

Diagnostisk nøyaktighet ved algoritmen til den norske versjonen av Strengths and Difficulties Questionnaire (SDQ)

Jon Sigurd Eidstuen¹ (jon.sigurd.eidstuen@ahus.no)

Hege Kornør² (hege.kornor@r-bup.no)

PsykTestBarn 2017, 1:2

ISSN 1893-9910

Mottatt dato 30. januar 2017

Publisert dato 7. mars 2017

Artikkelens URL <https://psyktestbarn.r-bup.no/no/artikler/sdq-algoritmen-strength-and-difficulties-questionnaire-algorithm>

DOI 10.21337/0049

Som alle artikler i PsykTestBarn, kan denne fagfellevurderte artikkelen lastes ned, skrives ut og distribueres fritt for alle slags formål på følgende vilkår: korrekt referanse skal oppgis (se under), ingen kommersiell bruk og ingen bearbeidelse av tekst eller innhold.

¹ Akershus universitetssykehus

² Regionsenter for barn og unges psykiske helse (RBUP Øst og Sør)

Denne artikkelen skal siteres på følgende måte:

Eidstuen, J. S. & Kornør, H. (2017). Diagnostisk nøyaktighet ved algoritmen til den norske versjonen av Strengths and Difficulties Questionnaire (SDQ). *PsykTestBarn*, 1:2

Sammendrag

Beskrivelse. Strengths and Difficulties Questionnaire (SDQ) er et sett spørreskjema med tre informantversjoner: foreldre, lærere og barn ≥ 11 år. Måleegenskapene ved de ulike informantversjonene av SDQ er vurdert i separate PsykTestBarn-artikler. Robert Goodman har utviklet en algoritme for å predikere hvorvidt en psykisk lidelse er tilstede. SDQ-algoritmen inneholder data fra minst to informanter, men kan også ha data fra samtlige tre informanter. Ved hjelp av algoritmen kan man estimere sannsynligheten for at et barn har en diagnose blant fire brede kategorier: emosjonell lidelse, atferdsforstyrrelse, hyperkinetisk-/oppmerksomhetsforstyrrelse og *noen* (minst én) psykisk lidelse. Sannsynligheten har tre terskelverdier: *unlikely*, *possible* og *probable*. I denne artikkelen undersøker vi den diagnostiske nøyaktigheten til SDQ-algoritmen i skandinaviske utvalg.

Litteratursøk. Av 541 unike treff vurderte vi 133 i fulltekst, hvorav 11 publikasjoner oppfylte inklusjonskriteriene. Fem publikasjoner var delstudier i kohortstudien Barn i Bergen, fem var andre norske studier og én var en dansk studie. Studiene omfattet både befolknings- og selekterte utvalg i alderen 5-18 år.

Psykometri. Seks studier rapporterte relevante data for vurdering av diagnostisk nøyaktighet, mens fem kun rapporterte forekomster. SDQ-algoritmen hadde generelt god sensitivitet, spesifisitet og negativ prediktiv verdi i selekterte utvalg, mens den positive prediktive verdien varierte. I befolkningsutvalg der kun ADHD var undersøkt hadde algoritmen utmerket spesifisitet og negativ prediktiv verdi, mens sensitiviteten og den positive prediktive verdien var utilfredsstillende for denne diagnosen. Ingen av referansestandardene i studiene av diagnostisk nøyaktighet hadde godt nok dokumentert kriterievaliditet.

Konklusjon. Med forbehold om usikkerheten knyttet til referansestandard og normgrunnlag i de inkluderte studiene, har SDQ-algoritmen god evne til å identifisere barn med og uten *noen* psykisk lidelse i selekterte utvalg. I befolkningsbaserte utvalg kan SDQ-algoritmen med stor sannsynlighet korrekt utelukke ADHD, mens den har dårlig evne til å påvise ADHD. For å få bedre kunnskap om SDQ-algorithmens diagnostiske nøyaktighet i norske utvalg trenger vi studier som bruker en referansestandard med dokumentert kriterievaliditet.

Abstract

Description. Strengths and Difficulties Questionnaire (SDQ) is a set of questionnaires with three informant versions: parents, teachers and youth aged ≥ 11 years. The psychometric properties of the various SDQ informant versions have been assessed in separate PsykTestBarn articles. Robert Goodman has developed an algorithm for the prediction of the presence of a mental health disorder. The SDQ algorithm encompasses data from at least two informants, but can also include data from all three categories of informants. The algorithm can estimate the likelihood of

a child's diagnoses within four broad categories: emotional disorder, behaviour disorder, hyperkinetic/attention disorder and *any* disorder. The likelihood is expressed at one of three thresholds; unlikely, possible and probable. The aim of this systematic review is to assess the diagnostic accuracy of the SDQ algorithm in Scandinavian samples.

Literature search. Out of 541 unique hits, we screened 133 in full text, whereof 11 publications met the inclusion criteria. Five publications were sub-studies in the Bergen Child Study, five were other Norwegian studies, and one was a Danish study. The studies included both population based and selected samples with an age range of 5 to 18 years.

Psychometrics. Six studies reported relevant diagnostic accuracy data, while five reported prevalence rates. The SDQ algorithm had generally good sensitivities, specificities and negative predictive values (NPV) in selected samples, but the positive predictive values (PPV) varied. In population-based samples where ADHD was the only target disorder, the algorithm had excellent specificity and NPV, while the sensitivity and PPV were poor. None of the reference standards used in the studies had sufficiently documented criterion validity.

Conclusion. The SDQ algorithm has a good ability to identify children with and without mental health disorders in selected samples. In population-based samples, the SDQ-algorithm can exclude ADHD correctly with great accuracy, but its ability to detect ADHD is poor. We need studies using reference standards with documented criterion validity to gain better knowledge of the diagnostic accuracy of the SDQ-algorithm.

