

Måleegenskaper ved den norske versjonen av Strengths and Difficulties Questionnaire, foreldrerapport (SDQ-P)

Hege Kornør¹ (hege.kornor@r-bup.no)

Sonja Heyerdahl¹ (sonja.heyerdahl@r-bup.no)

PsykTestBARN 2017, 1:1

ISSN 1893-9910

Mottatt dato 18. november 2016

Publisert dato 09. februar 2017

Artikkelens URL <https://psyktestbarn.r-bup.no/no/artikler/sdq-p-strengths-and-difficulties-questionnaire-foreldrerapport>

DOI 10.21337/0048

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.

Denne artikkelen skal siteres på følgende måte:

Kornør, H. & Heyerdahl, S. (2017). Måleegenskaper ved den norske versjonen av Strengths and Difficulties Questionnaire, foreldrerapport (SDQ-P). *PsykTestBarn*, 1:1. doi: 10.21337/0048

¹ RBUP Øst og Sør

Sammendrag

Beskrivelse. Strengths and Difficulties Questionnaire (SDQ) er et sett spørreskjema for utfylling av foreldre, lærere og barn ≥ 11 år. SDQ ble opprinnelig utgitt på engelsk av Robert Goodman i 1997. I 1999 kom de norske versjonene, basert på en oversettelse og tilbakeoversettelse ved Einar Heiervang og kolleger. Robert Goodman eier rettighetene til testen. Foreldreversjonen (SDQ-P) har 25 ledd fordelt på fem delskalaer. Det er et skjema som skal fylles ut av foreldre for å kartlegge psykiske helse, vennerelasjoner og prososial atferd hos barn i alderen 4-17 år. Det tar noen få minutter å fylle ut. Forskere, klinikere og pedagoger kan administrere og tolke SDQ-P.

Litteratursøk. Våre systematiske søk etter dokumentasjon for måleegenskapene ved den norske versjonen av SDQ-P førte til 408 unike treff, hvorav 66 publikasjoner fra 34 studier ble inkludert. I alt 35 av publikasjonene rapporterte fra store befolkningsstudier i Akershus, Bergen, Trondheim, Bodø, Østfold, Romsdal, Oslo og andre deler av landet.

Psykometri. De store befolkningsstudiene bidro med regionale normdata for barn og ungdom i alderen 4-19 år. I tre av studiene undersøkte forskerne faktorstrukturen til SDQ-P, og fire studier og to delstudier rapporterte verdier for diagnostisk nøyaktighet. Gruppesammenlikninger mellom selekterte utvalg og referansegrupper belyste også validiteten ved instrumentet, spesielt resultater for ulike diagnostiske grupper. Tre studier undersøkte også samsvaret mellom SDQ-P-skårer og skårer på andre, tilsvarende instrumenter. Vi fant mål på indre konsistens (Cronbachs alfa) i 21 av de inkluderte artiklene.

Konklusjon. SDQ-P har regionale normdata, men nasjonale data mangler. Forventede gruppeforskjeller og konfirmatoriske faktoranalyser støtter opp under instrumentets begrepsvaliditet. Tre av skalaene hadde utilfredsstillende indre konsistens. Evnen til å fange opp barn med en psykiatrisk diagnose synes å være ganske god. Likevel er ikke den diagnostiske nøyaktighet så god at SDQ-P kan anbefales som universelt screening-instrument. Den norske versjonen av SDQ-P kan være egnet til å screene for psykiatriske diagnoser i selekterte utvalg med risiko for psykiske helseproblemer.

Abstract

Description. Strengths and Difficulties Questionnaire (SDQ) is a set of parent-, teacher- and youth-reported questionnaires originally published in English by Robert Goodman in 1997. The Norwegian versions were published in 1999, based on a translation and back-translation by Einar Heiervang and colleagues. Robert Goodman holds the copyright. The parent version (SDQ-P) contains 25 items organised in five subscales. It is a parent report form to assess mental health, peer relations and prosocial behaviour in children aged 4-17 years. Completion takes a few minutes. Clinicians and teachers can administer and interpret SDQ-P scores.

Literature search. Our systematic searches for psychometric evidence for the Norwegian SDQ-P version identified 408 unique references, of which 66 publications from 34 studies were included. Thirtyfive publications were reports from large population-based studies in Akershus, Bergen, Trondheim, Bodø, Østfold, Romsdal, Oslo and other parts of the country.

Psychometrics. The large population-based studies contributed with regional norm data for children and adolescents aged 4 to 19. Researchers assessed the SDQ-P factor structure in three studies, and four studies and two substudies reported diagnostic accuracy data. Group comparisons between selected samples and reference groups also addressed the instrument's validity, especially data for various diagnostic groups. Three studies examined the agreement

between SDQ-P scores and scores from other, corresponding instruments. We identified measures of internal consistency (Cronbach's alpha) in 21 of the included publications.

Conclusions. The SDQ-P has regional norms, but there is a lack of national data. Expected group differences and confirmatory factor analyses support the construct validity of the instrument. Three scales have inadequate internal consistencies. The ability to detect children with a psychiatric diagnosis seems to be quite good. Still, the diagnostic accuracy is not sufficiently good to recommend the SDQ-P as a universal screening instrument. The Norwegian version of the SDQ-P may be appropriate for screening for psychiatric diagnoses in selected populations with high risk of mental health problems.

Innledning

Strengths and Difficulties Questionnaire (SDQ) (<http://www.sdqinfo.org/>) er utviklet av Robert Goodman (1997). SDQ ble oversatt til norsk av Einar Heiervang, Irene Elgen og Sonja Heyerdahl, og tilbakeoversatt av Melanie Young i 1999. På norsk heter instrumentet Sterke og svake sider, men vi vil i denne sammenhengen referere til det som SDQ.

SDQ er et kortfattet spørreskjema, laget for å kartlegge psykisk helse, vennerelasjoner og prososial atferd hos barn og unge. Det tar noen få minutter å fylle ut. Hovedskjemaet har 25 ledd (5 delskalaer med 5 ledd i hver). Det er fire problemskalaer: Atferdsproblemer, Hyperaktivitet – oppmerksomhetsproblemer, Emosjonelle symptomer og Venneproblemer. I tillegg er det en skala som kartlegger prososial atferd. Man angir om utsagnene i skjema stemmer for barnet (0-stemmer ikke, 1-stemmer delvis eller 2-stemmer helt). Skårene summeres for hver delskala, og de fire problemskalaene summeres til en skåre for totale vansker. Man kan også summere skala for emosjonelle symptomer og venneproblemer til en internaliserende skåre og skala for atferdsproblemer og hyperaktivitet-oppmærksomhetsproblemer til en eksternaliserende skåre (Goodman, Lamping & Ploubidis, 2010). Dette er særlig relevant i utvalg med relativt lite problemer. Når man undersøker barn med psykiske lidelser og/eller høye SDQ-skårer, kan det være en fordel å bruke de fem delskalaene (Goodman et al., 2010).

Omfang av psykiske vansker gir ikke alltid et godt bilde av plagenes alvorlighet og barnets funksjonsnivå. I et kort tilleggsskjema («impact supplement») kartlegges om barnet har vansker som man oppfatter som et problem, og hvilken

betydning disse har for barnet og familien. Man spør man om barnet har vansker med følelser, konsentrasjon, oppførsel eller med å komme overens med andre mennesker (Goodman, 1999). I så fall spørres om varighet av problemene og hvordan vanskene virker inn på barnets liv (fem spørsmål). Tre av spørsmålene som angir vanskenes innvirkning på barnets/ungdommens liv (0-ikke i det hele tatt, 0-bare litt, 1-en god del, 2-mye), kan summeres til en skåre for vanskenes innvirkning («Impact»).

Foredreversjonen (SDQ-P) var beregnet for barn/ungdom i alderen 4-16 år (Goodman, 1997). Aldersspennet er senere utvidet til 4-17 år (<http://www.sdqinfo.org/>). Det er også en versjon for barn i alderen 2-4 år. SDQ-P inngår i kartleggingspakken Development and Well-Being Assessment (DAWBA; Goodman, Ford, Richards, Gatward & Meltzer, 2000), som finnes både i papir- og elektronisk versjon. Spørreskjema foreligger også i lærerversjon (SDQ-T) for aldersgruppen 4-17 år, samt i selvrappportversjon (SDQ-S) for barn/ungdom 11-17 år (Goodman, 1999). Det foreligger også en egen versjon for oppfølging etter behandling. Denne har med to spørsmål om behandlingen (Ford, Hutchings, Bywater, Goodman & Goodman, 2009).

Oversettelser

SDQ er brukt i svært mange studier, både internasjonalt og i Norge. Pr 2015 er SDQ oversatt til 77 språk. Spørreskjema på norsk kan lastes ned fra <http://www.sdqinfo.org/>. SDQ foreligger også som internettversjon <http://www.youthinmind.info/>, men skjema på internett til bruk for alle, finnes ikke på norsk. En elektronisk norsk versjon er integrert i DAWBA-intervjuet. CheckWare, et kommersielt firma som har laget en web-basert løsning for innsamling, skåring og rapportering av testresultater, tilbyr sine

kunder SDQ som del av sin portefølje. Britiske normer brukes for skåring.

Papirversjonen kan brukes fritt for ikke-kommersielle formål, men man må ha tillatelse fra Robert Goodman for å benytte skjemaet elektronisk og betale en mindre avgift for dette.

Skåring

På <http://www.sdqinfo.org/> er det informasjon om både håndskåring og elektronisk skåring (skåringssyntaks og internettbasert skåring). Fra 1.3.2017 er det etablert et nytt nettsted for elektronisk skåring <http://admin.sdqscore.org/>. Det beregnes en avgift for hver skåring.

