

Måleegenskaper ved Klokketesten

Rapport fra Kunnskapssenteret nr 16–2015

Psykometrisk vurdering



 kunnskapssenteret

Nasjonalt kunnskapssenter for helsetjenesten skal, på oppdrag fra Helsedirektoratet vurdere måleegenskapene (reliabilitet og validitet) av psykologiske tester og psykiatriske måleinstrumenter brukt i helsetjenesten i Norge. Oppdraget i denne rapporten var å identifisere forskning om måleegenskapene ved klokketesten (KT) og gi en omtale av testen. • Klokketesten er et visuelt (non-verbal) screeninginstrument for kartlegging av milde til moderate kognitive vansker hos voksne og eldre. Klokketesten har eksistert siden 1950-tallet. Den første standardiserte skåringsversjonen ble utgitt av Freedman og Kaplan i 1994. Norsk versjon er utarbeidet av Strobel m. fl. (2012). Klokketestens utforming og oppgaveinstruks varierer noe med ulike skåringsystemer. Det vanligste er at den som blir testet skal tegne en rund klokkeskive, tegne inn tallene og tegne inn viserne på et angitt tidspunkt. Tidsbruken på testen varierer mellom to og ti minutter. Det er ingen spesielle kompetansekrav for å administrere klokketesten. • Vår oppsummering er basert på internasjonal forskning. Vi

Nasjonalt kunnskapssenter for helsetjenesten
Postboks 7004, St. Olavsplass
N-0130 Oslo
(+47) 23 25 50 00
www.kunnskapssenteret.no
Rapport: ISBN 978-82-8121-968-7 ISSN 1890-1298

nr 16–2015



kunnskapssenteret

utarbeidet søkestrategi for et systematisk litteratursøk og søkte etter nyere systematiske oversikter (publisert etter 2005). Søket ble utført i september 2014. Vi søkte også etter norske studier. To forskere har gjennomgått identifiserte publikasjoner, vurdert relevans i forhold til inklusjonskriteriene og ekstrahert relevante psyometridata. Alle medforfattere har bidratt til skriving av rapporten. • Vi inkluderte tre systematiske oversikter som oppsummerte måleegenskaper ved Klokketesten. Ved bruk av testen i utredning av mulig demens, er sensitivitet rapportert fra 67% til 98% og spesifisitet fra 69% til 95%, og en korrelasjon med Mini Mental State Examination på 0,79. Vi fant ingen relevante norske studier, men fant en norsk versjon som heter KT-NR2. • Klokketesten har overordnet sett adekvate psykometriske egenskaper som et screeninginstrument for kognitiv svikt. Klokketesten fanger bare opp enkelte aspekter ved kognitiv fungering.

Tittel	Måleegenskaper ved Klokketesten
English title	Psychometric assessment of the Clock Drawing Test
Institusjon	Nasjonalt kunnskapssenter for helsetjenesten
Ansvarlig	Magne Nylenna, direktør
Forfattere	Smedslund, Geir, seniorforsker, <i>Nasjonalt kunnskapssenter for helsetjenesten</i> Siqveland, Johan, seniorrådgiver, <i>Nasjonalt kunnskapssenter for helsetjenesten</i> Leiknes, Kari Ann, seniorrådgiver, <i>Nasjonalt Kunnskapssenter for helsetjenesten</i>
Antall sider	19 (23 inklusiv vedlegg)
Publikasjonstype	Psykometrisk vurdering
ISBN	978-82-8121-968-7
ISSN	1890 – 1298
Rapportnr.	16 – 2015
Emneord(MeSH)	Psychometrics; measurement precision; predictive value validity; clock drawing test
Prosjektnummer	9516
Oppdragsgiver	Helsedirektoratet
Sitering	Smedslund G, Siqveland J, Leiknes KA. Måleegenskaper ved klokketesten (KT). Rapport fra Kunnskapssenteret nr 16-2015. Nasjonalt kunnskapssenter for helsetjenesten, Oslo, juni 2015.
Utførelsessted	Kunnskapssenteret fremskaffer og formidler kunnskap om effekt av metoder, virkemidler, tiltak og om kvalitet innen alle deler av helsetjenesten. Målet er å bidra til gode beslutninger slik at brukerne får best mulig helsetjenester. Kunnskapssenteret er formelt et forvaltningsorgan under Helsedirektoratet, men har ingen myndighetsfunksjoner og kan ikke instrueres i faglige spørsmål.

Vi takker Ingvild Kirkehei, forskningsbibliotekar, Nasjonalt kunnskapssenter for helsetjenesten for utførelse av litteratursøkene.

Vi takker intern og ekstern fagfelle som har bidratt med sin ekspertise.

Nasjonalt kunnskapssenter for helsetjenesten,
Oslo, juni 2015.

Hovedfunn

Nasjonalt kunnskapssenter for helsetjenesten skal, på oppdrag fra Helse- direktoratet vurdere måleegenskapene (reliabilitet og validitet) av psykologiske tester og psykiatriske måleinstrumenter brukt i helsetjenesten i Norge.

Oppdraget i denne rapporten var å identifisere forskning om måleegenskapene ved klokketesten (KT) og gi en omtale av testen.

Beskrivelse av klokketesten

Klokketesten er et visuelt (non-verbal) screeninginstrument for kartlegging av milde til moderate kognitive vansker hos voksne og eldre. Klokketesten har eksistert siden 1950-tallet. Den første standardiserte skåringsversjonen ble utgitt av Freedman og Kaplan i 1994. Norsk versjon er utarbeidet av Strobel m. fl. (2012). Klokketestens utforming og oppgaveinstruks varierer noe med ulike skåringsystemer. Det vanligste er at den som blir testet skal tegne en rund klokkeskive, tegne inn tallene og tegne inn viserne på et angitt tidspunkt. Tidsbruken på testen varierer mellom to og ti minutter. Det er ingen spesielle kompetansekrav for å administrere klokketesten.