Innledning

Strengths and Difficulties Questionnaire (SDQ) er et sett spørreskjema for kartlegging av psykisk helse, vennerelasjoner og prososial atferd hos barn og unge fra 4 til 17 år (Goodman, 1997). Det er også en versjon for barn i alderen 2-4 år. Det finnes egne versjoner for ulike informanter: foreldrerapport (SDQ-P), selvrappport (SDQ-S) og lærerrapport (SDQ-T), samt eget follow-up-skjema. Alle papirutgavene er fritt tilgjengelig i originalversjoner og oversettelser til over 70 språk, inkludert norsk bokmål og nynorsk, på www.sdqinfo.org. Nettstedet drives gjennom det private britiske firmaet Youth In Mind Ltd, og administreres av SDQs opphavsmann Robert Goodman. I tillegg til SDQ-skjema som kan lastes ned og skrives ut, finnes det der flere relevante ressurser for SDQ: skåringsprogram, normer for ti land og artikkelliste over studier av SDQ. SDQ-skjemaene er også tilgjengelige på norsk i elektronisk versjon (krever betaling og avtale om bruk av Checkwares plattform) på www.checkware.com. Den norske versjonen av SDQ kom i 1999, og den er basert på en oversettelse av Einar Heiervang, Irene Elgen og Sonja Heyerdahl, samt tilbakeoversettelse ved

Melanie Young. De psykometriske egenskapene ved SDQ-S og SDQ-T ved bruk i Norge er undersøkt i to systematiske oversikter (Kornør & Heyerdahl, 2013, 2014).

Hoveddelen av SDQ består av 25 ledd som kan besvares med "stemmer", "stemmer delvis" eller "stemmer ikke". Svarene skåres henholdsvis 2, 1 og 0, (0, 1 og 2 ved reverserte ledd). Skårene kan summeres i fem skalaer (0-10), hvor hver skala rommer fem ledd. Disse fem skalaene er Atferdsproblemer, Emosjonelle symptomer, Hyperaktivitet- og oppmerksomhetsproblemer, Venneproblemer og Prososial atferd. Fire av skalaene (Hyperaktivitet- og oppmerksomhetsproblemer, Emosjonelle problemer, Atferdsproblemer og Venneproblemer) regnes som problemskalaer, og de kan summeres for å lage en skåre (0-40) for Totale problemer. Prososial atferd er ikke inkludert i denne totalen ettersom fravær av prososial atferd ikke er det samme som tilstedeværelse av psykiske problemer (Goodman, 1997). Skalaene kan brukes som kontinuerlige variabler, men grenseverdier kan også defineres. Det finnes i tillegg ulike måter å kombinere skalaer, for eksempel kan man summere skårer for Emosjonelle symptomer og

Venneproblemer til en internaliserende skåre, og skårer for Atferdsproblemer og Hyperaktivitet- og oppmerksomhetsproblemer til en eksternaliserende skåre (Goodman, Lamping & Ploubidis, 2010).

I et tilleggsskjema (*impact supplement*) spør man om informanten tror den unge personen har vansker/problemer, og hvis ja, om varigheten av problemet og hvordan vanskene påvirker barnet og familiens liv (sju ledd). Leddene besvares "ikke i det hele tatt" (0), "bare litt" (0), "en god del" (1) og "mye" (2). Fem av leddene kan summeres til en skåre som angir vanskenes innvirkning på barnets liv (*Impact*). Tanken bak bruk av *impact*-skåren er at et eksklusivt fokus på symptomer ikke er tilstrekkelig for å predikere om psykisk lidelse er tilstede. I tråd med diagnostiske kriterier i *International Classification of Diseases (ICD-10)* eller *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* er det også nødvendig med informasjon om i hvilken grad den unge personen opplever stress eller reduksjon av sosial funksjon (Goodman, 1999). Utover *impact*-tilleggsskjemaet, er det også utarbeidet en egen versjon for oppfølging. Denne rommer to spørsmål om behandling (Ford, Hutchings, Bywater, Goodman & Goodman, 2009).

Algoritmen

SDQ-algoritmen har som hensikt å predikere tre grader av sannsynlighet (*unlikely*, *possible* eller *probable*) for at en psykiatrisk diagnose er tilstede (Goodman, Renfrew & Mullick, 2000). For å få en så nøyaktig prediksjon som mulig kombinerer algoritmen symptom- og *impact*-skårer fra to eller flere informanter. Skårene legges inn i statistikkprogram som beregner hvorvidt et barn har en *unlikely*, *possible* eller *probable* diagnose innen fire brede diagnostiske kategorier (SDQ-indeks): atferdsforstyrrelser (ATF), emosjonelle lidelser (EMO), hyperaktivitet- og oppmerksomhetsforstyrrelser (ADHD) og *noen* (*minst én*) psykiatrisk diagnose (NOEN). Kriteriene for en *probable* ATF-diagnose er en symptomskåre over 95-persentilen på SDQ-skalaen Atferdsproblemer og en *Impact*-skåre på ≥ 2 . For ATF er det tilstrekkelig at disse kriteriene er oppfylt for én informant. For *probable* ADHD gjelder de samme kriteriene, men symptomskalaen er Hyperaktivitet- og oppmerksomhetsproblemer, og

kriteriene må oppfylles av minst to informanter. *Probable* EMO har de samme kriteriene som *probable* ADHD, men her er det symptomskårer på skalaen Emosjonelle problemer som gjelder. I tillegg kan ikke kriteriene for *probable* ATF eller ADHD være oppfylt. Kriteriet for *Probable* NOEN er at *probable*-kriteriene for minst én av de øvrige SDQ-indeksene skal være oppfylt. Kriteriene for *possible* SDQ-indeks er mer inkluderende, det vil si, lavere symptom- og *Impact*-skårer.

Algoritmen slik den er tilgjengelig fra www.sdqinfo.org er basert på britiske normer for hver informant. (Goodman, 1997). Norske normer foreligger ennå ikke, men sammenlikninger av gjennomsnittsskårer i regionale norske utvalg og britiske normer viser at norske barn generelt skårer noe lavere enn britiske (Kornør & Heyerdahl, 2013, 2014).

Algoritmens diagnostiske nøyaktighet

Å evaluere den diagnostiske nøyaktigheten til en test forutsetter at resultatene sammenliknes med en uavhengig etablert referansestandard. En helt korrekt «gullstandard» eksisterer sjelden og man forsøker å definere en tilstrekkelig referansestandard som er så nær en gullstandard som mulig. De fleste studier av diagnostisk nøyaktighet for SDQ eller SDQ-algoritmen tar utgangspunkt i en antagelse om at beste referansestandard for diagnoser av psykiske lidelser er strukturerte diagnostiske intervjuer (eller klinikers vurdering av informasjon fra strukturerte diagnostiske intervjuer). Eksempler på slike strukturerte intervjuer som er mye brukt av norske klinikere, er Development and Well-Being Assessment (DAWBA) (Goodman, Ford, Richards, Gatward & Meltzer, 2000) og Schedule for Affective Disorders and Schizophrenia for School Aged Children – Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997). DAWBA er en samling av utredningsinstrumenter i som består av et diagnostisk foreldreintervju, et diagnostisk intervju med ungdommer (11-17 år) og et spørreskjema for lærere. DAWBA er utviklet for å fange opp vanlige barne- og ungdomsdiagnoser i henhold til ICD-10 og DSM-IV (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). K-SADS-PL er et semi-strukturert diagnostisk intervju for kartlegging av nåværende og tidligere psykopatologi hos barn og unge mellom 6 og 18 år.