Det er angitt grenseverdier for antatt avvikende skårer. Skåringsgrupperingene er litt endret fra 2014 og også grenseverdiene, som er basert på et større britisk normalmateriale. Skåre over ca. 95. persentilen er angitt som «svært høy», mellom ca. 90. og 95. persentilen som «høy», og skåre mellom ca. 80. og 90. persentilen er kategorisert som «lett forhøyet». Robert Goodman (2000) har også utviklet en algoritme som kombinerer problemskåre og «Impact»-skåre og som sammenfatter resultater fra ulike informanter. Algoritmen predikerer om barnepsykiatrisk diagnose er sannsynlig på grunnlag av resultatene. Når man bruker oppfølgingsversjonen av SDQ-P, er det en egen skåringsprosedyre som beregner «added value» av intervensjonen ved å kontrollere for forventet endring ved spontan bedring (basert på epidemiologiske data) (Ford et al., 2009).

Måleegenskaper internasjonalt

Mange internasjonale studier har undersøkt reliabilitet og validitet ved SDQ-P. Goodman (2001) fant at femfaktorstrukturen ble bekreftet i data fra et britisk epidemiologisk utvalg. Han vurderte den interne konsistens som tilfredsstillende, spesielt for Totale vansker (0,82) og «Impact»-skåren (0,85). For subskalaene var Cronbachs alfa varierende, fra 0,57 for Venneproblemer til 0,77 for Hyperaktivitet-oppmerksomhetsproblemer. SDQ-P-skårer over 90. persentilen predikerte klart økt sannsynlighet for uavhengig diagnostisert psykiatrisk lidelse (gjennomsnittlig odds ratio 15,7). Spesifisitet og negativ prediktiv verdi for prediksjon av psykiatrisk lidelse var høye, mens sensitivitet og positiv prediktiv verdi var vesentlig lavere ved disse grenseverdiene. Generelt hadde SDQ-P

psykometriske egenskaper på nivå med SDQ-T (lærerrapport) og bedre enn SDQ-S (selvrapport).

Achenbach og kolleger (2008) har oppsummert funn på psykometriske egenskaper for både SDQ og tilsvarende skjemaer i Achenbach System of Empirically Based Assessment (ASEBA) i en oversiktsartikkel. Stone og medarbeidere (2010) har rapportert psykometriske egenskaper SDQ-P og SDQ-T fra 48 studier med barn fra 4 til 12 år. Konklusjonen er at de psykometriske egenskapene for SDQ-P og SDQ-T er sterke, spesielt for lærerversjonen.

Metode

Vi søkte etter dokumentasjon på testens psykometriske egenskaper i databasene NORA, Norart, BIBSYS, SveMed+, CRIStin.no, Cochrane Library, PsycINFO, Medline, Embase og PubMed. Søkedatoer: april 2013, januar 2015 og november 2015. Søkestrategiene er tilgjengelig på <http://www.psyktestbarn.no/CMS/ptb.nsf/pages/sdq-p>. Vi kontaktet også førsteforfattere av inkluderte artikler, samt oversetterne av SDQ, for å identifisere dokumentasjon som eventuelt ikke ble fanget opp av det systematiske søket.

Vi inkluderte alle publikasjoner av studier som har undersøkt og rapportert minst ett av følgende i norske utvalg:

- Normdata for testen
- Reliabilitet: indre konsistens, test-retest eller endringssensitivitet
- Validitet: samsvar med liknende testskåre, samsvar med referansestandard eller annet kriterium, og/eller faktorstruktur

I tillegg inkluderte vi publikasjoner som rapporterte gjennomsnittsskåre for henholdsvis generelle populasjoner og for kliniske og andre undergrupper.

Begge forfatterne 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 normering, validitet og reliabilitet, 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

De tre systematiske litteratursøkene etter dokumentasjon for den norske versjonen av SDQ-P identifiserte til sammen 408 unike publikasjoner. Av disse virket 160 relevante og ble bestilt i fulltekst. 64 publikasjoner oppfylte inklusjonskriteriene våre og ble inkludert. Vi ekskluderte de resterende 96 publikasjonene fordi de ikke rapporterte relevante data, rapporterte data fra ikke-norske utvalg, kun rapporterte data fra andre SDQ-versjoner enn foreldreversjonen, eller ikke hadde brukt SDQ i det hele tatt (Vedlegg 1).

Vi kontaktet også forskere knyttet til de største inkluderte studiene for å identifisere ytterligere relevant dokumentasjon. På den måten kom vi over to relevante publikasjoner, som vi inkluderte (Fevang, Hysing, Markestad & Sommerfelt, 2016; Van Roy, 2010). Vi inkluderte ikke dokumentasjon for svenske og danske versjoner av SDQ på grunn av det høye antallet inkluderte norske publikasjoner.

Inkluderte studier

De totalt 66 inkluderte publikasjonene rapporterte data fra til sammen 34 studier. 11 av de inkluderte studiene hadde generelle befolkningsutvalg (Tabell 1). De øvrige 23 studiene var mindre enkeltstudier fra selekterte utvalg (Tabell 2).

Studier med generelle befolkningsutvalg

Helseprofil for barn og ungdom i Akershus

Tverrsnittsundersøkelsen Helseprofil for barn og ungdom i Akershus ble inkludert med data fra fire artikler (Alfstad et al., 2011; Van Roy, Groholt,

Heyerdahl & Clench-Aas, 2010; Van Roy, Kristensen, Groholt & Clench-Aas, 2009; Van Roy, Veenstra & Clench-Aas, 2008) og en doktorgradsavhandling (Van Roy, 2010). Studiepopulasjonen var 43 248 skoleelever i alderen 8-19 år, hvorav 36 456 (84 %) ble inkludert. SDQ-P inngikk i datainnsamlingen i aldersgruppen 8-13 år, som inkluderte 18 622 deltakere. Av disse fikk forskerne inn 13 308 (71 %) komplette SDQ-P-skjemaer. De ulike artiklene rapporterte gjennomsnittsskårer fordelt på aldergrupper og kjønn, indre konsistens og faktoranalyser. Alle SDQ-P-skalaene ble brukt i dokumentasjonen. Forfatterne oppga ikke om det var mødre eller fedre som hadde fylt ut skjemaene. To delstudier sammenliknet SDQ-P-skårer i den generelle populasjonen med undergrupper (barn med henholdsvis epilepsi og sosial angst) (Alfstad et al., 2011; Van Roy et al., 2009).

Barn i Bergen

Tjuen av de inkluderte artiklene rapporterte SDQ-P-data fra den longitudinelle befolkningsbaserte kohortstudien Barn i Bergen (Ekornås, Heimann, Tjus, Heyerdahl & Lundervold, 2011; Heiervang, Goodman & Goodman, 2008; Heiervang et al., 2007; Helland, Lundervold, Heimann & Posserud, 2014; Hoekstra, Lundervold, Lie, Gillberg & Plessen, 2013; Hysing, Elgen, Gillberg, Lie & Lundervold, 2007; Hysing, Elgen, Gillberg & Lundervold, 2009; Hysing, Sivertsen, Stormark, Elgen & Lundervold, 2009; Iversen, Jakobsen, Havik, Hysing & Stormark, 2007; Lundervold, Heimann & Manger, 2008; Munkvold, Lundervold, Lie & Manger, 2009; Munkvold, Lundervold & Manger, 2011; Paap et al., 2013; Posserud et al., 2008; Sandvik, 2007; Sanne, Torsheim, Heiervang & Stormark, 2009; Sivertsen, Hysing, Elgen, Stormark & Lundervold, 2009; Sivertsen, Posserud, Gillberg, Lundervold & Hysing, 2012; Sørensen, Hugdahl & Lundervold, 2008; Sørensen, Plessen & Lundervold, 2012; Ullebø, Posserud, Heiervang, Gillberg & Obel, 2011).

Barn i Bergen har hatt fire runder. Datainnsamlingen til den første runden hadde samtlige 9430 2.-4. klassinger (7-9-åringer) i Bergen som målgruppe og fant sted skoleåret 2002/2003. Denne runden fikk forskerne inn SDQ-P-data for 6913 av 9430 (73 %) barn i målgruppen (M. Hysing et al., 2009). 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. SDQ-P ble innhentet for 5781 barn (83 %) (M. Hysing et al., 2009). I runde 3 og 4 var barna henholdsvis 14- 16 og 16-18 år.

De inkluderte artiklene rapporterte gjennomsnittsskårer, samt kjønnsforskjeller, grenseverdier, faktorstruktur, diagnostisk nøyaktighet, forhold mellom SDQ-P-skårer og andre skalaer, indre konsistens og stabilitet. Dataene omfattet alle SDQ-P-skalaene, men noen av artiklene fokuserte på spesifikke skalaer. Forfatterne oppga ikke om det var mødre eller fedre som hadde fylt ut skjemaene.

Til sammen ti Barn i Bergen-delstudier undersøkte forskjeller mellom ulike underutvalg og referansegruppe fra hovedstudien. Underutvalgene bestod av barn med emosjonelle problemer (Ekornås et al., 2011; Heiervang et al., 2008), atferdsproblemer (Heiervang et al., 2008; Helland, Lundervold, et al., 2014; Munkvold et al., 2011), autismspekterforstyrrelse (Heiervang et al., 2008; Sivertsen et al., 2012), ADHD (Heiervang et al., 2008), tics (Hoekstra et al., 2013), språkvansker (Lundervold et al., 2008), kronisk sykdom (M. Hysing et al., 2009; Sivertsen et al., 2009), og barn i barnevernet (Iversen et al., 2007).

Zippys venner

En annen stor studie var et klyngerandomisert kontrollert forsøk som inkluderte andreklassinger (gjennomsnittsalder 7,3 år) ved 35 skoler i Trondheim, Bodø og Østfold (Holen, Lervåg, Waaktaar & Ystgaard, 2012; Holen, Waaktaar, Lervåg & Ystgaard, 2012). I alt 18 skoler (745 andreklassinger) ble trukket ut til å delta i det universalforebyggende programmet Zippys venner, som skal forebygge psykiske problemer hos barn ved å gi dem økte mestringsferdigheter. De øvrige 17 skolene (738 andreklassinger) ble trukket ut til å utgjøre en kontrollgruppe som ikke fikk noe tiltak. Forskerne fikk inn SDQ-P-data for 1278 barn (86 %) i denne studien, og rapporterte data omfattet gjennomsnittsskårer, faktorstruktur, korrelasjoner og indre konsistens.

