gerelateerd naar functie-gericht keuren. TBV 2000; 8:110-115.
2. Dijk FJH van, Dormolen M van, Kompier MAJ et al: Herwaardering model belasting-belastbaarheid. TSG, 1990; 68: 3-10.
3. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
4. Raad J de, Reulings PGJ, Schouten HJA: De beperkingen van medische beperkingen. De reproduceerbaarheid van de beoordeling van de ABOHZIS-classificatie en de medische beperkingen bij periodiek geneeskundig onderzoek bij de Koninklijke Landmacht. Ned Mil Geneesk T; 1996; 49: 106-111.
5. Nijdam B, Buuren H. van: Statistiek voor de sociale wetenschap. Alphen aan de Rijn: Samson, 1979.

6. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Houten: Bohn, Stafleu en van Loghum 1995.
7. Altman DG: Practical statistics for medical research. London: Chapman and Hall, 1991.
8. Steinijans VW, Diletti E, Bömches B et al: Interobserver agreement: Cohens Kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J of Clin of Pharm And Ther.* 1997; 35; 3: 93 – 95.
9. Schouten HJA: Nominal scale agreement among observers. *Psychometrika* 1986; 51; 3: 453-466.

4 | De reproduceerbaarheid van de aanstellingskeuring op grond van de Basismedische eisen en het ABOHZIS-systeem

In belangrijke mate gebaseerd op: Raad J de, Slootman EJH, Groot RP de: Kwaliteitsverbetering bij de Koninklijke Landmacht. Onderzoek naar de kwaliteit van het beoordelen van aanstellingskeuringen voor de Koninklijke Landmacht. TBV 2002; 10: 266-272.

4.1 Inleiding en vraagstelling van dit onderzoek

Na ingebruikneming van het nieuwe keuringssysteem van de Basismedische Eisen van de Koninklijke Landmacht (BMEKL) is een onderzoek uitgevoerd met pensioenkeuringen (Hoofdstuk 3). Daaruit bleek dat voor wat betreft het eindoordeel er sprake was van een goede overeenstemming. Daarmee is althans een gedeelte van de vraagstelling uit hoofdstuk 1 beantwoord. De BMEKL worden echter niet alleen gehanteerd bij de pensioenkeuring maar ook bij de aanstellingskeuring. Omdat, zoals in de inleiding al is opgemerkt, het keuringssysteem zowel wordt gehanteerd bij de aanstellingskeuring als bij de pensioenkeuring moet ook voor beide situaties de reproduceerbaarheid worden onderzocht. De procedure van de aanstellingskeuring verloopt anders dan die van de pensioenkeuring, daarnaast is sprake van een andere populatie. Bij de pensioenkeuring heeft elke keurling (bepaalde) medische bijzonderheden, vaak leidend tot beperkingen, reden waarom de meerderheid van de personen in kwestie wordt afgekeurd. Bij de aanstellingskeuring daarentegen vertoont de meerderheid der kandidaten geen medische bijzonderheden en wordt dan ook geschikt verklaard. Verder wordt de pensioenkeuring altijd voor actief dienend of postactief militair personeel (dat derhalve in het verleden geschikt is verklaard na de aanstellingskeuring) uitgevoerd, terwijl de aanstellingskeuring voor nieuw aan te nemen personeel wordt uitgevoerd. Zoals reeds in hoofdstuk 3 is uiteengezet betekent dit dat bepaalde medische bijzonderheden in de praktijk bij actief dienend militair personeel niet voorkomen (zoals ernstige vormen van kleurenblindheid), omdat dit bij de aanstellingskeuring al tot afkeuring leidt. Daarnaast bestaat vermoedelijk ook een leeftijdsverschil tussen het nieuw aan te nemen personeel en met name dat actief dienend personeel dat wordt afgekeurd. Om een completer beeld te krijgen van de reproduceerbaarheid van het nieuwe keuringssysteem volstaat het onderzoek uit het voorgaande hoofdstuk 3 niet en dient daarnaast ook onderzoek te worden verricht naar de mate van overeenstemming bij de aanstellingskeuring.

Zoals in hoofdstuk 1 werd toegelicht kan voor dit onderzoek gebruik worden gemaakt van dezelfde populatie die dient voor de gerandomiseerde, gecontroleerde studie, weergegeven in de hoofdstukken 5, 6 en 7. In de studie van dit hoofdstuk zullen de uitkomsten van twee studiegroepen worden vergeleken, die gerandomiseerd middels het oude (ABOHZIS) ofwel het nieuwe keuringssysteem zijn gekeurd. Hierdoor wordt het mogelijk om de mate van overeenstemming van beide keuringssystemen, zoals zij voor de aanstellingskeuring worden gehanteerd, met elkaar te vergelijken.

De vraagstelling voor dit onderzoek luidt: "Hoe verhoudt zich de mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring?"

4.2 Achtergrondinformatie

Sinds de opschorting van de dienstplicht zijn er twee soorten verbintenissen voor militairen: een beroepscontract van onbepaalde tijd ("BOT") of een beroepscontract van bepaalde tijd ("BBT"), dit laatste kent over het algemeen een standaard looptijd van 2½ jaar. Werving, selectie en de medische aanstellingskeuring vinden centraal plaats bij het Instituut Keuring en Selectie (IKS) van de krijgsmacht te Amsterdam. Hierbij vinden achtereenvolgens plaats de administratieve selectie (kennis Nederlandse taal, minimaal vereiste opleiding) en de psychologische selectie (capaciteitentest [1, 2], respectievelijk een persoonlijkheidsvragentest [3, 4]). Hierna worden de kandidaten voor alle vier de krijgsmachtdelen volstrekt willekeurig ingedeeld voor de medische aanstellingskeuring. Bij die medische aanstellingskeuring speelt bij de KL ook de fysieke selectie een rol.

De medische aanstellingskeuring omvatte ten tijde van deze studie: urineonderzoek, audiometrisch onderzoek naar zowel de lucht- als de beengeleiding van 250 Hz tot 8kHz, optometrisch onderzoek ter bepaling van de gezichtsscherpte, het stereoscopisch diepte zien en het kleuronderscheidingsvermogen, een rust-ecg (ter preventie van cardiale problemen bij de bepaling van de VO₂max), de meting van lichaamsmaten (zoals het vetpercentage), onderzoek naar het bewegingsapparaat en de adnexusen door de fysiotherapeut, een anamnese en algemeen lichamelijk onderzoek door de keurend arts. De uitslag zelf wordt niet door de keurend arts gegeven, maar door de zogeheten "uitslaggevend arts". Dit is een seniorbedrijfsarts, met langdurige ervaring op het terrein van aanstellingskeuringen. Vooruitlopend op de invoering van het nieuwe keuringssysteem in november 1998, zijn alle betrokken (basis-, bedrijfs- en verzekeringsgeneeskundig) artsen van de KL vooraf bij wijze van introductie in een workshop getraind.

De fysieke selectie vindt plaats met behulp van lichaamsmaten (lichaamsgewicht, vetpercentage met de 4-punts huidplooiemeting), maximale kracht (Newtons) en uithoudingsvermogen (Watt/kg lichaamsgewicht). Op grond hiervan wordt een indeling gemaakt in bepaalde functieclusters van fysieke belastbaarheid.

Tijdens dit onderzoek was de minimumscore van de fysieke selectie "afgesteld" op de 10^e percentiel van de mannelijke kandidaten, alsmede op de 50^e percentiel van de vrouwelijke kandidaten. Daarmee bleef de ondergrens voor vrouwen in absolute zin lager dan die voor mannen.

4.3 Materiaal en methoden

4.3.1 Populatie en procedure

Vanaf 22 september tot en met 16 oktober 1998 – een periode van vier weken – zijn 863 kandidaten op de even weken volgens het ABOHZIS-systeem en op de oneven weken volgens de Basismedische Eisen gekeurd. Uit dit onderzoek zijn de 227 als eerste beschikbare keuringsformulieren, afkomstig uit de eerste weken, in drievoud gekopieerd (figuur 14). Vervolgens is aan de oorspronkelijke drie uitslaggevend bedrijfsartsen ongeveer anderhalf jaar later (2000) verzocht om opnieuw uitslag te geven inzake deze keuringen. De oorspronkelijke uitslag wordt hierbij gevormd door de gecombineerde uitslagen van de uitslaggevend artsen in 1998 (Beoordelaar “C”).

Figuur 14: De onderzoeksopzet van hoofdstuk 4. Met behulp van dossiers uit de gerandomiseerde, gecontroleerde studie wordt de reproduceerbaarheid van beide keuringssystemen voor wat betreft de aanstellingskeuring onderzocht. Hierbij worden zowel de nieuwe uitslagen onderling vergeleken (“Nieuw onderling”), als de nieuwe uitslagen met de oorspronkelijke uitslag (“Oud versus nieuw”, Arts C versus de Artsen 1, 2 en 3).

4.3.2 Statistische methoden

De reproduceerbaarheid kan met behulp van verschillende maten worden onderzocht[5, 6], af te leiden van tabel 16. Teneinde zoveel mogelijk informatie over de mate van overeenstemming te verkrijgen is voor dit onderzoek besloten

Tabel 17. De testuitslagen van twee verschillende beoordelaars (of tweemaal dezelfde beoordelaar) bij onderzoek in een populatie

		Beoordelaar 2	
Beoordelaar 1	Normale testuitslag	Normale testuitslag	Afwijkende testuitslag
	Afwijkende testuitslag	a	b
		c	d

om vier verschillende maten te hanteren. Twee maten, de kappa en het percentage discordante uitslagen, zijn reeds in hoofdstuk 2 beschreven. De overige twee te gebruiken maten worden aan de hand van tabel 17 beschreven.

Als met beoordelaar 1 en 2 twee verschillende onderzoekers worden bedoeld kan de intra-observers variatie worden berekend. Als met beoordelaar 1 en 2 dezelfde onderzoeker wordt bedoeld kan de intra-observers variatie worden berekend. Hieronder worden de twee nog niet behandelde maten voor overeenstemming weergegeven:

1. Het percentage overeenstemming. Deze maat wordt vooral berekend in geval onderzochte personen op basis van de testuitslag in een beperkt aantal categorieën wordt ingedeeld[5].

$$\text{waargenomen} = \frac{\text{aantal consistente waarnemingen}}{\text{totaal aantal waarnemingen}} \times 100\% = \frac{a + d}{a + b + c + d}$$

De op deze manier berekende mate van uniformiteit wordt beïnvloed door de door elk van de beide waarnemers geschatte prevalentie van afwijkende testuitslagen in de onderzochte populatie [5].

2. Men kan ook alleen de testuitslagen vergelijken door de testuitslagen die door één of door beide beoordelaars als afwijkend zijn bestempeld in de berekening te betrekken. Volgens Bouter en Van Dongen wordt zo een meer realistische indruk van de mate van overeenstemming verkregen[5]:

$$\frac{d}{b + c + d} \times 100\%$$

De oorspronkelijke testuitslagen zullen tegenover de nieuwe uitslagen van de drie artsen van de beide keuringssystemen worden beoordeeld. Omdat de

oorspronkelijke testuitslagen zijn verricht door dezelfde artsen die nu bij het onderzoek zijn betrokken, zou op deze manier de intra-waarnemers variatie kunnen worden beoordeeld. Helaas waren de dossiers dusdanig geanonimiseerd dat de individuele uitslaggevend arts niet meer te achterhalen viel. Hierdoor kan de intra-waarnemers variatie niet worden beoordeeld.

Ook de nieuwe testuitslagen zullen onderling worden vergeleken. Hieruit kan de inter-waarnemers variatie worden berekend.

4.4 Resultaten

4.4.1 Uitslagen

Van de 227 keuringsformulieren bleken 129 volgens het oude systeem (ABOHZIS-classificatie, "Systeem 1"), en 98 volgens het nieuwe systeem (Basismedische Eisen, "Systeem 2") goedgekeurd en beoordeeld te zijn. Met behulp van tabel 18 wordt voor de lezer inzichtelijk gemaakt hoe de resultaten over beide keuringssystemen verdeeld bleken te zijn.

De percentages ongeschikt verklaarde kandidaten, afkomstig uit beide keuringssystemen ($22/129=17\%$ versus $24/83=29\%$) verschillen niet statistisch significant ($p > 0.05$). Het percentage afgekeurde aspirant-militairen, afkomstig van beide keuringssystemen uit dit onderzoek is 20,3; dit getal correspondeert met het gemiddelde percentage jaarlijkse afkeuringen.

Tabel 18. De verdeling van het aantal keurlingen met en zonder medische beperkingen en de uitslag geschikt of ongeschikt over beide keuringssystemen (ABOHZIS en BMEKL).

	ABOHZIS-systeem 1 (n=129)			BMEKL-systeem 2 (n=98)		
	Medische bijzonderheden	Geen medische bijzonderheden	Totaal	Medische bijzonderheden	Geen medische bijzonderheden	Totaal
Geschikt	60	47	107	59	15	74
Ongeschikt	21	1	22	24	0	24
Totaal	81	48	129	83	15	98

Wat in tabel 18 verder opvalt is dat bij gebruik van het ABOHZIS-keuringssysteem bij minder keurlingen medische bijzonderheden ($81/129=63\%$) werden vastgelegd dan bij gebruik van het BMEKL-keuringssysteem ($83/98=85\%$). Bij nadere beschouwing blijkt dit verschil vooral te berusten op het voorkomen van de medische bijzonderheid "onvoldoende fysieke belastbaarheid" (9 van de 129 in geval van het ABOHZIS-keuringssysteem versus 21 van de 98 bij het BMEKL-keuringssysteem).

Als medische bijzonderheden – en potentiële afkeuringgronden – komen onder de keurlingen uit deze studie de volgende voor:

- onvoldoende fysieke belastbaarheid,
- aandoeningen van het bewegingsapparaat,
- gehoorbeperkingen,
- afwijkingen aan de visus,
- atopische constitutie.

De medische bijzonderheid “onvoldoende fysieke belastbaarheid” leidt relatief tot de grootste kans op afkeuring, namelijk in 24 van de in totaal 30 gevallen waarin daarvan sprake is.

Als opmerkelijk feit is er één kandidaat uit het ABOHZIS-keuringssysteem die geen medische bijzonderheden vertoont, maar die desondanks is afgekeurd. Deze persoon is vervolgens in bezwaar getreden, herkeurd en geschikt bevonden. Ook in de hernieuwde beoordeling wordt deze kandidaat telkens als geschikt beschouwd. Hier is sprake van een beoordelingsfout.

4.4.2 Overeenstemming

In tabel 19 worden de nieuwe beoordelingen van de drie artsen uit 2000 afgezet tegenover het oorspronkelijke oordeel in 1998. Hierbij is een onderverdeling gemaakt tussen het oude en het nieuwe systeem. Verticaal worden de te onderscheiden fracties weergegeven (GG, OO, GO en OG of a, d, b en c). Daarbij is de congruente fractie waarbij beide oordelen eensluidend zijn over de geschiktheid aangegeven als GG, respectievelijk a; de congruente fractie waarbij beide oordelen eensluidend zijn over de ongeschiktheid zijn aangegeven als OO, respectievelijk d; de discrepante fracties – er is geen overeenstemming – zijn achtereenvolgens aangeduid als OG, respectievelijk c (het oorspronkelijke geschiktheidsoordeel wordt achteraf veranderd in ongeschiktverklaring) en GO, respectievelijk b (het oorspronkelijke ongeschiktheidsoordeel wordt achteraf veranderd in geschiktverklaring). Daaronder worden de bij dit onderzoek gehanteerde maten voor overeenstemming weergegeven, met de gemiddeldes en standaardafwijkingen.

De 95%-betrouwbaarheidsintervallen van alle vermelde kappa's overlappen elkaar ruimschoots. Op grond van de onderlinge gemiddeldes en de daarbij vermelde standaardafwijkingen kan eveneens worden gezien dat de overige maten van overeenstemming onderling geen statistisch significante verschillen vertonen.

Opvallend is voorts het feit dat men veel minder afwijkt van het oorspronkelijke ongeschiktheidsoordeel OG (c) dan andersom GO (b).

In tabel 20 zijn de onderlinge uitslagen tussen de drie uitslaggevend seniorbedrijfsartsen zoals ze in 2000 beoordeelden, op dezelfde manier weergegeven

Tabel 19. De vergelijking tussen de gecombineerde uitslagen van de uitslaggevend artsen in 1998 (Beoordelaar "C") en de in 2000 verkregen uitslagen van de afzonderlijke uitslaggevend artsen (artsen 1, 2 en 3).

Uitslagen	ABOHZIS-systeem 1 (n=129)					BMEKL-systeem 2 (n=98)				
	C vs. Arts1	C vs. Arts2	C vs. Arts3	GEM	SD	C vs. Arts1	C vs. Arts2	C vs. Arts3	GEM	SD
GG (a)	95	97	96			67	67	69		
GO (b)	12	10	11			7	7	5		
OG (c)	1	1	5			1	2	1		
OO (d)	20	20	17			23	22	22		
Maten van Overeenstemming										
a+d/a+b+c+d (%)	90%	91%	88%	90%	1,9%	92%	91%	94%	92%	1,5%
d/b+c+d (%)	61%	65%	52%	59%	6,7%	74%	71%	79%	75%	3,8%
Kappa*)	0,69	0,73	0,60			0,80	0,76	0,84		
b+c/a+b+c+d (%)	10%	9%	12%	10%	1,9%	8%	9%	6%	8%	1,5%

Tabel 20. De vergelijking tussen de gecombineerde uitslagen van de uitslaggevend artsen zoals ze in 2000 beoordeelden (artsen 1, 2 en 3).

Uitslagen	ABOHZIS-systeem 1 (n=129)					BMEKLsysteem 2 (n=98)				
	Arts1 vs. Arts2	Arts1 vs. Arts3	Arts2 vs. Arts3	GEM	SD	Arts1 vs. Arts2	Arts1 vs. Arts3	Arts2 vs. Arts3	GEM	SD
	Arts2	Arts3	Arts3			Arts2	Arts3	Arts3		
GG (a)	95	94	95			67	68	68		
GO (b)	1	2	3			1	0	1		
OG (c)	3	6	5			2	2	2		
OO (d)	29	26	25			28	27	26		
Maten van Overeenstemming										
a+d/a+b+c+d (%)	97%	94%	94%	95%	1,8%	97%	98%	97%	97%	0,6%
d/b+c+d (%)	88%	76%	76%	80%	6,8%	90%	93%	90%	91%	1,8%
Kappa*)	0,92	0,83	0,82			0,93	0,95	0,92		
b+c/a+b+c+d (%)	3%	6%	6%	5%	1,8%	3%	2%	3%	3%	0,6%

O staat voor ongeschikt; G voor geschikt. De uitslagen GG, OO, OG en GO staan voor de uitslagen per artsenpaar. GEM staat voor gemiddelde, sd voor standaardafwijking. *) De kappa's in het oude en nieuwe keuringssysteem laten voor elk van de artsen geen significante verbetering zien (de 95%-betrouwbaarheidsintervallen overlappen elkaar ruimschoots).

zoals in tabel 18. Ook hierbij is geen onderscheid te vinden ten aanzien de maten van overeenstemming.

Anders dan in tabel 19 is er bij tabel 20 een minder groot verschil tussen de beide discongruente fracties OG (c) en GO (b).

De 95%-betrouwbaarheidsintervallen van alle vermelde kappa's overlappen elkaar wederom ruimschoots. Op grond van de onderlinge gemiddeldes en de daarbij vermelde standaardafwijkingen kan worden aangenomen dat de overige maten van overeenstemming onderling eveneens geen verschillen vertonen.

4.5 Beschouwing

Gelet op de kappa kan worden gesteld dat de overeenstemming "in het heden" (tabel 20) voortreffelijk is in het geval van beide systemen. De overeenkomst tussen de momentane uitslagen en die uit het verleden (tabel 19) kan op diezelfde grond als "voortreffelijk" worden gekenschetst. De mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring verschilt niet statistisch significant. De inter-waarnemers variatie is derhalve zeer aanvaardbaar, maar verschilt niet per keuringssysteem.

Idealiter waren alle 863 keuringen bij dit onderzoek betrokken. Door de eerste 227 ($n > 200$) keuringsdossiers te gebruiken is een voldoende hoeveelheid verkregen om verantwoorde, gegeneraliseerde uitspraken over te mogen doen. Bij het IKS wordt de uitvoering van de aanstellingskeuring voor kandidaten van de vier verschillende krijgsmachtdelen aselekt ingepland; de KL is slechts één van die krijgsmachtdelen. Gelet op de willekeurige wijze waarop kandidaten voor de aanstellingskeuring worden ingedeeld is er geen aanleiding om aan te nemen dat deze 227 kandidaten anders zijn samengesteld dan de overige bij de gerandomiseerde gecontroleerde studie betrokken kandidaten.

In de theorie is het bepalen dat er om de week op de ene dan wel op de andere wijze moet worden gekeurd niet het optimale design voor dit onderzoek. In de praktijk valt het gekozen design te verdedigen. Zo zou het dagelijks aanbieden van aselekt in de twee keuringssystemen ingedeelde kandidaten aan elke keurend arts tot diverse problemen leiden: het gevaar verstrikt te raken in verschillende diagnostische classificatieschema's [5], expectation bias [5], de hoge moeilijkheidsgraad van controle door derden op de correcte uitvoering en motivatieproblemen bij de keurend arts. Bij het IKS worden jaarlijks duizenden keuringen voor de vier krijgsmachtdelen (Koninklijke Marine, – Landmacht, – Luchtmacht en – Marechaussee) verricht. Om dit instituut tijdelijk te laten werken met twee keuringssystemen (onder meer met 10 keurend artsen en drie uitslaggevend artsen), bleek deze oplossing in de praktijk de optimaal haalbare werkwijze. De betrokken artsen waren dezelfde die in 1998 de oorspronkelijke

uitslag hadden gegeven; allen hadden dezelfde kennis van zaken van het nieuwe zowel als het oude keuringssysteem. Op grond hiervan lijkt selectiebias afdoende te zijn bestreden.

Het beoordelen van schriftelijke dossiers om de mate van overeenstemming tussen artsen te beoordelen is een hiervoor regelmatig gebruikte methode. Hierin schuilt wel een beperking: men keurt immers niet fysiek en mist wellicht (non-verbale en verbale communicatieve) signalen. Het voordeel van deze methode is dat de betrokken artsen per definitie uitgaan van dezelfde gegevens. Bovendien vindt het uitslag geven van de gebruikelijke aanstellingskeuring door de uitslaggevend arts ook plaats op grond van het door de keurend arts opgestelde dossier, al dan niet toegelicht door diezelfde keurend arts. In die zin wijkt de gehanteerde procedure voor de uitslaggevend artsen niet veel af van de praktijk situatie.

Door het gebruik van dossiers in plaats van onderzoekspersonen worden het Hawthorne-effect [5], het placebo-effect [5], en het Pygmalion-effect [5] juist vermeden. Omdat bij de herbeoordeling de artsen inmiddels al meer dan een jaar met het nieuwe systeem hadden gewerkt, kan er ook geen sprake meer zijn van een eventueel novelty-effect [5]. Op grond van het bovenstaande wordt aangenomen dat informatiebias afdoende is bestreden.

Er is sprake van een grote mate van overeenstemming, zowel wanneer de oude uitslagen met de nieuwe worden vergeleken, als de nieuw uitslagen onderling worden vergeleken. Tussen de oorspronkelijke uitspraak en de nieuwe is meer dan een jaar verstreken. Volgens Bouter en Van Dongen [5] kan een te lange tussentijd voor werkelijke veranderingen zorgen. Dit zou een verklaring kunnen zijn voor de overgang van het oordeel "goed" naar "voortreffelijk". Er zou sprake van een leereffect kunnen zijn, tenslotte hebben de betrokken artsen nu al meer dan een jaar in de praktijk met het nieuwe systeem gewerkt.

Gelet op de willekeurige volgorde waarin de medische aanstellingskeuring door het IKS in de tijd wordt gepland, berust het gesignaleerde verschil tussen beide keuringssystemen inzake de fysieke selectie op toeval.

De samenstelling van de gebruikte dossiers was representatief voor de gebruikelijke samenstelling van dergelijke dossiers. De verhoudingen tussen de geschikt verklaarden en ongeschikt verklaarden weerspiegelen het jaarlijks gemiddelde bij het IKS voor KL-kandidaten.

Het begrip "fysieke belastbaarheid" als medische bijzonderheid leidt wellicht tot vraagtekens. Enerzijds wordt die belastbaarheid door middel van instrumentarium gemeten zoals men ook de visus of het gehoor meet. Anderzijds kan

men zich afvragen of en in hoeverre iemand die louter enige statische of dynamische kracht op een door de KL voorgestane schaal tekort komt, een medische bijzonderheid heeft. Daar waar bijvoorbeeld perceptief verlies bij het gehoor niet meer te herstellen valt, kan men zich vervolgens afvragen of dat met onvoldoende conditie soms wel het geval kan zijn. Dat kracht en conditie voor een functie als KL-militair belangrijk zijn valt niet te ontkennen, maar behalve onvoldoende conditie of kracht mankeren deze personen medisch niets. Aangezien juist conditieproblemen relatief tot het hoogste percentage afkeuringen leidt, lijkt nader onderzoek hiernaar gewenst.

De artsen zijn het bijna altijd met elkaar eens inzake de geschiktheid. Deze eensgezindheid heeft te maken met het feit dat bij de meerderheid van deze beoordelingen ook geen medische bijzonderheden worden gevonden: niets aan de hand betekent geschikt.

Het hierboven vermelde "leereffect" kent overigens een prijs. Zijn er oorspronkelijk 46 afgekeurden in 1998, (tabel 18), anno 2000 worden dat er in totaal tussen de 55 en 62 (Tabel 20). Dat is letterlijk het effect van de gedaalde discongruentie aan de afkeuringzijde (OO), waardoor het totaal aan afkeuringen toeneemt.

Beoordeling van de reproduceerbaarheid van het nieuwe systeem bij de aanstellingskeuring diende, zoals in de inleiding vermeld, plaats te vinden. Maar heeft het, bij het ontbreken van een "gouden standaard", zin om de gevonden uitslagen voor beide keuringssystemen met elkaar te vergelijken? Een lagere reproduceerbaarheid van het nieuwe systeem zou kunnen betekenen dat het ten aanzien van die kwaliteitseis relatief als minder goed dan het vorige systeem kan worden gekwalificeerd. Daardoor zou verder onderzoek als minder relevant zou kunnen worden beschouwd.