Basert på klinisk intervju av barn og foreldre, samt informasjon fra andre kilder (som klinisk observasjon), genereres barnepsykiatriske diagnoser i henhold til DSM-IV. Måleegenskapene ved de norske versjonene av DAWBA og K-SADS-PL er undersøkt i hver sin PsykTestBarn-artikkel (Brøndbo & Martinussen, 2013; Kornør & Skarphedinsson, 2016). Datagenererte DAWBA-diagnoser hadde et visst samsvar (gjennomsnittlig $\kappa = 0,47$) med klinikergenererte diagnoser i én studie, mens andre studier viste at DAWBA underdiagnostiserte autisme og ADHD (Brøndbo & Martinussen, 2013). Kunnskapen om måleegenskapene ved den norske utgaven av K-SADS-PL begrenser seg til inter-raterreliabilitet og usikre data om begrepsvaliditet (Kornør & Skarphedinsson, 2016).

Ikke alle studier som inneholder informasjon om den diagnostiske nøyaktigheten til SDQ, bruker klinisk vurdering ved diagnostiske intervjuer som referansestandard. Aktuell referansestandard kan variere i studier etter hva som er formålet med den enkelte studien og studiens måldiagnoser. I studier hvor måldiagnosen er begrenset til en eller få diagnoser kan det for eksempel være tatt i bruk referansestandard bestående av diagnostiske tester eller kartleggingsverktøy rettet mot spesifikke diagnoser.

Metode

Vi søkte etter dokumentasjon på testens psykometriske egenskaper i databasene Norart, BIBSYS, SveMed+, CRISTin.no, Cochrane Library, NORA, PsycINFO, Medline, Embase, Libris og PubMed AHO. Søkedato: 23.01.2015. Søkestrategien er tilgjengelig på <http://www.psyktestbarn.no/CMS/ptb.nsf/pages/sdq-a>

Vi inkluderte alle publikasjoner av studier som har undersøkt og rapportert diagnostisk nøyaktighet for SDQ-algoritmen i skandinaviske utvalg, uavhengig av type referansestandard. I tillegg inkluderte vi publikasjoner som rapporterte forekomster basert på SDQ-algoritmen, for henholdsvis generelle populasjoner og kliniske undergrupper.

To forskere gikk gjennom sammendragene til alle identifiserte publikasjoner, uavhengig av hverandre, etter at dubletter var fjernet. Alle publikasjoner som kunne virke relevante ble bestilt inn i fulltekst, og prosessen over ble gjentatt for fulltekstrapportene.

To forskere vurderte diagnostisk nøyaktighet uavhengig av hverandre, ved hjelp av en tilpasset versjon av Test review form and notes for reviewers (European Federation of Psychologist Association (EFPA), 2008).

Resultater

Litteratursøk

Søket ga 637 treff, og 541 var igjen etter fjerning av duplikater. Vi vurderte 133 publikasjoner som potensielt relevante etter vurdering på tittel- og abstraktnivå. Disse 133 vurderte vi videre i fulltekst, hvorav 122 ble ekskludert (Vedlegg 1). De fleste ble ekskludert fordi de ikke hadde brukt algoritmen. Totalt inkluderte vi 11 publikasjoner, hvorav ti norske og én dansk. Ingen svenske publikasjoner ble inkludert.

Inkluderte studier

De ti norske inkluderte publikasjonene rapporterte resultater fra til sammen seks studier (Aasheim, 2012; Brøndbo et al., 2011; Ekornås, Heimann, Tjus, Heyerdahl & Lundervold, 2011; Hanssen-Bauer, Heyerdahl & Eriksson, 2007; Hysing, Elgen, Gillberg, Lie & Lundervold, 2007; Hysing, Elgen, Gillberg & Lundervold, 2009; Lehmann, Heiervang, Havik & Havik, 2014; Oellingrath, Svendsen & Hestetun, 2014; Posserud et al., 2014; Ullebø, Posserud, Heiervang, Gillberg & Obel, 2011) (Tabell 1). Tre av de inkluderte studiene rapporterte både forekomster og diagnostisk nøyaktighet, og tre rapporterte kun forekomster.

Barn i Bergen

Fem av de inkluderte artiklene rapporterte algoritmedata fra delstudier og underutvalg basert på den befolkningsbaserte kohortundersøkelsen Barn i Bergen (BiB) (Ekornås et al., 2011; Hysing et al., 2007; Hysing et al., 2009; Posserud et al., 2014; Ullebø et al., 2011). Datainnsamlingen til BiB gikk over fire runder i perioden 2002-2012 (<http://uni.no/nb/uni-helse/rkbu-vest/barn-i-bergen-flytdiagram/>). Den første runden hadde samtlige 9430 2.-4.-klassinger (7-9-åringer) i

Bergen som målgruppe og fant sted skoleåret 2002/2003. Runde 2 fant sted fire år senere, da barna var 11-13 år, og omfattet 7007 barn med foreldresamtykke for nye undersøkelser fra runde 1. I runde 3 og 4 var barna henholdsvis 14- 16 og 16-18 år.

Tre av de inkludert BiB-delstudiene rapporterte både forekomster og data som kunne brukes til å beregne diagnostisk nøyaktighet (Hysing et al., 2007; Posserud et al., 2014; Ullebø et al., 2011), mens de to øvrige kun rapporterte forekomster basert på SDQ-algoritmen (Ekornås et al., 2011; Hysing et al., 2009). Foreldre- og lærerrapporterte SDQ-data inngikk i algoritmen i samtlige delstudier.

Ullebø og medarbeidere brukte et utvalg fra runde 1 ($N = 6233$) med *både* foreldre- og lærerrapporterte SDQ-data og data fra Swanson, Nolan and Pelham Questionnaire-IV (SNAP-IV) (Swanson et al., 2001). Forskerne undersøkte SDQ-algorithmens evne til å identifisere barn med fenotyp ADHD-diagnose ved både *probable*- og *possible*-terskelen. Referansestandard for fenotyp ADHD-diagnose var gitt ved høye SNAP-IV-skårer. Algoritmebaserte forekomster av hyperaktivitet- og oppmerksomhetsforstyrrelser ble ikke oppgitt i artikkelen, men kunne beregnes på grunnlag av sensitivitet og N .