Metode

Vår oppsummering er basert på internasjonal forskning. Vi utarbeidet søkestrategi for et systematisk litteratursøk og søkte etter nyere systematiske oversikter (publisert etter 2005). Søket ble utført i september 2014. Vi søkte også etter norske studier. To forskere har gjennomgått identifiserte publikasjoner, vurdert relevans i forhold til inklusjonskriteriene og ekstrahert relevante psyometridata. Alle medforfattere har bidratt til skriving av rapporten.

Resultater

Vi inkluderte tre systematiske oversikter som oppsummerte måleegenskaper ved Klokketesten. Ved bruk av testen i utredning av mulig demens, er sensitivitet rapportert fra 67% til 98% og spesifisitet fra 69% til 95%, og en korrelasjon med Mini Mental State Examination på 0,79. Vi fant ingen relevante norske studier, men fant en norsk versjon som heter KT-NR2.

Konklusjon

Klokketesten har overordnet sett adekvate psykometriske egenskaper som et screeninginstrument for kognitiv svikt. Klokketesten fanger bare opp enkelte aspekter ved kognitiv fungering.

Tittel:

Måleegenskaper ved Klokketesten

Publikasjonstype:

Psykometrisk vurdering.

Oppsummering og vurdering av forskning om psykologiske måleinstrumenter, normdata, reliabilitet og validitet av Klokketesten (KT)

Svarer ikke på alt:

- Ingen kritisk vurdering av studienes kvalitet
- Ingen helseøkonomisk vurdering

Hvem står bak denne publikasjonen?

Kunnskapssenteret har gjennomført oppdraget etter forespørsel fra Helsedirektoratet

Når ble litteratursøket utført?

Søk etter studier ble avsluttet september 2014.

Fagfeller:

Intern fagfelle: seniorrådgiver Therese Kristine Dalsbø
Ekstern fagfelle: Anne Brækhus, Oslo Universitetssykehus

Main findings

The Norwegian Knowledge Centre for the Health Services is commissioned by the Norwegian Directorate of Health to undertake a review and assessment of research on psychometric properties of psychiatric measures used in Norway. In this report we were commissioned to identify research on the psychometric properties of the Clock Drawing Test (CDT) and provide an assessment of the test.

Description of Clock Drawing Test

The Clock Drawing Test is a visual (non-verbal) screening instrument for measuring mild to moderate cognitive impairment. The Test has been in use since the 1950's. The first published standardized scoring version was by Freedman and Kaplan in 1994. The Norwegian version is by Strobel et.al. (2012). Design and task instruction of the Clock drawing test varies somewhat according to the scoring system used. The most common is that the test person is instructed to draw a round clock circle with two clock arms facing towards a given time point. The time used is between 2 and 10 minutes. No special skills are required to administer the test.

Methods

We conducted a systematic literature search and searched for newer systematic reviews (published after 2005). We also searched for Norwegian sample studies. The search was conducted in September 2014. Two researchers have identified publications according to inclusion criteria and extracted relevant psychometric data from full text. All authors have contributed to writing the report under supervision by the last author.

Results

We included three systematic reviews. When screening for possible dementia, test sensitivity is reported from 67% to 98% and specificity from 69% to 95%, and correlation to the Mini Mental State Examination 0,79. Relevant Norwegian population studies were not identified, but a Norwegian test version named KT-NR2 was found.

Conclusion

The Clock Drawing Test has generally adequate psychometric properties as a screening instrument for mild to moderate cognitive impairment. The Clock Drawing Test encompasses only some of the aspects of cognitive functioning.

Title:

Psychometric assessment of the Clock Drawing Test

Type of publication:

Psychometric assessment

Review and assessment of research on psychometric properties of psychiatric measures used in Norway. Evaluation of norm data, reliability and validity of the Clock Drawing Test (CDT)

Limitations:

- No critical evaluation of study quality
- No health economic evaluation

Publisher:

Norwegian Knowledge Centre for the Health Services

Updated:

Last search for studies: September 2014.

Peer reviewers:

Senior adviser Therese Kristine Dalsbø at the Norwegian Knowledge Centre for the Health Services
And Anne Brækhus, Oslo University Hospital

Innhold

HOVEDFUNN	2
MAIN FINDINGS	3
INNHold	4
INNLEDNING	6
Beskrivelse	6
Problemstilling	7
METODE	8
Litteratursøk	8
Inklusjonskriterier	8
Eksklusjonskriterier	9
Artikkelutvelging	9
Dataauthenting og sammenstilling	9
RESULTAT	10
Litteratursøk	10
Inkluderte oversikter	11
Skåringssystemer	11
Normdata	13
Reliabilitet	13
Validitet	13
Sensitivitet og spesifisitet	14
DISKUSJON	15
Styrker og svakheter ved denne oversikten	16
KONKLUSJON	17
Behov for mer forskning	17
Implikasjoner for praksisfeltet	17
REFERANSER	18
VEDLEGG	20
Søkestrategier	20
Ekskluderte oversikter	22
Referanser ekskluderte	22

FORORD

Nasjonalt kunnskapssenter for helsetjenesten skal, på oppdrag fra Helsedirektoratet vurdere måleegenskapene (reliabilitet og validitet) av psykologiske tester og psykiatriske måleinstrumenter brukt i helsetjenesten i Norge. Dette oppdraget var å identifisere forskning om måleegenskapene ved klokketesten (KT), gi en vurdering av normdata, reliabilitet og validitet.