Tidlig trygg i Trondheim

To artikler rapporterte data fra den longitudinelle befolkningsbaserte studien Tidlig trygg i Trondheim,

med Trondheims rundt 3000 4-åringer som målgruppe (Sveen, Berg-Nielsen, Lydersen & Wichstrom, 2013; Wichstrom et al., 2012). Foreldre og barn ble rekruttert gjennom helsestasjonen, og 2475 familier samtykket til å delta. Til sammen 1250 av disse ble tilfeldig valgt ut fra fire strata av SDQ-skårer for å delta i det diagnostiske intervjuet Preschool Age Psychiatric Assessment (PAPA) (Egger et al., 2006) med en foresatt. Wichstrøm og kolleger rapporterte tverrsnittsdata for 995 (svarprosent 80 %) barn. I tillegg rapporterte Sveen og kolleger *Area Under the Curve* (AUC) for samsvar mellom SDQ-P Totale vansker og PAPA, samt indre konsistens og informantkorrelasjoner. For hvert barn var det enten mor (83 %) eller far (17 %) som hadde fylt ut SDQ-P.

Tracking Opportunities and Problems Project (TOPP)

Tre av de inkluderte publikasjonene stammet fra TOPP, en befolkningsbasert kohortstudie på Østlandet (Kjeldsen, Janson, Stoolmiller, Torgersen & Mathiesen, 2014; Mathiesen & Prior, 2006; Nilsen, Gustavson, Roysamb, Kjeldsen & Karevold, 2013). Studien omfattet 939 8 ½-åringer, hvorav SDQ-P-data fra mødre var innhentet for 401-535 deltakere (svarprosent 43 % - 57 % av de som var invitert ved T1, ved 18 mnd alder). De tre artiklene rapporterte standardiserte gjennomsnittsskårer, indre konsistens, kjønnsforskjeller, korrelasjoner og regresjonsanalyser for Totale vansker og for subskalaene Atferdsproblemer og Emosjonelle symptomer.

Romsdalstudien

Obel og kollegers artikkel var en sammenstilling av SDQ-data fra forskjellige nordiske studier, hvorav åtte norske (Obel et al., 2004). Sju av de norske studiene er inkludert som primærstudier i denne PsykTestBarn-oversikten, mens den åttende har vi kun funnet data fra i Obel-artikkelen. Den åttende studien var en befolkningsbasert tverrsnittstudie fra fire kommuner i Romsdal. Barn i alderen 4, 7, 12 og 16 utgjorde populasjonen (N=2077), og Obel-artikkelen rapporterte SDQ-P-skårer for 417 sjuåringer. Forfatterne oppga ikke om det var mødre eller fedre som hadde fylt ut skjemaene.

Alvik 2013

I Alvik, Aalen og Lindemanns (2013) befolkningsbaserte kohortstudie fylte 1324 Oslo-mødre ut SDQ-P da barna var 5,5 år gamle. Forskerne undersøkte sammenhengen mellom mors alkoholbruk tidlig i svangerskapet og barnas SDQ-P-skårer, rapportert som forekomster i de tre SDQ-kategoriene «normal», «grenseområde» og «antatt klinisk område», for skalaene Totale vansker, Emosjonelle symptomer, Atferdsproblemer og Hyperaktivitet. Grenseverdiene for «grenseområdet» og «antatt klinisk verdi» var satt ved henholdsvis 80. og 90. persentilen.

Den norske mor og barn-undersøkelsen (MoBa)

MoBa er en nasjonal kohortstudie i regi av Folkehelseinstituttet som følger mødre og deres barn. Deltakerne ble rekruttert i perioden 1999-2008 og omfatter over 100 000 barn. Vi har inkludert én artikkel fra en MoBa-delstudie hvor forskerne sammenliknet 3-åringer med medfødt hjertefeil (N=175) og den øvrige kohorten på samme alder (N=43929) med hensyn til ulike utviklingsområder (Brandlistuen et al., 2011). Det eneste resultatet fra denne artikkelen som var relevant for vår oversikt, var indre konsistens for SDQ-P-skalaen Prosocial atferd (som i artikkelen kalles «Prosocial impairment»).

Mindre befolkningsbaserte studier

Blant de tre mindre befolkningsbaserte studiene var en mastergradsoppgave om språk- og sosioemosjonelle problemer hos 4 år gamle barnehagebarn (N=175) (Kleven, 2008). Studenten oppga ikke om det var mødre eller fedre som hadde fylt ut SDQ-P, og indre konsistens var den eneste rapporterte datatypen vi kunne bruke. Vi kunne ikke bruke de rapporterte skårene fordi de var transformert til indeksskårer.

Den andre studien var Javo, Rønning, Handegård og Rudmins (2009) undersøkelse av samsvaret mellom foreldre- og lærerrapporterte data for samiske og etnisk norske barn på 11-12 år (N=173). De eneste dataene vi kunne inkludere fra denne studien, var «Impact»-skårer for samiske og etnisk norske barn.

Den siste studien var Walhovd og kollegers (2012) undersøkelse av forholdet mellom hjernebarkens tykkelse og atferdsregulering hos barn og ungdom i alderen 8-19 år (N=106). Vi inkluderte gjennomsnittsskårer for Totale vansker, Hyperaktivitet/oppmerksomhetsproblemer og Atferdsproblemer fra denne studien.

Tabell 1. Inkluderte studier med generelle befolkningsutvalg

Studie (navn)	Publikasjoner (førsteforfatter år)	N ¹	Svarprosent	Alder (år)	Informant (mor/far)	Underutvalg, delstudier
Helseprofil Akershus	Alfstad 2011, Van Roy 2008, Van Roy 2009, Van Roy 2010a, Van Roy 2010b	13308	71 %	8-13	Ukjent	Barn med epilepsi, sosial angst
Barn i Bergen	Ekornås 2011, Heiervang 2007, Heiervang 2008, Helland 2014b, Hoekstra 2013 ² , Hysing, 2007, Hysing 2009a, Hysing 2009b ² , Iversen 2007 ² , Lundervold 2008, Munkvold 2009, Munkvold 2011, Paap 2013, Posserud 2008, Sandvik 2007, Sanne 2009, Sivertsen 2009, Sivertsen 2012, Sørensen 2008, Sørensen 2012, Ullebø 2011	6913 5781	73 % 83 % ³	7-9 11-13 ²	Ukjent	Barn med emosjonelle problemer, atferdsprobleme autisme, ADHD, tics, språkvanske kronisk sykdom, barnevernsbarn
Zippys venner	Holen 2012a, Holen 2012b	1278	86 %	7	83 % mor, 17 % far	
Tidlig trygg i Trondheim	Sveen 2013, Wichstrøm 2012	995	80 %	4	84 % mor, 16 % far	
TOPP	Kjeldsen 2014, Mathiesen 2006, Nilsen 2013	535	57 % ⁴	8 ½	Mor	
Romsdalstudien	Obel 2004	417	Ukjent	7	Ukjent	
Alvik 2013	Alvik 2013	1138	66 % ⁴	5 ½	Mor	Alkoholbruk i svangerskapet
MoBa	Brandlistuen 2011	175	Ukjent	3	Mor	Barn med medfødt hjertefe
Kleven 2008	Kleven 2008	175	81 %	4	Ukjent	
Javo 2009	Javo 2009	173	58 %	11-12	Mor	Samiske og etnis norske barn
Walhovd 2012	Walhovd 2012	106	Ukjent	8-19	Ukjent	

Note: ¹Antall med SDQ-P-data; ²Den samme kohorten ved oppfølging; ³Av de med samtykke for nye undersøkelser; ⁴Av opprinnelig kohort; TOPP= Tracking Opportunities and Problems Project; MoBa= Den norske mor og barn-undersøkelsen

Studier med selekterte utvalg

Fire av de 23 studiene med selekterte utvalg fokuserte på barn med psykiske problemer (Helland, Helland & Heimann, 2014; Mathiassen et al., 2012; Nilsen, Handegard, Eisemann & Kvernmo, 2015; Tromop, 2011), fem studier fokuserte på barn med psykososiale risikofaktorer (Aasheim, 2012; Clucas, Skar, Sherr & Tetzchner, 2014; Freuchen, Kjelsberg, Lundervold & Groholt, 2012; Lehmann, Breivik, Heiervang, Havik & Havik, 2015; Lehmann, Heiervang, Havik & Havik, 2014; Sherr, Skar, Clucas, von Tetzchner & Hundeide, 2014; Skar, von Tetzchner, Clucas & Sherr, 2015; Vaage et al., 2011; Vaage et al., 2009), ti på barn med somatiske risikofaktorer (Bjørngaas, Elgen, Bøe & Hysing, 2013; Diseth, Tangeras, Reinfjell & Bjerre, 2011; Eilertsen, Rannestad, Indredavik & Vik, 2011; Feragen & Stock, 2014; Fjermestad et al., 2015; Gjertsen & Sanne, 2010; Hanssen-Bauer, Heyerdahl & Eriksson, 2007; Helgeland, Van Roy, Sandvik, Markestad & Kristensen, 2011; Kokkvoll,

Grimsgaard, Steinsbekk, Flaegstad & Njolstad, 2014; Ramstad, Jahnsen, Skjeldal & Diseth, 2012), og fire på barn født prematurt/med lav fødselsvekt (Elgen et al., 2012; Fevang et al., 2016; Indredavik, Vik, Heyerdahl, Kulseng & Brubakk, 2005; Nordhov, Rønning, Ulvund, Dahl & Kaarsen, 2012) (Tabell 2).

Elleve av studiene hadde kasus-kontrolldesign, hvor forskjellige kliniske utvalg (N=19-255) av barn og ungdom i alle aldersgrupper var sammenliknet med ikke-kliniske utvalg av matchede kontroller eller med data fra en av de store befolkningsundersøkelsene på gjennomsnittsskårer eller andel «høyyskårere» (Diseth et al., 2011; Eilertsen et al., 2011; Elgen et al., 2012; Feragen & Stock, 2014; Fevang et al., 2016; Freuchen et al., 2012; Gjertsen & Sanne, 2010; Helgeland et al., 2011; Helland, Helland, et al., 2014; Indredavik et al., 2005). Noen av artiklene rapporterte også indre konsistens.