Posserud og medarbeideres (2014) utvalg besto av 234 barn som hadde komplett datasett fra *både* foreldre- og lærerrapporterte SDQ, SNAP-IV, DAWBA og K-SADS-PL. Forskerne definerte 68 av barna som barn med en diagnose i henhold til DAWBA, og de resterende 186 som kontroller uten noen DAWBA-basert diagnose. I denne delstudien undersøkte forskerne SDQ-algorithmens evne til å identifisere barn med ADHD-diagnose ved *probable*-terskelen. ADHD-diagnose i henhold til K-SADS-PL var referansestandard.

I delstudien til Hysing og medarbeidere (2007) var utvalget barn i alderen 7-9 år med kronisk sykdom ($N = 109$). Forskerne undersøkte SDQ-algorithmens evne til å identifisere barn med *noen* psykisk lidelse ved både *probable*- og *possible*-terskelen, med DAWBA som referansestandard. Hysing og medarbeidere oppga også forekomster av *noen* psykisk lidelse for henholdsvis utvalget med

kronisk syke barn og for barn uten kronisk lidelse ($N = 931$).

De to delstudiene som kun rapporterte forekomster hadde relativt store utvalg, men fra hver sin runde av BiB (Ekornås et al., 2011; Hysing et al., 2009). Utvalget til Hysing og medarbeidere (2009) bestod av 6913 barn med *både* foreldre- og lærerrapporterte SDQ-data og data om kronisk sykdom, fra runde 1. Forskerne sammenliknet forekomster hos barn med kronisk sykdom ($N = 444$) og barn uten kronisk sykdom ($N = 6469$). Utvalget til Ekornås og medarbeidere (2011) omfattet barn med *både* foreldre- og lærerrapporterte SDQ-data og data fra The Self-Perception Profile for Children Revised (SPPC-rev) (Harter, 1985), fra runde 2 ($N = 5073$). Hysing og medarbeidere rapporterte forekomster ved både *probable*- og *possible*-terskelen, mens Ekornås og kolleger kun brukte *possible*-terskelen.

BUP-Nord

Brøndbo og medarbeidere (Brøndbo et al., 2011) inkluderte barn i alderen 5-18 år som var henvist til barne- og ungdomspsykiatrisk poliklinikk i Tromsø eller Alta ($N = 286$). Forskerne undersøkte SDQ-algorithmens evne til å identifisere barn med *noen* psykisk lidelse, ADHD, emosjonell lidelse og atferdsforstyrrelse. Både foreldre-, lærer- og selvrapporterte SDQ-data inngikk i algoritmen. Forskerne brukte både *probable*- og *possible*-terskelen, og DAWBA var referansestandard.

Lehmann 2014

Utvalget til Lehmann og medarbeidere (Lehmann et al., 2014) besto av fosterbarn i alderen 6-12 år fra Buskerud, Vestfold, Telemark, Øst-Agder og Vest-Agder ($N=141$). Forskerne undersøkte SDQ-algorithmens evne til å identifisere barn med *noen* psykisk lidelse, ADHD, emosjonell lidelse og atferdsforstyrrelse. Foreldre- og lærerrapporterte SDQ-data inngikk i algoritmen, og forskerne brukte både *probable*- og *possible*-terskelen. Også her var DAWBA referansestandard.

Oellingrath 2013

I den befolkningsbaserte studien til Oellingrath, Svendsen og Hestetun (Oellingrath et al., 2014) besto utvalget av 1095 12-13-åringene fra Telemark. Forskerne rapporterte algoritmebaserte forekomster av henholdsvis *noen* psykisk lidelse,

emosjonelle lidelser, atferdsforstyrrelser og ADHD. Terskelen var *possible*, og det var kun foreldrerapporterte SDQ-data som inngikk i algoritmen.

Mindre studier

De to øvrige studiene rapporterte kun forekomster av *noen* psykisk lidelse, ADHD, emosjonelle lidelser og atferdsforstyrrelser fra mindre, selekterte utvalg. Det ene utvalget besto av 45 barn (4-17 år) i en barnevernsinstitusjon i en kommune på Vestlandet (Aasheim, 2012), og det andre utvalget

besto av 73 barn i alderen 9-15 år med epilepsi (Hanssen-Bauer et al., 2007). Hanssen-Bauer, Heyerdahl og Eriksson brukte både foreldre-, lærer- og selvrapporterte SDQ-data i algoritmen, mens Aasheim brukte foreldre- og lærerrapporterte data.

Den danske versjonen av SDQ

Den danske studien hadde et befolkningsbasert utvalg fra København på 2315 barn i alderen 5-7 år (Rimvall et al., 2014). Forskerne undersøkte SDQ-algoritmens evne til å identifisere barn med ADHD-

Tabell 1. Inkluderte delstudier og studier

Referanse	N	Alder	Utvalg	SDQ-versjoner	SDQ-indeks	Terskel (Poss./prob.)	Mål-diagnoser	Referans-standar
Ullebø 2011 ¹	6233	7-9	Befolkningsbasert, Bergen	P+T	ADHD	Begge	ADHD	SNAP-IV
Posserud 2014 ¹	234	7-9	Barn med diagnose + kontroller, Bergen	P+T	ADHD	Prob	ADHD	K-SADS-F
Hysing 2007 ¹	109	7-9	Barn med kronisk sykdom, Bergen	P+T	NOEN ADHD EMO ATF	Prob	NOEN - - -	DAWBA - - -
Hysing 2009 ¹	6233	7-9	Befolkningsbasert, Bergen	P+T	NOEN	Begge	-	-
Ekornås 2011 ¹	5073	11-13	Befolkningsbasert, Bergen	P+T	NOEN	Begge	-	-
Brøndbo 2011 ²	286	5-18	Barn henvist til BUP, Nord-Norge	P+T+S	NOEN ADHD EMO ATF	Begge	NOEN ADHD EMO ATF	DAWBA
Lehmann 2014	141	6-12	Fosterbarn	P+T	NOEN ADHD EMO ATF	Begge	NOEN ADHD EMO ATF	DAWBA
Oellingrath 2014	789	12-13	Befolkningsbasert, Telemark	P	NOEN ADHD EMO ATF	Poss	- - - -	- - - -
Aasheim 2012	45	4-17	Barnevernsbarn, Vestlandet	P+T	NOEN ADHD EMO ATF	Poss	-	-
Hanssen-Bauer 2007	73	9-15	Barn med epilepsi	PTS	NOEN ADHD EMO ATF	Begge	-	-
Rimvall 2014 ³	2315	5-7	Befolkningsbasert, København	P+T	NOEN ADHD EMO ATF	Begge	- ADHD - -	- Register - -