Prosjektgruppen har bestått av:

- Geir Smedslund, seniorforsker, Kunnskapssenteret
- Johan Siqveland, seniorrådgiver, Kunnskapssenteret
- Kari Ann Leiknes, seniorrådgiver, Kunnskapssenteret

Gro Jamtvedt
Avdelingsdirektør

Marianne Klemp
Seksjonsleder

Kari Ann Leiknes
Prosjektleder

Innledning

Beskrivelse

Klokketesten (KT) er et screeninginstrument for kartlegging av milde til moderate kognitive vansker hos voksne og eldre (Tuokko, et al., 2008). Testen finnes i noe ulike formater men består grunnleggende av tre oppgaver: tegne en klokkeskive, tegne inn tallene på klokkeskiven og tegne viserne til å vise et spesifikt klokkeslett.

Testen ble opprinnelig laget for å kartlegge visuospatiale og visuokonstruktive ferdigheter. Klokketesten har vært i bruk siden 1950 årene (Peters & Pinto, 2008). Etersom testen er enkel å administrere (krever bare blyant og papir), spredte bruken av den seg fort utover i klinikken. I dag anbefales det at klokketesten inngår sammen med andre tester som bl.a. Mini Mental State Examination (MMSE) i utredningen av generell kognitiv funksjonsevne og demenslidelser (Peters & Pinto, 2008).

Klokketesten administreres ved at testdeltageren skal tegne en klokkeskive eller får presentert en sirkel og blir bedt om å sette inn tallene på klokken i korrekt rekkefølge. Deretter blir han eller hun bedt om å tegne inn visere slik at de viser et bestemt klokkeslett. Testen tar typisk 2-10 minutter å administrere, avhengig av hvilken versjon av testen man bruker og testdeltagerens kognitive fungering. Original skåringsversjon av testen ble først utgitt på engelsk i 1994 (Freedman, Leach, Kaplan, & al, 1994). En tidlig skåringsmanual ble utgitt av Tuokko og medarbeidere i 1995 (Tuokko, Hadjistavropoulos, Miller, & al, 1995). Norsk versjon er utarbeidet av Strobel m. fl. (2012).

Flere varianter for administrering av klokketesten er beskrevet (Burns, Lawlor, & Craig, 2004; Tuokko, et al., 2008):

- Testdeltageren skal tegne klokken på et blankt ark,
- Testdeltageren får en ferdig tegnet sirkel og blir bedt om å tegne tall og visere,
- Testdeltageren får en ferdig tegnet sirkel med tall og blir bedt om å tegne visere for å vise et oppgitt klokkeslett
- Testdeltageren får se en komplett klokke som skal kopieres.

Det finnes flere ulike administrasjonsprotokoller og skåringssystemer. Mange av de mest brukte skåringssystemene evaluerer tilstedeværelse og korrekthet av noen kritiske deler av klokkeproduksjonen (f.eks. sirkulær omkrets, plassering og nøyaktighet på tallene og plassering og nøyaktighet av visere). Den mest brukte skalaen er konstruert av Wolf-Klein (Wolf-Klein, Silverstone, Levy, & Brod, 1989). Denne skalaen går fra 0 poeng (irrelevant figur) til 10 poeng (korrekt klokke).

Andre beskrevne skåringssystemer er 3-punktsystemet til Freedman (Freedman, et al., 1994), 0-10 punktsskalaen foreslått av Sunderland (Sunderland, et al., 1989), og dens modifiserte utgave av Rouleau m. fl. (Rouleau, Salmon, Butters, Kennedy, & McGuire, 1992). Videre har Freedman en 0-15 poengskala (Freedman, et al., 1994) og Tuokko har en 0-31 poengskala (Tuokko, et al., 1995).

En norsk versjon for administrering og skåring fra 0 til 5 poeng har blitt utgitt av Aldring og helse (www.aldringoghelse.no) og heter KT-NR2 (Strobel, Johansen, Bekkhus-Wetterberg, & Engedal, 2014)

Problemstilling

Vi ønsket å gi en vurdering av måleegenskapene (normdata, reliabilitet og validitet) av klokketesten ved å se på systematiske oversikter av internasjonal forskning. Vi ønsket og å gi en vurdering av norsk forskning om den norske versjonen av klokketesten ved å se etter norske studier.

Metode

Litteratursøk

Vi har søkt i elektroniske kilder. Vi søkte systematisk etter litteratur i følgende databaser:

- Ovid Medline
- Ovid Embase
- Ovid PsycINFO
- Cochrane Database of Systematic Reviews
- Dare
- HTA
- Web of Science
- PubMed

Forskningsbibliotekar Ingvild Kirkehei utførte samtlige søk. Den fullstendige søkestrategien er beskrevet i vedlegget. Søk etter studier ble avsluttet september 2014. Fordi vi fant et stort antall primærstudier om klokke testen (525 i litteratursøket), bestemte vi oss til å bare inkludere systematiske oversikter i tråd med vår metodebok (www.kunnskapssenteret.no/verktoy/slik-oppsommerer-vi-forskning).

Vi har ikke søkt etter upubliserte studier eller oversikter.

Inklusjonskriterier

Vi inkluderte alle publikasjoner av studier som hadde følgende:

- | | |
|----------------------|---|
| Populasjon: | Voksne personer |
| Intervensjon: | Anvendelse av KT, sammenligning mot andre måleinstrumenter |
| Studiedesign: | Systematisk oversikter av høy eller middels kvalitet (vurdert med Kunnskapssenterets sjekklister) publisert 2005 eller senere med hovedfokus på KT |
| Utfall: | <ol style="list-style-type: none">1. Normdata for testen2. Reliabilitet (indre konsistens, test-retest, inter-rater og endringsensistivitet) |

3. Validitet (samsvar med lignende testskårer, sammenligning med “gull standard” eller referansestandard og/eller faktorstruktur)
4. Sensitivitet/ spesifisitet/ prediktiv verdi

Eksklusjonskriterier

Vi ekskluderte primærstudier og studier som rapporterte data utelukkende fra barn og ungdom (under 18 år).