De reproduceerbaarheid van het nieuwe systeem blijkt evenzeer te voldoen aan de volgens Bouter en Van Dongen [5] gebruikelijke mate van reproduceerbaarheid volgens kappa als het oude systeem. Ook bij de overige gebruikte maten voor reproduceerbaarheid valt geen verschil te ontdekken.

De voor dit onderzoek gebruikte casus zijn afkomstig uit twee studiegroepen die (zie hoofdstuk 5) in weinig van elkaar blijken te verschillen. Op grond hiervan kan worden aangenomen dat beide keuringssystemen inzake reproduceerbaarheid bij de aanstellingskeuring niet voor elkaar onderdoen.

Samengevat kan worden vastgesteld dat in de mate van overeenstemming tussen beide keuringssystemen geen statistisch significant verschil bestaat. In beide gevallen is sprake van een voortreffelijke overeenstemming.

Figuur 15: Het overzicht van de onderzoeken uit dit proefschrift en de daarbij behorende hoofdstukken. In de hoofdstukken 2 en 3 wordt de reproduceerbaarheid van het periodiek geneeskundig onderzoek, respectievelijk de pensioenkeuring onderzocht, analoog aan Figuur 5. Hoofdstuk 4 bevat soortgelijk onderzoek voor wat betreft de aanstellingskeuringen van dezelfde gerandomiseerde, gecontroleerde studie, van wie de twee studiegroepen in de hoofdstukken 5 en 6 gedurende hun plaatsing op functie worden gevolgd om de voorspellende waarde tussen beide keuringssystemen te kunnen vergelijken.

Daardoor lijkt bij het gebruik van de BMEKL het uitsluiten van gevaar voor de eigen veiligheid en gezondheid of die van anderen thans goed gewaarborgd bij een aanstelling als militair bij de KL, als louter wordt gekeken naar de mate van overeenstemming als kwaliteitsmaatstaf.

In hoofdstuk 2 is de mate van overeenstemming van het ABOHZIS-keurings-systeem in geval van het Periodiek Geneeskundig Onderzoek (PGO) ten tijde van een grote reorganisatie in 1994 – 1996. De uitkomsten (kappa's) van dat onderzoek sloten aan bij de bevindingen die De Kort in zijn onderzoek had gevonden. Dit hoofdstuk is omschreven als een systematische waarneming, die leidde tot de interesse in verder onderzoek. In de daarop volgende hoofdstukken is aandacht besteed aan één gedeelte van de vraagstelling (figuur 15): de mate van overeenstemming als kwaliteitsmaat.

In hoofdstuk 3 is de inter-waarnemers variatie bij de pensioenkeuring onderzocht voor het nieuwe keuringssysteem. De mate van overeenstemming bleek

*) exclusiecriteria

Figuur 16: De ideale onderzoeksopzet voor de bepaling van de voorspellende waarde.

Alle kandidaten, geschikt zowel als ongeschikt verklaard, worden na opkomst tot het einde van de arbeidsovereenkomst gevolgd.

hierbij, met name voor wat het definitieve eindoordeel betreft, kunnen worden omschreven als goed in de betekenis van de kappa.

In hoofdstuk 4 is aandacht besteed aan de mate van overeenstemming van beide keuringssystemen bij de aanstellingskeuring. Hieruit kwam naar voren dat deze mate van overeenstemming voor beide keuringssystemen als voortreffelijk kan worden beschouwd. Daarmee wordt het onderwerp overeenstemming voor wat dit proefschrift betreft afgesloten.

Een goede mate van overeenstemming is een noodzakelijke, maar op zichzelf staand onvoldoende voorwaarde voor de kwaliteit van een test. De meetuitkomsten moeten ook valide zijn. Om die validiteit te bepalen is ander onderzoek nodig. De resultaten van het nieuwe keuringssysteem konden niet aan een gouden standaard worden getoetst. Wel was het mogelijk om een simultane vergelijking te maken met een andere test met een onbekende validiteit: het oude ABOHZIS-keuringssysteem.

In de ideale onderzoeksopzet worden groeperingen kandidaat-militairen volgens één of beide systemen gekeurd en vervolgens aangenomen, opgeleid en

ingezet – onafhankelijk van goed- of afkeuring (figuur 16). Het te werk stellen van ongeschikt verklaarde militairen was echter geen haalbare optie. Om praktische en ethische redenen geeft de KL-organisatie hiervoor geen toestemming. Nu zou er toe kunnen worden overgegaan om de afgewezen kandidaten op dezelfde wijze te vervolgen gedurende dezelfde onderzoeksperiode als hun militaire, geschikt bevonden collega's. Hierbij zouden zich drie problemen voordoen. Ten eerste waren de mogelijkheden om data voor bij de KL te werk gestelde militairen te verzamelen aanmerkelijk uitgebreider dan die voor de externe onderzoeksgroep. Hierdoor zou selectiebias optreden. Ten tweede zouden de arbeidsomstandigheden en functie-eisen van die externe onderzoeksgroep hoogstwaarschijnlijk grote individuele variatie vertonen. Door de te verwachten selectiebias zou hiervoor niet kunnen worden gecorrigeerd. Als derde zou het problematisch kunnen zijn om voldoende respons te verkrijgen. Hierdoor resteerde alleen de mogelijkheid om onderzoek te verrichten naar verschillen tussen geschikt verklaarde militairen.

Om dit te beoordelen is gedurende de periode juni 1999 – juni 2001 een gecontroleerde gerandomiseerde studie verricht. Bij het begin van de studie zijn keurlingen gerandomiseerd gekeurd door middel van één van de twee keuringssystemen, ABOHZIS of BMEKL, hetgeen in de volgende hoofdstukken wordt weergegeven.

Geraadpleegde literatuur

1. Capaciteiten test: General Aptitude Test Battery, GATB versie B 1002-b. Unites States Employment Service, 1958. Voor Nederland bewerkt door de Stichting GATB-research. Swets Test Publishers, Lisse: 1994.
2. Flier H van der, Boomsma-Suerink JL: Handboek GATB B 1002-B. Utrecht, . Swets Test Publishers, Lisse: 1994.
3. Luteijn F, Starren J, Dijk H van: Nederlandse Persoonlijkheids Vragenlijst (NPV).. Swets Test Publishers, Lisse: 1974.
4. Luteijn F, Starren J, Dijk H van: Handleiding Nederlandse Persoonlijkheids Vragenlijst (herziene uitgave). Swets Test Publishers, Lisse: 1985.
5. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Bohn, Stafleu en van Loghum. Houten, 1995.
6. Kort WLAM, Post-Uiterweer HW: Project aanstellingskeuringen dossieronderzoek bij de RBB. Deel II: panelonderzoek. Ministerie van Sociale Zaken en Werkgelegenheid, S 53-2, 1991.
7. Altman DG: Practical statistics for medical research. London: Chapman and Hall, 1991.
8. Steijnans VW, Diletti E, Bönches B et al: Interobserver agreement: Cohens Kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J of Clin of Pharm And Ther.* 1997; 35; 3: 93 – 95.

5 | Een vergelijking tussen twee systemen voor de aanstellingskeuring bij de Koninklijke Landmacht met behulp van een gerandomiseerde, gecontroleerde studie

In belangrijke mate overgenomen uit: Raad, J de, Redekop WK: A comparison between two systems for pre-employment medical assessment in the RNLA by a randomised-controlled study. *Military Medicine* 2004; 169(6):437-443.

5.1 Inleiding en vraagstelling van dit onderzoek

In 1811 introduceerde Napoleon Bonaparte zowel de conscriptie als de medische keuring in het Nederlandse leger. Tot 1998 werd de medische aanstellingskeuring bij de Koninklijke Landmacht (KL) uitgevoerd met behulp van het daarvan afkomstige "Militair Keuringsreglement" (MKR) en de bijbehorende "Lijst van ziekten en gebreken".

Deze lijst van ziekten en gebreken werd van tijd tot tijd gemoderniseerd. Verder werd het aan de professionele autonomie van de keurend arts overgelaten om te beslissen of een keurling in voldoende ernstige mate aan een bepaalde ziekte of gebrek leed om te worden afgekeurd. Na de Tweede Wereldoorlog ontstond een groeiende behoefte tot verdere verfijning van het personele selectieproces in de KL. Daartoe adopteerde de Nederlandse overheid in 1946 het Canadese PULHEEMS-indelingssysteem, vertaald als ABOHZIS (Algemene toestand, Bovenste extremiteiten, Onderste extremiteiten, Horen, Zien, Intelligentie en Stabiliteit). Dit systeem werd gecombineerd met het MKR en de "Lijst van ziekten en gebreken". Door een verdere verbijzondering in 1967 werden praktisch alle toenmalig bekende ziekten en gebreken beschreven in termen van de ABOHZIS-classificatie. Na 1967 werd hierin geen verandering meer aangebracht.

Dit systeem vertoonde verschillende tekortkomingen [1]. Zo werden beperkingen opgelegd terwijl functie-eisen ontbraken. Voorts was aangetoond dat er een slechte overeenkomst bestond tussen de ABOHZIS-classificatie en de eisen die in de dagelijkse praktijk aan militairen worden gesteld [2]. Het systeem werd gehandhaafd zolang de dienstplicht van kracht bleef, of juist tot deze in 1996 werd opgeschort. Hierna ontstond een snel groeiende behoefte aan een ander keuringssysteem.

In 1998 werden de Basismedische Eisen KL (BMEKL) als nieuw keuringssysteem geïntroduceerd [3]. De te stellen eisen aan militairen waren in de vorm van 43 beoordelingspunten samengebracht, tezamen met keuringsprotocollen. Voor elk beoordelingspunt werd een te scoren grenswaarde gedefinieerd, de Basismedische Eis (hoofdstuk 1, tabel 1).

Dit systeem is door de overige krijgsmachtdelen overgenomen als nieuw keuringssysteem voor de aanstellingskeuring. Dit betekent dat een nieuw keuringssysteem bij de krijgsmacht is geïntroduceerd, gebaseerd op aan militairen te stellen eisen, dat bovendien niet strijdig is met de Wet op de Medische Keuringen uit 1998.

Het nieuwe keuringssysteem verschilt in belangrijke mate van het ABOHZIS-systeem. Het systeem houdt rekening met operationele inzetbaarheid. De eisen houden voorts rekening met zowel gedragswetenschappelijke elementen als taakspecifieke militaire vaardigheden. De keurend arts dient te

beoordelen of de rekrut op medische gronden in staat is om zonder beperkingen aan de gestelde eisen te voldoen. Is dit laatste het geval dan wordt de rekrut goedgekeurd, ongeacht de aanwezigheid van ziekte of gebrek. In het oude keuringssysteem bestonden geen functie-eisen, evenmin als een relatie met operationele inzetbaarheid. Er bestond weinig medische interesse in gedragswetenschappelijke elementen. De goedkeuring van rekruten voor de militaire dienst was gebaseerd op het afwezig zijn van ziekten en gebreken, ongeacht of zij in staat waren om de vereiste taken uit te voeren, omdat die taken niet waren gedefinieerd.

Dit nieuwe keuringssysteem wordt thans gebruikt voor de aanname van militairen, die momenteel allen vrijwillig dienend zijn. De meeste rekruten krijgen een arbeidsovereenkomst voor een bepaalde periode aangeboden, in de regel bedraagt dit periode 2 ½ jaar ("Beroeps Bepaalde Tijd", "BBT"). Daarnaast is tevens sprake van beroepsmilitairen met een arbeidsovereenkomst voor onbepaalde tijd ("Beroeps Onbepaalde Tijd", "BOT"). Militairen BBT ondergaan gedurende de eerste 6 maanden na opkomst de algemene militaire opleiding, gevolgd door een functiegericht opleiding. Manschappen worden opgeleid bij één van de vier Schoolbataljons, waarvan drie hetzelfde programma volgen. Het vierde Schoobataljon is bedoeld om manschappen op te leiden voor de Luchtmobiele Brigade en kent een zwaarder programma, dat 5 weken langer duurt. Officieren en onderofficieren worden op aparte opleidingsinstituten (Koninklijke Militaire Academie, KMA, respectievelijk Koninklijke Militaire School, KMS) opgeleid.

Sinds de omvorming van de op dienstplicht gebaseerde KL naar een beroepsleger ligt het percentage rekruten dat voortijdig de dienst verlaat in alle schoolbataljons tussen de 20 en 35. Meestal verlaten deze rekruten de dienst in de eerste 12 weken. Velen vertrekken omdat zij uiteindelijk niet de gewenste functie krijgen of omdat het militaire leven niet aan hun verwachtingen voldoet. Zelden is er een medische oorzaak voor hun vertrek.

Nadat het nieuwe keuringssysteem in werking was getreden, diende het te worden geëvalueerd. Inmiddels zijn de reproduceerbaarheid van het keuringssysteem bij zowel de pensioenkeuring als de aanstellingskeuring onderzocht. Beide onderzoeken leverden bevredigende resultaten op. Zo bleek bijvoorbeeld bij het vergelijken van de mate van overeenkomst tussen twee keurend artsen bij de pensioenkeuring de kappa van Cohen 0.74 (0.48 – 0.98) te zijn [4]. De mate van overeenkomst van de aanstellingskeuring tussen drie keurend artsen in de zin van $a+d/a+b+c+d(\%)$ bleek $97 \pm 0,6\%$ te zijn [5].

Het beoordelen van de validiteit van het nieuwe keuringssysteem is ingewikkelder dan het beoordelen van de reproduceerbaarheid. Weliswaar voldoen

zowel de omschrijving van de BMKEL als de keuringsprotocollen aan “face validity”, evenals aan “expert validity”. Ook lijkt ten gevolge van die keuringsprotocollen, door de constante werkomstandigheden en door de supervisie van seniorbedrijfsartsen, de correcte uitvoering van de keuring gegarandeerd. Maar nimmer is de specificiteit of sensitiviteit van het BMEKL-keuringssysteem onderzocht, evenmin bestaat er een “gouden standaard”.

Tengevolge van meerdere redenen, waaronder ethische, is het niet mogelijk om ongeschikte kandidaten aan te nemen en onder dezelfde (militaire) werkomstandigheden te volgen als de geschikt verklaarde soldaten. Voorts zou het erg moeilijk worden om een vergelijking te maken over een langere periode tussen de ongeschikt verklaarden, werkzaam in de civiele sector, en de geschikt verklaarden (selectie-bias, informatie bias, uiteenlopende werkomstandigheden, evenals een gebrekkige respons, waardoor voor die werkomstandigheden niet zou kunnen worden gecorrigeerd). Hetzelfde geldt ten aanzien van diegenen die weliswaar geschikt zijn verklaard, maar niet in dienst zijn getreden, evenals voor diegenen die reeds tijdens de algemene militaire opleiding de dienst alweer verlaten.

Om deze redenen is ervoor gekozen beide keuringssystemen te vergelijken voor wat betreft het vermogen om geschikte rekruten op te leveren. Daarbij worden twee uitkomstvariabelen gehanteerd, te weten het aantal dagen inzetbaarheid en de hoeveelheid medische consumptie. Hiervoor werd van juni 1999 tot en met juni 2001 een gerandomiseerde, gecontroleerde studie verricht (figuur 17). De vraagstelling voor dit onderzoek luidde: “Is er een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van de geschikt verklaarde militairen BBT op het gebied van de dagen inzetbaarheid en/of de genoten medische consumptie?”

5.2 Materiaal en methoden

5.2.1 Populatie en procedure

Twee studiegroepen werden samengesteld. Daarbij werden deelnemers uit groep één gekeurd volgens het ABOHZIS-systeem (“ABOHZIS”) en deelnemers uit groep twee volgens het BMEKL-systeem (“BMEKL”). Tussen 22 september en 16 oktober 1998 zijn 863 kandidaten gekeurd, waarbij randomisatie plaatsvond door iedere week van keuringssysteem te wisselen (figuur 17). Dit leidde tot de uiteindelijke samenstelling van twee onderzoeksgroepen van in totaal 352 personen, waarvan 166 in de ABOHZIS-studiegroep en 186 in de BMEKL-studiegroep.

Figuur 17: De onderzoeksopzet van de gerandomiseerde gecontroleerde studie.

Na gerandomiseerd te zijn gekeurd worden de geschikt verklaarde kandidaten na opkomst gedurende de periode van paraatstelling (twee jaar) tot het einde van de arbeidsovereenkomst gevolgd.

5.2.2 Informatie

Voor het nalopen van de te onderzoeken variabelen werd van de volgende middelen gebruik gemaakt: dossiers van de aanstellingskeuring, exitinterviews afgenomen bij dienstverlaters, centraal beheerde plaatsingsgegevens, het decentraal ingevulde, maar centraal gecontroleerde en beheerde verzuimregistratiesysteem, gegevens afkomstig van de afkeuring, ingediende claims, de interne geneeskundige verrichtingenadministratie zoals opgemaakt door de militaire zorgverleners en declaraties bij de ziektekostenverzekeraar.

5.2.3 Exclusie van verder onderzoek

De onderzoeksperiode liep van 1 juni 1999 tot 1 juni 2001. De redenen voor exclusie waren: ongeschikt verklaring, militairen BOT, geen opkomst respectievelijk vertrek reeds in de militaire opleiding, opkomst na 1 juni 1999 en atypische functies (bijvoorbeeld militaire muzikant).

5.2.4 Effectvariabelen

Inzetbaarheid. Per kandidaat is over de onderzoeksperiode het aantal (kalender-)dagen inzetbaarheid bepaald door aftrek van het aantal dagen ziekteverzuim of andere vormen van arbeidsverzuim, zoals beperkte operationele inzetbaarheid door blessures, vastgesteld door een arts. Hierbij bepaalde het klassieke ziekteverzuim 95% van het totaal en kon de zuivere inzetbaarheid (het aantal kalenderdagen dat een militair operationeel had kunnen worden ingezet) worden bepaald.

Medische consumptie. De KL bezit een eigen eerstelijns geneeskundig zorgsysteem (huisartsenzorg) en voorziet voor de belangrijkste klinisch specialismen ook in de tweede lijns gezondheidszorg. Die eerste lijns zorg geldt echter in eigen land slechts onder kantooruren. Tijdens operationele inzet en bij plaatsing in Duitsland wordt 24 uur per dag eerstelijns gezondheidszorg verstrekt. Buiten kantooruren kan een militair zich tot een civiele (eerste of tweedelijns) zorgverlener wenden; hiervoor is de militair verzekerd.

De verrichtingen van de eigen eerste lijn worden in een centrale database vergaard. Alle tweedelijns verrichtingen, alsmede extern geraadpleegde eerste lijns verrichtingen worden door de ziektekostenverzekeraar vergoed na declaratie. Om een volledig overzicht te verkrijgen over de totaal gemaakte medische kosten moeten zowel de ziektekostenverzekeraar als de database worden geraadpleegd.

De medische eerstelijns consumptie is berekend door alle eerste lijns consulten te rekenen naar het gemiddeld over de onderzoeksperiode geldende CTG-tarief (€ 17,92). De overige gemaakte en gedeclareerde ziektekosten zijn bij de ziektekostenverzekeraar opgevraagd.

5.2.5 Statistische methoden

Om de verschillen tussen beide studiegroepen te toetsen is voor continue variabelen gebruik gemaakt van zowel de Mann-Whitney-U test als de Kruskal-Wallis H test, voor dichotome variabelen de Pearson Chikwadraattest. Voorts zijn multiple lineaire regressie analyses uitgevoerd.

5.3 Resultaten

In totaal werden in de ABOHZIS-groep 399 kandidaten gekeurd, waarvan 304 geschikt werden bevonden (tabel 21). Daarvan kwamen uiteindelijk 243 beroeps bepaalde tijd (BBT) op, waarvan op grond van de exclusiecriteria uiteindelijk 166 in de studiegroep belandden. Voor de BMEKL-groep lag dit als volgt: 464 kandidaten werden gekeurd, waarvan 341 geschikt werden verklaard;

Tabel 21. Het verloop van de gekeurde militairen.

	ABOHZIS groep		BMEKL groep		p
Gekeurd	399	100%	464	100%	
Afgekeurd	-64	16%	-58	13%	0,14
Uitval door fysieke selectie	-31	8%	-65	14%	< 0,01
Geschied	304	76%	341	73%	
Beroeps Onbepaalde Tijd (BOT)	-8	2%	-17	4%	0,12
Niet bij KL maar elders	-15	4%	-18	4%	0,84
Nooit opgekomen	-36	9%	-33	7%	0,38
Anders	-2	1%	-3	1%	0,75
Beroeps Bepaalde Tijd (BBT)	243	61%	270	58%	0,76
Te kort in dienst	-75	19%	-80	17%	0,50
Te laat in dienst	-2	1%	-4	1%	
Onderzoeksgroep	166	42%	186	40%	40%

“Anders” staat voor afwijkende militaire functies, zoals geestelijk verzorger en militaire muzikant.

hiervan kwamen 270 BBT-ers op, hetgeen resulteerde in een studiegroep van 186 op grond van de exclusiecriteria.

Uit tabel 21 blijkt dat alleen een statistisch significant verschil op grond van de fysieke selectie tussen de keuringssystemen bestond. De persoonskenmerken van diegenen die geschikt verklaard zijn maar nooit opkomen, verschillen niet per keuringssysteem. Aanvullend kan worden vermeld dat de periode die tussen het moment van de aanstellingskeuring en het moment van de daadwerkelijke opkomst lag, 136 ± 34 kalenderdagen bedroeg.

Van 100 van de 155 militairen die binnen 12 weken alweer afzwaaien kon het exit-interview worden achterhaald. Van hen geven 4 een medische reden op, 11 geven aan de fysieke belasting toch te zwaar te vinden. De overige 85 geven andere, niet-medische redenen op. Geen van de in totaal 155 vroege dienstverlaters dient een claim in op grond van het Kaderbesluit Militaire Pensioenen, hetgeen de suggestie wekt dat de kans op serieuze medische problemen onder de vroege dienstverlaters als gering mag worden beschouwd.

Uit de beginsituatie (tabel 22) blijkt dat geen van de onderzochte variabelen statistisch significant verschilt tussen beide studiegroepen. Buiten de in tabel 22 vermelde variabelen is voorts nog gekeken naar de fysieke selectie ($p = 0,72$), medische bijzonderheden ($p = 0,54$), de schoolopleiding ($p = 0,20$) en de

Tabel 22. De persoonskenmerken aan het begin van de onderzoeksperiode.

Variabele		ABOHZIS N= 168		BMEKL N = 188		P
Leeftijd		20,8 ± 2,5		20,8 ± 2,7		0,95
Geslacht	Man	146	88,0	169	91,0	0,37
Demografisch:	Stedelijk	46	27,7	64	34,4	
> / < 50.000 inwoners	Platteland	103	62,0	104	55,9	0,40
Gezinssamenstelling	Compleet	115	69,3	133	71,5	
	Incompleet	34	20,5	37	19,9	0,85
Roker (sigaretten)	Ja	98	59,0	109	58,6	0,92
Burgerlijke Staat	Ongehuwd	156	94,0	180	96,8	
	Samenwonend	2	1,2	3	1,6	
	Gehuwd	6	3,6	3	1,6	
	Gescheiden	1	0,6	0		0,44
Recidivist (eerder in dienst)	Ja	8	4,8	14	7,4	0,30

Toelichting: Onder "Incompleet" bij gezinssamenstelling wordt verstaan: ouders gescheiden of één der ouders overleden.

verhouding tussen geschikt- en ongeschiktverklaring van de 10 verschillende keurend artsen ($p = 0,27$).

In tabel 23 staan de uitkomstvariabelen. Er is een statistisch significant verschil tussen beide studiegroepen inzake plaatsing naar gevechts- of gevechtsondersteunende functies ($p = 0,02$; de BMEKL-groep telt meer gevechtsfuncties, de ABOHZIS-groep meer gevechtsondersteunende functies), dagen inzetbaarheid ($p = 0,02$) en medische consumptie ($p = 0,05$). Het aantal dagen inzetbaarheid wordt voor 95% bepaald door ziekteverzuim en voor 5% door andere vormen van arbeidsverzuim, hoofdzakelijk door medische beperkingen ten aanzien van de operationele inzetbaarheid (blessures e.d.). De inzetbaarheid valt voor zowel gevechts- als gevechtsondersteunende functies in de BMEKL-groep hoger uit dan de ABOHZIS-groep.

De onderzoeksgroepen zijn ook per militaire opleidingsplaats naar de dagen inzetbaarheid en medische consumptie tussen de keuringssystemen vergeleken. Hierbij bleek dat de onderlinge verschillen alleen statistisch significant waren ten aanzien van het aantal dagen inzetbaarheid bij de opleiding van het Schoolbataljon Luchtmobiel ($p = 0,01$), gevolgd door het resultaat voor wat betreft Schoolbataljon Zuid ($p = 0,06$). Beoordeling van verschillen tussen de studiegroepen op grond van de plaats waar de Algemene Militaire Opleiding is gevolgd levert een p-waarde van 0,06 op.

Tabel 23. De uitkomstvariabelen gedurende de onderzoeksperiode.

Variabele		ABOHZIS		BMEKL		P
		N = 166	%	N = 186	%	
Militaire Opleiding	Noord	25	15,1	36	19,4	
(Opl manschappen	Centraal	75	45,2	64	34,4	
""	Zuid	28	16,9	23	12,4	
""	LMB	23	13,9	39	21,0	
""	Anders	8	4,8	11	5,9	
(Opl onderofficier)	KMS	7	4,2	9	4,8	
(Opl officier)	KMA	0	0,0	4	2,2	0,06
Plaatsing	11 LMB	16	9,6	31	16,7	
	13 Mechbrig	32	19,3	32	17,2	
	41 Mechbrig	27	16,3	30	16,1	
	43 Mechbrig	17	10,2	24	12,9	
	Divisietroepen	57	33,1	54	26,3	
	Overig	13	7,8	19	10,3	
	IBDKL	6	3,6	1	0,5	0,36
Uitzending	Wel uitgezonden	54	47,6	62	45,2	0,59
Indeling gevechtsfunctie	Gevechtsfunctie	71	41,5	105	57,5	
Ondersteunende functie	Ondersteunend	95	58,5	81	42,5	0,02
Dagen inzet, max 731		612 ± 170		648 ± 134		<0,01
Medische consumptie	€ / 731 dgn	746 ± 2170		396 ± 958		0,05

Toelichting: LMB staat voor Luchtmobiele Brigade, Mechbrig voor Gemechaniseerde brigade, IBDKL = plaatsing i.v.m. langdurig verminderde inzetbaarheid als militair. Opl staat voor opleiding. De p-waarden van dagen inzet en medische consumptie zijn berekend met behulp van de Mann-Whitney-U test, voor de overige p-waarden is de chi-kwadraat test gebruikt.