Sju av studiene hadde tverrsnittdesign uten kontrollgruppe med utvalgsstørrelser som varierte fra 50 til 223 (Aasheim, 2012; Fjermestad et al., 2015; Hanssen-Bauer et al., 2007; Lehmann et al., 2015; Lehmann et al., 2014; Mathiassen et al., 2012; Ramstad et al., 2012; Vaage et al., 2011; Vaage et al., 2009). Rapporterte SDQ-P-data var hovedsakelig gjennomsnittsskårer, men én studie

hadde også undersøkt den diagnostiske nøyaktigheten ved SDQ-P (Lehmann et al., 2014).

Fem studier evaluerte ulike tiltak: foreldreveiledningsprogrammet International Child Development Programme (ICDP) (Clucas et al., 2014; Sherr et al., 2014; Skar et al., 2015), tiltak for premature barn (Nordhov et al., 2012), tiltak for

Tabell 2. Inkluderte studier med selekterte utvalg (delstudier fra de befolkningsbaserte studiene ikke med)

Studie (førsteforfatter år)	Seleksjonsgrunnlag	N ¹	Svarprosent	Alder (år)	Informant (mor/far)	Referansegruppe (N)
	<i>Psykiske problemer</i>					
Helland 2014a ¹	ADHD	21	-	6-12	-	19
Mathiassen 2012	Henvist til BUP	132		11,5 ²	-	-
Nilsen 2015	Angst/depresjon (BUP)	77	92 %	5-18	Mor og far	-
Tromop 2011	Henvist til psykiatrisk akuttenhet	86	89 %	13-18	-	-
	<i>Psykososiale risikofaktorer</i>					
Lehmann 2014, 2015	Fosterhjem	223	56 %	6-12	-	-
Aasheim 2012	Barnevernstiltak	50	-	-	-	-
Clucas 2014, Sherr 2014, Skar 2015	Foreldreveiledning	184	Ukjent	0-16	75 % mor, 25 % far	
Freuchen 2012	Selv mord Dødsulykke	41 43	45 % 47 %	9-15	42 % mor, 10 % far, 48 % begge	410 ³
Vaage 2009, 2011	Foreldre vietnamesiske flyktninger	94	91 %	4-18	-	-
	<i>Somatiske tilstander</i>					
Diseth 2011	Transplantert nyre Leukemi	31 36	75 %	2-19	Mor	38
Eilertsen 2011	Kreftbehandlet	50	50 %	6-20	-	29
Gjertsen 2010	Synshemming	71	51 %	6-16	-	53
Helgeland 2011	Funksjonelle magesmerter	94		8-15	Mor	14000 ⁴
Bjergås 2013	Cerebral parese	47	48 %	7 ²	-	7007 ³
Feragen 2014	Leppe-kjeve-ganespalte	177	65 %	10	52 % mor, 22 % far, 26 % begge	-
Fjermestad 2015	Kjønnskromosomfeil Angstsymptomer	25 98	47 % -	2-18 7-15	80 % mor, 20 % far	-
Hanssen-Bauer 2007	Epilepsi	71	65 %	9-15	-	-
Kokkvoll 2014	Overvekt/fedme	88	91 %	6-12	-	-
Ramstad 2012	Cerebral parese	83	54 %	8-18	Mor	-
	<i>Prematur fødsel/lav fødselsvekt</i>					
Elgen 2012	Ekstremt prematur fødsel	255	69 %	5	-	1089
Fevang 2016	Ekstremt prematur fødsel/lav fødselsvekt	216	64 %	11	-	1893 ³
Indredavik 2005	Svært lav fødselsvekt Liten for gestasjonsalder	52 59	52 %	14	Mor og far	78
Nordhov 2012	Lav fødselsvekt	126	86 %	5	Mor og far	64

Note: ¹I tillegg hadde studien et utvalg med spesifikke språkvansker (N=19); ²Gjennomsnitt; ³Barn i Bergen; ⁴Helseprofil Akershus; - ikke oppgitt

barn med overvekt/fedme (Kokkvoll, 2014), behandling ved psykiatrisk poliklinikk (Nilsen et al., 2015) og i psykiatrisk akutenhet (Tromop, 2011). Utvalgsstørrelsene varierte fra 77 til 184, og relevante data for vår oversikt var gjennomsnittsskårer og indre konsistens.

Overlappende data

Der hvor vi har inkludert flere publikasjoner med overlappende data fra én og samme studie, har vi valgt å gjengi data fra de publikasjonene som hadde størst utvalg, og flest skalaer og psykometriske verdier.

Fra kohortstudier med to eller flere datainnsamlinger angir vi resultater fra de ulike bølgeene, med anmerkning om at dette er en longitudinell studie med samme populasjon.

I hovedtabellene gjengir vi SDQ-P-data rapportert av enten mor eller far, eller mor alene. I de fleste studiene var SDQ-P fylt ut av én av barnets foreldre, uten at det var spesifisert om det var mor eller far som hadde fylt ut. Noen studier hadde angivelser av andel skjemaer fylt ut av henholdsvis mor, far og eventuelt mor og far i fellesskap. Andel fedrerapporterte skjemaer varierte fra 10 % til 25 % i disse studiene (Tabell 1 og 2). I de studiene hvor mødre og fedre hadde fylt ut hvert sitt skjema, var tendensen at flere mødre enn fedre hadde fylt ut skjema. Derfor har vi valgt å gjengi mødrerapporterte data i hovedtabellene fra studier hvor forfatterne har oppgitt separate data mødre og fedre. I tillegg gjengir vi fedrerapporterte data i en egen tabell (Tabell 5).

Normer

Ingen av de inkluderte studiene rapporterte nasjonale normer, men Helseprofil Akershus, Barn i Bergen og Tidlig trygg i Trondheim bidro med store normdatasett for forskjellige aldersgrupper (Tabell 3). Vi har også inkludert gjennomsnittsskårer fra mindre befolkningsbaserte utvalg. Det foreligger ikke resultater for både totalutvalg og fordelt på kjønn fra alle de store studiene. Vi rapporterer derfor også data fra kontrollgruppen i delstudier som har undersøkt problemstillinger som barn med sosial angst (Van Roy et al., 2009), barn med

språkvansker (Lundervold et al., 2008) og barn med kronisk sykdom (M. Hysing et al., 2009). Disse kontrollgruppene må anses som «supernormale» fordi de er definert som «den generelle befolkningen minus barn med den undersøkte sykdom/tilstand». Fra disse studiene har vi estimert gjennomsnittsskåre for totalutvalget (angitt i fotnoter til Tabell 3).

Gjennomsnittsskårene var relativt like på tvers av utvalgene i de store befolkningsbaserte studiene ($N > 400$). Gjennomsnittsskårene for totale vansker for totalutvalgene varierte fra 5,3 til 6,2 (Tabell 3 med fotnoter for estimerte verdier): for 4-åringer 5,6 (Wichstrøm 2012); for aldersgruppen 7-10 år (5,9-6,2) (Heiervang 2008, Obel 2004, Van Roy 2009). For aldersgruppen 10-13 år var det høyere skåre for totale vansker i Akershus (van Roy 2010) enn i Bergen (Hysing 2009b): 6,2 versus 5,3. Resultater fra de befolkningsbaserte utvalgene/kontrollgrupper fra disse utvalgene fra 7 til 13 år viser variasjon i Emosjonelle symptomer fra 1,1 til 1,5; Atferdsproblemer fra 0,8 til 1,2; Hyperaktivitet fra 2,2 til 2,9; Problemer med venner fra 0,9 til 1,1; Prosocial atferd fra 8,0 til 8,6 og «Impact»-skåren fra 0,2 til 0,4.

Tabell 3. SDQ-P-skårer i generelle befolkningsutvalg, gjennomsnitt (standardavvik) om ikke annet er spesifisert

Studie (referanse)	Alder	N	TOT	EMO	ATF	HYP	VEN	PRO	Impact
Helseprofil Akershus									
(Van Roy 2009a) ¹	8-10	5410	6,1 (4,8)	1,2 (1,3)	1,1 (1,3)	2,8 (2,3)	1,1 (1,5)	8,3 (1,6)	
	10-13	7553	5,8 (4,7)	1,1 (1,6)	1,0 (1,2)	2,5 (2,2)	1,1 (1,6)	8,3 (1,6)	
(Van Roy 2010a) ²	10-13, jenter	4238	5,7 (4,8)	1,4 (1,8)	1,0 (1,2)	2,2 (2,0)	1,1 (1,6)	8,5 (1,5)	0,3 (1,0)
	10-13, gutter	4279	6,6 (5,2)*	1,2 (1,7)*	1,1 (1,4)*	3,0 (2,4)*	1,3 (1,7)*	8,0 (1,7)*	0,5 (1,3)*
Barn i Bergen									
(Heiervang 2008)	7-9	6658	5,9 (4,9)	1,3 (1,7)	1,0 (1,3)	2,7 (2,2)	1,0 (1,5)	8,5 (1,5)	0,5 (1,3)
(Lundervold 2008) ³	7-9, jenter	3048	5,1 (4,2)	1,3 (1,7)	0,8 (1,1)	2,2 (1,9)	0,8 (1,4)	8,8 (1,3)	0,2 (1,0)
	7-9, gutter	2949	5,9 (4,8)*	1,2 (1,6)	1,0 (1,3)*	2,9 (2,2)*	0,9 (1,4)*	8,3 (1,6)*	0,3 (1,4)*
(Hysing 2009b) ⁴	11-13	5285	5,1 (5,0-5,2)	1,1 (1,2-1,2)	0,8 (0,8-0,8)	2,2 (2,2-2,3)	1,0 (0,9-1,0)		
Tidlig trygg i Trondheim (Wichstrøm 2012)									
	4	995	5,6 (0,3) ⁵						
Zippys venner (Holen 2012a)	7 ⁶	684		1,5 (1,7)	1,1 (1,2)	2,9 (2,1)	1,0 (1,4)	8,0 (1,3)	0,2 (0,8)
Romsdalstudien (Obel 2004)	7	417	5,9 (4,4)	1,2 (1,6)	1,0 (1,3)	2,6 (2,1)	0,9 (1,3)	8,4 (1,6)	
TOPP (Nilsen 2013)	8 ½, jenter	?		1,2 (0,3)	1,2 (0,2)				
	8 ½, gutter			1,2 (0,3)	1,2 (0,2)				
Javo 2009	11-12, samiske	71							0,1 (0,5)
	11-12, etnisk norske	77							0,5 (1,5)
Walhovd 2012	8-11	35	4,1 (3,4)		0,7 (0,8)	2,1 (2,1)			
	12-15	37	4,8 (3,9)		0,8 (0,9)	1,6 (1,6)			
	16-19	34	4,1 (3,3)		0,9 (0,8)	1,5 (1,7)			

Note: *Kjønnsforskjeller $p < 0,01$, for venneproblemer i Lundervolds studie $p < 0,05$; ¹Gjelder barn uten sosial angst. Totale vansker for hele utvalget 8-10 år estimert til 6,2, 10-13 år 5,9; ²Totale vansker for gutter og jenter 10-13 år samlet estimert til 6,2; ³Gjelder barn uten språkvansker. Totale vansker for hele utvalget rapportert å være 5,8 (4,8), gutter 6,3 (5,1), jenter 5,3 (4,4); ⁴Oppfølgingsdata, barn uten kronisk sykdom. Totale vansker for hele utvalget 11-13 år estimert til 5,3; ⁵95 % konfidensintervall i parentesene; ⁶Standardfeil (SE); ⁷Pretestskårer, intervensjonsgruppe.