Artikkelutvelging

Alle referanser fra litteratursøket ble samlet i en database. To forskere (Smedslund og Siqveland) gikk gjennom alle titler og sammendrag for å vurdere relevans i henhold til inklusjonskriteriene. Mulige relevante referanser ble bestilt i fulltekst. Endelig avgjørelse om inklusjon ble avgjort etter å ha lest referansen i fulltekst. Vurderingene ble foretatt uavhengig av hverandre og sammenlignet i etterkant. Der det var uenighet om vurderingene, ble inklusjon eller eksklusjon avgjort ved konsensus. Kvaliteten på oversikten ble vurdert med Kunnskapssenteret sin sjekklister for systematiske oversikter (<http://www.kunnskapssenteret.no/verktoy/sjekklistere-for-vurdering-av-forskningsartikler>).

Dataauthenting og sammenstilling

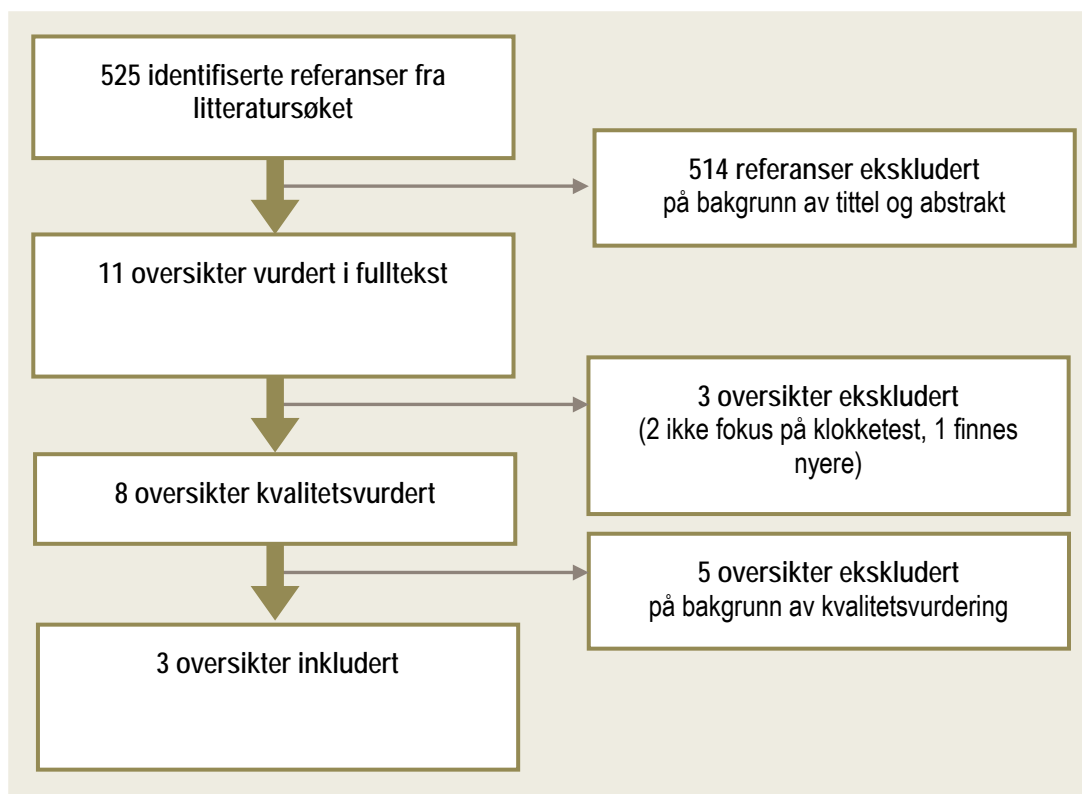
Relevant psykometridata (bl.a. mean, middlverdier, inter-rater reliability, korrelasjoner, sensitivitet, spesifisitet, prediktiv verdi) ble hentet ut fra de inkluderte publikasjonene. Dette ble gjort av Smedslund og sjekket av Siqveland. Leiknes har bidratt til sammenstillingen. Data som ble hentet ut er presentert og sammenstilt i egne tabeller i resultatdelen. En tilpasset versjon av European Federation of Psychologist Association (EFPA) skjema «The revised EFPA 2013 Test Review Model Version 4.2.6» (EFPA, 2013) ble anvendt for psykometrivurderingen.

Resultat

Litteratursøk

Søket resulterte i 716 referanser. Etter dublettkontroll satt vi igjen med 525 referanser. Vi vurderte tittel og abstrakt i 11 av de identifiserte referansene til å være relevante i henhold til inklusjonskriteriene. Vi innhentet 11 oversikter i fulltekst og vurderte dem i forhold til våre inklusjonskriterier. Vi inkluderte tre systematiske oversikter av middels eller høy kvalitet (Se Figur 1).

Hovedårsaken til eksklusjon var at publikasjonen ikke hadde hovedfokus på klokketesten eller hadde lav metodisk kvalitet (Se Vedlegg, Ekskluderte oversikter).



Figur 1. Flytskjema over identifisert litteratur

Vi ønsket også å gi en omtale av norsk forskning og vurdering av måleegenskapene (normdata, reliabilitet og validitet) for den norske versjonen av klokketesten anvendt i Norge.

Vi fant ingen slik forskning og valgte å presentere resultater fra internasjonale systematiske oversikter om klokketesten. Det at forskningen i denne sammenheng ikke stammer fra norske populasjoner vurderes av underordnet betydning ettersom testen ikke inneholder språklige oppgaver.