Soms volgt na plaatsing op het definitieve onderdeel nogmaals een overplaatsing. Uiteindelijk blijkt dat in totaal 19 militairen naar de speciale eenheid voor langdurig verminderd inzetbare militairen (Individuele Begeleidingsdienst KL, IBDKL) zijn overgeplaatst: 13 uit de ABOHZIS-groep en 6 uit de BMEKL-groep. Het verschil wordt echter niet statistisch significant ($p = 0,17$). Het aantal uitgezonden militairen is evenmin verschillend tussen beide onderzoeksgroepen.

Na het vinden van deze verschillen in inzetbaarheid en medische consumptie tussen de studiegroepen is onderzocht of een associatie kon worden gevonden tussen deze uitkomsten en andere gegevens over deze rekruten. Daarvoor

Tabel 24. Multiple lineaire regressie naar keuringssysteem en militaire opleiding(-plaats) met achtereenvolgens dagen inzetbaarheid en medische consumptie als afhankelijke variabelen.

Onafhankelijke variabelen	Afhankelijke variabele: dagen inzetbaarheid				Afhankelijke variabele: Medische consumptie			
	Bèta coëfficiënt	Standaard fout	P-waarde	R ²	Bèta coëfficiënt	Standaard fout	P-waarde	R ²
Keuringssysteem (BMEKL vs ABOHZIS)	43	16	0,01		-376	171	0,03	
Type opleiding versus Overige typen opleiding								
Opleiding officier	-86	46	0,06		-415	477	0,39	
Opleiding onderofficier	80	78	0,30		-251	812	0,76	
Schoolbataljon LMB	71	42	0,09		-316	437	0,47	
Schoolbataljon Zuid	-39	27	0,14		404	277	0,15	
Schoolbataljon Centraal	20	28	0,47		410	293	0,16	
Schoolbataljon Noord	15	22	0,52		-103	234	0,66	
				0,06				0,04

Toelichting: Schoolbataljons leiden manschappen op; kaderleden worden opgeleid op aparte instituten. R² staat voor verklaarde variantie.

bestonden twee redenen. In de eerste plaats om de grootte van het effect van het keuringssysteem beter in te kunnen schatten, in de tweede plaats om het mechanisme te ontdekken waarom het BMEKL-systeem tot hogere uitkomsten had geleid. Hiervoor zijn per afhankelijke variabele telkens twee lineaire regressiemodellen gebruikt.

Bij multiple lineaire regressie analyse met inzetbaarheid als afhankelijke variabele en het keuringssysteem ($p = 0,02$), de militaire opleidingsplaats ($p = 0,00$) en gevechts- of gevechtsondersteunende functie ($p = 0,88$) als onafhankelijke variabelen bleek een associatie tussen de eerste twee onafhankelijke variabelen en inzetbaarheid. Bij een tweede regressie model met inzetbaarheid als afhankelijke variabele tegenover het keuringssysteem en alle militaire opleidingsplaatsen afzonderlijk als onafhankelijke variabelen (tabel 24) bleek alleen een associatie tussen inzetbaarheid en het type keuringssysteem aanwezig ($p = 0,01$). De p -waarde van de opleiding tot officier is weliswaar ook slechts 0,06 maar slechts 4 personen uit beide studiegroepen werden opgeleid tot officier.

Bij multiple lineaire regressie analyse met medische consumptie als afhankelijke variabele en het keuringssysteem, de militaire opleidingsplaats en gevechts- of

gevechtsondersteunende functie als onafhankelijke variabelen bleek alleen het keuringssysteem als onafhankelijke variabele statistisch significant van invloed te zijn ($p = 0,01$). Bij een tweede regressie model met medische consumptie als afhankelijke variabele tegenover het keuringssysteem en alle militaire opleidingsplaatsen afzonderlijk als onafhankelijke variabelen (tabel 24, tweede deel) wordt dit beeld bevestigd.

5.4 Beschouwing

In deze studie bleken militairen, gekeurd volgens het nieuwe keuringssysteem (BMEKL) en gevolgd gedurende een periode van twee jaar na de opleiding, statistisch significant meer dagen (648 versus 612) inzetbaar te zijn en minder medisch te consumeren (€ 396 versus € 746) dan militairen die terzelfder tijd volgens het oude keuringssysteem (ABOHZIS) zijn gekeurd.

Hierboven is reeds vermeld dat juist tussen manschappen uit beide keuringsystemen een statistisch significant verschil in inzetbaarheid bestond indien zij afkomstig waren van de zwaarste opleiding. Het voorkomen van meer gevechtsfuncties in de BMKEL-studiegroep heeft wellicht te maken met verschillen in trainingsresultaten tussen beide studiegroepen.

Er was echter geen verschil in de absolute aantallen geschikt verklaarde kandidaten, evenmin als in het percentage vroege drop-outs om niet-medische redenen. Het nieuwe systeem is niet strenger, maar brengt ook geen verlaging van het aantal vroege drop-outs. De verklaring voor deze laatste bevinding is dat de oorzaken voor dienstverlating (exit-interviews, geen claims) van niet-medische aard zijn en door een keuringssysteem hoe dan ook niet worden beïnvloed. Het bovenstaande roept drie vragen op. Beïnvloedt het keuringssysteem het type militaire opleiding dat voor en door de kandidaat wordt gekozen? En als dat zo is, beïnvloedt deze militaire opleiding vervolgens zelf de inzetbaarheid? En als dat niet zo is, welke factor kan anders de gevonden verschillen verklaren?

Wat de eerste vraag betreft: naar welke opleiding de kandidaat gaat is in de praktijk afhankelijk van de eigen wensen van de kandidaat en van de momentane vullingbehoefte bij de KL. Dit maakt het onaannemelijk dat de keuze voor het type opleiding wordt beïnvloed door het keuringssysteem.

Of de training zelf iemands inzetbaarheid kan beïnvloeden blijft onduidelijk. Wel is sprake van een statistisch significant verschil in inzetbaarheid tussen beide onderzoeksgroepen voor zover ze afkomstig zijn van de zwaarste (luchtmobiele) opleiding. Als nu onder die kandidaten de verschillen het kleinst zouden zijn geweest, zou een relatie tussen het type militaire opleiding en de inzetbaarheid aannemelijk zijn: de meest gemotiveerde en best inzetbare militairen zouden in dat geval bewust voor de zwaarste functies opteren. Nu echter onder die

zwaarste functies het verschil in inzetbaarheid het grootst is tussen beide keuringssystemen, lijkt het er eerder op dat de verschillen tussen beide keuringssystemen juist bij die zwaarste functies tot uiting komen.

Wat de laatste vraag betreft kan het antwoord niet uit dit onderzoek worden verkregen. Na het vaststellen van het verschil in zowel inzetbaarheid als medische consumptie tussen beide keuringssystemen is zonder resultaat onderzocht of andere factoren konden worden gevonden die met de twee uitkomstmaten waren geassocieerd. Er is derhalve een associatie tussen het keuringssysteem en de twee uitkomstmaten, waarvan het mechanisme op grond van de gepleegde analyses niet kon worden achterhaald. Voorlopig wordt dit mechanisme verklaard op basis van het verschil tussen beide keuringssystemen.

Het verschil in uitval op grond van fysieke selectie heeft te maken met het feit dat de BMEKL-groep op basis van toeval meer vrouwelijke kandidaten bevatte, voor wie de fysieke selectie relatief zwaarder is. In de uiteindelijke studiegroepen is geen verschil meer naar geslacht.

De voornaamste reden om niet op te komen na geschikt verklaring lijkt de vertraging van vier maanden die tussen keuring en opkomst ligt. De vroegtijdige uitval van relatief veel geschikte kandidaten vindt niet op medische gronden plaats. De uitslagen van de in deze studie gebruikte exit-interviews sluiten aan bij uitslagen van alle exit-interviews sinds deze in 1996 worden afgenomen, in die zin dat daarin eveneens zelden medische oorzaken voor de vroege "drop-out" naar voren komen.

In deze studie is ervoor gekozen om beide keuringssystemen te vergelijken voor wat betreft hun vermogen om inzetbare militairen op te leveren, tevens kijkend naar de medische consumptie. Weliswaar had follow-up van de ongeschikten mogelijk nog meer informatie kunnen verschaffen, maar het ontbreken hiervan doet geen afbreuk aan de resultaten van deze studie.

Fouten in de interne validiteit zijn voorkomen door randomisatie en het gebruik van data van goede kwaliteit centraal beheerd en gecontroleerd binnen de KL). Ten gevolge van die toevallige samenstelling en omdat de groepen goed vergelijkbaar zijn met de samenstelling van de gehele populatie aan KL-rekruten, kan worden gesteld dat in de externe validiteit geen fouten zijn opgetreden. Ten gevolge van de juiste groeps grootte, de gebruikte statistische methoden en betrouwbaarheidsintervallen was het mogelijk om fouten in de precisie te vermijden.

Tussen het keuringssysteem en de twee uitkomstmaten is een associatie vastgesteld, waarvan het mechanisme niet kon worden achterhaald. Voorlopig wordt dit mechanisme verklaard op basis van het verschil tussen beide keuringssystemen. Nader onderzoek wordt aanbevolen om dit mechanisme te achterhalen. In dit onderzoek is gebruik gemaakt van gegevens over de onderzoekspopulatie.

Volgens Broersen [6] beïnvloedt het beleven van de gezondheid, respectievelijk de gezondheidsrisico's het effect dat deze gezondheidsrisico's op de werknemer uitoefenen. Navraag bij de onderzoeksgroep naar hun persoonlijk beleven had relevante informatie kunnen opleveren teneinde het mechanisme achter de relatie tussen het keuringssysteem en de uitkomstvariabelen te achterhalen. Een vragenlijst zou derhalve wellicht kunnen helpen om verder na te gaan waar de in dit onderzoek gevonden verschillen op berusten.

Geraadpleegde literatuur

1. Raad J de, Reulings PGJ, Schouten HJA: De beperkingen van medische beperkingen. Ned Mil Geneesk T; 1996; 49: 106-111.
2. Kessel JGFM van: Consequenties van onjuist keuren. Ned Mil Geneesk T; 1984; 37: 176-228.
3. Raad J de, Peters JHM, Groot RP de et al: Modernisering van het keuringssysteem bij de Koninklijke Landmacht. TBV 2000; 8: 110-115.
4. Raad J de, Peters JHM., Nijhuis FJN: De kwaliteit van een keuringssysteem. TBV 2000; 8:105-109.
5. Raad J de, Slootman EJH, Groot RP de: Kwaliteitsverbetering bij de Koninklijke Landmacht. Onderzoek naar de kwaliteit van het beoordelen van aanstellingskeuringen voor de Koninklijke Landmacht. TBV 2002; 10: 266-272.
6. Broersen JPJ, Weel ANH, Dijk FJH van: Periodiek Bedrijfsgeneeskundig Onderzoek; middel en maatstaf. Amsterdam, 1989.

6 | Analyse van het verschil in prognostische waarde tussen beide keuringssystemen

In belangrijke mate overgenomen uit: Raad, J de, Redekop WK: Analysis of health factors as predictors for the functioning of military personnel. *Military Medicine* 2005; 170(1): 14-20.

6.1 Inleiding en vraagstelling voor dit onderzoek.

In november 1998 is bij de Koninklijke Landmacht (KL) een nieuw keuringssysteem in gebruik genomen, de Basismedische Eisen KL[1] (BMEKL), ter vervanging van het vorige systeem (ABOHZIS). Het nieuwe systeem is gebaseerd op het model belasting-belastbaarheid en gaat uit van functie-eisen, het oude was diagnosegericht en liet de arts toetsen op de aanwezigheid van ziekten en gebreken.

Om te onderzoeken of tussen beide keuringssystemen verschillen bestaan qua vermogen om geschikte rekruten te onderscheiden, is in de periode juni 1999 – juni 2001 een gerandomiseerde gecontroleerde studie verricht. Daarbij werd een groep van 352 militairen BBT gedurende een periode van twee jaar gevolgd, nadat zij gerandomiseerd via één van beide keuringssystemen waren goedgekeurd en vervolgens opgeleid. Als maatstaven voor het vermogen om geschikte rekruten te onderscheiden werden de variabelen medische consumptie en het aantal dagen inzetbaarheid gehanteerd [2].

Uit deze studie bleek dat militairen die waren gekeurd volgens het nieuwe keuringssysteem statistisch significant meer dagen (648 versus 612) inzetbaar waren en minder medisch consumeerden (€ 396 versus € 746) dan militairen die terzelfder tijd volgens het oude keuringssysteem waren gekeurd. Hierna is onderzocht of andere factoren konden worden gevonden die met de twee uitkomstmaten waren geassocieerd. Weliswaar werd daarbij een associatie tussen het keuringssysteem en de twee uitkomstmaten vastgesteld, maar op grond van de gepleegde analyses kon het mechanisme daarvan niet worden achterhaald. Voorlopig is dit mechanisme verklaard op basis van het verschil tussen beide keuringssystemen.

Eerder onderzoek naar het effect van aanstellingskeuringen door De Kort [2] had aangetoond dat aanstellingskeuringen slechts weinig effect op werknemers uitoefenden. Hij onderzocht de relatie tussen het uitvoeren van aanstellingskeuringen en het optreden van arbeidsongeschiktheid door het optreden van beroepsziekten. De Kort voegt aan zijn hoofdconclusie toe dat het gebruik van een aanstellingskeuring aanvaardbaar kan zijn als het te vermijden risico groot wordt geacht. De risico's voor militairen bij de KL kunnen inderdaad groot zijn, met name gedurende operationele inzet. Het nieuwe BMEKL-systeem is afgeleid van de dagelijkse te verrichten militaire werkzaamheden – inclusief dit operationeel optreden – en vanuit die optiek gezien zou dit systeem in theorie dan ook tot betere resultaten moeten leiden dan het oude ABOHZIS-systeem, in de zin van langduriger inzetbare militairen en/of lagere medische consumptie. Dit laat de vraag open of de in het vorige onderzoek gevonden verschillen tussen beide studiegroepen werkelijk moet worden toegeschreven

aan een verschil in het keuringssysteem of dat (ook nog) andere factoren van invloed een rol spelen.

Er zijn veel factoren mogelijk die de inzetbaarheid van werknemers beïnvloeden [3], zonder dat die invloed direct tot uitval wegens arbeidsongeschiktheid door beroepsziekte leidt. Die invloed kan bijvoorbeeld beperkt blijven tot een algemeen gevoel van onbehagen, frequenter artsbezoek en kortdurend ziekteverzuim, waardoor zowel inzetbaarheid als medische consumptie worden beïnvloed zonder dat er sprake is van een beroepsziekte die tot arbeidsongeschiktheid leidt. Daaraan kan worden toegevoegd dat uit de studie van Broersen, Weel en Van Dijk blijkt dat de eigen beleving van gezondheidsrisico's invloed uitoefent op het effect dat gezondheidsrisico's werkelijk uitoefenen op de betreffende werknemer [4]. De persoonlijke beleving kan aldus mede de onderzoeksuitkomsten (ziekmelding, uitval) beïnvloeden. Het gebruik van bijvoorbeeld een periodieke vragenlijst zou op grond hiervan een toegevoegde waarde kunnen hebben bij verder onderzoek, met name om na te gaan welk mechanisme achter de associatie tussen het keuringssysteem en de uitkomstvariabelen schuilt.

Gedurende de onderzoeksperiode zoals vermeld in het vorige onderzoek werd halfjaarlijks aan de bij het onderzoek betrokken kandidaten een vragenlijst aangeboden, met als doel om inzicht te krijgen in het beleven van de gezondheid, respectievelijk de gezondheidsrisico's door die onderzoekspersonen zelf. De vraagstelling voor dit onderzoek luidde: "Indien een verschil in prognostische waarde tussen beide keuringssystemen wordt gevonden, kan het mechanisme daarvan worden verklaard door analyse van de determinanten persoonskenmerken, de eigen beleving van gezondheid, inzetbaarheid en operationele inzet?"

6.2 Materiaal en methoden

6.2.1 Populatie en procedure

Twee studiegroepen werden samengesteld. Daarbij werden deelnemers uit groep één gekeurd volgens het ABOHZIS-systeem ("ABOHZIS") en deelnemers uit groep twee volgens het BMEKL-systeem ("BMEKL"). Tussen 22 september en 16 oktober 1998 zijn 863 kandidaten gekeurd, waarbij randomisatie plaatsvond door iedere week van keuringssysteem te wisselen (figuur 18). Dit leidde tot de uiteindelijke samenstelling van twee onderzoeksgroepen van in totaal 352 personen, waarvan 166 in de ABOHZIS-studiegroep en 186 in de BMEKL-studiegroep. Uit figuur 18 blijkt de onderzoeksopzet, evenals de exclusiecriteria (een aanvullend, niet in figuur 18 vermeld exclusie criterium was een arbeidsovereenkomst van onbepaalde duur, de zgn. militair "BOT").

Figuur 18: De onderzoeksopzet voor dit hoofdstuk.

Kandidaten, gerandomiseerd gekeurd en in twee studiegroepen verdeeld na opkomst, worden gedurende twee jaar (periode van paraatstelling) gevolgd. Daarbij wordt halfjaarlijks gebruik gemaakt van een vragenlijst.

6.2.2 Informatie

Gegevens werden gebruikt uit de keuringsdossiers van de studiegroepen van de centraal beheerde personeelsgegevens. Gegevens inzake de medische consumptie werden ontleend aan zowel de interne (binnen de KL) gebruikte verrichtingenadministratie als bij de ziektekostenverzekeraar.

In totaal werd gebruik gemaakt van de volgende gegevens: keuringsgegevens, exitinterviews van vroegtijdige dienstverlaters, plaatsings- en persoonsgegevens, verzuimgegevens, gegevens afkomstig van eventuele pensioenkeuringen, overplaatsingen op sociale of medische indicatie, de interne geneeskundige verrichtingenadministratie en declaraties bij de ziektekostenverzekeraar. Voorts werd aan iedere deelnemer gevraagd om halfjaarlijks een vragenlijst in te vullen en terug te sturen, vanaf het begin van de onderzoeksperiode en vervolgens elke 6 maanden gedurende twee jaar. De gehanteerde vragenlijst was gebaseerd op de klassieke "groene vragenlijst" die lang is gebruikt voor het periodiek arbeidsgeneeskundig onderzoek (PAGO) en vrijwel op dezelfde wijze geclusterd [5, 6] (tabel 26). Die groene vragenlijst bestond zelf uit twee onderdelen: het onderdeel gezondheidsbeleven (in deze vragenlijst clusters 3, 4, 5 en 6) en het

onderdeel werkbeleven (in deze vragenlijst clusters 8 tot en met 13). Hieraan waren voor de vragenlijst van dit onderzoek 8 clusters toegevoegd: twee ter controle van elders verkregen gegevens (clusters 1 en 2), één voor het beoordelen van de medische consumptie (cluster 7), en vier clusters om de uit het FIS ontleende BMEKL-vragen in te plaatsen[7] (14 tot en met 17). Het achtste nieuwe cluster – 18 – bevatte de omzetting van de specifieke militaire beoordelingspunten naar vragen. Alle vragen vanaf cluster 3 waren dichotoom. Positieve reacties telden voor 1 punt, negatieve voor 0 punten. Bij de vragen uit de clusters 3 t/m 7 werd specifiek naar de situatie van het laatste half jaar gevraagd.

6.2.3 Effectvariabelen

Inzetbaarheid. Per kandidaat is over de onderzoeksperiode het aantal (kalender-)dagen inzetbaarheid bepaald door aftrek van het aantal dagen ziekteverzuim of andere vormen van arbeidsverzuim, zoals beperkte operationele inzetbaarheid door blessures, vastgesteld door een arts. Hierbij bleek dat het klassieke ziekteverzuim 95% van het totaal uitmaakte. Op deze wijze kon de zuivere inzetbaarheid (het aantal kalenderdagen dat een militair operationeel had kunnen worden ingezet) worden bepaald.

Medische consumptie. De medische eerstelijns consumptie is berekend door alle eerste lijns consulten te rekenen naar het gemiddeld over de onderzoeksperiode geldende CTG-tarief (€ 17,92). De overige gemaakte en gedeclareerde ziektekosten zijn bij de ziektekostenverzekeraar opgevraagd.

6.2.4 Statistiek

Om relaties tussen de effectvariabelen en de overige variabelen te onderzoeken zijn multiple lineaire regressieanalyses uitgevoerd. Omdat sprake was van twee effectvariabelen zijn twee series analyses uitgevoerd. Eerst zijn multiple lineaire regressie analyses verricht met dagen inzetbaarheid als afhankelijke variabele en keuringssysteem als vaste onafhankelijke variabele, achtereenvolgens samen met de volgende onafhankelijke variabelen: leeftijd, gezinsstructuur, geslacht, demografische herkomst (stedelijk of platteland), rookgewoontes (sigaretten of niet), civiele schoolopleiding, fysieke opleidbaarheid, vetpercentage, maximaal vermogen bij de fietsergometrie in watt/kg lichaamsgewicht, de ter keuring gevonden medische bijzonderheden en een aantal uitkomstvariabelen zoals plaatsing, soort functie en het feit of men al dan niet operationeel voor (vrede-)missies is uitgezonden.

Deze regressie analyses zijn vervolgens herhaald, maar nu met medische consumptie als onafhankelijke variabele in plaats van dagen inzetbaarheid.

Tabel 25. De clusters uit alle vragenlijsten, de respons per vragenlijst per studiegroep en de gemiddelde score van de clusters per studiegroep.

Vragenlijst nr.		1	2	3	4	5						
		Nulmeting	6 ^e maand	12 ^e maand	18 ^e maand	24 ^e maand						
Respons / studiegroep	#	Score		Score		Score		Score		Score		
(A)BOHZIS / (B)MEKL		A	B	A	B	A	B	A	B	A	B	
Respons		75%	83%	65%	75%	55%	62%	44%	57%	42%	48%	
Nr	Cluster											
3	Algemene gezondheid	3	2,9 3,0	2,8 2,8	2,7 2,8	2,7 2,6	2,7 2,6	2,7 2,6	2,7 2,6	2,7 2,6	2,8 2,8	
4	Bewegingsapparaat	8	7,5 7,6	7,0 7,0	6,8 7,0	6,6 6,8	6,6 6,8	6,6 6,8	6,6 6,8	6,8 7,0	7,0 7,0	
5	Keel, neus en oor	3	2,5 2,5	2,6 2,6	2,6 2,6	2,6 2,6	2,6 2,6	2,6 2,6	2,6 2,6	2,6 2,6	2,6 2,6	
6	Stress	4	3,8 3,8	3,9 3,9	3,7 3,8	3,7 3,9	3,7 3,9	3,7 3,9	3,7 3,9	3,8 3,8	3,8 3,8	
7	Medische consumptie	8	5,7 5,8	5,6 5,9	5,8 5,8	5,7 5,8	5,7 5,8	5,7 5,8	5,7 5,8	5,6 5,8	5,8 5,8	
8/9	Taakomschrijving / Werkorganisatie	10	8,8 9,1	7,9 7,8	7,6 7,4	6,9 6,8	6,9 6,8	6,9 6,8	6,9 6,8	7,4 7,3	7,3 7,3	
10	Samenwerking	5	4,3 4,4	4,3 4,1	4,0 4,0	3,8 3,8	3,8 3,8	3,8 3,8	3,8 3,8	3,8 3,9	3,9 3,9	
11	Toekomstperspectief	5	4,2 4,2	4,0 4,0	3,7 3,8	3,6 3,2	3,6 3,2	3,6 3,2	3,6 3,2	3,7 3,7	3,7 3,7	
12	Duidelijkheid	3	2,7 2,7	2,4 2,5	2,4 2,5	2,3 2,3	2,3 2,3	2,3 2,3	2,3 2,3	2,2 2,3	2,3 2,3	
13	Privé	3	2,4 2,6	2,5 2,6	2,5 2,6	2,4 2,5	2,4 2,5	2,4 2,5	2,4 2,5	2,4 2,5	2,5 2,5	
14	FIS-bewegingsapparaat	16	15,1 15,3	15,0 15,2	14,8 15,0	14,8 15,2	14,8 15,2	14,8 15,2	14,8 15,2	14,9 15,4	15,4 15,4	
15	Fysisch-chemisch, buiten	8	7,5 7,6	7,4 7,7	7,4 7,7	7,6 7,7	7,6 7,7	7,6 7,7	7,6 7,7	7,6 7,8	7,8 7,8	
16	Zintuigen / veiligheid	4	3,8 3,9	3,8 3,8	3,7 3,7	3,7 3,7	3,7 3,7	3,7 3,7	3,7 3,7	3,7 3,7	3,7 3,7	
17	FIS-Psychisch	13	12,1 12,4	12,2 12,4	12,0 12,5	12,1 12,3	12,1 12,3	12,1 12,3	12,1 12,3	12,2 12,8	12,8 12,8	
18	Militaire beoordelingspunten	15	14,2 14,3	14,0 14,3	13,8 14,1	13,8 14,2	13,8 14,2	13,8 14,2	13,8 14,2	13,8 14,3	14,3 14,3	

#: aantal vragen / cluster, A: ABOHZIS, B: BEMKL, FIS: vragen ontleend aan het FIS.

Van alle vragenlijsten is per studiegroep de respons berekend, van alle clusters uit elke vragenlijst is per studiegroep de gemiddelde uitslag berekend, zoals weergegeven in tabel 25.

Voorafgaand aan verder onderzoek is Crohnbachs' alfa van elk cluster uit de vragenlijsten bepaald. Uit tabel 26 blijkt dat de meeste clusters voldeden aan de eis dat de alfa tenminste 0,50 moet zijn, behoudens enkele clusters uit de eerste vragenlijst en cluster 5 (KNO) uit de tweede. Alleen de clusters 12 "Duidelijk-

heid" (drie vragen) en 16 "Zintuigen/veiligheid" (vier vragen) voldeden stelselmatig niet aan deze eis. Uit vier clusters (3, 4, 7 en 17) is stelselmatig dezelfde enkele vraag verwijderd om tot een $\alpha \geq 0,50$ te komen, om diezelfde reden zijn de clusters 8 en 9 gecombineerd. Om de clusters met statistisch significante associatie met inzetbaarheid en/of medische consumptie te selecteren werden bivariate lineaire regressie analyses uitgevoerd met achtereenvolgens inzetbaarheid en medische consumptie als afhankelijke variabele, waarbij het keurings-systeem samen met telkens een enkel cluster de onafhankelijke variabelen vormen.