¹ Delstudie fra Barn i Bergen; ² Delstudie fra BUP-Nord; ³ Dansk SDQ

P=foreldreversjon; T=lærerversjon; S=selvrapport; NOEN=noen psykisk lidelse; EMO=Emosjonell lidelse;

ATF=Atferdsforstyrrelse; Poss=*possible*; Prob=*probable*; SNAP-IV=Swanson, Nolan og Pelhams spørreskjema for lærere og foreldre - versjon IV; Kiddie-SADS-PL=Schedule for Affective Disorders and Schizophrenia for School Aged Children - Present and Lifetime Version; DAWBA=Development and Well-Being Assessment

Tabell 2a. Forekomster av algoritmebaserte diagnosekategorier i befolkningsbaserte utvalg. Terskel: *possible* hvis ikke annet er oppgitt

Referanse	N	Alder	NOEN	ADHD	EMO	ATF
Ullebø 2011 ¹	6233	7-9		5 %		
Hysing 2009 ¹	6469	7-9	14 %			
Ekornås 2011 ¹	5073	11-13	14 %			
Oellingrath 2014	789	12-13	9 %	6 %	4 %	2 %
Rimvall 2014 ²	2315	5-7	21 %	4 %	6 %	17 %

¹ Delstudie fra Barn i Bergen; ² Dansk SDQ; NOEN=*noen* psykisk lidelse; EMO = Emosjonell lidelse; ATF = Atferdsforstyrrelse.

diagnose ved både *probable*- og *possible*-terskelen. Referansestandard var gitt ved registerdata. Kriteriet for å bli klassifisert som barn med ADHD-diagnose var at barnet var registrert med en ADHD-diagnose og/eller med en resept på et sentralstimulerende legemiddel i en 7-årsperiode etter utfylling av SDQ.

Forekomster av algoritmebaserte diagnoser

I befolkningsbaserte utvalg varierte forekomsten av *noen* diagnose (terskel: *possible*) fra ni prosent blant 12-13-åringer i Telemark til 21 % blant 5-7-åringer i København (Tabell 2a). Den høye forekomsten i København ser ut til å kunne forklares av en høy forekomst av atferdsforstyrrelser. Forekomsten av henholdsvis ADHD og emosjonelle lidelser (terskel: *possible*) i befolkningsbaserte utvalg varierte fra 4 % til 6 %, og forekomsten av atferdsforstyrrelser varierte fra 2 % til 17 %.

Forekomsten av *noen* diagnose (terskel: *probable*) varierte fra 5 % blant barn med «annen kronisk sykdom» til 73 % blant barn i psykisk helsevern (Tabell 2b). Blant fosterbarn og barn med epilepsi var forekomsten av *noen* psykisk lidelse 45 % - 47 %. Forekomsten av ADHD varierte fra 18 % blant barn med epilepsi til 30 % blant barn i BUP. Forekomsten av emosjonelle lidelser varierte fra 16 % blant fosterbarn til 40 % blant barn i psykisk helsevern. Forekomsten av atferdsforstyrrelser varierte fra 12 % blant fosterbarn til 25 % blant barn i psykisk helsevern.

Kriterievaliditet: diagnostisk nøyaktighet

Diagnostiske instrumenter har forskjellige egenskaper i selekterte og befolkningsbaserte utvalg. Vi har derfor valgt å skille mellom de to typene utvalg i vår gjengivelse av data for SDQ-algoritmens diagnostiske nøyaktighet.

Selekterte utvalg

De selekterte utvalgene bestod av barn med

Tabell 2b. Forekomster av algoritmebaserte diagnosekategorier i selekterte utvalg. Terskel: *probable* hvis ikke annet er oppgitt

Referanse	N	Alder	Utvalg	NOEN	ADHD	EMO	ATF
Posserud 2014 ¹	234	7-9	Barn med DAWBA-diagnose + kontroller		21 %		
Hysing 2007 ¹	109	7-9	Barn med kronisk sykdom	11 %			
Hysing 2009 ¹	86	7-9	Barn med neurologisk sykdom	27 %			
	284		Barn med astma	9 %			
	74		Barn med annen kronisk sykdom	5 %			
Brøndbo 2011 ²	286	5-18	Barn henvist til BUP	73 %	30 %	40 %	25 %
Lehmann 2014 ³	141	6-12	Fosterbarn	47 %	20 %	16 %	12 %
Aasheim 2012	45	4-17	Barnevernsbarn	44 % ⁵	24 % ⁵	31 % ⁴	22 % ⁴
Hanssen-Bauer 2007	73	9-15	Barn med epilepsi	45 %	18 %	21 %	26 %

¹ Delstudie fra Barn i Bergen; ² Delstudie fra BUP-Nord; ³ Forekomster beregnet utfra data i artikkelen; ⁴ Terskel: *possible*; NOEN = *noen* psykisk lidelse; EMO = Emosjonell lidelse; ATF=Atferdsforstyrrelse.

Tabell 3: SDQ-algorithmens diagnostiske nøyaktighet for *noen* psykisk lidelse i selekterte utvalg

Referanse	Sensitivitet		Spesifisitet		PPV		NPV		Forekomst
	Prob	Pos	Prob	Pos	Prob	Pos	Prob	Pos	
Brøndbo 2011	0,85	0,96	0,52	0,24	0,77	0,71	0,65	0,74	66 %
Lehmann 2014	0,68	0,84	0,82	0,58	0,83	0,73	0,65	0,73	57 %
Hysing 2007	0,77	0,95	0,85	0,62	0,57	0,39	0,94	0,98	20 %

PPV = positiv prediktiv verdi; NPV = negativ prediktiv verdi; Forekomst basert på referansestandard.

kronisk sykdom (Hysing et al., 2007), barn i psykisk helsevern (Brøndbo et al., 2011) og fosterbarn (Lehmann et al., 2014). Vi gjengir her SDQ-algorithmens evne til å identifisere barn med henholdsvis *noen* psykisk lidelse, ADHD, atferdsforstyrrelser og emosjonelle lidelser. Vi rapporterer kun diagnostisk nøyaktighet for SDQ-indeks og diagnosekategorier som korresponderer direkte. Det vil si at den rapporterte diagnostiske nøyaktigheten for *noen* psykisk lidelse er basert på SDQ-indeks *noen* psykisk lidelse, for ADHD er den basert på SDQ-indeks hyperkinetiske- og oppmerksomhetsforstyrrelser, og så videre.