Inkluderte oversikter

Beskrivelsen av klokketesten i boken «Handbook of Psychiatric Measures» (Tuokko, et al., 2008) og tre systematiske oversikter danner grunnlaget for vurderingen (Se Tabell 1)

Tabell 1: Inkluderte oversikter (N= 3)

Referanse	Antall relevante studier	Populasjon	N	Metodisk kvalitet	Rapporterte egenskaper, verdier
Davis (2013)	1	Personer med normal MMSE skåre	68	Middels	Administrasjonstid, antall ledd, skala, sensitivitet, spesifisitet
Pinto (2009)	41	Personer i alderen 65 år og eldre	Ikke oppgitt	Middels	Sensitivitet, spesifisitet, inter-rater reliabilitet, test-retest reliabilitet, validitet
Lin (2013)	7	Eldre (alder ikke spesifisert)	2509	Høy	Sensitivitet, spesifisitet

Skåringssystemer

Tabell 2 gir en oversikt over skåringssystemer for klokketesten utformet av Pinto (Pinto & Peters, 2009). I noen systemer inngår tegning av klokke, mens i andre skal man kopiere en tegning av en klokke. I noen systemer er sirkelen laget ferdig, mens i andre skal sirkelen tegnes. Tidspunktet som skal markeres på klokken er oftest klokken 11.10, men i noen systemer er det andre tidspunkter og ikke alle systemer krever et tidspunkt. Skåringsmetodene og hvilket nivå («cut-off») som skiller mellom bestått og ikke bestått varierer også.

Tabell 2: Egenskaper ved ulike skåringsystemer for klokketesten etter Pinto (Pinto & Peters, 2009).

Referanse	Test	Ferdig sirkel	Tidspunkt	Skåringsmetode	Cut-off
Shulman	Tegning	Ja	11.10	5 punkts modifisert/6 punkts hierarkisk (1-5, 0-5)	>2 (1986) <3 (1993)
Wolf-Klein	Tegning	Ja	Nei	1-10 hierarkisk	<7
Sunderland	Tegning	Nei	2:45	10 hierarkisk punktskala (1-10)	<6
Mendez	Tegning	Nei	11:10	20-punkts skala basert på feil skåret 0/1 (0-20)	<19
Rouleau	Tegning og kopiering	Nei	11:10	10-punkt skala og 6-punkts kvalitativ skala (0-10)	≤7
Tuokko	Tegning, klokkesetting, klokkelesning	Ja	11:10	25 definerte feil i 7 feilkategorier (0 til ingen maks)	>2
Watson	Tegning	Ja	Nei	Klokke delt i kvadranter: feil i første til fjerde kvadrant = 1 poeng; feil i fjerde kvadrant = 4 poeng (0-7)	>3
Manos og Wu	Tegning	Ja	11:10	Klokke delt i åttendedeler, poeng gitt for tall og visere på riktig plass (0-10)	<8
Cahn	Tegning	Nei	11:10	Kvantitativ skala etter Rouleau og også 8 kvalitative feiltyper (0-10 poeng, 1-8 feil)	Kvantitativ ≤7 Kvalitativ ≥1
Lam	Tegning, lesing og tidssetting	Ja	3:00	10-punktskala (0-10)	>3
Borson	Tegning	Nei	8:20	4-punkts med anker (0-3)	>0
Royall	Tegning og kopiering	Nei	1:45	15-punkt skala (0-15)	<10
Cacho	Tegning og kopiering	Nei	11:10	10-punkt (system modifisert etter Rouleau og Sunderland)	≤6
Jitapunkul	Tegning	Ja	11:10	Klokke delt i kvadranter, 5 ledd markert 0-2 poeng (0-10)	<7
Lin	Tegning og kopiering	Ja	10:10	Klokke delt i kvadranter, 16 ledd markert 0/1 (0-16); kort versjon bare 3 ledd (0-3)	<11 Kort versjon <3

Normdata

De tre inkluderte oversiktene inneholdt ikke norske normdata. Freedman (Freedman, et al., 1994) har publisert normdata fra et stort canadisk utvalg i aldersgruppen 20-90 år (Green, 2000). Vi antar imidlertid at normene for klokketesten ikke er vesentlig forandret over tid, ettersom klokkene våre med skivevisning ikke har forandret seg siden 1990-tallet. Med normer må det her menes det som skiller normale skårer fra patologiske og det er like relevant for ikke-verbale som for verbale tester. Pinto og Peters (Pinto & Peters, 2009) sin Tabell 2 lister opp en rekke ulike skåringssystemer med tilhørende nivå («cut-off») punkt for å skille mellom «normal» og «ikke-normal» skår.

Reliabilitet

Pinto (Pinto & Peters, 2009) har en tabell hvor inter-rater reliability (IRR) for 14 studier er rapportert. Her varierer IRR mellom 0,75 og 0,97 med median 0,94. I samme tabell varierer test-retest reliabiliteten mellom 0,70 og 0,94 i 5 studier. Her er medianen 0,82.

Validitet

Tabell 3 gir en oversikt over korrelasjoner mellom klokketesten og andre skåringsinstrumenter etter Pinto (Pinto & Peters, 2009).

Tabell 3: Korrelasjoner mellom klokketesten og andre tester (Pinto & Peters, 2009).

Test	Korrelasjon med KT
BDRS	+0,51
Block Design Test	+0,56
CAMCOG	+0,77
CAMCOG memory subscale	+0,7
CAMCOG praxis	+0,71
CASI	-0,8
CASI (kinesisk versjon)	+0,73
CMT	+0,66
DRS	+0,59
EXIT25	-0,78
GDS	-0,32 til +0,56

Test	Korrelasjon med KT
MMSE	-0,79 til +0,77
Rey Complex Figure Test	+0,66
SPMSQ	-0,66 til +0,59
Symbol Digit Modalities Test	+0,65
Trail Making Test Part A	-0,48
WAIS	+0,76

BDRS = Blessed Dementia Rating Scale; CAMCOG = Cambridge Cognitive Examination Score; CASI = Cognitive Abilities Screening Instrument; CMT = Chula Mental Test; DRS = Dementia Rating Scale; EXIT25 = Executive Interview; GDS = Global Deterioration Scale; MMSE Mini Mental State Examination; SPMSQ = Short Portable Mental Status Questionnaire; WAIS = Wechsler Adult Intelligence Scale.