TOT=Totale vansker; EMO=Emosjonelle symptomer; ATF=Atferdsproblemer; HYP=Hyperaktivitet; VEN=Problemer med venner; PRO=Prososial atferd

Akershus-jenter i alderen 10 til 13 år hadde statistisk signifikant lavere skårer enn guttene på alle problemskalaer, bortsett fra Emosjonelle symptomer, hvor jentene hadde noe høyere gjennomsnittsskåre enn guttene (Tabell 3). Guttene hadde også høyere «Impact»-skåre. Jentene skåret høyere enn gutter for Prososial atferd. Kjønnsforskjellmønsteret var nesten likt blant 7-9-åringene i Bergen (Lundervold et al., 2008).

Heiervang og kolleger (2008) sammenliknet SDQ-P-skårer hos britiske barn ($N = 4898$; 8-10 år) og Barn

i Bergen-deltakere ($N = 6658$; 7-9 år). De britiske barna skåret høyere enn de norske på samtlige skalaer. Den gjennomsnittlige forskjellen for Totale vansker var 2,5 (95 % konfidensintervall 2,2 – 2,7); Gjennomsnittsskåre for de norske barna var 5,9 ($SD = 4,9$); for de britiske barna 8,4 ($SD = 4,9$). Alle forskjellene var statistisk signifikante ($p < 0,001$ – justert for «clustered sample design»). Studien brukte både spørreskjema (SDQ-P og SDQ-T) og intervju (DAWBA) og undersøkte om forskjeller i SDQ-skårer mellom landene reflekterte reelle

Tabell 4a. Totale vansker i utvalg med psykiske problemer (sammenliknet med referansegrupper der data foreligger)

Studie (førsteforfatter år)	Seleksjonsgrunnlag	Alder (år)	Utvalg		Referansegruppe	
			N	M (SD)	N	M (SD)
Van Roy 2009 ¹	Sosial angst	8-9	135	10,5 (5,7)	5410	6,1 (4,8)**
		10-13	210	10,9 (6,1)		
Heiervang 2008 ^{2,3,4}	Emosjonell forstyrrelse	7-9		10,5 (8,2-12,8)		5,3 (5,1-5,6) ⁵
		Atferdsforstyrrelse	7-9	18,2 (16,4-19,9)		
		ADHD	7-9	20,9 (19,2-22,6)		
		Autismespekterforstyrrelse	7-9	20,7 (17,2-24,2)		
Helland 2014a ⁶	ADHD	6-12	19	20,3 (5,1)*	6235	5,9 (4,9)**
Helland 2014b ²	ADHD/språkvansker	7-9	40	14,1 (7,1)		
Fjermestad 2015	Angstsymptomer	2-18	98	16,9 (6,9)		
Nilsen 2015	Angst/depresjon (BUP)	5-18	77	13,1 (0,9)		
Tromop 2011	Henvist til psykiatrisk akuttenhet	13-18	86	18,8 (6,0)		
Hoekstra 2013 ²	Tics	7-9	38	11,2 (8,2)	3847	5,3 (4,4)**
		11-13 ⁷	38	10,5 (8,0)		

Note: M=gjennomsnitt; SD=standardavvik; **p < 0,001 for forskjell utvalg/referansegruppe; ¹Delstudie i Helseprofil Akershus; ²Delstudie i Barn i Bergen; ³N ikke angitt for utvalgene, total N i intervjuundersøkelsen som er grunnlag for diagnosene er 1024; ⁴95 % konfidensintervall i parentesene; ⁵p-verdi ikke angitt; ⁶Referansegruppedata fra Barn i Bergen; ⁷Utvalg som deltok i både bølge 1 og 2

forskjeller i mental helse eller forskjeller i rapporteringsstil.

Gjennomsnittsskårer i selekterte utvalg

30 studier eller delstudier rapporterte gjennomsnittsskårer for Totale vansker i selekterte utvalg (Tabell 4a-d). Deltakerne hadde psykiske problemer (Tabell 4a), psykosiale risikofaktorer eller språkvansker (Tabell 4b), somatiske sykdommer eller risikofaktorer (Tabell 4c) og prematur fødsel/lav fødselsvekt (Tabell 4d). Totale vansker varierte fra 4,3 (barn som omkom i ulykke; Freuchen et al., 2012) til 20,9 (7-9-åringer med ADHD; Heiervang et al., 2008).

Gjennomsnittsskårer på SDQ-P-subskalaer i selekterte utvalg er gjengitt i Vedlegg 2.

Mange av studiene og delstudiene sammenliknet skårene i de selekterte utvalgene med skårer i referansegrupper fra den generelle befolkningen. Vi kommer nærmere tilbake til disse forskjellene i avsnittet om validitet.

Tabell 4b. Totale vansker i utvalg med psykososiale risikofaktorer eller språkvansker (sammenliknet med referansegrupper der data foreligger)

Studie (førsteforfatter år)	Seleksjonsgrunnlag	Alder (år)	Utvalg		Referansegruppe	
			N	M (SD)	N	M (SD)
Lehmann 2014	Fosterhjem	6-12	141	14,7 (7,8)		
Iversen 2007 ¹	Barnevernstiltak	11-13	82	14,5 (7,5)	4072	5,3 (4,6)**
Aasheim 2012	Barnevernstiltak	-	50	10,4 (6,3)		
Clucas 2014	Foreldreveiledning	-	86-91	8,8 (4,7)		
Freuchen 2012 ²	Selvmord	< 16	41	6,3 (4,7)	410	5,7 (4,8) ³
		Dødsulykke	< 16	43		
Vaage 2009, 2011	Foreldre vietnamesiske flyktninger	4-18	88	9,0 (5,7)		
Helland 2014a	Spesifikke språkvansker	6-12	21	12,2 (2,1)		
Lundervold 2008 ¹	Språkvansker	7-9	237	12,0 (6,7)	5997	5,5 (4,5)**

Note: M=gjennomsnitt; SD=standardavvik; **p < 0,001 for forskjell utvalg/referansegruppe; ¹Delstudie i Barn i Bergen;

²Referansegruppedata fra Barn i Bergen; ³p-verdi ikke angitt; - ikke oppgitt

Tabell 4c. Totale vansker i utvalg med somatiske sykdommer/ risikofaktorer (sammenliknet med referansegrupper der data foreligger)

Studie (førsteforfatter år)	Seleksjonsgrunnlag	Alder (år)	N	Utvalg		Referansegruppe	
				N	M (SD)	N	M (SD)
Diseth 2011	Transplantert nyre	2-19	31		10,7 (6,3)	38	4,2 (3,6)**
	Leukemi	2-19	36		7,6 (5,2)		
Eilertsen 2011	Kreftbehandlet	6-20	50		9,4 (6,8)	26	4,9 (4,2)*
Helgeland 2011	Funksjonelle magesmerter	8-15	94		8,8 (5,3)		
Hysing 2009b ^{1,2}	Kronisk sykdom	11-13	496		7,2 (6,7-7,8)	5285	5,1 (5,0-5,2)**
Sivertsen 2009 ^{1,2}	Kronisk sykdom	7-9	295		7,8 (7,0-8,5)	3730	5,2 (5,0-5,3)**
Feragen 2014	Leppe-kjeve-ganespalte	10	109		5,8 (4,3)		
	Spalte og tilleggstilstand	10	69		12,4 (6,8)	8154	6,2 (5,0)**
Gjertsen 2010	Synshemming	6-16	71		8,3 (5,7)	53	5,7 (4,5)*
Fjermestad 2015	Kjønnskromosomfeil	2-18	25		18,8 (8,0)		
Hanssen-Bauer 2007	Epilepsi, gutter	9-15	40		16,7 (7,2)		
	Epilepsi, jenter	9-15	32		17,5 (6,4)		
Bjørgås 2013	Cerebral parese	6-8	47		13,9 (5,3)	7007	5,6 (5,0)**
Ramstad 2012	Cerebral parese	8-18	83		11,0 (6,6)		
Kokkvoll 2014	Overvekt/fedme	6-12	88		10,1 (5,8)		

Note: M=gjennomsnitt; SD=standardavvik; ** $p < 0,001$ for forskjell utvalg/referansegruppe; * $p < 0,01$ for forskjell utvalg/referansegruppe;
¹Delstudie i Barn i Bergen; ²95 % konfidensintervall i parentesene

Fedrerapporterte gjennomsnittsskårer

Kun Clucas og kollegaer (2014) testet forskjeller mellom mødre- og fedrerapporterte skårer. Forskerne fant ingen forskjeller i data innhentet før foreldreveiledningstiltaket ICDP, men de fant en forskjell for data innhentet etter tiltaket (Cohens $d=0,67$; $p=0,003$). De oppgir imidlertid at denne forskjellen ble ikke-signifikant etter justering for konfundere (justeringsvariabler).

grenseområde eller lett fornøyet skåre, ca 10 % og antatt klinisk område eller høy skåre, ca 10 % (van Roy, 2010). Områdene er definert i tråd med Robert Goodmans kategorier fra det britiske normdatasettet: <http://SDQinfo.org/>. Goodman har senere endret kategoriene slik at «antatt klinisk område» er delt i høy skåre (ca 5 %) og svært høy (ca 5 %).