Noen psykisk lidelse

Ved terskel *probable* i selekterte utvalg identifiserte SDQ-algoritmen 68 - 85 av 100 barn med *noen* psykisk lidelse, og 52 - 85 av 100 barn uten *noen* psykisk lidelse (Tabell 3). Andel testpositive barn som hadde *noen* psykisk lidelse i henhold til referansestandard (positiv prediktiv verdi - PPV) varierte fra 55 % til 83 % av, og 65 % - 94 % av testnegative barn hadde ikke *noen* psykisk lidelse (negativ prediktiv verdi - NPV).

Ved terskel *possible* i selekterte utvalg identifiserte

SDQ-algoritmen 84 - 96 av 100 barn med *noen* psykisk lidelse, og 24 - 62 av 100 barn uten *noen* psykisk lidelse (Tabell 3). Andel testpositive barn som hadde *noen* psykisk lidelse i henhold til referansestandard (PPV) varierte fra 39 % til 73 %, og 73 % - 98 % av testnegative barn hadde ikke *noen* psykisk lidelse (NPV).

ADHD

Ved terskel *probable* i selekterte utvalg identifiserte SDQ-algoritmen 68 - 77 av 100 barn med ADHD, og 80 - 88 av 100 barn uten ADHD (Tabell 4). Andel testpositive barn med en ADHD-diagnose i henhold til referansestandard (PPV) varierte fra 46 % til 51 %, og 92 % - 94 % av testnegative barn hadde ikke en ADHD-diagnose (NPV).

Ved terskel *possible* i selekterte utvalg identifiserte SDQ-algoritmen 82 - 100 av 100 barn med ADHD, og 45 - 55 av 100 barn uten ADHD (Tabell 4). Andel testpositive barn med ADHD-diagnose i henhold til referansestandard (PPV) varierte fra 28 % til 31 %, og 92 % - 100 % av testnegative barn hadde ikke en ADHD-diagnose (NPV).

Tabell 4: SDQ-algorithmens diagnostiske nøyaktighet for ADHD i selekterte utvalg

Referanse	Sensitivitet		Spesifisitet		PPV		NPV		Forekomst
	Prob	Pos	Prob	Pos	Prob	Pos	Prob	Pos	
Brøndbo 2011	0,77	1,00	0,80	0,45	0,46	0,28	0,94	1,00	18 %
Lehmann 2014	0,71	0,82	0,80	0,55	0,47	0,31	0,92	0,92	25 %
Posserud 2014	0,68		0,88		0,51		0,94		16 %

PPV=positiv prediktiv verdi; NPV=negativ prediktiv verdi; Forekomst basert på referansestandard.

Tabell 5: SDQ-algorithmens diagnostiske nøyaktighet for atferdsforstyrrelser i selekterte utvalg

Referanse	Sensitivitet		Spesifisitet		PPV		NPV		Forekomst
	Prob	Pos	Prob	Pos	Prob	Pos	Prob	Pos	
Brøndbo 2011	0,83	0,97	0,75	0,58	0,59	0,51	0,91	0,98	0,31
Lehmann 2014	0,73	1,00	0,84	0,64	0,55	0,43	0,92	1,00	0,21

PPV = positiv prediktiv verdi; NPV = negativ prediktiv verdi; Forekomst basert på referansestandard

Atferdsforstyrrelser

Ved terskel *probable* i selekterte utvalg identifiserte SDQ-algoritmen 73 - 83 av 100 barn med atferdsforstyrrelse, og 75 - 84 av 100 barn uten atferdsforstyrrelse (Tabell 5). Andel testpositive barn med atferdsforstyrrelse i henhold til referansestandard (PPV) varierte fra 55 % til 59, og 91 % - 92 % av testnegative hadde ikke en atferdsforstyrrelse (NPV).

Ved terskel *possible* i selekterte utvalg identifiserte SDQ-algoritmen 97 - 100 av 100 barn med atferdsforstyrrelse, og 58 - 64 av 100 barn uten atferdsforstyrrelse (Tabell 5). Andel testpositive barn med atferdsforstyrrelse i henhold til referansestandard (PPV) varierte fra 43 % til 51 %, og 98 % - 100 % av testnegative barn hadde ikke en atferdsforstyrrelse (NPV).

Emosjonelle lidelser

Ved terskel *probable* i selekterte utvalg identifiserte SDQ-algoritmen 30 - 47 av 100 barn med emosjonell lidelse, og 87 - 94 av 100 barn uten emosjonell lidelse (Tabell 6). Andel testpositive barn emosjonell lidelse i henhold til referansestandard (PPV) varierte fra 65 % til 66 %, og 76 % - 80 % av testnegative hadde ikke en emosjonell lidelse (NPV).

Ved terskel *possible* i selekterte utvalg identifiserte

Tabell 6: SDQ-algorithmens diagnostiske nøyaktighet for emosjonelle lidelser i selekterte utvalg

Referanse	Sensitivitet		Spesifisitet		PPV		NPV		Forekomst
	Prob	Pos	Prob	Pos	Prob	Pos	Prob	Pos	
Brøndbo 2011	0,47	0,81	0,87	0,55	0,66	0,48	0,76	0,84	34 %
Lehmann 2014	0,30	0,69	0,94	0,72	0,65	0,46	0,80	0,87	26 %

PPV=positiv prediktiv verdi; NPV=negativ prediktiv verdi; Forekomst basert på referansestandard

SDQ-algoritmen 69 - 81 av 100 barn med emosjonell lidelse, og 55 - 72 av 100 barn uten emosjonell lidelse (Tabell 6). Andel testpositive barn med emosjonell lidelse i henhold til referansestandard (PPV) varierte fra 46 % til 48 % av testpositive barn hadde en emosjonell lidelse i henhold til referansestandard, og 84 % - 87 % av testnegative barn hadde ikke en emosjonell lidelse (NPV).

Befolkningsbaserte utvalg

Diagnostiske nøyaktighetsdata i de to befolkningsbaserte studiene som rapporterte slike tall, var begrenset til SDQ-algorithmens evne til å identifisere barn med ADHD på grunnlag av SDQ-indeks hyperkinetiske- og oppmerksomhetsforstyrrelser.