Korrelasjon mellom klokketesten og andre kognitive tester var lavest i forhold til BDRS ($r=0,51$) og høyest til EXIT25 ($r=0,78$) og MMSE ($0,79$).

Sensitivitet og spesifisitet

Sensitivitet

Lin (Lin, O'Connor, Rossom, Perdue, & Eckstrom, 2013) inkluderte seks studier med 2170 deltakere og rapporterte en sensitivitet på mellom 67 og 97,9 (95 prosent konfidensintervall: 39-100) for diagnostisering av demens blant pasienter i primærhelsetjenesten (primary care). Davis (Davis & Allen, 2013) rapporterer sensitivitet på 59% for å identifisere kognitiv svikt hos pasienter med hjertefeil (en studie med 68 deltakere). Det er ikke oppgitt i Lin sin oversikt hvilke instrumenter som ble brukt i denne studien.

Spesifisitet

Lin (Lin, et al., 2013) inkluderte seks studier med 2170 deltakere og rapporterte en spesifisitet på mellom 69 og 95,2 for diagnostisering av demens blant pasienter i primærhelsetjenesten (primary care) (95 prosent konfidensintervall: 54-97,1). Davis (Davis & Allen, 2013) rapporterer spesifisitet på 70 % for å identifisere kognitiv svikt hos pasienter med hjertefeil (en studie med 68 deltakere). Det er ikke oppgitt i Lin sin oversikt hvilke diagnostiske instrumenter som ble brukt i denne studien.

Prediktiv verdi

Ingen av de tre inkluderte oversiktene rapporterte resultater på positiv prediktiv verdi (PPV) eller negativ prediktiv verdi (NPV). En oversikt av Peters og Pinto (Peters & Pinto, 2008) som omtaler prediksjon av klokketesten i forhold til Mini Mental State Examination (MMSE) ble imidlertid ekskludert ut fra lav kvalitet.

Diskusjon

Vi inkluderte tre systematiske internasjonale (engelskspråklige) oversikter for å vurdere dokumentasjon på klokketestens psykometriske egenskaper.

Klokketesten synes overordnet sett å ha adekvate psykometriske egenskaper som et screeninginstrument for demens. Siden den bare fanger opp enkelte aspekter ved kognitiv fungering bør den suppleres med andre tester i en mer omfattende utredning.

Det er imidlertid noe usikkerhet knyttet til bruk av klokketesten i norsk sammenheng. Vi fant ingen normgivende norske studier basert på norske populasjoner. Ettersom klokketesten er utviklet for å teste non-verbale kognitive funksjoner er norske studier, og studier med norsk pasientutvalg trolig av underordnet betydning. Det foreligger dog en norsk klokketesten versjon med 5 poengskala (5 best/høyest og 0 dårligst/lavest skår) og foreslått nivå («cut-off»). Godkjent klokke er 4-5 poeng, og ikke godkjent er 0-3 poeng (Strobel, et al., 2014). Internasjonalt, som vist i denne rapporten finnes det flere klokketesten varianter med flere skåringssystemer. Med bakgrunn i våre funn og den sparsomme psykometriske dokumentasjonen, er det ikke mulig å dokumentere at den norske varianten (Strobel, et al., 2014) kan anbefales fremfor andre klokketesten skåringsvarianter. Hvorvidt denne norske klokketesten skåringsvarianten (Strobel, et al., 2014) er den mest utbredte i Norge er heller ikke mulig å vurdere ut fra denne litteraturgjennomgangen. Utfra eksisterende mangfold av klokketesten skåringsvarianter, som rapportert her fra oversikten til Pinto (Pinto & Peters, 2009), vil det trolig finnes flere skåringsvarianter av klokketesten i bruk i norsk klinisk praksis.

Når det gjelder normdata er det vanskelig å si noe generelt fordi det eksisterer så mange ulike skåringsystemer med tilhørende «cut-off» punkter for å definere hva som er et positivt funn for diagnosen demens. Publikasjonen om normdata fra 1994 av Freedman og medarbeidere (Freedman, et al., 1994) er og av eldre dato. Det ble her ikke funnet noe nyere.

Klokketestens validitet er vurdert ut fra korrelasjonstallene som er presentert av Pinto (Pinto & Peters, 2009). Spredningen og variasjonen i korrelasjonstallene som angitt (jfr. Tabell 3) vil også være påvirket av anvendelse av forskjellig skåringsystemer (0 til 10 eller 0 til 5) for klokketesten. Vi stiller oss også noe skeptiske til at mange av skalaene som Pinto og Peters har brukt til å vurdere validitet ikke er ment

å måle de samme egenskaper som klokketesten. De gir altså ikke noe særlig godt grunnlag for å vurdere validiteten til klokketesten.

Sensitivitet og spesifisitet var svært varierende, men dette er ikke overraskende siden det er så mange ulike skåringssystemer som er i bruk, samt at testen har blitt brukt på mange ulike populasjoner over mange år.

Klokketesten er lett å administrere og anses som et relativt effektivt kognitivt screeninginstrument. Når det gjelder bruk og anvendelse i forskjellige aldersgrupper og pasientpopulasjoner, har klokketesten for det meste blitt brukt blant eldre for å kartlegge kognitiv svikt og demens.