Forhøyede skårer – grenseverdier for avvik

Den store Helseprofil-undersøkelsen fra Akershus (N=14 786) rapporterer grenseverdier for tre områder: normalområdet, som tilsvare ca 80 %,

Tabell 4d. Totale vansker i utvalg med prematur fødsel/lav fødselsvekt (sammenliknet med referansegrupper der data foreligger)

Studie (førsteforfatter år)	Seleksjonsgrunnlag	Alder (år)	N	Utvalg		Referansegruppe	
				N	M (SD)	N	M (SD)
Elgen 2012	Ekstremt prematur fødsel	5	255		8,7 (5,8)	1089	5,0 (3,8)**
Indredavik 2005	Svært lav fødselsvekt	14	52		8,6 (6,5)	78	3,4 (2,9)**
	Liten for gestasjonsalder	14	59		5,7 (5,7)		
Nordhov 2012	Lav fødselsvekt, med tiltak	5	67		6,7 (5,3)	64	4,9 (3,3) ¹
	Lav fødselsvekt, uten tiltak	5	59		9,0 (5,5)		

Note: M=gjennomsnitt; SD=standardavvik; ** $p < 0,001$ for forskjell utvalg/referansegruppe; * $p < 0,01$ for forskjell utvalg/referansegruppe;
¹ p -verdi ikke angitt

Tabell 5. Fedrerapporterte SDQ-P-skårer, gjennomsnitt (standardavvik) om ikke annet er spesifisert

Studie (referanse)	Utvalg	N	TOT	EMO	ATF	HYP	VEN	PRO	Impact
Clucas 2014	0-16-åringer før ICDP	31	9,7 (5,2)	-	-	-	-	6,9 (2,3)	0,7 (1,2)
	0-16-åringer etter ICDP	31	10,2 (5,3)	-	-	-	-	6,9 (2,3)	0,4 (1,1)
Indredavik 2005	14-åringer med svært lav fødselsvekt	47	7,0 (4,6)	1,4 (1,6)	0,8 (0,9)	3,2 (1,9)	1,7 (2,0)	7,6 (1,8)	0,9 (1,8)
	14-åringer, liten for svangerskapsalder	55	5,8 (5,6)	0,8 (1,6)	0,9 (1,5)	2,6 (2,2)	1,4 (2,0)	7,8 (1,7)	0,6 (1,6)
	14-åringer, referansegruppe	72	3,8 (3,6)	0,5 (1,0)	0,5 (0,9)	1,7 (1,7)	1,1 (1,5)	8,2 (1,4)	0,1 (0,4)
Nilsen 2015 ¹	5-18-åringer med angst/depresjon	50	12,3 (1,1)	4,8 (0,4)	-	-	-	-	2,9 (0,4)
Nordhov 2012	For tidlig fødte 5-åringer med tiltak	60	6,1 (3,6)	1,5 (1,4)	0,8 (0,9)	2,7 (2,0)	1,2 (1,4)	7,9 (1,7)	-
	For tidlig fødte 5-åringer uten tiltak	51	8,0 (5,1)	1,6 (1,6)	1,4 (1,5)	3,8 (2,4)	1,4 (1,7)	7,5 (2,0)	-
	5-åringer, referansegruppe	59	4,9 (3,7)	0,7 (1,1)	0,8 (0,8)	2,4 (1,8)	1,0 (1,5)	8,0 (1,6)	-

Note: ¹ Standardfeil (SE) i parentesene; TOT=Totale vansker; EMO=Emosjonelle symptomer; ATF=Atferdsproblemer; HYP=Hyperaktivitet; VEN=Problemer med venner; PRO=Prososial atferd; ICDP=International Child Development Program

Validitet

Begrepsvaliditet

Ved hjelp av faktoranalyser, korrelasjoner med andre instrumenter og gruppesammenlikninger fant vi data som kunne brukes til vurdering av begrepsvaliditeten ved SDQ-P-skalaene.

Faktorstruktur

Tre store studier: Barn i Bergen (N=3 166, barn 7-9 år), Helseprofil Akershus (N=6 645, barn 10-13 år) og Zippys venner (N=1 272, barn 7-8 år), har undersøkt SDQ-P's faktorstruktur ved hjelp av konfirmatoriske faktoranalyser (Holen, Lervåg, et al., 2012; Sanne et al., 2009; Van Roy et al., 2008). De konkluderte alle med at Goodman's femfaktormodell hadde akseptabel eller god modelltilpasning. RMSEA (root mean square error of approximation)-indeksen varierte fra 0,043 til 0,051 (median = 0,050). Tilsvarende CFI (comparative fit index) varierte fra 0,886 til 0,98 (median = 0,90). Funnene bekrefter de fem skala-dimensjonene, men enkelte ledd og subskalaene er ikke veldig distinkte. Ved tolkning av resultater bør det ikke legges stor vekt på enkeltspørsmål og de enkelte subskalaene (Sanne et al., 2009).

Lehmann og medarbeidere (2015) undersøkte primært tilknytningsforstyrrelser hos barn, men de studerte også om tilknytningsforstyrrelser er forskjellig fra andre dimensjoner av psykopatologi. I denne sammenheng brukte de SDQ i en studie av 122 fosterbarn, 6-10 år gamle. Konfirmatorisk faktoranalyse viste god tilpasning for en modell med to ulike dimensjoner av tilknytningsforstyrrelse og de fire problemskalaene fra SDQ (RMSEA 0,04 og CFI 0,95). Dvs. funnene støtter at tilknytningsforstyrrelsene er distinkt fra problemområdene som måles ved SDQ, og analysene støtter SDQ's modell med fire problemskalaer.

Korrelasjoner mellom SDQ-P- og tilsvarende skalaer fra andre instrumenter

Indredavik og kolleger (2005) studerte samsvar mellom klassifisering i klinisk område for to spørreskjema: Child Behaviour Checklist (CBCL; Achenbach, 1991; Achenbach & Rescorla, 2001) og SDQ-P i sin studie av barn med lav fødselsvekt. Kappa var 0,66 (95% CI 0,46-0,86) for barna med veldig lav fødselsvekt, 0,72 (95% CI 0,50-0,94) for barna som var små for gestasjonsalder og 0,69 (95% CI 0,40-0,98) for kontrollgruppen.

Posserud og kolleger (2008) undersøkte forholdet mellom SDQ-P-subskalaene og de tre skalaene

Tabell 6. Grenseverdier for Totale vansker og subskalaer i britisk normdata og norske befolkningsutvalg.

	Normal	Grense- område (Lett forhøyet)	Antatt klinisk område (Høy skåre)
Totale vansker			
Britiske normer ¹	0-13	14-16	17-40
Akershus ²	0-10	11-13	14-40
Emosjonelle symptomer			
Britiske normer ¹	0-3	4	5-10
Akershus ²	0-2	3	4-10
Barn i Bergen ³			4-10
Atferdsproblemer			
Britiske normer	0-2	3	4-10
Akershus	0-2	3	4-10
Hyperaktivitet			
Britiske normer	0-5	6	7-10 ⁴
Akershus	0-4	5	6-10
Barn i Bergen			6-10
Problemer med venner			
Britiske normer	0-2	3	4-10
Akershus	0-2	3	4-10
Barn i Bergen			3-10
Prososial atferd			
Britiske normer	6-10	5	0-4
Akershus	7-10	6	0-5

¹ <http://SDQinfo.org/>, ²Van Roy 2010, ³ Munkvold et al. 2011, ⁴Nye grenseverdier er angitt på <http://SDQinfo.org/> i 2014, basert på et større datamateriale. Her er antatt klinisk område/høy skåre for hyperaktivitet 8-10.

For «Impact» er skåre 0 antatt å være normalt, 1 er grenseområde, mens ≥ 2 er antatt å være avvikende skåre, (<http://SDQinfo.org/>). I Akershus-studien ($N=8\ 018$) var «Impact»-skåren 0 for 87 %, 1 for 5 % og ≥ 2 for 9 % (Van Roy, 2010). Tilsvarende britiske tall er 83 %, 8 % og 4 %.

Social difficulties, Motor/tics/OCD og Autistic style i Autism Spectrum Screening Questionnaire (ASSQ; Ehlers, Gillberg & Wing, 1999). Utvalget besto av 6229 deltakere i Barn i Bergen med komplett utfylte ASSQ, både lærer- og foreldreversjon. Samtlige korrelasjoner var statistisk signifikante ($p < 0,001$), og korrelasjonene var høyest mellom ASSQ-skalaen Social Difficulties og SDQ-P-subskalaene (r : -0,31 til 0,66) (Tabell 7).

Mathiassen og kolleger (2012) studerte sammenhenger mellom IQ, SDQ-P og to mål for psykisk helse/barnets funksjon, vurdert av klinikere: Children's Global Assessment Scale (CGAS; Shaffer et al.) og Health of the Nation

Outcome Scales for Children and Adolescents (HoNOSCA; Gowers et al., 1999). Utvalget var barn henvist til barne- og ungdomspsykiatrisk poliklinikk. De rapporterer korrelasjoner mellom SDQ-P-subskalaene og CGAS og HoNOSCA Total skåre og også IQ: fullskala IQ (FSIQ), verbal IQ (VIQ) og utførings-IQ (performance IQ; PIQ) i Wechsler Intelligence Scale for Children, Third Edition (WISC-III; Wechsler, 1992).