Ved terskel *probable* i selekterte utvalg identifiserte SDQ-algoritmen 16 - 23 av 100 barn med ADHD, og samtlige barn uten ADHD (Tabell 7). 55 % - 93 % av testpositive barn hadde en ADHD-diagnose i henhold til referansestandard, og 96 % - 98 % av testnegative barn hadde ikke en ADHD-diagnose.

Ved terskel *possible* i selekterte utvalg identifiserte SDQ-algoritmen 46 - 52 av 100 barn med ADHD, og 97 - 98 av 100 barn uten ADHD (Tabell 7). 33 % - 59 % av testpositive barn hadde en ADHD-diagnose i

Tabell 7: SDQ-algoritmens evne til å predikere hyperkinetiske- og oppmerksomhetsforstyrrelse på grunnlag av positivt utfall på HYP befolkningsbaserte utvalg (norsk og dansk)

Referanse	Sensitivitet		Spesifisitet		PPV		NPV		Forekomst
	Prob	Pos	Prob	Pos	Prob	Pos	Prob	Pos	
Rimvall 2014 ¹	0,16	0,46	1,00	0,97	0,55	0,33	0,98	0,98	3 %
Ullebø 2011 ²	0,23	0,52	1,00	0,98	0,93	0,59	0,96	0,97	5 %

¹Dansk SDQ/utvalg; ²NPV estimert på grunnlag av oppgitte data. PPV = positiv prediktiv verdi; NPV = negativ prediktiv verdi; Forekomst basert på referansestandard

henhold til referansestandard, og 97 % - 98 % av testnegative barn hadde ikke en ADHD-diagnose.

Diskusjon og konklusjon

I selekterte utvalg hadde SDQ-algoritmen i de fleste studiene sensitivitet på over 70 % for *noen* psykisk lidelse, ADHD og atferdsforstyrrelser ved terskel *probable*. Dette betyr at SDQ-algoritmen fanger opp de fleste barn med slike diagnoser. Sensitiviteten for emosjonelle lidelser var bare 30 % og 47 % ved terskel *probable*, som betyr at de fleste barn med denne typen diagnoser *ikke* ville bli fanget opp av SDQ-algoritmen. Spesifisiteten lå godt over 70 % for alle diagnosekategorier i de fleste studiene. Dette betyr at SDQ-algoritmen i de fleste tilfeller korrekt identifiserer barn som *ikke* har *noen* diagnose. Den positive prediktive verdien varierte både mellom studier og diagnosegrupper, med høyest verdier for *noen* psykisk lidelse og lavest verdier for ADHD. Mindre enn halvparten av barna i de selekterte utvalgene som ifølge SDQ-algoritmen hadde en hyperkinetisk-/oppmerksomhetsforstyrrelse, hadde faktisk en ADHD-diagnose. Den negative prediktive verdien var høy, rundt 90 %, i de fleste studier og for de fleste diagnoser. Dette betyr at SDQ-algoritmen fører til få falske negative. Vi har her valgt å legge vekt på den høyeste terskelverdien for SDQ-indeksene, *probable*, fordi de selekterte utvalgene har høy pre-test sannsynlighet for diagnose (forekomsten av psykiske lidelser er høyere enn i den generelle befolkningen). Med *possible*-terskelen var sensitiviteten jevnt over høyere, og spesifisiteten lavere.

I befolkningsbaserte utvalg med lav pre-test sannsynlighet for diagnose har vi valgt å framheve

den laveste terskelverdien for SDQ-indeksene, *possible*. Vi fant kun data for ADHD-diagnose i de inkluderte studiene. SDQ-algoritmen hadde lav sensitivitet og positiv prediktiv verdi (rundt 50 %) og tilnærmet perfekt spesifisitet og negativ prediktiv verdi. Dette betyr at SDQ-algoritmen har en svært god evne til å utelukke ADHD i befolkningsbaserte utvalg, men en dårlig evne til å identifisere barn som har ADHD. *Probable*-terskelen ga betydelig lavere sensitivitet og 100 % spesifisitet.

Referansestandard er et viktig metodisk element i studier av diagnostisk nøyaktighet. I denne sammenhengen var DAWBA referansestandard i tre studier, K-SADS-PL i én studie, SNAP-IV i én studie og registerdata i én studie. Måleegenskapene ved DAWBA, K-SADS-PL og SNAP-IV er undersøkt i hver sin PsyktTestBarn-artikkel (Brøndbo & Martinussen, 2013; Kornør & Bøe, 2011; Kornør & Skarphedinnsson, 2016). Basert på hva vi vet om den diagnostiske nøyaktigheten for DAWBA er det grunn til å stille spørsmålsteget ved hvor egnet dette instrumentet er som referansestandard. Vår begrensede kunnskap om den diagnostiske nøyaktigheten ved K-SADS-PL-baserte diagnoser gjør at psykomestriestudier med K-SADS-PL som referansestandard også må tolkes med varsomhet. Måleegenskapene ved SNAP-IV er overhodet ikke dokumentert for den norske versjonen (Kornør & Bøe, 2011). SNAP-IV bør derfor ikke brukes som referansestandard. Registerdataene i den danske studiene er også problematiske som referansestandard fordi vi ikke vet hvor gode de er. For å oppsummere har vi mangelfull kunnskap om disse instrumentene i en norsk sammenheng.

Den største styrken ved SDQ-algoritmen er at den omfatter data fra to eller flere informanter. Barn oppfattes forskjellig av forskjellige personer rundt dem, og de viser ofte forskjellige sider av seg i ulike sammenhenger. Ved å innhente informasjon fra så mange kilder som mulig får vi et mer helhetlig bilde av barnets symptomer. SDQ-algoritmen er imidlertid basert på britiske normer, som avviker noe fra norske gjennomsnittsskårer (Kornør & Heyerdahl, 2013, 2014). Det er vanskelig å uttale seg om hvilke konsekvenser det får for den diagnostiske nøyaktigheten ved SDQ-algoritmen.

Med forbehold om usikkerheten knyttet til referansestandarder som er benyttet i studiene og normgrunnlaget, har SDQ-algoritmen god evne til å identifisere barn med og uten *noen* psykisk lidelse, ADHD og atferdsforstyrrelse i selekterte utvalg. Algoritmen har dårlig evne til å identifisere barn med emosjonelle forstyrrelser. I befolkningsbaserte utvalg kan SDQ-algoritmen med stor sannsynlighet korrekt utelukke ADHD, mens den har dårlig evne til å påvise ADHD. Vi vet ikke hvordan algoritmen fungerer på andre diagnosekategorier i befolkningsbaserte utvalg. For å få bedre kunnskap om SDQ-algorithmens diagnostiske nøyaktighet trenger vi studier som bruker en referansestandard med gode måleegenskaper.