Klokketesten anses derimot som lite sensitiv ved mild kognitiv svikt. Ettersom klokketesten ikke omfatter de kognitive funksjonene som er affisert tidlig ved Alzheimers lidelse, bl.a. episodisk hukommelse, taleferdighet (verbal «fluency») og cognitive set shifting (evne til å skifte mellom ulike kategorier, regler eller tankesett), kan klokketesten overse klinisk signifikant nedsatte kognitive funksjoner i en tidlig fase for denne pasientgruppen (Tuokko, et al., 2008).

Visuelle forhold (visuell neglekt) og motoriske utfall kan også påvirke testresultatet (Bystad, Skjerve, & Strobel, 2013). Vi anser det som rimelig å anta at jo mer testen legger opp til at personen selv må ta initiativ for å utføre oppgavene på en måte som krever mental energi jo mer påvirket er de av depresjon.

Man bør også være oppmerksom på at resultatet på klokketesten kan påvirkes av depressive plager. En fordel med klokketesten sammenlignet med MMSE hevdes å være at den er mindre influert av nedsatt stemningsleie, utdanningsnivå og verbale språkferdigheter (Peters & Pinto, 2008).

Styrker og svakheter ved denne oversikten

Det er en styrke ved denne oppsummeringen at vi har gjort et systematisk litteratursøk. Vi er sikre på at vi har funnet de mest oppdaterte oversiktene av høy og middels kvalitet som er publisert. En svakhet er at det likevel kan finnes primærstudier som har blitt publisert etter at de inkluderte oversiktene gjorde sine litteratursøk. En slik oversikt over oversikter kan heller ikke bli bedre enn rapporteringen i de inkluderte oversiktene (som igjen ikke kan bli bedre enn det som er rapportert i primærstudiene). Det kan også være at det finnes norskspråklige studier som de inkluderte oversiktene ikke har funnet.

Konklusjon

Klokketesten har overordnet sett adekvate psykometriske egenskaper som et screeninginstrument for demens, mild kognitiv svikt og Alzheimers. Selv om vi ikke har funnet normdata for norske populasjoner, er klokketesten ikke en verbal test og har derfor sannsynligvis forholdsvis like normer på tvers av land. Klokketesten fanger opp enkelte aspekter ved kognitiv fungering og bør derfor suppleres med andre tester i en full screeningundersøkelse.

Behov for mer forskning

Det er et problem at det finnes så mange ulike typer klokketester og tilhørende skåringssystemer. Derfor er det behov for sammenliknende studier for å forsøke å finne fram til den varianten av klokketesten som har de beste psykometriske egenskapene slik at det på sikt kan bli mulig å komme fram til en enighet om hvordan testen skal administreres og hvordan den skal skåres.

Implikasjoner for praksisfeltet

Bystad og medforfattere skriver: «Som en enkeltstående test har klokketesten begrenset nytteverdi og bør derfor brukes som et supplement sammen med andre tester der også hukommelse og orientering blir vektlagt (Bystad, et al., 2013).» Den kan for eksempel brukes sammen med MMSE.

Referanser

Burns, A., Lawlor, B. & Craig, S. (2004). *Assessment scales in old age psychiatry* (Second Edition utg.). London: Martin Dunitz.

Bystad, M., Skjerve, A. & Strobel, C. (2013). Assessing dementia: A presentation of MMSE-NR, Clock-Drawing Test and informant scales. *Tidsskrift for Norsk Psykologforening*, 50(1), 7-11.

Davis, K. K. & Allen, J. K. (2013). Identifying cognitive impairment in heart failure: A review of screening measures. *Heart & Lung*, 42(2), 92-97. doi: 10.1016/j.hrtlng.2012.11.003

EFPA. (2013 13.07.2013). *European Federation of Psychologists' Associations (EFPA) Test review form and notes for reviewers, Version 4.2.6* Lokalisert 11.06.2014, på [http://www.psychtesting.org.uk/download\\$.cfm?file_uuid=696852C1-985D-B6B0-7E55-29D144984AF3&siteName=ptc](http://www.psychtesting.org.uk/download$.cfm?file_uuid=696852C1-985D-B6B0-7E55-29D144984AF3&siteName=ptc)

Freedman, M., Leach, L., Kaplan, E. & al, e. (1994). *Clock drawing: a neuropsychological analysis*. New York: Oxford University Press.

Green, J. (2000). *Neuropsychological evaluation of the older adult: a clinician's guidebook*. London: Academic Press.

Lin, J. S., O'Connor, E., Rossom, R. C., Perdue, L. A. & Eckstrom, E. (2013). Screening for cognitive impairment in older adults: a systematic review for the US Preventive Services Task Force. *Annals of Internal Medicine*, 159(9), 601-612.

Peters, R. & Pinto, E. M. (2008). Predictive value of the Clock Drawing Test. A review of the literature. *Dement Geriatr Cogn Disord*, 26(4), 351-355. doi: <http://dx.doi.org/10.1159/000162261>

Pinto, E. & Peters, R. (2009). Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dementia and Geriatric Cognitive Disorders*, 27(3), 201-213. doi: <http://dx.doi.org/10.1159/000203344>

Rouleau, I., Salmon, D. P., Butters, N., Kennedy, C. & McGuire, K. (1992). Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain Cogn*, 18(1), 70-87.

Strobel, C., Johansen, H., Bekkhus-Wetterberg, P. & Engedal, K. (2014). Norsk revidert klokke-test (KT-NR2). I Aldring og helse: Helsedirektoratet. <http://www.aldringoghelse.no/startside/utredningsverktoy>

Sunderland, T., Hill, J. L., Mellow, A. M., Lawlor, B. A., Gundersheimer, J., Newhouse, P. A. et al. (1989). Clock drawing in Alzheimers-disease - a novel measure of dementia severity. *Journal of the American Geriatrics Society*, 37(8), 725-729.

Tuokko, H., Hadjistavropoulos, T., Miller, J. & al, e. (1995). *The Clock Test: Administration and Scoring Manual*. Toronto, ON, Canada: Multi-Health Systems.