Vervolgens werden twee multiple lineaire regressie analyses uitgevoerd met één van beide effectvariabelen als afhankelijke variabele en de geselecteerde clusters plus keuringssysteem als onafhankelijke variabelen. Een p-waarde van minder dan 0,05 werd als statistisch significant beschouwd. Alle analyses zijn uitgevoerd met behulp van SPSS versie 10.0.

6.3 Resultaten

6.3.1 Uitkomsten van het voorgaande onderzoek op hoofdlijnen

In totaal werden in de ABOHZIS-groep 399 kandidaten gekeurd, waarvan 304 geschikt werden bevonden (tabel 27). Hiervan kwamen 243 als Beroepsmilitair Bepaalde Tijd (BBT) op. Op grond van de exclusiecriteria bleef daarvan een studiegroep van 166 kandidaten over. Voor de BMEKL-groep werden 464 kandidaten gekeurd, waarvan 341 geschikt werden verklaard; hiervan kwamen 270 BBT-ers op, op grond van de exclusiecriteria resulterend in een studiegroep van 186 kandidaten.

Van de vroege drop-outs werd op grond van verschillende informatiebronnen (uitslagen van de exitinterviews, het achterwege laten van ingediende claims) aangenomen dat hun vertrek op niet-medische gronden berust. In het begin van de onderzoeksperiode bleken persoonskenmerken (leeftijd, geslacht, gezins-samenstelling, demografische herkomst, schoolopleiding, rookgewoonten) noch de keuringsgegevens (gewicht, lengte, medische bijzonderheden, vetpercentage, fysiek prestatievermogen) statistisch significant te verschillen tussen beide studiegroepen. Tijdens de onderzoeksperiode bleken statistisch significante verschillen te zijn opgetreden tussen beide studiegroepen qua indeling naar gevechts- of gevechtsondersteunende functie ($p = 0,02$), het aantal dagen inzetbaarheid (gemiddeld 648 dagen voor de BMEKL-groep tegenover 612 voor de ABOHZIS-groep; $p = 0,02$) en de medische consumptie (gemiddeld € 396 tegenover € 746; $p = 0,05$).

Tabel 26. De clusterindeling van de gebruikte vragenlijst, tezamen met de herkomst van de vragen, respons en Cronbachs' alfa van de clusters van alle vragenlijsten.

Respons		Enquête	1	2	3	4	5
			70%	69%	58%	49%	46%
		Aantal vragen	Alfa	Alfa	Alfa	Alfa	Alfa
Nr	Cluster						
3	Algemene gezondheid *	3	0,61	0,63	0,64	0,61	0,74
4	Bewegingsapparaat *	8	0,62	0,66	0,62	0,71	0,62
5	KNO *	3	0,46	0,48	0,55	0,55	0,53
6	Stress *	4	0,29	0,73	0,56	0,71	0,61
7	Medische Consumptie **	8	0,62	0,65	0,72	0,70	0,70
8/9 (1)	Taakhoud/ Werkorganisatie *	10	0,52	0,56	0,58	0,66	0,65
10	Samenwerking *	5	0,54	0,68	0,54	0,60	0,62
11	Toekomstperspectief *	5	0,50	0,50	0,65	0,70	0,68
12	Duidelijkheid *	3	0,34	0,34	0,31	0,41	0,36
13	Privé *	3	0,52	0,59	0,44	0,44	0,51
14	FIS-bewegingsapparaat **	16	0,60	0,87	0,86	0,89	0,88
15	Fysisch-chemisch, buiten **	8	0,35	0,81	0,81	0,59	0,67
16	Zintuigen / veiligheid **	4	0,28	0,09	0,12	0,21	0,08
17	FIS-psychisch **	13	0,44	0,79	0,55	0,79	0,76
18	Militaire beoordelingspunten **	15	0,12	0,69	0,59	0,61	0,66

Toelichting: De grijs gearceerde clusters voldoen niet aan de eis dat de alfa tenminste 0,50 moet zijn. (1): Op grond van de alfa's zijn de clusters 8 en 9 gecombineerd. De met * gemarkeerde clusters zijn afkomstig uit het PAGO en de met ** gemarkeerde zijn door de onderzoeker ontwikkeld, respectievelijk ontleend aan het FIS.

6.3.2 Analyse persoonskenmerken en uitkomstvariabelen

Uit multiple lineaire regressie analyses met dagen inzetbaarheid en medische consumptie als afhankelijke variabelen en persoonskenmerken en uitkomstvariabelen als onafhankelijke variabelen bleek dat uitgezonden militairen statistisch significant meer dagen inzetbaar waren dan niet uitgezonden militairen (tabel 27). Schoolopleiding bleek geen invloed te hebben ($p = 0,06$). De p-waarde van het keuringssysteem bleef $< 0,05$.

Tabel 27. Multiple lineaire regressie analyse naar keuringssysteem en de eventuele uitzending, waarvan er twee mogelijk waren: één in Bosnië-Herzegovina (SFOR) en één in Kosovo (KFOR).

Onafhankelijke variabelen	Afhankelijke variabele: Dagen inzetbaarheid			
	Bèta Coëfficiënt	Standaard fout	P-waarde	R ²
Keuringssysteem (BMEKL vs ABOHZIS)	118	16	<0,01	
Type uitzending versus Overige typen uitzending				
Geen uitzending	510	46	<0,01	
SFOR	280	48	0,01	
KFOR	250	51	0,04	
				0,16

Toelichting: R² staat voor de verklaarde variantie.

6.3.3 Enquêtes

Alleen de volledig ingevulde vragenlijsten werden meegeteld. Dit resulteerde in een respons van gemiddeld 2,9 van de 5 lijsten ($2,9 \pm 1,6$). In totaal reageerden 28 personen nooit, 61 reageerden eenmalig (vooral aan het begin), 60 reageerden tweemaal, 48 driemaal, 79 viermaal en 76 vijfmaal. Tussen beide studiegroepen zat qua respons per vragenlijst eenmaal een statistisch significant verschil (vragenlijst 4, tabel 25), de respons daalde van 70% naar 46% (tabel 25).

Bij lineaire regressie analyses met inzetbaarheid als afhankelijke variabele en het keuringssysteem met telkens één van de clusters als onafhankelijke variabelen komt een statistisch significante associatie ($p < 0,05$) naar voren tussen inzetbaarheid en de clusters Algemene gezondheid (3), Bewegingsapparaat (4), Medische consumptie (7) en Militaire beoordelingspunten (18).

Vervolgens is multiple lineaire regressie analyse uitgevoerd met al deze clusters per vragenlijst tezamen en het keuringssysteem als onafhankelijke variabelen en met dagen inzetbaarheid als afhankelijke variabele. Daaruit bleek (tabel 28) dat het cluster Militaire beoordelingspunten (18) bij drie enquêtes statistisch significante associatie vertoont met de inzetbaarheid, voor de clusters Medische consumptie (7) en Algemene gezondheid (3) is dit elk twee maal het geval, voor de beide clusters over het bewegingsapparaat (4 en 14) is dit elk eenmaal het geval. De p-waarde van het keuringssysteem bleef bij deze toetsing boven de 0,05.

Bij herhaling van de lineaire regressie analyse zoals hierboven vermeld, maar nu met medische consumptie in plaats van dagen inzetbaarheid als afhankelijke

Tabel 28. Resultaten van de multiple lineaire regressie analyse met dagen inzetbaarheid als afhankelijke variabele en het keuringssysteem en de clusters als onafhankelijke variabelen.

Enquête	Cluster	Beta- Coëfficiënt (dagen)	Standaardfout (dagen)	P-waarde	R ²
2	Keuringssysteem	-23	15	0,07	0,14
	Cluster 7: medische consumptie 2	-15	4	0,02	
	Cluster 18: militaire punten 2	-43	6	<0,01	
3	Keuringssysteem	-29	15	0,06	0,23
	Cluster 3: algemene gezondheid 3	-35	11	<0,01	
	Cluster 4: bewegingsapparaat 3	-26	6	<0,01	
	Cluster 18: militaire punten 3	-37	7	<0,01	
4	Keuringssysteem	-18	19	0,41	0,21
	Cluster 3: algemene gezondheid 4	-60	7	0,01	
	Cluster 18: militaire punten 4	-37	7	<0,01	
5	Keuringssysteem	-7	20	0,54	0,20
	Cluster 7: medische consumptie 5	-21	6	0,01	
	Cluster 14: FIS-bewegingsapparaat 5	-20	9	<0,01	

Alleen de clusters met een lage p-waarde zijn weergegeven. R² staat voor de verklaarde variantie.

variabele bleek vaak sprake van een statistisch significante associatie tussen de clusters Stress (6), Samenwerking (10), FIS-bewegingsapparaat (14), FIS-fysisch-chemisch (15), FIS-psychisch (17) en Militaire beoordelingspunten (18) en medische consumptie.

Hierna is opnieuw multiple lineaire regressie analyse uitgevoerd zoals eveneens hierboven weergegeven, maar met medische consumptie in plaats van dagen

Tabel 29. De resultaten van de multiple lineaire regressie analyse met medische consumptie als afhankelijke variabele en het keuringssysteem en de clusters als onafhankelijke variabelen.

Enquête	Cluster	Bèta-coëfficiënt	Standaardfout	P-waarde	R ²
		€	€		
1	Keuringssysteem	-358	217	0,10	0,10
	Cluster 9: samenwerking 1	-470	114	<0,01	
	Cluster 14:	-184	85	0,03	
	Fis-bewegingsapparaat 1				
2	Keuringssysteem	-351	216	0,10	0,10
	Cluster 15:	-431	107	<0,01	
	FIS-fysisch-chemisch 2				
3	Keuringssysteem	-190	168	0,29	0,15
	Cluster 6: Stress 3	-396	197	0,03	
	Cluster 15:	-257	133	0,01	
	FIS-fysisch-chemisch 3				
	Cluster 17: FIS-psychisch 3	-212	92	0,01	
4	Keuringssysteem	-136	222	0,33	0,31
	Cluster 4: bewegingsapparaat 4	-220	97	0,03	
	Cluster 5: KNO 4	-387	74	0,01	
	Cluster 18: militaire punten 4	-407	113	<0,01	
5	Keuringssysteem	-350	271	0,61	0,28
	Cluster 17: FIS-psychisch 5	-520	150	<0,01	
	Cluster 18: militaire punten 5	-340	113	<0,01	

Alleen de clusters met een lage p-waarde zijn weergegeven. R² staat voor de verklaarde variantie.

inzetbaarheid als afhankelijke variabele (tabel 29). Uit deze analyse bleek dat de volgende clusters alle in twee van de vijf vragenlijsten een statistisch significante associatie ($p < 0,05$) met medische consumptie vertoonden: Stress (6), FIS-fysisch-chemisch (15), FIS-psychisch (17) en Militaire beoordelingspunten (18).

6.3.4 (Non) respons

Uit non-responsie analyse bleek dat kandidaten niet responderen indien zij zijn afgekeurd, niet zijn opgekomen of reeds vroeg tijdens de militaire opleiding zijn vertrokken. Ook nadien bleek dat diegenen die uit dienst vertrekken, daarna niet langer reageren. Qua leeftijd, schoolopleiding of keuringssysteem zit er geen statistisch significant verschil tussen de groepen respondenten en non-respondenten, wel naar geslacht (meer vrouwen non-respondent).

Nadere (multiple lineaire regressie) analyse (inzetbaarheid als afhankelijke variabele en het aantal malen respons als onafhankelijke variabele) toonde aan dat diegenen die nooit reageren op de vragenlijst ($p < 0,01$) of slechts eenmalig ($p = 0,02$) statistisch significant minder dagen inzetbaar zijn geweest dan de overige respondenten. Tijdens deze analyse bleef de p -waarde van het keuringssysteem relatief gering ($p = 0,09$). Voor de medische consumptie bleek het aantal malen response niet van belang ($p = 0,40$). Op grond hiervan is vervolgens nagegaan of tussen de beide te onderscheiden groepen (non-responders plus eenmalige responders versus de overigen) ten aanzien van vermelde persoonskenmerken, keuringsgegevens en uitkomstvariabelen sprake was verschillen. Alleen de schoolopleiding van diegenen die meerdere malen op de vragenlijst reageerden, bleek statistisch significant hoger te liggen dan die van de anderen ($p = 0,02$).

6.4 Beschouwing

In de door de keurend arts geregistreerde persoonskenmerken zijn geen factoren van invloed te vinden op de inzetbaarheid of medische consumptie. Wel werd een aantal andere factoren gevonden, geassocieerd met de inzetbaarheid of medische consumptie.

Het al (positief) dan niet (negatief) operationeel worden uitgezonden is een met inzetbaarheid geassocieerde factor. Een hogere respons op de vragenlijst – gekoppeld aan een hogere schoolopleiding – is eveneens een factor van invloed. Ook het eigen beleven van de gezondheid, het bewegingsapparaat en de mate waarin gebruik wordt gemaakt van medische instanties is van invloed; deze invloed is vermoedelijk tijdelijk (voor periodes van telkens een half jaar, gelet op de vraagstelling). Hoe meer men zichzelf in staat acht tot het uitvoeren van specifieke militaire taken, des te groter is het aantal dagen inzetbaarheid.

Anders dan bij inzetbaarheid wordt medische consumptie uitsluitend door de eigen beleving beïnvloed. Factoren die daarbij een rol spelen zijn de mate waarin men zichzelf in staat acht tot het ondergaan van meer extreme klimatologische omstandigheden, de mate waarin men zichzelf psychisch belastbaar acht en – gelijk bij inzetbaarheid – de mate waarin men zichzelf in staat acht tot het verrichten van specifieke militaire taken.

De directe verklaring voor de relatie tussen respons en inzetbaarheid is het niveau van de schoolopleiding: hoe hoger die opleiding, des te groter de inzet. Tegelijkertijd ontbreekt de relatie tussen response en medische consumptie. Omdat behalve schoolopleiding geen enkel verschil tussen beide groepen kan worden aangetoond, is de meest waarschijnlijke indirecte verklaring voor de bovenstaande gegevens dat (een gebrek aan) motivatie op de achtergrond speelde, bijvoorbeeld door een hoger verzuim te vertonen waarbij geen extra medische kosten werden gemaakt. Omdat non-respons echter het primaire kenmerk van de ene groep is, kan dit niet met behulp van de vragenlijst worden aangetoond.

Operationele uitzending voor buitenlandse missies gedurende 6 maanden – een kerntaak van de KL – is geassocieerd met grotere inzetbaarheid, c.q. een lager verzuim. De meest plausibele verklaringen hiervoor zijn de specifieke verblijfsomstandigheden en de sociale controle tijdens uitzending. Voor de medische consumptie maakt het niet uit of men al dan niet is uitgezonden. Dit wordt niet veroorzaakt door de verrichtingenregistratie, welke tijdens uitzending gecontinueerd wordt. Bij ernstiger medische problemen tijdens uitzending volgt bovendien repatriëring, waarna verdere ziektekosten in eigen land worden gedeclareerd. Het kan niet worden uitgesloten dat het niet worden uitgezonden een rol heeft gespeeld bij inzetbaarheid: teleurstelling omdat men niet wordt uitgezonden, met meer verzuim als gevolg.

De invloed van clusters uit de vragenlijst op inzetbaarheid en medische consumptie is beperkt van omvang: op een potentiële inzetbaarheid van 731 dagen is de gezamenlijke bèta-coëfficiënt van de clusters uit de tweede vragenlijst bijvoorbeeld 58 ± 10 dagen. Ook bij multiple lineaire regressie analyses met medische consumptie als afhankelijke variabele zijn de bèta-coëfficiënten van de betrokken clusters bescheiden van omvang.

In deze studie is onderzoek verricht met als doel het vinden van kenmerken, geassocieerd met het functioneren als militair. Dergelijke kenmerken zijn ook gevonden (figuur 19). Daarnaast is een verklaring gezocht voor het feit dat het BMEKL-keuringssysteem rekruten selecteert die een grotere inzetbaarheid en lagere medische consumptie vertonen dan het ABOHZIS-keuringssysteem. Aanwezige blessures worden in beide keuringssystemen door anamnese en lichamelijk onderzoek stellig opgemerkt. Of veel blessures zullen optreden valt niet te voorspellen, evenmin of gebruik van medische zorg zal worden gemaakt. Internationaal wordt te intensieve training beschouwd als een hoofdoorzaak van blessureleed [8, 9, 10], maar door de fysieke selectie vooraf bij de KL zou de

* : De factor (hogere motivatie) is gebaseerd op een aanname.

Figuur 19: ter keuring meetbare en inzetbaarheid bevorderende (+), respectievelijk niet ter keuring meetbare, de inzetbaarheid verlagende (-) factoren van invloed gedurende de contractduur van militairen BBT.

noodzaak voor te intensieve training worden vermeden. Van schoolopleiding was door interne KL-studies al langer een relatie met uitval bekend; bovendien is het vaststellen van de daarvoor aangewezen ondergrens geen medische aangelegenheid. Of militairen al dan niet voor vredesmissies zullen worden uitgezonden wordt evenmin door de medische keuring bepaald.

Het beoordelen of een kandidaat voldoet aan taakspecifieke (militaire) eisen vormt, in tegenstelling tot bij het oude ABOHZIS keuringssysteem, een vast onderdeel van het nieuwe BMEKL keuringssysteem.

Volgens deze redenatie heeft één factor uit de keuring conform het BMEKL-systeem de meeste invloed gehad op het verschil in inzetbaarheid. Dat betreft de gehanteerde werkwijze (toetsing aan de taakspecifieke eisen) voor het beoordelen van de geschiktheid. Voor medische consumptie speelt daarnaast de psychische belastbaarheid een rol. Ook deze belastbaarheid wordt alleen met behulp van het BMEKL-systeem getoetst.

Deze onderzoeksbevindingen komen niet overeen met de theorie dat de invloed van aanstellingskeringen op werknemers gering is, een theorie die vooral is ontstaan naar aanleiding van het onderzoek van De Kort [2]. Voor het feit dat deze bevindingen niet met de heersende theorie overeenkomen, kunnen twee verklaringen worden gegeven.

Zelfs De Kort erkent dat het gebruik van een aanstellingskeuring aanvaardbaar kan zijn als het te vermijden risico groot wordt geacht. Dit is nu precies het geval bij het optreden als militair. Een militair wordt dan ook niet voor een bepaalde functie gekeurd, maar wordt gekeurd om een uiteenlopend scala van taken uit te voeren, met inbegrip van operationele inzet. Het verschil tussen de theorie en de onderzoeksbevindingen zou kunnen zijn gebaseerd op het feit dat dit soort werk – het militaire – uitermate gevarieerd is en veel verschillende risico's kent, geestelijk zowel als lichamelijk. In geval er sprake zou zijn van minder veeleisende functies zou de vigerende theorie van toepassing blijven. Een tweede verklaring kan zijn dat de eisen hoger liggen dan de werkelijk vereiste belastbaarheid, waardoor veel feitelijk geschikte kandidaten onterecht worden afgewezen. Deze verklaring lijkt minder waarschijnlijk dan de eerste. Weliswaar bestaat geen gouden standaard, zodat de mogelijkheid niet kan worden uitgesloten, maar tussen beide keuringssystemen zit geen statistisch significant verschil inzake de percentages afgewezen kandidaten. Desalniettemin zou het alsnog zo kunnen zijn dat beide keuringssystemen "te streng keuren".

Samenvattend kan worden geconcludeerd dat in deze studie een aantal factoren is gevonden die van invloed zijn op de inzetbaarheid en de medisch consumptie: schoolopleiding, blessures, operationele inzet en het BMEKL-keuringssysteem zelf. De betere prestatie van het BMEKL-keuringssysteem lijkt deels te kunnen worden toegeschreven aan het vermogen om op taakspecifieke eisen te beoordelen. Maar het primaire mechanisme waarop dit is gebaseerd, blijft verborgen.

Geraadpleegde literatuur

1. Raad J de, Peters JHM, Groot RP de et al: Modernisering van het keuringssysteem bij de Koninklijke Landmacht; keuringsprotocollen en functioneel beoordelen. TBV 2000; 8: 110-115.
2. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
3. Klein Hesselink DJ, Kruidenier HJ, Veerman TJ et al: Afwezigheid verklaard. Een literatuurstudie naar determinanten van ziekteverzuim en arbeidsongeschiktheid. NIA. Amsterdam, 1993.
4. Broersen JPJ, Weel ANH, Dijk FJH van: Periodiek Bedrijfsgeneeskundig Onderzoek; middel en maatstaf. Amsterdam, 1989.
5. Putten DJ van, Marcelissen FGH, Winter CJ de: De vragenlijst voor PBGO 1-4. Achtergronden voor het gebruik ervan. Het wijzigen van antwoorden door bedrijfsartsen. Methoden voor gebruik van referentiegegevens. TSG 1988; 66: 115-120, 143-147.
6. Putten DJ van: De vragenlijst voor PBGO, voorstellen tot wijziging van de vragenlijst. Leiden, NIPG/TNO, 1989.
7. Raad J. de, Peters JHM., Nijhuis FJM: De kwaliteit van een keuringssysteem. Een pilot-onderzoek naar de kwaliteit van het keuringssysteem volgens de basismedische eisen van de Koninklijke Landmacht. TBV 2000; 8:105-109.

8. Jordaan G, Schweltnus MP: The incidence of overuse injuries in military recruits during basic military training. *Mil Med* 1994; 159(6):421-426.
9. Almeida SA, Williams KM, Shaffer RA et al: Epidemiological patterns of musculoskeletal injuries and physical training *Med Sci Sports Exerc* 1999; 31(8):1176-1182.
10. Pope RP, Herbert R, Kirwan JD et al: Predicting attrition in basic military training. *Mil Med* 1999; 164(10): 710-714.

7 | Verschillen in inzetbaarheid tussen uitgezonden militairen

In belangrijke mate overgenomen uit: Raad J de, Nijhuis FJN, Willems JHBM: Difference in fitness for duty between soldiers on a mission. Can these be explained by a difference in the pre-employment assessment? In september 2005 door Military Medicine voor publicatie ingepland.

7.1 inleiding en vraagstelling voor dit onderzoek

In november 1998 is bij de Koninklijke Landmacht (KL) een nieuw keuringssysteem in gebruik genomen, de Basismedische Eisen KL [1] (BMEKL), ter vervanging van het vorige systeem (ABOHZIS). Het nieuwe systeem is gebaseerd op het model "belasting-belastbaarheid" [2]. De functie-eisen voor de militair zijn geoperationaliseerd in 43 beoordelingspunten. Elk beoordelingspunt (zoals "staan" en "lopen", maar ook "vuurwapen hanteren", "scherfwerend vest dragen" en "graven") is voorzien van een norm. Tezamen vormen deze normen de BMEKL. Ze zijn afkomstig van het belastbaarheidpatroon van het FIS en specifieke militaire activiteiten. Het oude – diagnosegerichte – keuringssysteem verplichtte de keurend arts om ziekten en gebreken te determineren, waarna de uitslag werd bepaald op grond van een vastgestelde lijst met aandoeningen. Er bestonden geen functie-eisen. Bekend was dat dit systeem geen relatie met de feitelijke militaire werkzaamheden vertoonde [3]; er werd vooral geen rekening gehouden met operationeel optreden tijdens (vrede-)missies. Bij het nieuwe keuringssysteem wordt gekeken of de gevonden medische bijzonderheden tot beperkingen leiden ten opzichte van functie-eisen. Deze functie-eisen zijn geoperationaliseerd in een minimaal vereist operationele functioneringsniveau, dat wel alle feitelijke militaire werkzaamheden omvat. In het ABOHZIS-systeem kon men op grond van een aandoening worden afgekeurd, ongeacht de mate van belastbaarheid; in het BMEKL-systeem kan men geschikt worden verklaard aan de hand van de mate van belastbaarheid, ongeacht de aandoening.

Na de introductie van het nieuwe systeem is begonnen met de evaluatie ervan. Inmiddels is al onderzoek verricht naar de mate van overeenstemming [4], [5]. Hierbij was ter illustratie de kappa van Cohen [6] ¹bij pensioenkeuringen tussen de twee bij dit onderzoek betrokken artsen 0,74; bij de aanstellingskeuringen werd tussen de verschillende bij dat onderzoek betrokken paren artsen een kappa van Cohen gevonden die varieerde tussen 0,69 en 0,95.

Voorafgaande aan de formele invoering van het nieuwe systeem deed zich de gelegenheid voor om beide keuringssystemen naast elkaar te gebruiken bij de aanstellingskeuring, in de wetenschap dat een gouden standaard ontbreekt. Als gevolg daarvan kon een onderzoek worden opgezet om de (lange termijn) effecten van beide keuringssystemen op het aldus gekeurde militair personeel te vergelijken. De hieruit voortgekomen studies onderscheiden zich in tweeërlei

1 $Kappa = P_o - P_e / 1 - P_e$, waarbij P_o de waargenomen overeenstemming en P_e de toevallige overeenstemming is.

opzicht van wat tot op heden in Nederland aan onderzoek naar de relevantie van medische aanstellingskeuringen is verricht [7]. Tot nu toe is de relatie tussen keuring enerzijds en uitval op het werk door ziekte of uitstroom naar arbeidsongeschiktheid anderzijds onderzocht. In deze studie is de tegenovergestelde relatie onderzocht: tussen de keuring en de mate waarin de werknemer adequaat functioneert. Daarnaast is in deze studie ook aandacht besteed aan de beleving van de arbeidsomstandigheden door de werknemers zelf.

Om te onderzoeken of tussen beide keuringssystemen verschillen in dagen inzetbaarheid en medische consumptie onder het aangenomen personeel zouden optreden, is in de periode 1999 – 2001 een gerandomiseerde gecontroleerde studie verricht. Daarbij werd een groep van 352 militairen met een verbintenis van bepaalde duur (Beroeps Bepaalde Tijd, BBT) gedurende een periode van twee jaar gevolgd, nadat zij gerandomiseerd via één van beide keuringssystemen waren goedgekeurd en in twee studiegroepen waren verdeeld [8]. Als maatstaven voor het meten van de effecten op het personeel werden de uitkomstvariabelen medische consumptie en het aantal dagen inzetbaarheid gehanteerd.

Gedurende diezelfde onderzoeksperiode werd halfjaarlijks aan de bij het onderzoek betrokken kandidaten een vragenlijst aangeboden. Het doel daarvan was om inzicht te krijgen in het beleven van de gezondheid(-srisico's) door de respondent. Aan die beleving wordt invloed toegeschreven op het effect dat gezondheidsrisico's uitoefenen op de werknemer [9]. Een vragenlijst zou aldus de mogelijkheden uitbreiden om na te gaan waar de in het vorige onderzoek gevonden verschillen op berusten.