Gruppeforskjeller

18 av de inkluderte studiene og delstudiene rapporterte om forskjeller i Totale vansker mellom selekterte utvalg og referansegrupper (Tabell 4a-d). De selekterte utvalgene hadde gjennomgående

Tabell 7. Korrelasjoner mellom SDQ-P-subskalaer og andre, relevante skalaer

Førsteforfatter	Referanse-instrument	Referanseskala	EMO	ATF	HYP	VEN	PRO
Posserud 2008	ASSQ	<i>Social difficulties</i>	0,44**	0,50**	0,48**	0,66**	-0,31**
		<i>Motor/tics/OCD</i>	0,26**	0,25**	0,26**	0,31**	-0,14**
		<i>Autistic style</i>	0,30**	0,27**	0,23**	0,35**	-0,06*
Mathiassen 2012	HoNOSCA Total skåre CGAS		0,07	0,25*	0,24*	0,26*	-0,24*
			-0,05	-0,23*	-0,20	-0,14	0,18
	WISC-III	FSIQ	0,12	-0,04	-0,16	-0,05	-0,11
		VIQ	0,14	-0,04	-0,15	0,04	-0,05
		PIQ	0,08	-0,03	-0,13	-0,12	-0,16

Note: * $p < 0,01$; ** $p < 0,001$; EMO=Emosjonelle symptomer; ATF=Atferdsproblemer; HYP=Hyperaktivitet; VEN=Problemer med venner; PRO=Prososial atferd; ASSQ= Autism Spectrum Screening Questionnaire; OCD=obsessive compulsive disorder (tvangsforstyrrelse); HoNOSCA= Health of the Nation Outcome Scales for Children and Adolescents; CGAS=Children's Global Assessment Scale; WISC-III=Wechsler Intelligence Scale for Children, Third Edition; FSIQ= full-scale IQ; VIQ=verbal IQ; PIQ=performance IQ

statistisk signifikant høyere gjennomsnittsskåre på Totale vansker enn referansegruppene. Se Vedlegg 2 for gruppeforskjeller i subskalaskåre.

Noen studier rapporterer ikke gjennomsnittsskåre, men forekomst av forhøyede skåre i spesielle utvalg versus kontrollgruppe. Alfstad og kollegaer (2011) fant at 110 barn med epilepsi, 8-13 år, i Helseprofil-undersøkelsen i Akershus hadde større andel forhøyede skåre for Totale vansker (gutter 42 % versus 20 %; jenter 33 % versus 14 %), «Impact» (gutter 32 % versus 16 %; jenter 30 % versus 10 %) og alle problemskalaene, unntatt Atferdsproblemer for gutter. Prososial atferd ikke var signifikant forskjellig fra kontrollgruppen ($N=13\ 564$). Munkvold og kollegaer (2011) studerte forekomst av andre problemer enn atferdsproblemer hos barn med opposisjonell atferdsforstyrrelse (ODD) ved 6-9 år i Barn i Bergen-studien. De fant signifikant høyere forekomst av skåre over 90. persentilen for Emosjonelle symptomer, Hyperaktivitet og Venneproblemer hos barn med ODD ($N=163$) enn barn uten ODD ($N=6\ 844$). F.eks. hadde 46 % av jentene ($N=41$) og 49 % av guttene ($N=122$) med ODD forhøyet skåre for Emosjonelle symptomer versus 10 % for barna som ikke hadde ODD. En tredje studie ($N=996-1001$) undersøkte om det var sammenheng mellom mors bruk av alkohol de første seks uker i svangerskapet og barnas SDQ-

skåre ved 5,5 års alder (Alvik et al., 2013). Forfatterne fant signifikante sammenhenger mellom alkoholbruk og SDQ Totale vansker, Emosjonelle symptomer, Atferdsproblemer og Hyperaktivitet. Der mor rapporterte at hun hadde drukket ≥ 5 alkoholenheter mer enn en gang i uken i de første 6 ukene av svangerskapet ($N=100$), hadde 19 % forhøyede og 36 % lett forhøyede skåre for Totale vansker, i forhold til henholdsvis 13 % og 18 % i gruppen som rapporterte slike episoder sjeldnere enn en gang i uken ($N=220$), og henholdsvis 8 % og 15% i gruppen som ikke rapporterte episoder hvor de hadde drukket ≥ 5 alkoholenheter ($N=745$).

Kriterievaliditet

Fire studier (Bjørngaas et al., 2013; Indredavik et al., 2005; Lehmann et al., 2014; Sveen et al., 2013) og to delstudier fra Barn i Bergen (Hysing et al., 2007; Ullebø et al., 2011) undersøkte SDQ-P-skårenes diagnostiske nøyaktighet ved å sammenholde dem med ulike referansestandarder (Tabell 8). Samtlige studier, bortsett fra én, brukte et diagnostisk intervju som referansestandard. Av de fem studiene som brukte diagnostiske intervju, var det kun én (Sveen et al., 2013) som oppga at intervjuerne var blindet for SDQ-P-skåre. Utvalgene i de fem studiene var forskjellige både med hensyn til størrelse, alder og forekomst av psykiske lidelser. Fire av studiene undersøkte SDQ-

Tabell 8. Studier av diagnostisk nøyaktighet

Førsteforfatter år	Utvalg	Alder (år)	N	Referanse- standard	SDQ-P-skala(er)	Diagnose(r)	Pre- valens
Sveen 2013 ¹	Befolkning	4	845	PAPA	TOT/Impact	Minst én psykiatrisk diagnose Emosjonell forstyrrelse Atferdsforstyrrelse Emosjonell og atferdsforstyrrelse	7 % 3 % 3 % 5 %
Bjørgås 2013	Cerebral parese	7-9	47	Kiddie- SADS-PL	TOT/VEN/Impact EMO ATF HYP	Minst én psykiatrisk diagnose Emosjonell forstyrrelse Atferdsforstyrrelse/ODD ADHD/ADD	57 % 11 % 9 % 51 %
Hysing 2007 ²	Kronisk sykdom	7-9	109	DAWBA	Algoritme	Minst én psykiatrisk diagnose	10 %
Ullebø 2011 ²	Befolkning/ADHD	7-9	6233	SNAP IV	HYP	ADHD	5 %
Lehmann 2014	Fosterbarn	6-12	223	DAWBA	TOT/Impact	Minst én psykiatrisk diagnose	57 %
Indredavik 2005	VLBW SGA Kontroll	14	44 51 72	Kiddie- SADS-PL	TOT	Minst én psykiatrisk diagnose	25 % 10 % 7%

Note: ¹ Tidlig trygg i Trondheim; ² Barn i Bergen; TOT=Totale vansker; VEN=Problemer med venner; EMO=Emosjonelle symptomer; ATF=Atferdsproblemer; HYP=Hyperaktivitet; ODD=Opposisjonell atferdsforstyrrelse; ADD=Attention deficit disorder; VLBW=Very Low Birthweight; SGA=Small for Gestational Age; PAPA=Preschool Age Psychiatric Assessment; DAWBA=Development and Well-Being Assessment; 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

P Totale vansker og /eller «Impact»-skalaens evne til å predikere minst én psykiatrisk diagnose (Tabell 8). I tillegg undersøkte to studier diagnostisk nøyaktighet for undergrupper av psykiatriske diagnoser, og en studie så spesifikt på ADHD. Hysing og medarbeidere (2007) benyttet en algoritme utviklet av Goodman (<http://sdqinfo.org/>) for å sammenfatte resultater for SDQ-symptomskalaer og «Impact» i en skåre for å predikere «sannsynlig», «mulig» og «lite sannsynlig» psykiatrisk diagnose.

En tests screeningegenskaper, spesielt testens prediktive evne, varierer med prevalens. Det var relativt lav forekomst av psykiske lidelser i alle utvalgene med unntak av studiene av henholdsvis fosterbarn (Lehmann et al., 2014) og barn med cerebral parese (Bjørgaas et al., 2013).

Screeningegenskaper

Et overordnet mål for en tests screeningegenskaper, dvs evnen til korrekt klassifisering av individer i forhold til det aktuelle kriteriet (ofte en diagnose) er Areal under kurven -

Area Under the Curve (AUC) i Receiver Operator Characteristics- (ROC-) analyser. Perfekt klassifisering gir AUC 1, mens AUC 0,5 ikke er bedre enn tilfeldig klassifisering. AUC for minst én diagnose predikert av Totale vansker varierte fra 0,76 til 0,84 (Tabell 9). For emosjonell forstyrrelse var AUC 0,80 når predikert av Totale vansker blant 4-åringer, og 0,82 når predikert av Emosjonelle symptomer blant fosterbarn. AUC for Atferdsforstyrrelse var 0,89 predikert av Totale vansker blant 4-åringer og også 0,89 når predikert av Atferdsproblemer i fosterbarnutvalget. AUC var 0,83 for kategorien emosjonell og atferdsforstyrrelse hos 4-åringer. For ADHD var AUC 0,91 og 0,81 i henholdsvis befolknings- og fosterbarnutvalg.

Fra «Tidlig trygg i Trondheim» rapporterte Sveen og medarbeidere (2013) også AUC for «Impact»-skalaen. AUC var lavere enn for problemskalaene, 0,62 (95% konfidensintervall 0,57-0,67) for prediksjon av «minst en psykiatrisk diagnose» og varierte fra 0,64 til 0,71 for de ulike

Tabell 9. Evne til å skille mellom barn med og uten psykiatrisk diagnose. «Area Under the Curve» - AUC (95 % konfidensintervall)

Førsteforfatter år	Minst en psykiatrisk diagnose	Emosjonell forstyrrelse	Atferdsforstyrrelse	Emosjonell og atferdsforstyrrelse	ADHD
Sveen 2013 ¹	0,76 (0,68-0,83)	0,80 (0,72-0,89)	0,89 (0,81-0,97)	0,83 (0,76-0,90)	
Ullebø 2011					0,91 (0,90-0,92) ²
Lehmann 2014	0,83 (0,78-0,88) ¹	0,82 (0,76-0,88) ³	0,89 (0,84-0,93) ⁴		0,81 (0,74-0,87) ²
Indredavik 2005 ¹					
VLBW	0,81 (0,67-0,94)				
SGA	0,84 (0,70-0,98)				
Kontroll	0,79 (0,60-0,99)				

Note: ¹ SDQ-P-prediktor: Totale vansker; ² SDQ-P-prediktor: Hyperaktivitet; ³ SDQ-P-prediktor: Emosjonelle symptomer; ⁴ SDQ-P-prediktor: Atferdsproblemer; VLBW=Very Low Birthweight; SGA=Small for Gestational Age

diagnosegruppene. I sin studie av fosterbarn vurderte Lehmann og medarbeidere (2014) screeningegenskapene for Totale vansker, symptomskalaer (Tabell 9) og «Impact»-skalaen. AUC for «Impact»-skåren var 0,80 (95% CI 0,75-0,86).