Referanser

- Aasheim, R. (2012). *Psykisk helse blant barnevernsbarn: en studie av psykiske vansker målt med kartleggingsverktøyet Strengths and Difficulties Questionnaire*. [Bergen]: HEMIL-senteret, Det psykologiske fakultet, [Universitetet i Bergen].
- Brøndbo, P. H. & Martinussen, M. (2013). Måleegenskaper ved den norske versjonen av Development and Well-Being Assessment (DAWBA). *PsykTestBARN*, 3(2).
- Brøndbo, P. H., Mathiassen, B., Martinussen, M., Heiervang, E., Eriksen, M., Moe, T. F. et al. (2011). The strengths and difficulties questionnaire as a screening instrument for norwegian child and adolescent mental health services, application of UK scoring algorithms. *Child and Adolescent Psychiatry and Mental Health*, 5(32). doi: <http://dx.doi.org/10.1186/1753-2000-5-32>
- Ekornås, B., Heimann, M., Tjus, T., Heyerdahl, S. & Lundervold, A. J. (2011). Primary school children's peer relationships: Discrepancies in self-perceived social acceptance in children with emotional or behavioral disorders. *Journal of Social and Clinical Psychology*, 30(6), 570-582. doi: <http://dx.doi.org/10.1521/jscp.2011.30.6.570>
- European Federation of Psychologist Association (EFPA). (2008). EFPA Review model for the description and evaluation of psychological tests: Test review form and notes for reviewers (3.42 utg.): EFPA.
- Ford, T., Hutchings, J., Bywater, T., Goodman, A. & Goodman, R. (2009). Strengths and Difficulties Questionnaire Added Value Scores: evaluating effectiveness in child mental health interventions. *The British Journal of Psychiatry*, 194(6), 552-558. doi: 10.1192/bjp.bp.108.052373
- Goodman, A., Lamping, D. L. & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179-1191. doi: 10.1007/s10802-010-9434-x
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-586. doi: 10.1111/j.1469-7610.1997.tb01545.x
- Goodman, R., Ford, T., Richards, H., Gatward, R. & Meltzer, H. (2000). The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 41(5), 645-655.
- Goodman, R., Renfrew, D. & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, 9(2), 129-134.
- Hanssen-Bauer, K., Heyerdahl, S. & Eriksson, A. S. (2007). Mental health problems in children and adolescents referred to a national epilepsy center. *Epilepsy and Behavior*, 10(2), 255-262. doi: <http://dx.doi.org/10.1016/j.yebeh.2006.11.011>
- Harter, S. (1985). *Manual for the self-perceptin profile for children*. Denver: University of Denver.
- Hysing, M., Elgen, I., Gillberg, C., Lie, S. A. & Lundervold, A. J. (2007). Chronic physical illness and mental health in children. Results from a large-scale population study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48(8), 785-792. doi: <http://dx.doi.org/10.1111/j.1469-7610.2007.01755.x>
- Hysing, M., Elgen, I., Gillberg, C. & Lundervold, A. J. (2009). Emotional and behavioural problems in subgroups of children with chronic illness: Results from a large-scale population study. *Child: Care, Health and Development*, 35(4), 527-533. doi: <http://dx.doi.org/10.1111/j.1365-2214.2009.00967.x>
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P. et al. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(7), 980-988. doi: 10.1097/00004583-199707000-00021
- Kornør, H. & Bøe, T. (2011). Måleegenskaper ved den norske versjonen av Swanson, Nolan og Pelham vurderingsskala for lærere og foreldre - versjon IV (SNAP-IV). *PsykTestBarn*(1), 4.

- Kornør, H. & Heyerdahl, S. (2013). Måleegenskaper ved den norske versjonen av Strengths and Difficulties Questionnaire, selvrappport (SDQ-S). *PsyktTestBARN*(2), 13.
- Kornør, H. & Heyerdahl, S. (2014). Måleegenskaper ved den norske versjonen av Strengths and Difficulties Questionnaire, lærerrappport (SDQ-T). *PsyktTestBARN*(2), 14.
- Kornør, H. & Skarphedinsson, G. Á. (2016). Måleegenskaper ved den norske versjonen av Schedule for Affective Disorders and Schizophrenia for School-Aged Children: Present and Lifetime Version (6-18) (Kiddie-SADS-PL). *PsyktTestBARN*(1), 12.
- Lehmann, S., Heiervang, E. R., Havik, T. & Havik, O. E. (2014). Screening foster children for mental disorders: Properties of the strengths and difficulties questionnaire. *PLoS ONE*, 9(7). doi: <http://dx.doi.org/10.1371/journal.pone.0102134>
- Oellingrath, I. M., Svendsen, M. V. & Hestetun, I. (2014). Eating patterns and mental health problems in early adolescence--a cross-sectional study of 12-13-year-old Norwegian schoolchildren. *Public Health Nutrition*, 17(11), 2554-2562. doi: <http://dx.doi.org/10.1017/S1368980013002747>
- Posserud, M. B., Ullebø, A. K., Plessen, K. J., Stormark, K. M., Gillberg, C. & Lundervold, A. J. (2014). Influence of assessment instrument on ADHD diagnosis. *European Child and Adolescent Psychiatry*, 23(4), 197-205. doi: <http://dx.doi.org/10.1007/s00787-013-0442-6>
- Rimvall, M. K., Elberling, H., Rask, C. U., Helenius, D., Skovgaard, A. M. & Jeppesen, P. (2014). Predicting ADHD in school age when using the Strengths and Difficulties Questionnaire in preschool age: A longitudinal general population study, CCC2000. *European Child & Adolescent Psychiatry*, 23(11), 1051-1060. doi: <http://dx.doi.org/10.1007/s00787-014-0546-7>
- Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B. et al. (2001). Clinical Relevance of the Primary Findings of the MTA: Success Rates Based on Severity of ADHD and ODD Symptoms at the End of Treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(2), 168-179. doi: 10.1097/00004583-200102000-00011
- Ullebø, A. K., Posserud, M. B., Heiervang, E., Gillberg, C. & Obel, C. (2011). Screening for the attention deficit hyperactivity disorder phenotype using the strength and difficulties questionnaire. *European Child and Adolescent Psychiatry*, 20(9), 451-458. doi: <http://dx.doi.org/10.1007/s00787-011-0198-9>