Eén van de belangrijkste redenen om naar een nieuw keuringssysteem over te gaan was dat het oude systeem onvoldoende rekening hield met de operationele taakstelling. Die taakstelling manifesteert zich in de participatie van internationale missies in het buitenland. Sinds 1992 is de bijdrage van de Nederlandse krijgsmacht aan internationale missies een steeds grotere plaats gaan innemen in de algehele taakstelling. Voor dergelijke missies worden eenheden samengesteld die gedurende 4 tot 6 maanden achtereen in het aangegeven inzetgebied operationeel actief zijn. De arbeidsomstandigheden ter plekke zijn afhankelijk van de lokale (militaire, politieke, klimatologische en geografische) omstandigheden en zijn meestal anders dan die in het eigen land.

De militair wordt niet voor een bepaalde functie aangesteld, maar moet alle opgedragen werkzaamheden en diensten verrichten en kan daarbij overal ter wereld onder uiteenlopende arbeidsomstandigheden worden ingezet. De wetgever heeft aangegeven dat deze “multifunctionaliteit” zware fysieke en

geestelijke eisen aan de militair stelt. Juist daarom stelt de wetgever de krijgsmacht in de gelegenheid om een aanstellingskeuring uit te voeren bij aspirant-militairen. De verzwarende arbeidsomstandigheden daarbij zijn de training en de opleiding voor de operationele taakstelling, de oefeningen en de uitzendingen zelf. Op de eigen werkplek – de kazerne – verschillen de arbeidsomstandigheden weinig van die van andere Nederlandse werknemers.

De aanstellingskeuring wordt gebruikt voorafgaande aan de indiensttreding van militair beroepspersoneel, waar de KL sinds de opschorting van de dienstplicht (1996) nog louter uit bestaat. Naast kaderleden met een verbintenis voor onbepaalde tijd ("BOT") bestaat dit personeel in toenemende mate uit manschappen en kaderleden met een verbintenis voor bepaalde tijd, in de regel voor 2½ jaar ("BBT"). Het eerste half jaar van die verbintenis bestaat uit verschillende opleidingen², waarna een plaatsing op de definitieve functie volgt. Deze functies kunnen worden onderverdeeld in zogenaamde gevechts- dan wel gevechtsondersteunende functies. Uit intern onderzoek was reeds bekend dat gedurende de opleidingsperiode een uitstroom plaatsvindt van 20 – 35%. Hiervoor bestaan diverse oorzaken van niet-medische aard: dat men niet de functie heeft gekregen die men wenste, of dat de bedrijfscultuur achteraf toch tegenviel, ofwel dat men het vele buitenwerk (veldoefeningen etc.) onaantrekkelijk vindt. Na de opleiding is het verloop zeer gering.

Met name het aanstellen van de hierboven vermelde militairen BBT is op participatie aan operationele missies in het buitenland gericht. Niet elke militair BBT wordt echter gedurende de verbintenis effectief operationeel ingezet.

Zowel bij een vergelijking van de mate van overeenstemming als van de voorspellende waarde tussen beide keuringssystemen zal er sprake zijn van groepen waarbij een gedeelte operationeel is ingezet en een ander gedeelte waarbij dit niet het geval is. Juist onder operationele omstandigheden zou het verschil tussen beide keuringssystemen in theorie moeten opvallen. Om die reden is nader onderzoek naar verschillen tussen louter uitgezonden militairen uit de hierboven gememoreerde studiegroepen relevant.

Uit het bovenstaande volgt de vraagstelling voor dit onderzoek: "Bestaat er ook een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van uitsluitend de uitgezonden militairen en zo ja, kan dat worden verklaard

2 Voor manschappen bestaan er qua fysieke belasting twee soorten opleiding, de reguliere en de zwaardere (en iets langere) voor de Luchtmobiele Brigade. Kaderleden krijgen een opleiding aan eigen instituten, fysiek ongeveer vergelijkbaar met de zwaarste opleiding voor manschappen.

door analyse van de determinanten persoonskenmerken, het beleven van de eigen gezondheid, de eigen beleving van gezondheid en inzetbaarheid?"

7.2 Materiaal en methoden

7.2.1 Populatie en procedure

Twee studiegroepen werden samengesteld. Daarbij werden deelnemers uit groep één gekeurd volgens het ABOHZIS-systeem ("ABOHZIS") en deelnemers uit groep twee volgens het BMEKL-systeem ("BMEKL"). Tussen 22 september en 16 oktober 1998 zijn 863 kandidaten gekeurd, waarbij randomisatie plaatsvond door iedere week van keuringssysteem te wisselen. Dit leidde tot de uiteindelijke samenstelling van twee onderzoeksgroepen van in totaal 352 personen, waarvan 166 in de ABOHZIS-studiegroep en 186 in de BMEKL-studiegroep (figuur 20). Uit deze studiegroepen zijn voor dit onderzoek die militairen geselecteerd die volgens hun persoonsgegevens zijn uitgezonden tijdens de onderzoeksperiode. Uiteindelijk bleken van de ABOHZIS-groep 77 (46%) uitgezonden te zijn geweest, tegenover 84 (45%) uit de BMEKL-groep (Tabel 30).

Teneinde voor het volgen van de onderzoekspersonen gebruik te kunnen maken van hun persoons- en keuringsgegevens verzochten de keurend artsen alle kandidaten om hun schriftelijke toestemming; die toestemming werd zonder uitzondering verkregen. De kandidaten waren niet op de hoogte van het onderzoek of het feit dat zij gerandomiseerd werden gekeurd. De onderzoeksperiode liep van 1 juni 1999 – 1 juni 2001. De redenen voor exclusie waren: ongeschikt verklaring, militairen BOT, plaatsing buiten de KL, geen opkomst / vertrek reeds in de militaire opleiding, opkomst na 1 juni 1999 en atypische functies.

7.2.2 Informatie

Informatie werd gehaald uit de keuringsdossiers van de studiegroepen, alsmede van de centraal beheerde personeelsgegevens inzake de studiegroepen. Gegevens over de medische consumptie werden ontleend aan zowel de interne (binnen de KL) gebruikte verrichtingenadministratie als bij de ziektekostenverzekeraar.

In totaal werd gebruik gemaakt van de volgende gegevens: keuringsgegevens, exitinterviews van vroegtijdige dienstverlaters, plaatsings- en persoonsgegevens, verzuimgegevens, gegevens afkomstig van eventuele pensioenkeuringen, overplaatsingen op sociale of medische indicatie, de interne geneeskundige verrichtingenadministratie en declaraties bij de ziektekostenverzekeraar. Voorts werd

Figuur 20: De onderzoeksopzet voor hoofdstuk 7.
In deze studie worden uitsluitend de uitgezonden militairen, gerandomiseerd gekeurd en in twee studiegroepen verdeeld, gevolgd.

Tabel 30. Het verloop van de gekeurde militairen.

	ABOHZIS groep		BMEKL groep	
Gekeurd	399	100%	464	100%
Afgekeurd	-64	16%	-58	13%
Uitval door fysieke selectie	-31	8%	-65	14%
Geschild	304	76%	341	73%
Beroeps Onbepaalde Tijd (BOT)	-8	2%	-17	4%
Niet bij KL maar elders	-15	4%	-18	4%
Nooit opgekomen	-36	9%	-33	7%
Anders *)	-2	1%	-3	1%
Beroeps Bepaalde Tijd (BBT)	243	61%	270	58%
Te kort in dienst	-75	19%	-80	17%
Te laat in dienst	-2	1%	-4	1%
Studiegroepen	166	42%	186	40%
Uitgezonden	77	46%	84	45%

*) "Anders" staat voor afwijkende militaire functies, zoals geestelijk verzorger en militaire muzikant.

Tabel 31. De clusterindeling van de gebruikte vragenlijst, tezamen met de herkomst van de vragen, respons en Crohnbachs' alfa van de clusters van alle vragenlijsten.

Nr	Cluster	Enquête	2	3	4
		Aantal vragen	69% Alfa	58% Alfa	49% Alfa
3	Algemene gezondheid *	3	0,63	0,64	0,61
4	Bewegingsapparaat *	8	0,66	0,62	0,71
5	KNO *	3	0,48	0,55	0,55
6	Stress *	4	0,73	0,56	0,71
7	Medische Consumptie **	8	0,65	0,72	0,70
8/9 (1)	Taakhoud/ Werkorganisatie *	10	0,56	0,58	0,66
10	Samenwerking *	5	0,68	0,54	0,60
11	Toekomstperspectief *	5	0,50	0,65	0,70
12	Duidelijkheid *	3			
13	Privé *	3	0,59	0,44	0,44
14	FIS-bewegingsapparaat **	16	0,87	0,86	0,89
15	Fysisch-chemisch, buiten **	8	0,81	0,81	0,59
16	Zintuigen / veiligheid **	4			
17	FIS-psychisch **	13	0,79	0,55	0,79
18	Militaire beoordelingspunten **	15	0,69	0,59	0,61

De grijs gearceerde clusters voldoen niet aan de eis dat de alfa tenminste 0,50 moet zijn. (1): Op grond van de alfa's zijn de clusters 8 en 9 gecombineerd. De met * gemarkeerde clusters zijn afkomstig uit het PAGO en de met ** gemarkeerde zijn door de onderzoeker ontwikkeld, respectievelijk ontleend aan het FIS.

aan iedere deelnemer gevraagd om halfjaarlijks een vragenlijst in te vullen en terug te sturen, vanaf het begin van de onderzoeksperiode en vervolgens elke 6 maanden gedurende twee jaar.

De gehanteerde vragenlijst was gebaseerd op de klassieke "groene vragenlijst" die lang is gebruikt voor het periodiek arbeidsgezondheidskundig onderzoek (PAGO) en vrijwel op dezelfde wijze geclusterd [10, 11] (tabel 31). Die groene vragenlijst bestond zelf uit twee onderdelen: het onderdeel gezondheidsbeleven (in deze vragenlijst clusters 3, 4, 5 en 6) en het onderdeel werkbeleven (in deze vragenlijst clusters 8 t/m 13). Hieraan waren voor de vragenlijst van dit onderzoek 8 clusters toegevoegd: twee ter controle van elders verkregen gegevens (clusters 1 en 2), één voor het beoordelen van de medische consumptie (cluster

7), en vier clusters om de uit het FIS ontleende BMEKL-vragen in te plaatsen[11] (14 tot en met 17). Het achtste nieuwe cluster-18-bevatte de omzetting van de specifieke militaire beoordelingspunten naar vragen. Alle vragen vanaf cluster 3 waren dichotoom. Bij de vragen uit de clusters 3 tot en met 7 werd specifiek naar de situatie van het laatste half jaar gevraagd.

Vanaf het begin van de onderzoeksperiode ontving de kandidaat deze vragenlijst halfjaarlijks, in totaal vijf maal. Positieve reacties telden voor 1 punt, negatieve voor 0 punten.

7.2.3 Effectvariabelen

Inzetbaarheid. Per kandidaat is over de onderzoeksperiode het aantal (kalender-)dagen inzetbaarheid bepaald door aftrek van het aantal dagen ziekteverzuim of andere vormen van arbeidsverzuim, zoals beperkte operationele inzetbaarheid door blessures, vastgesteld door een arts. Hierbij bleek dat het klassieke ziekteverzuim 95% van het totaal uitmaakte. Op deze wijze kon de zuivere inzetbaarheid (het aantal kalenderdagen dat een militair operationeel had kunnen worden ingezet) worden bepaald.

Medische consumptie. De medische eerstelijns consumptie is berekend door alle eerste lijns consulten te rekenen naar het gemiddeld over de onderzoeksperiode geldende CTG-tarief (€ 17,92). De overige gemaakte en gedeclareerde ziektekosten zijn bij de ziektekostenverzekeraar opgevraagd.

7.2.4 Statistische methoden

Voor de beoordeling van standaardafwijkingen en betrouwbaarheidsintervallen is voor continue variabelen gebruik gemaakt van de Mann Whitney U test, voor niet-continue variabelen de Pearson Chikwadraattest.

Voorts zijn lineaire regressie analyses (voor de selectie van clusters die inzetbaarheid en/of medische consumptie beïnvloedden) en multiple lineaire regressie analyses (met die geselecteerde clusters) verricht. Voor het analyseren van de relaties tussen de effectvariabelen en de overige variabelen zijn lineaire regressie analyses en multiple lineaire regressie analyses gepleegd. Omdat er sprake was van twee effectvariabelen (inzetbaarheid en medische consumptie) zijn ook twee sets analyses gepleegd. Omdat dit onderzoek zich beperkt tot uitgezonden militairen en omdat deze uitzendingen plaatsvonden tussen het moment van de tweede en de vierde vragenlijst, zijn alleen deze tweede tot en met de vierde vragenlijst in dit onderzoek gebruikt.

Voorafgaande aan de analyse van de vragenlijsten is Crohnbachs' alfa van elk cluster bepaald. Uit tabel 31 blijkt dat de meeste clusters voldeden aan de eis dat deze alfa tenminste 0,50 moet zijn. Alleen de clusters 12 "Duidelijkheid" (drie

vragen) en 16 “Zintuigen/veiligheid” (vier vragen) voldeden stelselmatig niet aan deze eis en zijn derhalve buiten beschouwing gelaten. Uit vier clusters (3, 4, 7 en 17) is stelselmatig een vraag verwijderd om tot een $\alpha \geq 0,50$ te komen; om diezelfde reden zijn de clusters 8 en 9 gecombineerd.

Om de clusters, geassocieerd met de effectvariabelen, te selecteren zijn lineaire regressie analyses gepleegd waarbij de relatie tussen elk cluster en de effectvariabele één voor één is nagegaan. Hierna werd multiple lineaire regressie analyse gepleegd met alle clusters, waaruit in voorgaand onderzoek (de lineaire regressie analyse) op individuele basis een statistisch significante associatie was gebleken met de effectvariabele. Een p-waarde van minder dan 0,05 werd als statistisch significant beschouwd. Alle analyses zijn met behulp van SPSS 10.0 verricht.

7.3 Resultaten

7.3.1 Uitkomsten van het voorgaande onderzoeken op hoofdlijnen

In het eerste onderzoek [8] werd een ABOHZIS-studiegroep samengesteld van 166 personen en een BMEKL-studiegroep van 186 personen (tabel 30). Van de vroege drop-outs kon op grond van verschillende informatiebronnen (uitslagen van de exitinterviews, het achterwege laten van ingediende claims) worden aangenomen dat hun vertrek op niet-medische gronden berustte.

In het begin van de onderzoeksperiode bleken persoonskenmerken noch de keuringsgegevens statistisch significant te verschillen tussen beide studiegroepen. Tijdens de onderzoeksperiode bleken statistisch significante verschillen te zijn opgetreden tussen beide studiegroepen qua indeling naar gevechts- of gevechtsondersteunende functie, het aantal dagen inzetbaarheid en de medische consumptie.

Teneinde meer inzicht te krijgen in de oorzaak van de gevonden verschillen werd de associatie tussen deze gegevens en informatie inzake de deelnemers aan het onderzoek (zoals schoolopleiding, uitzending) nader onderzocht [12]. Daarnaast ontvingen de participanten halfjaarlijks een vragenlijst, die grotendeels was gebaseerd op een (“groene vragen-”) lijst die voor het periodiek bedrijfsgezondheidkundig onderzoek in ons land werd gehanteerd. Uit dit onderzoek bleek dat de volgende factoren van invloed waren op inzetbaarheid en medische consumptie: opleiding, blessures, operationele inzet (uitzending) en het keuringssysteem zelf.

7.3.2 Analyse persoonskenmerken en uitkomstvariabelen

Bij multiple lineaire regressie analyse met dagen inzetbaarheid als afhankelijke variabele tegenover het al dan niet uitgezonden zijn als onafhankelijke variabele,

Tabel 32. Multiple lineaire regressie analyse naar keuringssysteem en de eventuele uitzending, hetzij naar Bosnië-Herzegovina (SFOR), dan wel naar Kosovo (KFOR).

Onafhankelijke variabelen	Afhankelijke variabele: dagen inzetbaarheid			R^2
	Bèta coëfficiënt	Standaard Fout	P-waarde	
Keuringssysteem (BMEKL vs ABOHZIS)	118	16	<0,01	
Type uitzending versus Overige typen uitzending				
Geen uitzending	-510	46	<0,01	
SFOR	280	48	0,01	
KFOR	250	51	0,04	
				0,16

R^2 staat voor de verklaarde variantie.

blijkt dat uitgezonden militairen statistisch significant meer dagen inzetbaar zijn geweest dan diegenen die niet zijn uitgezonden (tabel 32). Uit verdere analyses bleek dat op het gebied van de geregistreerde persoonskenmerken, keuringsgegevens en personeelsgegevens verder geen statistisch significante verschillen tussen wel en niet uitgezonden militairen bestonden.

Vervolgens is onderzocht of tussen de fracties uitgezonden militairen uit beide studiegroepen statistisch significante verschillen bestonden inzake de persoonskenmerken, keuringsgegevens, persoonsgegevens en uitkomstvariabelen. Dit bleek niet het geval met betrekking tot leeftijd ($p = 0,51$), geslacht ($p = 0,53$), demografische herkomst (stad of platteland, $p = 0,43$), ouderlijke gezinssamenstelling (compleet of niet; $p = 0,55$), rookgewoonten (altijd sigaretten; $p = 0,17$), schoolopleiding ($p = 0,53$), burgerlijke staat ($p = 0,57$), medische bijzonderheden uit de keuringsgegevens ($p = 0,37$), de fysieke selectietest ($p = 0,28$), definitieve plaatsing na de militaire opleiding ($p = 0,27$) en eventuele overplaatsingen op sociaal-medische indicaties ($p = 0,93$). In tabel 33 worden de enige drie variabelen weergegeven waarin tussen beide fracties uitgezonden militairen wel een statistisch significant verschil bestond. Uitgezonden militairen uit de BMEKL-fractie bleken statistisch significant meer dagen inzetbaar te zijn geweest. Voorts werden in de uitgezonden fractie van de BMEKL-studiegroep statistisch significant meer gevechtsfuncties uitgeoefend dan in de uitgezonden fractie van de ABOHZIS-studiegroep. In de derde plaats zaten in de uitgezonden fractie van de BMKEL-studiegroep statistisch significant meer

Tabel 33. Uitkomstvariabelen gedurende de onderzoeksperiode.

Variabele		ABOHZIS		BMEKL		p
		N = 77	%	N = 84	%	
Indeling op functie	Gevechtsfunctie	27	35,0	46	54,7	0,01
	Gevechtsondersteunend	50	65,0	38	45,3	
Militaire Opleiding	Regulier	72	93,5	69	82,1	
a) Manschappen	LMB	1	1,3	11	13,1	
	Anders	2	2,6	2	2,4	
b) Kaderleden	KMS	2	2,6	2	2,4	<0,05
Dagen inzet, max 731		682 ± 65		693 ± 75		<0,05
Medische consumptie	€ / 731 dgn	767 ± 1514		494 ± 746		0,23

De p-waarden (p) van dagen inzet en medische consumptie zijn berekend met behulp van de Mann-Whitney U test, voor de overige p-waarden is de chi-kwadraattest gebruikt. De reguliere opleiding voor manschappen vindt op drie centra met een verschillende locatie plaats, waarbij een identiek programma wordt gehanteerd. LMB staat voor een zwaardere en langere opleiding.

manschappen die een zwaardere (luchtmobiele) militaire opleiding hadden gevolgd dan in de uitgezonden fractie van de ABOHZIS-studiegroep.

Bij onderscheid van de uitgezonden militairen naar gevechtsfunctie dan wel gevechtsondersteunende functie bleek daartussen geen statistisch significant verschil in het aantal dagen inzetbaarheid te bestaan ($p = 0,35$).

Om de onderlinge effecten juister te kunnen beoordelen en om het mechanisme te vinden op grond waarvan de uitgezonden militairen uit de BMEKL-studiegroep beter presteerden, zijn vervolgens multiple lineaire regressie analyses gepleegd.

Bij de eerste is het aantal dagen inzetbaarheid als afhankelijke variabele gehanteerd tegenover het keuringssysteem, gevechts- of gevechtsondersteunende functie en de soort militaire opleiding als onafhankelijke variabelen. Hierbij bleek alleen de onafhankelijke variabele (soort) militaire opleiding statistisch significant te verschillen ($p < 0,05$). Uit een tweede regressie analyse met het keuringssysteem en de verschillende opleidingsplaatsen als afhankelijke variabelen (tabel 34) bleek een associatie tussen zowel het keuringssysteem als de zwaarste opleiding voor manschappen en het aantal dagen inzetbaarheid.

Tabel 34. Multiple lineaire regressie naar keuringssysteem en militaire opleiding, respectievelijk opleidingsplaats met achtereenvolgens dagen inzetbaarheid en medische consumptie als afhankelijke variabelen.

Onafhankelijke variabelen	Afhankelijke variabele: dagen inzetbaarheid			R^2
	Bèta Coëfficiënt	Standaard fout	P-waarde	
Keuringssysteem (BMEKL vs ABOHZIS)	-54	27	<0,05	
Type opleiding versus Overige typen opleiding				
Opleiding officier	-90	97	0,23	
Opleiding onderofficier	-360	131	0,07	
Opleiding manschappen Noord	-24	39	0,55	
Opleiding manschappen Zuid	-37	28	0,50	
Opleiding manschappen Centraal	-44	42	0,31	
Opleiding manschappen LMB	-199	64	<0,01	
				0,13

Opleiding manschappen LMB (Luchtmobiel) is langduriger en zwaarder, de opleidingen op de schoolbataljons Noord, Zuid en Centraal. Deze laatste drie zijn qua programma identiek, maar verschillen qua locatie. Kaderleden worden opgeleid op aparte opleidingsinstituten. R^2 staat voor de verklaarde variantie.

7.3.3 Enquêtes

Voor nadere analyse telden alleen de geheel ingevulde exemplaren van de vragenlijst mee. Tussen de uitgezonden militairen uit beide studiegroepen bestond geen verschil in respons inzake de tweede tot en met de vierde vragenlijst. De totale respons van de uitgezonden daalde tussen de tweede en de vierde vragenlijst van 69% naar 49%.

De antwoorden op de vragen werden geclusterd. Vervolgens is behulp van lineaire regressie analyse de associatie onderzocht tussen de separate clusterscores en de uitkomstvariabele dagen inzetbaarheid. Bij deze analyse vertoonden de clusters Bewegingsapparaat (4) en Militaire punten (18) telkens een statistisch significante associatie met de uitkomstvariabele. Voor het cluster Algemene Gezondheid (3) werd deze associatie tweemaal vastgesteld (3^e en 4^e vragenlijst); ten aanzien van de clusters Bewegingsapparaat (7; 2^e vragenlijst) en FIS-bewegingsapparaat (14; 4^e vragenlijst) was dit elk eenmaal het geval.

Vervolgens is multiple lineaire regressie analyse uitgevoerd met de hierboven vermelde clusters tezamen en het keuringssysteem als onafhankelijke variabelen,

Tabel 35. De resultaten van de multiple lineaire regressie analyse met dagen inzetbaarheid als afhankelijke variabele en de clusters die op basis van lineaire regressie analyse associatie vertoonden met inzetbaarheid, tezamen met het keuringssysteem als onafhankelijke variabelen.

Enquête	Onafhankelijke variabelen	Afhankelijke variabele: dagen inzetbaarheid			
		Bèta-coëfficiënt	Standaardfout	P-waarde	R ²
		(dagen)	(dagen)		
2	Keuringssysteem	-37	24	0,08	0,18
	Cluster 18: militaire punten2	-46	10	0,01	
4	Keuringssysteem	-29	26	0,13	0,36
	Cluster 3: algemene gezondheid4	-73	19	<0,01	
	Cluster 4: bewegingsapparaat4	-14	9	0,03	
	Cluster 18: militaire punten4	-21	8	0,09	

R² staat voor de verklaarde variantie.

met dagen inzetbaarheid als afhankelijke variabele. Uit deze analyse bleek (tabel 35) dat de clusters Algemene Gezondheid (3), Bewegingsapparaat (4) en Militaire beoordelingspunten (18) elk eenmaal een statistisch significante associatie vertoont met de inzetbaarheid. Bij deze analyse bleef de p-waarde van het keuringssysteem > 0,05, terwijl de verklaarde variantie in de tijd toenam.

7.4 Beschouwing

Uitgezonden militairen blijken statistisch significant meer dagen inzetbaar te zijn dan niet uitgezonden militairen, zonder dat dit tot een verschil in medische consumptie tussen beide groepen leidde.

Dit verschil in inzetbaarheid kan op meerdere gronden berusten. Zo zal iemand die zich heeft ziek gemeld niet mee op uitzending gaan. Er kan sprake zijn van selectie van de meest fitte en/of geschikte kandidaten. Gedurende een uitzending is de sociale controle groter dan op de eigen kazerne, behoudens tijdens eventueel verlof kan niemand zich gedurende een uitzending ziek thuis melden. Andersom kan het niet worden uitgezonden tot teleurstelling, demotivatatie en leegloop leiden gedurende de resterende periode van de arbeidsovereenkomst. Deze laatste verklaring wordt ondersteund door het feit dat er geen statistisch significant verschil bestaat tussen uitgezonden en niet uitgezonden militairen

ten aanzien van de medische consumptie. Niet uitgezonden militairen vertonen een hoger verzuim zonder bijkomende meerkosten.

Dergelijke redenen kunnen het gevonden verschil in inzetbaarheid tussen de uitgezonden militairen uit beide studiegroepen niet verklaren. De uitgezonden fractie van de BMEKL-studiegroep is statistisch significant meer dagen inzetbaar geweest dan de uitgezonden fractie van de ABOHZIS-studiegroep. Dit sluit aan bij wat in theorie in de inleiding werd verondersteld.

De fracties verschillen onderling verder qua opleidingsplaats en wat betreft de verhouding tussen het aantal gevechtsfuncties en gevechtsondersteunende functies. Bij nadere analyse blijkt er een associatie te bestaan tussen het gesignaleerde verschil in inzetbaarheid en de zwaarste opleiding voor manschappen, maar dat verklaart het gevonden verschil in inzetbaarheid niet.

Het ontbreken van een verschil in inzetbaarheid tussen gevechts- respectievelijk gevechtsondersteunende functies komt overeen met het verwachtingspatroon: het verschil tussen deze twee soorten functies komt pas tot uiting bij missies waarbij de vrede moet worden afgedwongen of onder pure oorlogsomstandigheden. Daarvan was tijdens deze studie geen sprake.