Tuokko, H., Hadjistavropoulos, T., Miller, J. A., Freedman, M., Leach, L., Kaplan, E. et al. (2008). Clock Drawing Test. I A. J. Rush Jr, M. B. First & D. Blacker (Red.), *Handbook of Psychiatric Measures* (Second Edition utg., s. 413-416). Washington DC: American Psychiatric Publishing.

Wolf-Klein, G. P., Silverstone, F. A., Levy, A. P. & Brod, M. S. (1989). Screening for Alzheimer's disease by clock drawing. *Journal of the American Geriatrics Society*, 37(8), 730-734.

Vedlegg

Søkestrategier

Klokketesten

Dato: 29.09.2014

Databaser: Ovid MEDLINE, Ovid Embase, Ovid PsycINFO, Cochrane Database of Systematic Reviews, DARE, HTA, Web of Science, PubMed

Resultat: 525 referanser (716 inkludert dubletter)

MEDLINE, Embase, PsycINFO

Søketreff

Embase 1980 to 2014 Week 39: 157

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) 1946 to Present: 151

PsycINFO 1806 to September Week 4 2014: 211

1. (clock test* or clocktest* or clock draw*).mp,cm,tm,nt.
2. psychometrics/ use pmoz,psych or pschometry/ use emez
3. "reproducibility of results"/ use pmoz or dimensional measurement accuracy/ use pmoz
4. exp measurement precision/ use emez
5. exp validity/ use emez or exp reliability/ use emez
6. test validity/ use psych or test reliability/ use psych
7. predictive value*.mp.
8. or/2-7
9. 1 and 8
10. (news or comment or editorial).pt. or comment on.cm. or comment reply.dt. or letter.dt.
11. 9 not 10
12. remove duplicates from 11

Cochrane Library

Søketreff

Cochrane Database of Systematic Reviews: 27

DARE: 7

CENTRAL: 6

- #1 ((clock next test*) or clocktest* or (clock next draw*))
- #2 MeSH descriptor: [Psychometrics] explode all trees
- #3 MeSH descriptor: [Reproducibility of Results] explode all trees

- #4 (Consistenc* or dependabilit* or psychometr* or reliabilit* or validit* or reproducibilit*)
- #5 predictive next value
- #6 MeSH descriptor: [Predictive Value of Tests] explode all trees
- #7 #2 or #3 or #4 or #5 or #6
- #8 #1 and #7

PubMed

Søketreff: 16

(pubstatusaheadofprint) AND ((clock draw* or clock test* or clocktest*[Title/Abstract]))

ISI Web of Science

Dato: 29.9.2014

Søketreff: 253

- # 5 #2 AND #1 Refined by: Databases: (WOS) AND DOCUMENT TYPES: (ARTICLE OR ABSTRACT OR CLINICAL TRIAL OR REVIEW)
- # 4 #2 AND #1
- # 3 #2 AND #1
- # 2 TOPIC: ("cock test*" or "clock draw*" or "clocktest*")
- # 1 TOPIC: ("Consistenc*" or "dependabilit*" or "psychometr*" or "reliabilit*" or "validit*" or "reproducibilit*" or "predictive value*")

CRD Databases

Søketreff DARE: 7

Søketreff HTA: 0

Any field: ("clock test*" or "clock draw*" or "clocktest*")

Ekskluderte oversikter

Tabell over ekskluderte oversikter innhentet i fulltekst (N=8)

Referanse Førsteforfatter	Grunn
Appels (2010)	Handler ikke om klokketesten
Aprahamian (2009)	Systematisk oversikt av lav kvalitet
Cullen (2007)	Ikke systematisk oversikt
Harvan (2006)	Systematisk oversikt av lav kvalitet
Ismail (2010)	Systematisk oversikt av lav kvalitet
Mainland (2014)	Systematisk oversikt av lav kvalitet
Peters (2008)	Vi har inkludert en nyere versjon
Shulman (2000)	Publisert før 2005

Referanser ekskluderte

Appels, B. A. & Scherder, E. (2010). The Diagnostic Accuracy of Dementia-Screening Instruments With an Administration Time of 10 to 45 Minutes for Use in Secondary Care: A Systematic Review. *American Journal of Alzheimers Disease and Other Dementias*, 25(4), 301-316. doi: 10.1177/1533317510367485

Aprahamian, I., Martinelli, J. E., Neri, A. L. & Yassuda, M. S. (2009). The Clock Drawing Test: A review of its accuracy in screening for dementia. *Dementia & Neuropsychologia*, 3(2), 74-80.

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F. & Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. *Journal of Neurology Neurosurgery and Psychiatry*, 78(8), 790-799. doi: 10.1136/jnnp.2006.095414

Harvan, J. R. & Cotter, V. T. (2006). An evaluation of dementia screening in the primary care setting. *Journal of the American Academy of Nurse Practitioners*, 18(8), 351-360. doi: 10.1111/j.1745-7599.2006.00137.x

Ismail, Z., Rajji, T. K. & Shulman, K. I. (2010). Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry*, 25(2), 111-120. doi: <http://dx.doi.org/10.1002/gps.2306>

Mainland, B. J., Amodeo, S. & Shulman, K. I. (2014). Multiple clock drawing scoring systems: simpler is better. *International Journal of Geriatric Psychiatry*, 29(2), 127-136. doi: <http://dx.doi.org/10.1002/gps.3992>

Peters, R. & Pinto, E. M. (2008). Predictive value of the Clock Drawing Test. A review of the literature. *Dement Geriatr Cogn Disord*, 26(4), 351-355. doi: <http://dx.doi.org/10.1159/000162261>

Shulman, K. I. (2000). Clock-drawing: is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, 15(6), 548-561.