Ondanks het ontbreken van dergelijke zware omstandigheden gedurende de uitzendingen in deze studie blijkt dat uitgezonden militairen die gekeurd zijn volgens het BMEKL-systeem beter presteren ten aanzien van gezondheidsklachten en het voldoen aan specifieke militaire eisen dan de uitgezonden militairen die gekeurd zijn volgens het ABOHZIS-systeem. Aangezien het BMEKL-keuringssysteem, in tegenstelling tot het ABOHZIS-keuringssysteem, bestaat uit het beoordelen of iemand aan die specifieke militaire functie-eisen voldoet, sluit deze bevinding aan bij de verwachting.

Kunnen deze bevindingen op een artefact berusten?

Door randomisatie is selectie-bias voorkomen. De fracties uitgezonden militairen zijn representatief voor wat gemiddeld wordt uitgezonden van alle militairen BBT: ze worden over de vier verschillende uit te zenden eenheden verdeeld. Van deze eenheden worden gedurende de actieve diensttijd van de studiegroepen achtereenvolgens twee voor uitzending aangewezen. Weliswaar verschillen beide fracties ten aanzien van de militaire opleidingsplaats en in hun verhouding tussen gevechts- en gevechtsondersteunende functies, maar deze verschillen zijn niet van doorslaggevende invloed op de inzetbaarheid. Bij de multiple lineaire regressie analyses blijkt immers dat de p-waarde van het keuringssysteem niet tot boven de 0,05 stijgt. Een deel van de conclusies berust op de analyse van de aan de PBGO-vragenlijst toegevoegde clusters vragen. Die vragen zijn nimmer gevalideerd; ze zijn door de onderzoeker zelf ontwikkeld. Alle vragen zijn eenvoudig en helder geformuleerd, alle betreffen navraag naar

een enkele handeling ("acht u uzelf in staat tot het 8 uur per dag dragen van een helm?"). Bovendien is bij de analyse niet gebruik gemaakt van de enkele vragen, maar van de clusters waartoe ze behoorden. Van de hiervoor gebruikte clusters is in elk geval aangetoond dat de alfa boven de vereiste 0,50 ligt over alle drie voor dit onderzoek gebruikte vragenlijsten.

Samengevat komt uit dit onderzoek naar voren dat uitgezonden militairen, gekeurd volgens het functiegericht BMEKL-systeem, een grotere inzetbaarheid vertoonden dan de uitgezonden militairen, gekeurd volgens het diagnosegericht ABOHZIS-systeem. Het BEMKL-keuringssysteem voorspelt beter ten aanzien van het functioneren als militair in het algemeen en specifiek op basis van uitzending, gezondheid en bewegingsapparaat dan het ABOHZIS-systeem.

Geraadpleegde literatuur

1. Raad J de, Peters JHM, Groot RP de et al: Modernisering van het keuringssysteem bij de Koninklijke Landmacht; keuringsprotocollen en functioneel beoordelen. TBV 2000; 8: 110-115.
2. Dijk FJH van, Dormolen M van, Kompier MAJ et al: Herwaardering model belasting-belastbaarheid. TSG, 1990; 68: 3-10.
3. Kessel JGFM van: De consequenties van onjuist keuren. Ned Mil Geneesk T; 1984 (37): 176-228.
4. Raad J. de, Peters JHM., Nijhuis FJM: De kwaliteit van een keuringssysteem. TBV 2000; 8:105-109.
5. Raad J de, Slootman EJH, Groot RP de: Onderzoek naar de consistentie van het beoordelen van aanstellingskeuringen voor de Koninklijke Landmacht. TBV 2002; 10: 266-272.
6. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Bohn, Stafleu en van Loghum. Houten, 1995.
7. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
8. Raad, J de, Redekop WK: A comparison between two systems for pre-employment medical assessment in the RNLA by a randomised-controlled study. Military Medicine 2004; 169(6):437-443.
9. Broersen JPJ, Weel ANH, Dijk FJH van: Periodiek Bedrijfsgeneeskundig Onderzoek; middel en maatstaf. Amsterdam, 1989.
10. Putten DJ van, Marcelissen FGH, Winter CJ de: De vragenlijst voor PBGO 1-4. Achtergronden voor het gebruik ervan. Het wijzigen van antwoorden door bedrijfsartsen. Methoden voor gebruik van referentiegegevens. TSG 1988; 66: 115-120, 143-147.
11. Putten DJ van: De vragenlijst voor PBGO, voorstellen tot wijziging van de vragenlijst. Leiden, NIPG/TNO, 1989.
12. Raad, J de, Redekop WK: Analysis of health factors as predictors for the functioning of military personnel. In februari 2004 door Military Medicine voor publicatie aanvaard.

8 | Slotbeschouwing

8.1 Inleiding

In de vorige hoofdstukken is onderzoek beschreven naar het effect van aanstellingskeuringen bij militairen. Allereerst zijn niet-systematische en systematische waarnemingen weergegeven. Vanuit die waarnemingen is een theorie inzake de kwaliteit van keuringen geïnduceerd. Mede op grond van die theorie is een nieuw keuringssysteem voor de KL ontwikkeld. Hierna is een hypothese gededuceerd, waarbij een verschil in kwaliteit tussen het nieuwe (BMEKL) en het oude (ABOHZIS) systeem werd verondersteld. Op basis van deze hypothese is een aantal onderzoeksvragen geformuleerd. Met behulp van deze vragen zijn vervolgens twee kwaliteitskenmerken van het meetinstrument keuring onderzocht.

Bouter en van Dongen [1] stellen dat de resultaten van één enkel onderzoek een theorie zelden kunnen maken of breken. Meestal is veel onderzoek nodig voordat een theorie voldoende in de empirie is verankerd. In feite, zo stellen zij [1], is een *empirische spiraal* nodig om de werkelijkheid steeds nauwkeuriger te beschrijven. De toegevoegde waarde van het enkele onderzoek aan deze spiraal is afhankelijk van de hoeveelheid bewijslast, de interne en externe validiteit en de omvang van de onderzoekspopulatie [1].

In de hoofdstukken 2 tot en met 7 zijn al meer opmerkingen vermeld inzake de toegevoegde waarde van de betreffende onderzoeken. Thans kan echter de som der delen worden overzien. Daarbij valt niet uit te sluiten dat bij deze onderzoeken, apart of in onderlinge samenhang, nog kanttekeningen moeten worden geplaatst. Een nabeschuiving is noodzakelijk.

Na deze nabeschuiving kan, met inachtneming van eventuele kanttekeningen, de vraagstelling worden beantwoord. Met behulp van die antwoorden kan vervolgens de hypothese worden getoetst. Daarna kan worden nagegaan of en in hoeverre het onderzoek de theorie inzake keuringen heeft beïnvloed. Als laatste kan worden nagegaan wat de betekenis van het onderzoek is geweest voor de empirische cyclus.

8.2 Nabeschuiving van het verrichte onderzoek

Hoofdstuk 2 valt als systematische waarneming buiten de eigenlijke vraagstelling van het onderzoek. Het onderzoek vormde echter de aanleiding tot het operationaliseren van de BMEKL en was op deze wijze van grote invloed op het verdere onderzoek. Bij dit onderzoek valt de volgende kanttekening te maken. De mate van overeenkomst – gemeten in kappa – ligt in dit onderzoek lager dan in het onderzoek zoals beschreven in hoofdstuk 4. Kan dit worden verklaard? Bij het ABOHZIS-systeem, zoals dat bij het PGO werd gehanteerd,

was sprake van veel meer keuzemogelijkheden in de uitspraken (A1 of A2, H2 of H3 etc.) dan bij de vaststelling van de aanstellingskeuring. Deze laatste vaststelling heeft een dichotoom karakter: men is geschikt of ongeschikt, ongeacht of van het nieuwe dan wel oude keuringssysteem sprake was. In hoofdstuk 3 is reeds opgemerkt dat de kappa onder meer afhankelijk is van het aantal beoordelingscategorieën (strata): hoe meer strata des te kleiner de kappa [1].

In hoofdstuk 3 is de reproduceerbaarheid van 30 pensioenkeuringen onderzocht volgens het BMEKL-systeem. De consistentie van het eindoordeel – gemeten in kappa – is goed. Ook de meeste, in dit onderzoek meetbare beoordelingspunten komen er goed vanaf. De gebruikte hoeveelheid van 30 keuringsdossiers is echter een minimum voor onderzoeksdoeleinden. Het gevolg is dat maar een beperkt scala van medische bijzonderheden voorkomt, zodat niet elk beoordelingspunt kan worden getoetst. Voor een aantal van die beoordelingspunten maakt de gekozen hoeveelheid dossiers overigens niet uit. De constatering dat iemand beperkt is op sommige beoordelingspunten leidt al bij de aanstellingskeuring tot afwijzing (zoals een onvoldoende kleuronderscheidend vermogen). Dergelijke beperkingen komen bij de pensioenkeuring feitelijk nooit voor. Al met al moet worden erkend dat dit onderzoek maar een beperkt inzicht geeft in de mate van overeenstemming bij de pensioenkeuring. Deze constatering doet echter geen afbreuk aan de uitspraken over die beoordelingspunten die wel te beoordelen vielen. Als laatste moet worden vastgesteld dat in dit hoofdstuk wel inzicht is verkregen in de inter-waarnemers variatie, maar niet in de intra-waarnemers variatie.

In hoofdstuk 4 wordt de betrouwbaarheid van beide keuringssystemen bij de aanstellingskeuring onderzocht. Hierbij worden vier verschillende overeenstemmingsmaten gebruikt. Als tekortkoming uit het tweede onderzoek geldt dat de oorspronkelijke beoordelaar de samengestelde is van de drie verschillende oorspronkelijke uitslaggevend artsen. Daardoor valt intra-waarnemers variatie niet te bepalen. Deze constatering doet echter geen afbreuk aan de bevindingen op het gebied van de inter-waarnemers variatie.

In de hoofdstukken 5, 6 en 7 wordt telkens gebruik gemaakt van gegevens afkomstig van dezelfde gerandomiseerde gecontroleerde studie. In die studie zijn twee onderzoeksgroepen geschikt verklaarde militairen (gekeurd volgens één van beide keuringssystemen) twee jaar gevolgd. Bij deze hoofdstukken kunnen de volgende kanttekeningen worden vermeld.

Uit de laatste hoofdstukken blijkt dat het nieuwe keuringssysteem een betere voorspeller is ten aanzien van inzetbaarheid, maar dat het mechanisme waarop

dit is gebaseerd niet kan worden achterhaald. Weliswaar kan gewezen worden op bepaalde verschillen tussen beide keuringssystemen, die mogelijk tot een verklaring zouden kunnen leiden, maar wetenschappelijk gezien kan het mechanisme niet worden aangetoond.

Ondanks het voorkomen van selectiebias door randomisatie zou men kritiek kunnen hebben op de wijze van randomisatie (om de week van systeem wisselen). De indeling van gegadigden voor een aanstellingskeuring vindt echter volstrekt willekeurig plaats, waardoor een eventueel verschil tussen kandidaten die op de even en oneven weken worden gekeurd op toeval moet berusten. Het hanteren van een werkweek als termijn lijkt daarbij optimaal om dit onderzoek te verrichten. Het dagelijks aselekt aanbieden van op verschillende wijze te keuren kandidaten aan keurend artsen zou immers risico's opleveren: het gevaar verstrikt te raken in verschillende diagnostische classificatieschema's [1], expectation bias [1], motivatieproblemen bij de uitvoering en het bemoeilijken van de controle door derden op het gebruik van het juiste keuringssysteem. Als aanvullend bewijs dat deze vorm van randomisatie aanvaardbaar is wordt gewezen op de goede onderlinge verhouding qua onder meer leeftijd en geslacht in beide studiegroepen.

Het verzamelen van de gegevens voor het berekenen van de medische consumptie zou ook tot selectiebias kunnen leiden, met name ten aanzien van de "externe" consumptie (zorgafname buiten het eigen zorgsysteem waarvoor declaraties bij de ziektekostenverzekeraar worden ingediend). Vrijwel in al deze gevallen wordt de betreffende declaratie niet door de verzekerde, maar door de zorgverlener ingediend bij de verzekeraar. Bovendien zou bij het niet declareren de militair de rekening zelf moeten betalen. Op basis van voornoemde argumenten wordt aangenomen dat de kans dat hierin selectiebias is opgetreden gering is.

Opgemerkt zou kunnen worden dat in de totstandkoming van de studiegroepen – die immers uit geschikt verklaarden bestonden – sprake kan zijn geweest van inter-doktersvariatie. De individuele interpretatie van de begrippen "geschikt" en "ongeschikt" zou tussen de tien betrokken keurend artsen tot (te grote) verschillen in de uiteindelijk geschikt verklaarde rekruten hebben kunnen leiden. Hierdoor zouden de uitkomsten kunnen zijn beïnvloed. Als onderdeel van de introductie van het nieuwe keuringssysteem zijn onder meer al deze artsen tijdens een workshop voorbereid. Zoals ook in hoofdstuk 1 is beschreven, vormt het gebruik maken van keuringsprotocollen onderdeel van de methodiek, waardoor in theorie de eenduidigheid wordt bevorderd. Blijkens paragraaf 5.3 zit ook geen verschil in de onderlinge verhouding tussen geschikt- en ongeschiktverklaring ($p = 0,27$) tussen de artsen. Bovendien wordt, zoals in

paragraaf 2 van hoofdstuk 4 al werd uiteengezet, de uitslag niet door de keurend artsen vastgesteld maar door drie vaste uitslaggevend seniorbedrijfsartsen, van wie het oordeel volgens datzelfde hoofdstuk 4 in goede tot voortreffelijke mate reproduceerbaar bleek te zijn. Op grond van de hierboven weergegeven redactie wordt aangenomen dat een eventuele inter-doktersvariatie de resultaten niet in ernstige mate heeft beïnvloed.

De onderzoeksgroep had aanzienlijk uitgebreid kunnen worden indien gekozen was om ook die militairen te volgen, die al in de militaire opleiding (AMO) de dienst alweer verlieten. Het meenemen van de reeds tijdens de AMO vertrekkende militairen had de onderzoeksgroep weliswaar optisch aanmerkelijk groter gemaakt, maar door de uiterst korte periode waarin zij effectief bij de KL aanwezig waren (niet werkzaam, maar slechts in opleiding), zou dit weinig waarde aan het onderzoek hebben toegevoegd.

Ook bij de clustering kunnen kanttekeningen worden gemaakt. Bij het stellen van een enkele vraag naar een bepaald onderwerp kan de foutieve interpretatie van de respondent een verstorende rol spelen. Dit effect kan worden bestreden door meer vragen over hetzelfde onderwerp te stellen: clustering – reden waarom clustering voor het onderzoek uit de hoofdstukken 6 en 7 heeft plaatsgevonden. Voorwaarde voor clustering is dat er voldoende correlatie tussen die vragen bestaat. Clustering heeft volgens Swanborn [2] en Meerling [3] zowel voor- als nadelen. De voordelen zijn onder meer dat de clusterscore betrouwbaarder is dan de score op een enkele vraag en dat de informatie kort en overzichtelijk kan worden gepresenteerd. Als nadelen worden onder meer aangehaald dat men zo signalen kan missen en dat de informatie op een hoger abstractieniveau komt te liggen. Het abstractieniveau vormde voor deze clusters geen nadelig effect. Voor zover ze tot de oorspronkelijke PAGO-vragenlijst behoorden, waren ze al gedefinieerd en werden ze door verschillende onderzoekers als zodanig gebruikt [4, 5]. Dat mogelijk signalen, afkomstig van een enkele vraag, kunnen zijn gemist, kan nadelig zijn, maar doet aan de conclusies inzake de clusters geen afbreuk.

Diezelfde onderzoekers [4, 5] wijzen bovendien op het feit dat de reactie op de PAGO-clusters altijd assymmetrisch is: de meerderheid van de werknemers rapporteert weinig tot geen klachten, sommigen daarentegen rapporteren juist veel klachten op vragen uit een cluster. Dit zou volgens deze auteurs de lijst vooral geschikt maken voor het opsporen van knelpunten. Aangezien het de intentie was om verschillen te zoeken behoeft ook dit niet als verstorend voor het onderzoek te worden beschouwd.

Van de hierboven vermelde kanttekeningen is een aantal van belang voor het beantwoorden van de vraagstelling in de volgende paragraaf. Deze worden hieronder samengevat.

In algemene zin valt het onderzoek uiteen in twee delen: in het eerste deel wordt onderzoek verricht naar de mate van overeenstemming (werkt het systeem kwalitatief?), in het tweede deel wordt onderzoek verricht naar de prognostische waarde van beide keuringssystemen.

Wat het eerste gedeelte betreft, is het onderzoek beperkt gebleven tot het beoordelen van de inter-waarnemers variatie. Dit kan weliswaar als een tekortkoming worden opgevat, maar doet geen afbreuk aan de conclusies inzake de gevonden inter-waarnemers variatie.

Het tweede gedeelte is beperkt gebleven tot de in dienst tredende en niet al tijdens de opleiding weer vertrekkende keurlingen. Daarmee valt geen uitspraak te doen inzake de validiteit. Er is weinig reden om te twijfelen aan de uitkomsten van het rekenmodel zoals De Kort [6] dat in 1997 publiceerde. Daarom kan niet worden uitgesloten dat ook in deze studie meer fout-positieven tegelijk met een enkele risicodragers worden geweerd. Daaraan kan ter verdediging van het keuringssysteem worden toegevoegd dat de nauwkeurige definiëring van de meeste Basismedische Eisen (Tabel 1, hoofdstuk 1) in elk geval theoretisch de variatie in de besluitvorming zal beperken.

Tevens blijkt uit de hoofdstukken 5, 6 en 7 dat het nieuwe BMEKL-keuringssysteem een betere voorspeller is ten aanzien van inzetbaarheid dan het oude ABOZIS-systeem. Het lijkt aannemelijk dat het nieuwe keuringssysteem, gebaseerd op de feitelijke taakstelling (met inbegrip van operationele inzet), onder operationele omstandigheden ook beter voorspelt, omdat het nu juist voor die omstandigheden is ontworpen.

Maar het mechanisme waarop dit is gebaseerd kan wetenschappelijk niet worden aangetoond.

Met inachtneming van deze bemerkingen kan thans de vraagstelling worden beantwoord.

8.3 Beantwoording van de vraagstelling

In paragraaf 1.5.7. zijn voor het onderzoek naar de kwaliteit van het BMEKL-keuringssysteem ten opzichte van het ABOHZIS-keuringssysteem vijf vragen gesteld. Nadat in hoofdstuk 2 een systematische waarneming is beschreven, waaruit de lage reproduceerbaarheid van het oude ABOHZIS-keuringssysteem naar voren kwam, hebben die vijf vragen als onderzoeksvraag deel uitgemaakt van de hoofdstukken 3 tot en met 7. In de vorige paragraaf is aandacht besteed aan kanttekeningen die bij deze hoofd-

stukken kunnen worden geplaatst. Hierdoor wordt het mogelijk deze vijf vragen thans te beantwoorden.

De eerste onderzoeksvraag was: "Voldoet de reproduceerbaarheid van het nieuwe keuringssysteem bij de pensioenkeuring aan de normering die gebruikelijk is voor screeningstesten?"

Volgens Bouter en van Dongen [1] suggereren gepubliceerde onderzoeksgegevens dat een kappa tussen de 0,40 en 0,75 voor diagnostische testen in de klinische diagnostiek gebruikelijk is. In hoofdstuk 3 wordt de beoordeling volgens het BMEKL-systeem van 30 aselekt gekozen dossiers van recente pensioenkeuringen door twee seniorbedrijfsartsen onderzocht. Dit levert ten aanzien van het definitieve geschiktheidoordeel een Kappa op van 0,74 (0.48 – 0.98) en een PD van 13,3%.

Op grond hiervan kan de gestelde vraag bevestigend worden beantwoord voor wat betreft de inter-waarnemers variatie.

De tweede onderzoeksvraag was: "Hoe verhoudt zich de mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring?"

In hoofdstuk 4 wordt de reproduceerbaarheid van het geschiktheidoordeel bij de aanstellingskeuring van beide keuringssystemen vergeleken. Er wordt gebruik gemaakt van 227 keuringsdossiers van de aanstellingskeuring; bij het onderzoek zijn op drie dit terrein ervaren seniorbedrijfsartsen betrokken. Elke vergelijking levert een zeer hoge mate van onderlinge overeenstemming op, ongeacht welk keuringssysteem wordt beoordeeld.

Op grond van dit onderzoek kan worden geconstateerd dat ten aanzien van de medische aanstellingskeuring beide keuringssystemen, voor wat betreft de inter-waarnemers variatie, een even hoge mate van overeenstemming vertonen.

De derde onderzoeksvraag was: "Is er een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van de geschikt verklaarde militairen BBT op het gebied van de dagen inzetbaarheid en/of de genoten medische consumptie?"

Uit een gerandomiseerde, gecontroleerde studie, beschreven in hoofdstuk 5 blijkt dat militairen, gekeurd volgens het nieuwe keuringssysteem (BMEKL) gedurende een periode van twee jaar na de opleiding statistisch significant meer dagen (648 versus 612) inzetbaar waren en minder medisch consumeerden (€ 396 versus € 746) dan militairen die terzelfder tijd volgens het oude keuringssysteem (ABOHZIS) waren gekeurd. Daarmee is de inzetbaarheid van 84% tot 89% toegenomen. Niet alleen de prognostische waarde, maar ook de construct validiteit is toegenomen.

Het antwoord op de derde vraag is derhalve dat er een verschil in prognostische waarde tussen beide keuringssystemen bestaat, waarbij keuring volgens het nieuwe BMEKL-systeem zowel meer dagen inzetbaarheid als lagere medische consumptie oplevert dan keuring volgens het ABOHZIS-systeem.

De vierde onderzoeksvraag was: "Indien een verschil in prognostische waarde tussen beide keuringssystemen wordt gevonden, kan het mechanisme daarvan worden verklaard door analyse van de determinanten persoonskenmerken, de eigen beleving van gezondheid, inzetbaarheid en operationele inzet?"

In hoofdstuk 6 wordt een analyse verricht van gegevens, afkomstig van dezelfde studie als in hoofdstuk 5, met inbegrip van de resultaten van een halfjaarlijks afgenomen vragenlijst. Uit die analyse blijkt dat een aantal factoren wordt gevonden die van invloed zijn op de inzetbaarheid en de medisch consumptie. De betere prestatie van het BMEKL-keuringssysteem lijkt deels te kunnen worden toegeschreven aan het vermogen om op taakspecifieke eisen te beoordelen. Maar het primaire mechanisme waarop dit is gebaseerd, blijft verborgen. Het antwoord op de vraag hierboven is derhalve dat er weliswaar vermoedens zijn dat het BMEKL-keuringssysteem beter voorspelt omdat het ook beter is gericht op het functioneren als militair, maar dat dit niet kan worden aangetoond.

De vijfde en laatste onderzoeksvraag was: "Bestaat er ook een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van uitsluitend de uitgezonden militairen en zo ja, kan dat worden verklaard door analyse van de determinanten persoonskenmerken, het beleven van de eigen gezondheid, de eigen beleving van gezondheid en inzetbaarheid?"

In hoofdstuk 7 zijn de uitgezonden militairen uit beide studiegroepen aan nadere analyse onderworpen.

De uitgezonden militairen, gekeurd volgens het BMEKL-systeem, zijn meer dagen inzetbaar geweest, vertonen minder gezondheidsklachten en achten zichzelf meer in staat tot het voldoen aan specifieke militaire eisen dan de uitgezonden militairen die gekeurd zijn volgens het ABOHZIS-systeem.

Het antwoord op de gestelde vraag is dat inderdaad een verschil in prognostische waarde wordt gevonden in het voordeel van het BMEKL-keuringssysteem; dit verschil lijkt te zijn gebaseerd op het beleven van de eigen gezondheid en de mate waarin men zich bekwaam acht tot het verrichten van de vereiste taken.

8.4 Toetsing van de hypothese

In hoofdstuk 1 is gesteld dat het nieuwe BMEKL-keuringssysteem, anders dan het oude ABOHZIS-keuringssysteem, in hoofdzaak aan het gestelde in zowel de WMK als de ARA voldoet. Daaruit werd geconcludeerd dat dit hypothetisch zou moeten betekenen dat de kwaliteit van het BMEKL-keuringssysteem beter zou moeten zijn dan die van het ABOHZIS-systeem. Op grond van deze hypothese zijn de verschillende onderzoeksvragen geformuleerd. Op deze onderzoeksvragen is in de vorige paragraaf antwoord gegeven.

Uit de beantwoording van de vragen kan worden opgemaakt dat de reproduceerbaarheid van het nieuwe keuringssysteem in alle onderzochte gevallen volstaat. Echter, anders dan in tabel 6 in hoofdstuk 1 het geval leek te zijn, blijkt uit hoofdstuk 4 de reproduceerbaarheid van het ABOHZIS-systeem even goed te volstaan wanneer het om de aanstellingskeuring gaat. De vermoedelijke verklaring daarvoor is hierboven in paragraaf 8.2. uiteengezet. Het woord "situationeel" lijkt derhalve op de reproduceerbaarheid van het ABOHZIS-systeem van toepassing. Omdat keuringen zoals vermeld in hoofdstuk 2 echter niet meer voorkomen en dit onderzoek de aanstellingskeuring betrof, kan niet worden gesteld dat de kwaliteit van het nieuwe systeem inzake de reproduceerbaarheid beter is dan het oude.

Anders is dit wat de prognostische waarde bij geschikt verklaarde militairen betreft. Uit drie achtereenvolgende onderzoeken blijkt dat deze prognostische waarde voor het BMEKL-systeem hoger ligt, met name onder operationele en andere zware omstandigheden. Weliswaar kan het mechanisme waarop dit is gebaseerd niet worden achterhaald, maar de constatering uit de laatste drie onderzoeken valt daarmee niet weg. Er zijn zelfs aanwijzingen, op grond van de vragenlijst, dat het verschil te herleiden zou zijn naar de eigen beleving van gezondheid en taakbekwaamheid. Het feit dat de eisen van het nieuwe BMEKL-keuringssysteem waren afgeleid van de feitelijke taakstelling – anders dan het ABOHZIS-systeem – deed vooraf vermoeden dat het BMEKL-systeem beter zou functioneren. Nu dit laatste, althans wat de prognostische waarde bij geschikt verklaarde militairen betreft, ook is aangetoond, wordt dit opgevat als een bevestiging van het feit dat de hypothese juist is geweest.

In tabel 36 wordt, analoog aan tabel 6 uit hoofdstuk 1, een overzicht gegeven van de stand van zaken voor beide keuringssystemen ten opzichte van de kwaliteitscriteria, zoals de onderzoeker deze zelf na afloop van het onderzoek beschouwt.

Tabel 36. De stand van zaken met betrekking tot de kwaliteitskenmerken van de BMKEL-keuring en ABOHZIS-keuring na afloop van het verrichte onderzoek.

Kwaliteitskenmerk Keuring	ABOHZIS	BMEKL	Toelichting
Relevantie	Vermoedelijk gelijkwaardig		Gemeten kenmerk moet betekenis hebben gelet op het met de test beoogde doel
Toepasbaarheid	Vermoedelijk gelijkwaardig		Toegesneden op situatie, proportioneel, subsidiar, voor alle partijen acceptabel (implementatie)
Efficiëntie	Vermoedelijk gelijkwaardig		Minimalisering van inspanningen en uitgaven
Validiteit	Gedeeltelijk bekend; lager	Gedeeltelijk bekend; hoger	Operationalisatie, uitvoering, te meten concept: sensitiviteit/specificiteit, diagnostische waarde, gouden standaard
Reproduceerbaarheid	Situationeel	Altijd hoog	Inter-waarnemers variatie (Intra-waarnemers variatie niet beoordeeld)
	Hoog bij aanstellingskeuringen		

Met "gedeeltelijk bekend" achter "validiteit" wordt de uitkomst van het onderzoek naar de voorspellende waarde uit de hoofdstukken 5, 6 en 7 aangeduid, waarin BMEKL-gekeurde militairen een hogere inzetbaarheid vertoonden dan de ABOHZIS-gekeurden. Met "situationeel" wordt het verschil in reproduceerbaarheid van het ABOHZIS-keuringssysteem uit de hoofdstukken 2 (Periodiek Geneeskundig Onderzoek) en 4 (aanstellingskeuringen) aangeduid.

8.5 Aanpassing van de theorie?

De theorie is dat de kwaliteit van aanstellingskeuringen als meetinstrument te wensen over laat. Aan de hand van de hypothese en de operationalisatie van de onderzoeksvragen uit hoofdstuk 1 heeft dat geleid tot de stelling dat de kwaliteit van aanstellingskeuringen voor militairen dusdanig gering is, dat aanstellingskeuringen geen invloed op de inzetbaarheid van het personeel uitoefenen. In deze studie is aangetoond dat aanstellingskeuringen wel degelijk invloed op de inzetbaarheid van het personeel uitoefenen. Voorts is hieruit gebleken dat het keuringssysteem, dat is afgeleid van de feitelijke taakstelling van de (althans militaire) werknemer en dat op hoofdlijnen voldoet aan de eisen van WMK en ARA, een hogere kwaliteit oplevert dan het keuringssysteem dat niet aan de vermelde twee voorwaarden (afgeleid van de taak, voldoen aan WMK en ARA) voldoet. Anders dan bij voorgaande onderzoeken zoals die van De Kort [7] is in deze studie voor een lange onderzoeksperiode gekozen.

Moet de theorie nu worden aangepast? Voor het beantwoorden van deze vraag worden hieronder de volgende argumenten gehanteerd.

Weliswaar is aangetoond dat aanstellingskeuringen invloed op de inzetbaarheid van het personeel uitoefenen, maar het mechanisme waarop die invloed berust kan niet worden aangetoond. Hooguit kan het vermoeden worden geuit dat dit mechanisme niet op fysieke gronden, maar op mentale gronden berust, zoals in het onderzoek van Carbone [8].

Het percentage afgekeurden ligt in deze studie voor beide keuringssystemen (13-16% exclusief uitval door fysieke selectie) hoger dan de percentages die in het onderzoek van De Kort [7] worden vermeld. Het argument van De Kort [7] dat een keuring weinig toevoegt aan het wervingsproces op grond van een laag percentage afgekeurden, is in deze studie derhalve niet van toepassing.

De medische bijzonderheden blijken bij geschikt verklaarden niet maatgevend voor de inzetbaarheid of de medische consumptie, evenmin als welk ander persoonskenmerk. Net als in het proefschrift van De Kort [7] ontbreekt in deze studie de relatie tussen medische bijzonderheden en de oorzaak van de uitval op het werk. Tijdens de aanstellingskeuring wordt veel onderzocht dat geen voorspellende waarde voor geschikt verklaarden blijkt te bezitten. Daarmee is overigens nog niet aangetoond dat medische bijzonderheden ten opzichte van ongeschikt verklaarden geen waarde hebben. Maatgevend bij geschikt verklaarden lijkt immers hoezeer men zichzelf tot het verrichten van taakspecifieke militaire eisen in staat acht. Dit kan inhouden dat instrumenten om zelfoverschatting te bewaken op hun plaats zouden kunnen zijn, maar dat is in deze studie geen onderwerp van onderzoek geweest.

Voor deze studie zijn andere effectvariabelen gebruikt dan die welke De Kort [7] hanteerde. Voor militairen van de KL is bijvoorbeeld met name inzetbaarheid een zeer aanvaardbare effectvariabele, maar daarmee is nog niet aangetoond dat deze effectvariabele ook voor de overige in ons land voorkomende beroepen toegepast kan worden.

De bijzondere taakstelling van de militair speelt een rol in dit onderzoek. De militair wordt niet voor een bepaalde, welomschreven functie gekeurd, maar voor het kunnen uitoefenen van een breed scala aan opdrachten, waaraan de nodige uiteenlopende risico's zijn verbonden – multifunctioneel. Juist die brede taakstelling rechtvaardigt een keuring met een breed spectrum, samen met administratieve, psychologische en fysieke selectie. De meeste beroepen in ons land zijn aanmerkelijk minder pluriform van aard, zeker wat betreft de eraan verbonden risico's. Als bij dergelijke beroepen al een aanstellingskeuring mag plaatsvinden, dan kan daarbij vermoedelijk de door De Kort [7] gekozen effectvariabele zeer adequaat worden toegepast, terwijl de in deze studie gekozen effectvariabelen wellicht minder invloed zullen hebben. Al met al lijken de resultaten van dit onderzoek niet eenvoudig naar andere werksituaties te extrapoleren.

De weergegeven argumenten leiden tot de volgende afweging. Dat in deze studie de keuring invloed uitoefent kan op meerdere factoren berusten, zoals de gebruikte effectvariabelen, het feit dat is voldaan aan wettelijke normen, waardoor deze keuring een afspiegeling vormt van de werkelijke te verwachten belasting en de lange onderzoeksperiode. Daarnaast kan niet worden uitgesloten dat het resultaat uit deze studie (te) sterk samenhangt met de specifieke werkzaamheden van de onderzoekspopulatie.

Ook kan niet concreet worden vastgesteld op grond waarvan de keuring invloed uitoefent. Die invloed ligt in elk geval niet aan een associatie tussen keuringsbevindingen en redenen voor uitval. Sterker, men kan zich afvragen of de mogelijke reden (zoals motivatie) wel in een medische keuring thuishoort. Vanuit deze redenatie kunnen vier mogelijke verklaringen worden onderscheiden, of wellicht een combinatie daarvan:

- De voor deze studie gekozen variabelen vormen een concrete aanvulling voor onderzoek naar het effect van keuringen, waarmee bij toekomstig onderzoek rekening moet worden gehouden.
- Keuringen die voldoen aan de eisen, gesteld in WMK en ARA, hebben effect op werknemers, dit in tegenstelling tot keuringen die daaraan niet voldoen.
- Het resultaat van deze studie kan uitsluitend worden bereikt bij een onderzoek naar een populatie die zwaar en gevarieerd werk verricht.
- Het effect op werknemers uit deze studie is gebaseerd op een niet-medisch mechanisme.

Het bovenstaande kan worden opgevat als vier hypothesen ter mogelijke verklaring van de gevonden variaties in uitslagen. Het valt echter buiten het bestek van dit onderzoek om deze hypothesen te toetsen. Uitsluitend wanneer één van de eerste twee verklaringen juist zou zijn (of beide), zou de theorie worden beïnvloed. Over de voor- en nadelen van de in de eerste verklaring vermelde effectvariabelen is zowel hierboven als in hoofdstuk 1 voldoende opgemerkt. In het voordeel van de tweede verklaring kan worden gesteld dat de invoering van de WMK ongetwijfeld ook tot doel had dat de toegestane keuringen effect zouden hebben. De derde verklaring beïnvloedt de theorie niet of nauwelijks; de laatste verklaring ondersteunt de heersende theorie. Alle verklaringen zijn het gevolg van één studie. Zelfs indien één van de eerste twee verklaringen juist zou zijn, dan nog blijft van toepassing wat in de inleiding is gesteld: de resultaten van één enkel onderzoek kunnen een theorie zelden maken of breken. Meestal is veel onderzoek nodig voordat een theorie voldoende in de empirie is verankerd [1].

Er is geen aanleiding om op grond van deze bevindingen de theorie te wijzigen. Wel kunnen deze resultaten worden beschouwd als een aansporing tot verder onderzoek. In de zin van de empirische spiraal zal nog meer onderzoek moeten worden verricht vooraleer revisie van de theorie moet worden overwogen.

8.6 Over keuringen: de betekenis van dit onderzoek voor de empirische spiraal

In tabel 36 is de stand van zaken van het BMEKL-systeem ten opzichte van de vijf kwaliteitskenmerken aan het einde van deze studie weergegeven, naar de visie van de onderzoeker. In de vorige paragraaf zijn verklaringen aangedragen voor de bevindingen uit dit onderzoek. Welke onderwerpen voor verder onderzoek levert dat tezamen op? Zonder de pretentie uitputtend alle mogelijkheden weer te geven, volgt hieronder een opsomming op hoofdlijnen op grond van de bevindingen uit deze studie.

Wat de relevantie betreft kan de werkelijke betekenis van het gemeten kenmerk pas worden vastgesteld indien de validiteit of de gehele voorspellende waarde van het systeem wordt vastgesteld. Dan pas wordt immers bekend hoe relevant de gemeten kenmerken feitelijk zijn.

Inzake de toepasbaarheid en efficiëntie is al opgemerkt dat bij de keuring veel kenmerken worden onderzocht die in deze studie geen prognostische waarde bleken te bezitten. Ook bleek het aantal verschillende categorieën medische bijzonderheden die een rol spelen bij de aanstellingskeuring beperkt te zijn. Deze constatering nodigen uit tot het verrichten van een studie naar de mogelijkheden om het aantal gehanteerde keuringsinstrumenten te reduceren. Deze bemerking sluit aan op de kanttekeningen die in paragraaf 1.4.2. bij toetsing van het BMEKL-systeem aan de ARA werden gemaakt en die hier tevens als aanbevelingen voor nader onderzoek kunnen worden weergegeven: er moet nog nader onderzoek worden verricht naar de validiteit, proportionaliteit en/of relevantie van bepaalde ter keuring gehanteerde meetinstrumenten, evenals naar het uiteindelijke effect van het Convenant Fysieke Belasting tussen de Ministeries van SZW en Defensie).

Inzake de reproduceerbaarheid kan nader onderzoek naar de intra-waarnemers variatie worden aanbevolen.

Of het te meten concept valide is, is slechts gedeeltelijk aangetoond. Weliswaar is er nog steeds geen gouden standaard, maar naast de in paragraaf 1.4.1. aangenomen gemaakte "face validity" [1] en "expert validity" [1] is nu ook de prognostische waarde voor geschikt verklaarden aangetoond. Hierdoor kan dit onderzoek de basis vormen voor een studie naar de prognostische waarde, waarin tevens de afgekeurden worden gevolgd.

Door in een dergelijke studie zowel de effectvariabele van De Kort [7] te gebruiken als de effectvariabelen uit deze studie, zouden meer associaties tegelijk kunnen worden onderzocht. In de eerste plaats kan het belang van alle effectvariabelen worden herbeoordeeld, waarbij ook de onderlinge verhouding qua relevantie kan worden vastgesteld (welke associatie is het sterkst?). Zowel herbevestiging van de associaties uit dit onderzoek, als het mogelijk vinden van een associatie tussen keuringsbevindingen en uitval wegens arbeidsongeschiktheid kunnen voorts worden beschouwd als bewijs dat functiegericht keuren conform de in WMK en ARA weergegeven eisen wel invloed op de theorie heeft. Gegeven de resultaten uit de in dit proefschrift weergegeven studie zou hiervoor mogelijk kunnen worden volstaan met het hanteren van een vragenlijst. In deze lijst moeten – naast de in deze studie gehanteerde – ook de onderwerpen verzuim, arbeidsongeschiktheid en motivatie worden opgenomen. Een knelpunt in zulk onderzoek kan worden gevormd door de (te lage) respons bij niet actief dienenden, zoals ook in deze studie het geval was.

Om het mechanisme achter het effect te achterhalen is ook nader onderzoek geïndiceerd. Bij dat onderzoek kan bij het determineren van te meten variabelen rekening worden gehouden met de in Figuur 17 uit hoofdstuk 6 weergegeven mogelijke factoren (zelfinzicht, psychische belastbaarheid, opleiding, motivatie, blessures, gezondheid en operationele inzet). Van de keurend arts kan worden gevraagd om bij bepaalde factoren, zoals motivatie of algemene gezondheid, een gradatie aan te geven bij de kandidaten, bijvoorbeeld door middel van een schaalverdeling.

Ook zou kunnen worden beoordeeld of de in deze studie gevonden onderzoeksresultaten afhankelijk zijn van de aard van de werkzaamheden. Daarvoor zou een soortgelijke studie als in de vorige alinea's is weergegeven moeten worden herhaald bij een niet-militaire werknemerspopulatie, waarvoor eveneens een aanstellingskeuring is geïndiceerd.

Kortom, over keuringen kan ook in de toekomst nog voldoende onderzoek worden gepubliceerd.

Geraadpleegde literatuur

1. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Bohn, Stafleu en van Loghum. Houten, 1995.
2. Swanborn PG: Methoden van sociaal-wetenschappelijk onderzoek. Meppel, Boom, 1987.
3. Meerling: Methoden en technieken van psychologisch onderzoek. Meppel, Boom, 1989.

4. Putten DJ van, Marcelissen FGH, Winter CJ de: De vragenlijst voor PBGO 1-4. Achtergronden voor het gebruik ervan. Het wijzigen van antwoorden door bedrijfsartsen. Methoden voor gebruik van referentiegegevens. TSG 1988; 66: 115-120, 143-147.
5. Vrijhof BJ, Broersen JPJ: De structuur van de vragenlijst voor periodiek bedrijfsgezondheidkundig onderzoek (PBGGO). Tijdschrift voor toegepaste arbowetenschap 1993; 6: 14-17.
6. Kort WLAM de, Dijk FJH van: Preventive effectiveness of pre-employment medical assessments. Occupational Environmental Medicine 1997; 54: 1-6.
7. Kort WLAM de : Personnel selection through pre-employment medicals. Amsterdam, 1993.
8. Carbone EG, Cigrang JA, Todd SL et al: Predicting outcome of military basic training for individuals referred for psychological evaluation. J Pers Assess 1999; 72(2):256-265.

Bijlage:
Verklaring van begrippen en
afkortingen

ABOHZIS Classificatie	Selectiesysteem, waarbij ziekten en gebreken worden verdeeld naar: Algemene fysieke toestand, Bovenste extremiteiten, Onderste extremiteiten, Horen, Zien, Intelligentie en Stabiliteit.
AMAR	Algemeen Militair Ambtenaren Reglement, waarin de rechtspositie is omschreven.
AMO	Algemene Militaire Opleiding, die elke militair initieel ontvangt.
ARA	Algemene Richtlijn Aanstellingskeuringen.
Arbo-dienst KL	Interne Arbo-dienst ten behoeve van zowel de Koninklijke Landmacht als de Koninklijke Marechaussee.
BBT	Beroepsmilitair Bepaalde Tijd, militair ambtenaar met een contractuele verbintenis voor een beperkte tijd, vaak 2 ½ jaar.
BLS	De Bevelhebber der Landstrijdkrachten, de werkgever (tot september 2005).
BMEKL	Basismedische Eisen KL, keuringssysteem gebaseerd op het model belasting- belastbaarheid.
BOT	Beroepsmilitair Onbepaalde Tijd, militair ambtenaar met een contractuele verbintenis tot aan het functioneel leeftijdsontslag (thans 55 jaar of ouder).
Brigade	Resultaat Verantwoordelijke Eenheid (RVE), bestaande uit meerdere bataljons. Grootste eenheid die op grond van de Defensienota buiten de eigen landsgrenzen operationeel kan worden ingezet voor peace-keeping tot en met peace-enforcing missies.
CTG	College Tarieven in de Gezondheidszorg.
Dienstongeschikt	Medisch ongeschiktheidoordeel op grond van verschillende artikelen uit het AMAR, zoals door middel van de pensioenkeuring of de aanstellingskeuring IKS.

DP&O	Directie Personeel & Organisatie KL; centrale personeelsdienst, onderdeel van de "Haagse Staven". Sinds juni 2004 Personeelscommando geheten.
Expert-validity	Validiteit, gebaseerd op het oordeel van deskundigen (bijvoorbeeld audiometrie-criteria laten vaststellen door KNO-artsen).
Face-validity	Validiteit "op het oog" (bijvoorbeeld "timmermans-oog").
FIS	Functie Informatie Systeem. Met behulp van het belastbaarheidpatroon van 28 beoordelingspunten is in de vorige eeuw voor de Nederlandse marktsector de belasting van bestaande arbeid in kaart is gebracht, waarna voor elke functie of beroep het specifieke patroon in een database is opgenomen. Op deze wijze konden bij de WAO-keuring geconstateerde beperkingen worden gekoppeld aan functies die, met inachtneming van die beperkingen, nog konden worden verricht door de betrokken werknemer. Bij veel van deze beoordelingspunten kon gedifferentieerd worden tussen meerdere prestatieniveaus.
IBDKL	Individuele Begeleidingsdienst KL, IBDKL. Speciale eenheid voor de reïntegratie van langdurig verminderd inzetbare militairen.
IGO	Incidenteel Geneeskundig Onderzoek. Voor KL en KMar uitgevoerd door de VGDKL op grond van AMAR Artikel 99, meestal voorafgaande aan eventuele overplaatsing naar de IBDKL voor militairen.
IKS	Instituut Keuring en Selectie voor de Krijgsmacht. Verricht voor de krijgsmachten de selectie en medische aanstellingskeuring.
Keurend arts	Bedrijfsarts (in opleiding) in dienst van het IKS die de medische aanstellingskeuring effectief verricht.
KFOR	Peace-keeping missie in Kosovo 1999-2000.
KL	Koninklijke Landmacht.
KLu	Koninklijke Luchtmacht.

KM	Koninklijke Marine.
KMA	Koninklijke Militaire Academie, opleidingsinstituut voor officieren KL.
KMar	Koninklijke Marechaussee, militaire politie.
KMS	Koninklijke Militaire School, opleidingsinstituut voor onderofficieren KL.
MGO	Zie pensioenkeuring.
Milicien	Frans, "soldaat van een lichting".
MKR	Militair Keuringsreglement. Dit reglement vormt de grondslag voor de uitvoering van zowel de aanstellingskeuring als de pensioenkeuring of MGO voor militairen, evenals dit voorheen voor het PGO het geval was.
Natres	Korps Nationale Reserve. Parttime vrijwillig dienenden, die louter binnen de landsgrenzen worden ingezet in geval van een grootschalig conflict.
PAGO	Periodiek Arbeidsgezondheidkundig Onderzoek.
Paraatstelling	Parate plaatsing. Het op de eigen functie bij een bepaald onderdeel plaatsen van de militair na het doorlopen van de opleiding(-en).
PBGO	Periodiek Bedrijfsgezondheidkundig Onderzoek.
Peace-enforcing missie	Militaire operatie om de vrede in een gebied af te dwingen met militaire middelen.
Peace keeping missie	Militaire operatie met als doel om de vrede in een gebied te handhaven.
Pensioenkeuring	(commissoriaal) Geneeskundig onderzoek volgens het Besluit Procedure Geneeskundig onderzoek blijvende dienstongeschiktheid en pensioenkeuring militairen, Staatsblad 1997, 67.
PGO	Periodiek Geneeskundig Onderzoek. Periodieke (om de vijf jaar) beoordeling van de dienstgeschiktheid op medische gronden voor beroepsmilitairen. Het PGO werd bij de KL uitgevoerd van 1960 t/m/ 1995.

Proportionaliteit-criterium	Een onderzoek is alleen gerechtvaardigd indien de prevalentie van de te onderzoeken aandoening bij onderzoekspopulatie dit rechtvaardigt .
PULHEEMS	Indeling na aanstellingskeuring bij het Canadese Leger in 1945 naar: Physical capacity, Upper limbs, Locomotion, Hearing, Eyesight, Emotional and Mental.
RBB	Rijksbedrijfsgezondheids- en Bedrijfsveiligheidsdienst. Naam van de toenmalige Arbo-dienst voor overheids-personeel uit de tijd van het onderzoek van De Kort (1992).
Relevantiecriterium	Het te verrichten onderzoek moet relevant zijn voor het verwezenlijken van het doel van de keuring.
Remedial Platoon	Speciale opleidingseenheid om leerlingen die op medische of sociale gronden de AMO tijdelijk niet geheel kunnen volgen, op te vangen.
SFOR	Tot op heden doorlopende peace keeping missie in Bosnië-Herzegowina.
Subsidiariteit-criterium	Een onderzoeksmethode moet zo weinig mogelijk belastend of ingrijpend voor de keurling zijn.
SZW	Ministerie van Sociale Zaken en Werkgelegenheid.
Uitslaggevend arts	Seniorbedrijfsarts in dienst van één van de Arbo-diensten van de krijgsmachtdelen, die niet de aanstellingskeuring zelf uitvoert, maar nadien de uitslag vaststelt.
UNFICYP	Peace keeping missie in Cyprus t/m 2001.
UNMEE	Peace keeping operatie in Eritrea 2001.
VGDKL	Eenheid waar onder meer voor militairen van zowel de KL als de KMar de pensioenkeuring en het IGO worden uitgevoerd.
VO ₂ max	Maximale zuurstofopname, gemeten in ml / kg lichaamsgewicht / minuut.
WMK	Wet op de Medische Keuringen.

Samenvatting

Vanuit de behoefte om een goede voorspelling te doen van de inzetbaarheid van militairen, is een valideringsonderzoek [1] uitgevoerd. In dit onderzoek is de kwaliteit van een nieuw ontwikkeld keuringssysteem vergeleken met de kwaliteit van het voorheen gebruikte keuringssysteem.

In hoofdstuk 1 worden achtereenvolgens weergegeven: de niet-systematische en systematische waarnemingen, de ontwikkeling van een nieuw instrument voor keuringen, de theorie en de daaruit gededuceerde onderzoekshypothese, de onderzoeksvragen en het de hoofdlijnen van het materiaal en de methoden die gebruikt zijn om deze onderzoeksvragen te beantwoorden.

De niet-systematische waarnemingen bestonden uit de ontdekking dat er kwaliteitscriteria voor het instrument "keuringen" bleken te bestaan [2] en de confrontatie met de vraag of keuringen bij de Koninklijke Landmacht ooit aan die criteria waren getoetst. Op grond van deze niet-systematische waarnemingen was de interesse in het onderwerp, te weten de kwaliteit van (aanstellings-) keuringen zoals uitgevoerd bij de KL, gewekt. Deze interesse leidde tot systematische waarnemingen. Eerst werd nagegaan waarop de werkwijze voor het uitvoeren van keuringen bij de KL was gebaseerd en of die werkwijze tekortkomingen vertoonde.

Nadat Holland in 1810 bij het Franse keizerrijk was ingelijfd en Napoleon Bonaparte ook hier in 1811 de conscriptie had laten invoeren, werd alhier het eerste (Franse) reglement op het geneeskundig onderzoek van miliciens van kracht. Dit reglement vormt de grondslag van het Militair Keuringsreglement (MKR), dat sedertdien bij de Koninklijke Landmacht als basis dient voor zowel de medische aanstellingskeuring als de pensioenkeuring voor militair personeel. Bij dat MKR behoorde van oudsher een "Lijst van Ziekten en Gebreken". De keurend arts besliste op grond van professionele autonomie of een rekrut in voldoende ernstige mate aan één der in deze lijst vermelde ziekten of gebreken leed, wat inhield dat deze moest worden afgekeurd. Na de Tweede Wereld-

oorlog werd door de Nederlandse regering de ABOHZIS-classificatie¹ van het Canadese leger overgenomen. Dit classificatiesysteem ter nadere verfijning van het selectieproces werd vervolgens aan het MKR en de "Lijst van ziekten en gebreken" gekoppeld. Hierdoor werd de professionele autonomie van de keurend arts sterk ingeperkt. Met dit keuringssysteem is tot november 1998 gewerkt.

De tweede systematische waarneming bestond uit literatuuronderzoek naar de kwaliteit van keuringen. Uit dit onderzoek kwam naar voren dat de kwaliteit van het instrument keuringen over het algemeen te wensen over liet. Vermoedelijk gold dit ook voor het in de KL gebruikte systeem. Dit keuringssysteem was diagnosegericht en niet gebaseerd op het model "belasting-belastbaarheid" [3]. Er bestonden geen functie-eisen. Het classificatiesysteem was niet aangepast, alhoewel de taakstelling van de krijgsmacht was uitgebreid met operationele uitzendingen naar het buitenland. Op deze wijze voldeed het systeem niet aan de eisen die de Wet op de Medische Keuringen (WMK, 1998) stelt.

De derde systematische waarneming bestond uit het uitvoeren van onderzoek naar de reproduceerbaarheid van keuringen binnen de KL, waaraan in hoofdstuk 2 aandacht wordt besteed.

Vanuit deze waarnemingen werd de theorie geïnduceerd dat de kwaliteit van bestaande aanstellingskeuringen niet voldeed. De Bevelhebber der Landstrijdkrachten (BLS) werd in 1997 geadviseerd om een ander keuringssysteem te laten ontwikkelen. Als basis hiervoor dienden de recent geformuleerde werkgevers-eisen, minimaal te stellen aan alle militairen. Deze eisen werden geformuleerd op grond van de concrete militaire taakstelling, met inbegrip van operationele uitzendingen naar het buitenland. Dit leidde tot een lijst met 43 beoordelingspunten. Elk beoordelingspunt werd voorzien van een grenswaarde: de Basismedische Eis KL (BMEKL). Tijdens de keuring wordt onderzocht of de kandidaat beperkt is ten opzichte van deze eisen. Om de keurings situatie te structureren zijn keuringsprotocollen ontwikkeld.

In tegenstelling tot het oude ABOHZIS-systeem voldeed het nieuwe BMEKL-systeem aan de eisen, gesteld door de WMK. Hypothetisch zou dit moeten betekenen dat de kwaliteit van het BMEKL-keuringssysteem beter zou moeten zijn dan die van het ABOHZIS-systeem.

1 De letters ABOHZIS stonden voor **A**lgemene fysieke toestand, **B**ovenste extremiteiten, **O**nderste extremiteiten, **H**oren, **Z**ien, **I**ntelligentie en **S**tabiliteit, de terreinen waarop werd ingedeeld, met een gradatie van 1 (geen ziekte of gebrek) tot 5 (ongeschikt op grond van...).

Omdat voor de beoordeling van de kwaliteit van keuringssystemen geen gouden standaard voorhanden is, resteerde alleen de mogelijkheid om het ene keuringssysteem te vergelijken ten opzichte van het andere. Het BMEKL-keuringssysteem zou per 1 november 1998 het ABOHZIS-systeem gaan vervangen. Dat bood de gelegenheid om deze beide keuringssystemen kwalitatief met elkaar te vergelijken.

Met inachtneming van deze kennis en de hypothese werden hierna de volgende onderzoeksvragen geoperationaliseerd:

1. Voldoet de reproduceerbaarheid van het nieuwe keuringssysteem bij de pensioenkeuring aan de normering die gebruikelijk is voor screeningstesten?
2. Hoe verhoudt zich de mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring?
3. Is er een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van de geschikt verklaarde militairen BBT op het gebied van de dagen inzetbaarheid en/of de genoten medische consumptie?
4. Indien een verschil in prognostische waarde tussen beide keuringssystemen wordt gevonden, kan het mechanisme daarvan worden verklaard door analyse van de determinanten persoonskenmerken, de eigen beleving van gezondheid, inzetbaarheid en operationele inzet?
5. Bestaat er ook een verschil in prognostische waarde tussen beide keuringssystemen ten aanzien van uitsluitend de uitgezonden militairen en zo ja, kan dat worden verklaard door analyse van de determinanten persoonskenmerken, het beleven van de eigen gezondheid, de eigen beleving van gezondheid en inzetbaarheid?

Om deze vragen te beantwoorden is een gerandomiseerde gecontroleerde studie uitgevoerd. In deze studie zijn 352 soldaten gedurende twee jaar gevolgd, nadat ze waren goedgekeurd volgens één van beide keuringssystemen.

Als effectvariabelen werd gekozen voor zowel inzetbaarheid (het aantal dagen dat een militair geen arbeidsverzuim vertoont) als de medische consumptie gedurende de onderzoeksperiode. Alleen voor het beantwoorden van de eerste onderzoeksvraag diende een andere onderzoekspopulatie te worden gebruikt.

In hoofdstuk 2 wordt nader ingegaan op de in de inleiding vermelde derde systematische waarneming. In dit hoofdstuk wordt de mate van overeenstemming onderzocht in de eerste plaats tussen de ABOHZIS-classificaties, vervolgens tussen de medische beperkingen, en als laatste tussen die classificatie en de beperking. Voor dat doel werden 24 casus afkomstig van het periodiek geneeskundig onderzoek (PGO) geanonimiseerd. Vervolgens werd aan twee artsen verzocht om zowel een ABOHZIS-classificatie op te maken als in voorkomend

geval medische beperkingen op te leggen. Alhoewel de uitkomsten niet altijd als onvoldoende kunnen worden gekenschetst, corresponderen de resultaten toch met eerdere teleurstellende resultaten uit studies van anderen [2]. De conclusie is dat door de ABOHZIS-classificatie te handhaven niet langer aan de huidige behoeften wordt voldaan.

In hoofdstuk 3 wordt het onderzoek naar de reproduceerbaarheid van het nieuwe keuringssysteem tijdens de pensioenkeuring weergegeven. 30 geanonimiseerde keuringsdossiers met een evenredige verdeling tussen geschikt en ongeschikt verklaarden werden ter beoordeling voorgelegd aan twee seniorbedrijfsartsen. Voor het meten van de overeenstemming werd gebruik gemaakt van twee maten: de kappa [1] en PD% [2]. Het resultaat is dat er, voor wat betreft het geschiktheidoordeel, sprake is van een goede overeenstemming [1]. De conclusie is dat het nieuwe keuringssysteem bij de pensioenkeuring voldoet aan de normering die gebruikelijk is voor diagnostische testen.

In hoofdstuk 4 wordt het onderzoek naar de mate van overeenstemming tussen beide keuringssystemen bij de aanstellingskeuring weergegeven. Hiervoor werd aan dezelfde bedrijfsartsen die voor de oorspronkelijke keuringsuitslag verantwoordelijk waren, na ongeveer 1½ jaar gevraagd om 227 keuringsdossiers te herbeoordelen. De dossiers waren geanonimiseerd en afkomstig van de gerandomiseerde, gecontroleerde studie waarvan in de hoofdstukken 5, 6 en 7 sprake is. Voor het meten van de overeenkomst werd van 4 verschillende maten [1, 2] gebruik gemaakt. Op grond van de gevonden resultaten wordt geconcludeerd dat de inter-waarnemers overeenstemming bij de aanstellingskeuring voor beide keuringssystemen als voortreffelijk [1] kan worden beschouwd.

In hoofdstuk 5 worden beide keuringssystemen met elkaar vergeleken op hun vermogen om geschikte kandidaten op te leveren. Voor dit laatste worden de effectvariabelen dagen inzetbaarheid en medische consumptie gehanteerd. Er werd een gerandomiseerde, gecontroleerde studie verricht, waarbij 352 soldaten gedurende twee jaar zijn gevolgd, na geschikt te zijn verklaard en de militaire opleiding te hebben gevolgd.

De aanstellingskeuring blijkt de dominante factor te zijn bij het voorspellen van zowel het aantal dagen inzetbaarheid als de medische consumptie. Degenen, gekeurd volgens die door het BMEKL-systeem, blijken gemiddeld statistisch significant een hoger aantal dagen inzetbaar te zijn geweest (648 tegenover 612) en tegelijkertijd een lagere medische consumptie te vertonen (396 tegenover 746). In deze studie bleek het niet mogelijk om het mechanisme op grond waarvan het BMEKL-systeem betere uitkomsten laat zien, te ontdekken. Voor-

lopig wordt dit mechanisme verklaard op basis van het verschil tussen beide keuringssystemen.

Hoofdstuk 6 wordt analyse gepleegd, teneinde het mechanisme te ontdekken waarop de associatie tussen het soort keuringssysteem en de twee uitkomstmaten uit het vorige onderzoek berust. Gedurende het onderzoek zijn uiteenlopende data omtrent de deelnemers aan de studie verzameld (opleiding, uitzending etc.) Gedurende de onderzoeksperiode is tevens vijf maal een vragenlijst aan de onderzoekspersonen voorgelegd. Deze vragenlijst was inhoudelijk voor een deel gebaseerd op de vragenlijst die lang in Nederland is gebruikt bij het Periodiek Bedrijfsgezondheidkundig Onderzoek. De volgende factoren bleken een statistisch significante associatie met inzetbaarheid en medische consumptie te vertonen: schoolopleiding, blessures, operationeel ingezet worden en de mate waarin men zichzelf geschikt acht tot het uitvoeren van specifieke militaire vaardigheden. De betere prestatie van het BMEKL-systeem lijkt deels te kunnen worden toegeschreven aan het betere vermogen om op geschiktheid voor militaire taakspecifieke eisen te keuren. Desalniettemin blijft opnieuw het primaire mechanisme verborgen.

Omdat het feit of men al dan niet wordt uitgezonden invloed uitoefent op de inzetbaarheid, worden in hoofdstuk 7 dezelfde analyses gepleegd als in de voorgaande hoofdstukken 5 en 6, maar nu alleen gericht op militairen uit beide studiegroepen die operationeel zijn ingezet in het kader van een uitzending. De betreffende militairen werden gedurende twee jaar gevolgd. Als effectvariabelen werden opnieuw het aantal dagen inzetbaarheid en de medische consumptie gehanteerd. Daarnaast werd gebruik gemaakt van de halfjaarlijks toegezonden vragenlijsten. Uit dit onderzoek komt naar voren dat uitgezonden militairen, gekeurd volgens het functiegerichte BMEKL-systeem, een grotere inzetbaarheid vertoonden dan de uitgezonden militairen, gekeurd volgens het diagnosegerichte ABOHZIS-systeem. Het BMEKL-keuringssysteem voorspelt beter dan het ABOHZIS-systeem ten aanzien van het functioneren als militair in het algemeen, en specifiek op het gebied van uitzending, gezondheid en bewegingsapparaat.

In hoofdstuk 8 wordt de slotbeschouwing weergegeven. Nabeschouwing van de onderzoeken levert een aantal kanttekeningen op, sommige daarvan hebben invloed op het uiteindelijke antwoord op de onderzoeksvragen. Ten eerste is het onderzoek beperkt gebleven tot het beoordelen van de inter-waarnemers variatie. Ten tweede is het nieuwe BMEKL-keuringssysteem een betere voorspeller dan het ABOHZIS-systeem, maar het mechanisme waarop dit is gebaseerd kan

niet worden aangetoond. Dit impliceert dat delen van sommige vragen (reproduceerbaar voor zover het de inter-waarnemers variatie betreft; factoren die het opgetreden verschil beïnvloeden kunnen niet worden verklaard) niet kunnen worden beantwoord. Tegelijk wordt de bewijslast, bijeengebracht door achtereenvolgende onderdelen van deze studie, als voldoende beschouwd om te stellen dat de hypothese juist is geweest. Het feit dat de hypothese wordt aangenomen, leidt tot de vraag of de theorie inzake keuringen moet worden aangepast. Er kunnen verschillende verklaringen worden aangedragen om het – wellicht alleen ogenschijnlijke – verschil tussen hypothese en theorie aannemelijk te maken. Voor alle verklaringen geldt evenwel dat zij alleen op grond van deze studie worden weergegeven. Bouter en van Dongen [1] stellen dat de resultaten van één enkel onderzoek een theorie zelden kunnen maken of breken. In feite, zo stellen zij, is er een *empirische spiraal* nodig om de werkelijkheid steeds nauwkeuriger te beschrijven. Ten gevolge van deze constatering lijkt een aanpassing van de theorie prematuur. Om in het kader van de empirische spiraal te onderzoeken of de theorie in de toekomst wellicht toch moet worden aangepast, wordt besloten met enige aanbevelingen voor verder onderzoek.

Geraadpleegde literatuur

1. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Bohn, Stafleu en van Loghum. Houten, 1995.
2. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
3. Dijk FJH van, Dormolen M van, Kompier MAJ et al: Herwaardering model belasting-belastbaarheid. TSG, 1990; 68: 3-10.

Abstract

The necessity for a good prediction of the fitness for duty of soldiers from the Royal Netherlands Army (RNLA) has led to a validation study [1]. In this study the quality of a new developed medical pre-employment assessment is compared to the quality of the previous used system.

Chapter 1 contains: the non-systematic and the systematic observations, the development of a new instrument for medical assessment, a theory and, deduced from that theory, a hypothesis, the study questions and the main features of both material and methods, used to answer these study questions. The non-systematic observations were the discovery that criteria existed with which to measure the quality of the instrument of medical assessment [2], as well as the confrontation with the question as to whether the quality of the medical assessments as conducted in the RNLA had ever been measured against those criteria. As a result of these non-systematic observations, the interest in the quality of medical (pre-employment) assessments, as conducted within the RNLA, was evoked. That interest led to systematic observations being made. First, the method of conducting medical assessments within the RNLA was studied, as well as the question as to whether this method showed any shortcomings.

Following the incorporation of the Kingdom Holland into the French Empire in 1810, and the introduction by Napoleon Bonaparte of conscription in 1811, the first (French) regulation on medical assessment of conscripts came into force also here. This regulation is the foundation of the "Regulations for Military Medical Examinations", which from that day on have been used by the RNLA as a basis for pre-employment medical assessments as well as medical assessments in the event of a pension claim.

Ever since their introduction, the "Regulations for Military Medical Examinations" have been accompanied by a "List of Diseases and Infirmities". It was left to the professional autonomy of the physician performing the examination to decide whether a recruit was suffering from any disease or infirmity mentioned in this list severely enough to be declared unfit for military service.

After the Second World War the Dutch government adopted the PULHEEMS¹ classification method used by the Canadian army, known in Dutch as the ABOHZIS system. This classification system for further refinement of the personnel selection process was combined with the "Regulations for Military Medical Examinations" and the accompanying "List of Diseases and Infirmities". The professional autonomy of the physician performing the examination was thus reduced. This system remained in place until November 1998.

The second systematic observation was a literature study regarding the quality of medical assessments. The conclusion from that research was that the quality of the instrument "medical assessments" in general was poor. This was presumably also the case with the system used within the RNLA. This system was diagnosis-related and not based on the "workload-capability" model [3]. Job requirements did not exist. The classification system was not changed even though the tasks of the armed forces had recently been expanded to include missions abroad. Thus the system was not in keeping with the Medical Examinations Act of 1998.

As a third systematic observation a study was performed to investigate the reproducibility of medical assessments in the RNLA, as will be explained in further detail in Chapter 2.

From these observations the theory was induced that the quality of existing medical pre-employment assessments was insufficient.

The Commander-in-Chief of the Army was advised in 1997 to have a new system for medical assessment developed. As a basis for that further development the recently defined employer's requirements were used, which applied minimally to every soldier. These requirements were developed according to the actual military tasks, including deployment on missions abroad. Thus a list of 43 assessment criteria for physical examination was developed. For each criterion we defined a minimum, which we called the "RNLA Basic Medical Requirements" (BMEKL in Dutch). During the examination the physician checks whether the candidate has a restriction with regard to these requirements. In addition, we developed protocols to give structure to the medical assessments. Unlike the old ABOHZIS system, the new BMEKL system was in keeping with the Medical Examinations Act of 1998. Hypothetically, this should lead to a better quality of the BMEKL system in comparison with the old system. Since

1 The letters of the Dutch acronym ABOHZIS stand for: A = physical capacity, B = upper limbs, O = lower limbs, H = hearing, Z = eyesight, I = mental state, S = emotional state, which are the terms used for classification, graded from 1 (no disease or infirmity) to 5 (unfit due to...).

there is no "golden standard" for pre-employment medical assessment to compare it with, the only remaining possibility was to compare the quality of the two systems mentioned above. The BMEKL system was to replace the ABOHZIS system in November 1998, which provided the opportunity to compare the quality of both systems. Based on this knowledge and the hypothesis, the following study questions were developed:

1. Does the reproducibility of the Basic Medical Requirements meet the standards commonly used for diagnostic tests, when used for assessments related to pension claims?
2. How does the reproducibility of both systems compare for pre-employment medical assessments?
3. Is there a difference in the prognostic value between the two systems for pre-employment medical assessment when looking at the amount of time fit-for-duty and/or the health care costs incurred, once a candidate has been declared fit for active service?
4. If a difference in the prognostic value between the two systems for pre-employment medical assessment is found, can this be explained by analysing the following determinants: personal characteristics, the personal experience of health and fitness for duty as well as being deployed or not?
5. Is there also a difference in the prognostic value between the two systems for pre-employment medical assessment when looking only at those persons who have actually been deployed on a mission abroad, and if so, can that difference be explained by analysing the following determinants: personal characteristics, the personal experience of health and fitness for duty?

To answer these questions, a randomised-controlled study was performed. In this study 352 soldiers were monitored for two years after being declared fit by one of the pre-employment medical assessment systems in question. Both the fitness for duty of military personnel (the number of days a soldier is not absent from work) and the health care costs incurred during the research period were chosen as the effect variables. Only for the first question was it necessary to use another study population.

Chapter 2 discusses the third systematic observation, as mentioned in the introduction. The reproducibility was investigated, first of the ABOHZIS classification, then of the medical restrictions and finally the reproducibility of the combination of both classification and restrictions. For that purpose 24 cases as used at the time for the periodical medical assessment were made anonymous. Two occupational health physicians were asked to apply the ABOHZIS classification and if needed the medical restrictions to each case. Although the results

of this study cannot always be defined as insufficient, they correspond well to earlier disappointing results of other studies [2]. The conclusion is that the ABOHZIS classification no longer meets the demands of our time.

Chapter 3 studies the reproducibility of the Basic Medical Requirements with regard to the new system of medical assessments when used for pension claims. Thirty medical files were made anonymous. In 15 cases the candidates were declared to be fit-for-duty, in the other 15 cases the candidates were declared to be unfit. Then two occupational health physicians of the RNLA were asked to re-examine these cases on the basis of the medical files. To analyse the agreement two measures were used: the agreement according to kappa [1] and PD% [2]. In the case of the fit-for-duty declaration the reproducibility appears to be good [1]. Therefore the conclusion was that the reproducibility of the Basic Medical Requirements, used for medical assessment regarding a pension claim, meets the standards commonly used for diagnostic tests.

In chapter 4 the reproducibility of the results of the pre-employment medical assessment was examined. The three occupational health physicians who were responsible for the original declarations were asked to re-examine 227 cases, approximately 1½ years later. These cases were 227 anonymous medical files from the randomised-controlled study as described in chapters 5 and 6. To analyse the agreement four different measures [1, 2] were used.

According to the results of this study, the inter-observers variation can be described as excellent [1] for both pre-employment medical assessment systems.

In chapter 5 the two different examination systems were compared for their ability to identify suitable recruits. For that latter purpose we used the two operational measures of days fit-for-duty and health care costs. A randomised-controlled study was performed, in which 352 soldiers were monitored for two years after being declared fit by one of the pre-employment medical assessment systems in question and having passed their General Military Training. The pre-employment medical assessment system was found to be the dominant factor for predicting the number of days fit-for-duty as well as the health care costs incurred. Those declared fit by the new BMEKL system showed a statistically significant higher mean number of days fit-for-duty (648 compared to 612) and incurred significantly lower mean health care costs (€ 396 compared to € 746). In this study it was not possible to uncover the mechanism by which the BMEKL examination system led to an improvement in outcome. For the time being this mechanism has been explained on the basis of the differences between the two systems.

In chapter 6 an analysis was performed to uncover the mechanism on which the association between the two systems of assessment and the two measures of outcome depends. During the study, divergent data about the subjects of the study have been collected (training, on mission abroad etc.). Also, the soldiers were asked five times to complete a questionnaire, its content based in part on a periodic occupational health examination questionnaire commonly used in the Netherlands. The following factors showed a statistically significant association with fitness for duty and health care costs: education, injuries, actual operational deployment and the extent to which the soldiers considered themselves able to perform specific military skills.

The superior performance of the new BMEKL system seems partly due to the better ability of assessing to meet the military task-specific requirements. However, the primary mechanism remains as yet undiscovered.

Because being deployed on a mission is associated with fitness for duty, in chapter 7 the same analyses are performed as in chapters 5 and 6, but this time restricted to those soldiers from both study groups who had been deployed on a mission abroad. In this study those soldiers were monitored for two years, using the two operational measures: number of days being fit for duty and health care costs. In addition, during the study period the candidates were given a questionnaire twice a year. The study revealed that the soldiers who were deployed on a mission and assessed using the job-based BMEKL system, displayed a greater fitness for duty than those assessed using the diagnosis-based ABOHZIS system. The BMEKL assessment system is a better predictor, with regard to the ability to function as a soldier in general and with regard to deployment, health and the locomotor apparatus in particular, than the ABOHZIS system.

In chapter 8 the concluding remarks are made. A final consideration of the studies leads to a number of remarks, some of which may influence the final answers to the study questions. Firstly, only the inter-observers variation was studied. Secondly, the new BMEKL system is a better predictor than the old ABOHZIS system, but the primary mechanism remains undiscovered. This implies that parts of some study questions (reproducibility measured only in terms of the inter-observers variation; no explanation as to why some factors are responsible for the differences) cannot be answered. Yet the evidence, gathered from the various parts of this study, is considered to be sufficient to accept the hypothesis. Accepting the hypothesis raises the question as to whether the theory on the subject of assessments should be adapted. Several explanations can be found to make the – possibly only apparent – difference between this hypothesis and the theory acceptable. All those explanations, however, only apply to

the present study. According to Bouter and Van Dongen [1] the results of one single study seldom change a theory. In fact they state that an empirical spiral is necessary to describe reality more accurately each time. Therefore, changing the theory would seem premature. It would require further research to find out whether, according to the empirical spiral, the theory should be adapted in the future. This chapter concludes with several recommendations for such research.

References

1. Bouter LM, Dongen MCJM van: Epidemiologisch onderzoek, opzet en interpretatie. Bohn, Stafleu en van Loghum. Houten, 1995.
2. Kort WLAM de: Personnel selection through pre-employment medicals. Amsterdam, 1993.
3. Dijk FJH van, Dormolen M van, Kompier MAJ et al: Herwaardering model belasting-belastbaarheid. TSG, 1990; 68: 3-10.

Nawoord

Bij de totstandkoming van dit proefschrift heb ik van veel mensen ondersteuning mogen ontvangen. Die ondersteuning was uiteenlopend van vorm; telkens heb ik daaraan motivatie en inspiratie ontleend. Op het gevaar af dat ik iemand vergeet te vermelden, wil ik daarom hieronder iedereen bedanken die mij de afgelopen jaren heeft ondersteund.

Ik bedank de collega's waarmee ik de Basismedische Eisen heb ontwikkeld. In de eerste plaats Jeroen Peters – medeauteur van het artikel waarop hoofdstuk 3 is gebaseerd – en voorts Ed Hendriks, Hans de Boer, Tjeerd Homsma, Erik Peperkamp en John van Meeteren, voor de prettige en succesvolle samenwerking in 1997.

Voorts bedank ik de collega's in de periode 1998 – 2000 werkzaam waren als uitslaggevend arts bij het Instituut Keuring en Selectie (IKS), voor hun medewerking: Ronald de Groot, Theo van der Snoek[†] en Remco Holst. Evenzeer gaat mijn dank uit naar alle keurend artsen van het IKS uit die periode, voor hun medewerking.

Mijn dank gaat verder uit naar Evert van Ankum, de toenmalige Commandant van het Militair Geneeskundig Facilitair Bedrijf, voor het beschikbaar stellen van financiële ondersteuning. Ook wil ik Adri Blom, Wim Melger en alle medewerkers van het bureau Medisch en Paramedisch Personeel bedanken, voor hun behulpzaamheid bij het raadplegen van de personeelsbestanden. Ook Paulien Burema bedank ik voor al haar zoekwerk. Tevens gaat mijn dank uit naar Maarten Hoeienbos, die als geen ander iemand tot publiceren weet aan te zetten, evenals naar Dirk Kramer, voor zijn medewerking en morele steun.

In deze opsomming mogen beslist Joop de Lange en André Sondeijker niet ontbreken, voor hun gastvrijheid bij de redactie van het Nederlands Militair Geneeskundig Tijdschrift. Joop bovendien nog in het bijzonder bedankt voor zijn redactionele bijdrage. Ook bedank ik Thea van Maanen en Onno Mulder, voor hun enthousiaste inzet waardoor telkens weer alle vragenlijsten halfjaarlijks op hun bestemming terecht zijn gekomen.

Voorts bedank ik Ken Redekop, voor zijn onmisbare bijdrage. Evenzeer bedank ik Evert-Jan Slootman voor zijn steun en inspiratie. Ook mijn superieuren gedurende de onderzoeksperiode, Wil van der Water en Rob van der Meer, ben ik dank verschuldigd voor het beschikbaar stellen van “de baas zijn tijd”.

Tot slot gaat mijn dank in het bijzonder uit naar mijn promotoren, Frans Nijhuis en Han Willems, die zich beiden leermeester in de ware zin des woords hebben getoond.

 

 

 

Curriculum Vitae

John de Raad is op 9 november 1956 in Dordrecht geboren. Na het Johan de Witt-gymnasium en het vervullen van de militaire dienstplicht heeft hij de studie geneeskunde gevolgd aan de Erasmus Universiteit Rotterdam. Als assistent werkte hij vervolgens twee jaar bij het St. Ignatius Ziekenhuis te Breda. In 1989 werd hij officier-arts bij de Koninklijke Landmacht. Tot op heden vervulde hij hier de volgende functies:

- Bataljonsarts in de rang van eerste-luitenant-arts bij het 41e Pantserinfanteriebataljon in Ermelo, 1989-1990.
- Hoofd Geneeskundige Instructiegroep in de rang van kapitein-arts aan het Opleidingscentrum Militair Geneeskundige Dienst te Hollandsche Rading, 1990-1992.
- Plaatsvervangend Hoofd Hygiëne Inspectie Orgaan in de rang van majoor-arts in de Luitenant-Generaal Knoopkazerne te Utrecht, 1992-1994.
- Seniorbeleidsmedewerker bedrijfs- en verzekeringsgeneeskunde bij de Directie Personeel en Organisatie, Cluster Gezondheidszorg, Afdeling Gezondheidszorg Beleid en Planning in Den Haag, in de rang van luitenant-kolonel-arts, 1994-2004.

Momenteel is hij Hoofd van de Verzekeringsgeneeskundige Dienst KL te Utrecht in de rang van kolonel-arts.

Van 1990 tot 1993 volgde hij de postacademische opleiding in de Arbeids- en bedrijfsgeneeskunde, gevolgd door de postacademische opleiding in de Verzekeringsgeneeskunde van 1994 tot 1996, beide bij TNO/Preventie en Gezondheid te Leiden.

Gedurende het gehele onderzoeksperiode maakte John deel uit van de Koninklijke Landmacht. Het onderzoek zoals weergegeven in hoofdstuk 2 van dit proefschrift is in 1996 uitgevoerd en gepubliceerd. De opzet voor het onderzoek, weergegeven in de overige hoofdstukken, is in 1998 ontwikkeld en uitgevoerd.

In 1999 is hij als Senior Medical Officer in het kader van SFOR6 met het 12e Luchtmobiele Bataljon (Regiment Van Heutsz) naar Bosnië-Herzegowina uitgezonden.

De overige in dit proefschrift vermelde publicaties zijn in de jaren 1999 – 2004 tot stand gekomen.

Samen met Jeroen Peters en Frans Nijhuis ontving hij in 2000 de Talmapenning voor het artikel, weergegeven in hoofdstuk 3 van dit proefschrift.

John de Raad is gehuwd en vader van een dochter.