Evne til å identifisere barn med psykisk lidelse

Sensitivitet og spesifisitet karakteriserer testens evne til å identifisere de som har og de som ikke har et problem eller en diagnose. Ved en gitt grenseverdi for testen angir sensitivitet andelen med positiv testskåre av de som har en diagnose, mens spesifisitet angir andelen med negativ testskåre av de som ikke har en diagnose (Tabell 10).

I studien av 4-åringer i Tidlig trygg i Trondheim fant Sveen og medarbeidere (2013) at grenseverdi for Totale vansker som ga best kombinasjon av sensitivitet og spesifisitet, var ≥ 10 . Ved denne grenseverdien var sensitiviteten mellom 0,54 (minst én diagnose) og 0,79 (atferdsforstyrrelse),

mens spesifisiteten lå stabilt på 0,88-0,89 (Tabell 11). Blant 4-åringer i befolkningsutvalg ser det altså ut til at Totale vansker ≥ 10 kun identifiserer 54 % av dem med minst én psykiatrisk diagnose, men 79 % av dem med en atferdsforstyrrelse. På den annen side vil nesten 90 % av dem uten noen diagnose skåre under grenseverdien. Den positive prediktive verdien viser dessuten at det kun var 26 % (95% konfidensintervall 20 % - 32 %) av de med Totale vansker ≥ 10 som faktisk hadde noen psykiatrisk diagnose. Dette kan også formuleres som at tre fjerdedeler av de som skårer over grenseverdien ikke har noen diagnose (falske positive). Sveen og medarbeidere gjorde beregninger av PPV i utvalg med høyere forekomst av psykiske lidelser. Ved 10 % prevalens var PPV 0,35, og ved 20 % prevalens 0,55.

Bjørgås og medarbeidere (2013) brukte 90. persentilen som grenseverdi for alle skalaer. Også i denne studien var det stor variasjon i sensitivitet - fra 0,13 (ADHD/ADD) til 1,0 (minst én diagnose predikert av Problemer med venner) (Tabell 11).

Tabell 10. 2 x 2 tabell som viser hvordan en test identifiserer de som har et problem eller en diagnose.

	Diagnose/Problem +	Diagnose/Problem -	
Test +	SP - Sanne positive	FP - Falske positive	SP+FP
Test -	FN - Falske negative	SN - Sanne negative	FN+SN
	SP+FN	FP+SN	

Sensitivitet: $SP/(SP+FN)$; Spesifisitet: $SN/(FP+SN)$; Positiv prediktiv verdi: $SP/(SP+FP)$; Negativ prediktiv verdi: $SN/(FN+SN)$

Spesifisiteten varierte fra 0,25 (minst én diagnose predikert av Problemer med venner) til 0,87 (ADHD/ADD). I dette utvalget identifiserte altså Hyperaktivitet-skårer over 90. persentilen kun 13 % av dem med en ADHD-diagnose, og tilsvarende skårer på Problemer med venner identifiserte samtlige deltakere med minst én psykisk lidelse. Kun 25 % av barna uten noen psykiatrisk diagnose skåret under 90. persentilen på Problemer med venner, mens 89 % av dem uten ADHD-diagnose lå under 90. persentilen på Hyperaktivitet-skalaen.

Hysing og medarbeidere (2007) brukte en algoritme som kombinerer symptomer og «Impact» (Goodman, Renfrew & Mullick, 2000), og rapporterte sensitivitet (0,77) og spesifisitet (0,89) for kategorien «sannsynlig diagnose» («probable») (Tabell 11). Dette betyr at kategorien «sannsynlig diagnose» korrekt identifiserer henholdsvis 77 % og 89 % av dem som har og ikke har noen diagnose. Når algoritmen bare var basert på symptomer, var sensitiviteten 0,82 og spesifisiteten 0,69.

Ullebø og medarbeidere (2011) fant at en relativt lav grenseverdi for Hyperaktivitet-skalaen, 4 eller 5, var nødvendig for å oppnå god sensitivitet. Ved grenseverdi 5 var sensitiviteten 0,79, spesifisiteten 0,86 og PPV 0,24 (ved prevalens: 5 %) (Tabell 11). Hyperaktivitet ≥ 5 fanger altså opp 79 % av dem med ADHD og utelukker diagnose hos 86 % av dem som ikke har ADHD. Andelen falske positive blir imidlertid relativt høy: det er kun 24 % av dem med Hyperaktivitet ≥ 5 som faktisk har ADHD.

Basert på ROC-analysene i studien av fosterbarn anså Lehmann og medarbeidere grenseverdi 13 for Totale vansker som den mest optimale verdi i forhold til sensitivitet og spesifisitet. Sensitiviteten og spesifisiteten på henholdsvis 0,83 og 0,74 (Tabell 11) viser at Totale vansker ≥ 13 korrekt fanger opp 83 % av fosterbarn med en diagnose, og utelukker diagnose hos 74 % som ikke har noen. PPV på 0,81 betyr at hele 81 % av fosterbarn med Totale vansker ≥ 13 har en diagnose. For «Impact» var 2 den optimale grenseverdien, og den

Tabell 11. Evne til å identifisere barn med psykisk lidelse. Sensitivitet, spesifisitet, og positiv og negativ prediktiv verdi (PPV/NPV).

Førsteforfatter år	SDQ-P-skala	Grense-verdi	Diagnose	Sensitivitet	Spesifisitet	PPV	NPV
Sveen 2013	TOT	≥ 10	Minst én diagnose	0,54	0,89	0,26	0,97
			Emosjonell forstyrrelse	0,59	0,88	0,14	0,98
			Atferdsforstyrrelse	0,79	0,88	0,17	0,99
			Emosjonell og atferdsforstyrrelse	0,65	0,89	0,24	0,98
Björgås 2013	TOT	90. persentilen	Minst én diagnose	0,85	0,55	0,71	0,73
	VEN		1,00	0,25	0,63	1,00	
	Impact		0,74	0,65	0,74	0,65	
	EMO		1,00	0,79	0,36	1,00	
	ATF		0,50	0,67	0,13	0,93	
Hysing 2007	Algoritme		Atferdsforstyrrelse/ODD	0,13	0,87	0,50	0,49
			ADHD/ADD	0,77	0,89		
Ullebø 2011	Hyperaktivitet ¹	≥ 5	ADHD	0,79	0,86	0,24	
Lehmann 2014	Totale vansker	≥ 13	Minst én diagnose	0,83	0,74	0,81	0,76
	Impact	≥ 2		0,80	0,70	0,78	0,72
Indredavik 2005	Totale vansker	90. persentilen	Minst én diagnose				
	VLBW		0,83	0,58			
	SGA		0,67	0,79			
	Kontroll		0,40	0,93			

Noter: EMO=Emosjonelle symptomer; ATF=Atferdsproblemer; HYP=Hyperaktivitet; VEN=Problemer med venner; ODD=Opposisjonell atferdsforstyrrelse; ADD=Attention deficit disorder; PPV=Positiv prediktiv verdi; NPV=Negativ prediktiv verdi; VLBW=Very Low Birthweight; SGA=Small for Gestational Age; ¹Algoritmekategori «Sannsynlig diagnose» versus «Mulig diagnose» og «Lite sannsynlig diagnose» (basert på Totale vansker og Impact).

diagnostiske nøyaktigheten noe lavere enn for Totale vansker ≥ 13 (Tabell 11). Når *enten* Totale vansker var ≥ 13 *eller* «Impact» var ≥ 2 , økte sensitiviteten til 0,89, mens spesifisiteten ble dårligere (0,62). Når *både* Totale vansker var ≥ 13 og «Impact» ≥ 2 , ble spesifisiteten bedret (0,81), mens sensitiviteten ble noe lavere (0,73).

Indredavik og medarbeidere (2005) brukte også 90. persentilen for Totale vansker som grenseverdi for «minst en psykiatrisk diagnose». Utvalget i denne studien var delt i tre grupper som hver for seg var relativt små: barn med svært lav fødselsvekt (VLBW), barn som var små for gestasjonsalder (SGA) og kontrollgruppe. Funnene for sensitivitet og spesifisitet varierte mellom gruppene. Sensitiviteten var høyest for VLBW-barna, og

spesifisiteten var høyest for kontrollgruppen (Tabell 11).

Reliabilitet

21 av artiklene rapporterte indre konsistens (Cronbachs α) for SDQ-P-skalaene (Tabell 12). For delskalaene varierte median α fra 0,55 for Atferdsproblemer til 0,77 for Hyperaktivitet.

Diskusjon og konklusjon

Litteratursøket identifiserte 66 publikasjoner fordelt på 34 studier som hadde brukt SDQ-P. Det har vært omfattende norsk forskning på SDQ-P, med fokus på psykometriske egenskaper, data fra befolkningsutvalg og mange grupper med

Tabell 12. Indre konsistens (Cronbachs alfa)

Førsteforfatter år	N	TOT	EMO	ATF	HYP	VEN	PRO
Brandlistuen 2011	44104					0,75	
Feragen 2014	182	0,84	0,65	0,58	0,80	0,64	0,62
Fjermestad 2015	98	0,73					
	25	0,76					
Gjertsen 2010	124	0,79	0,67	0,57	0,76	0,47	0,72
Helgeland 2011	94	0,79	0,68	0,45	0,79	0,56	0,65
Holen 2012a	1272		0,70	0,56	0,77	0,57	0,68
Kjeldsen 2014	535			0,48			
Kleven 2008	175		0,45			0,43	
Kvalevaag 2015	28695						0,77
Lehmann 2015	122	0,88	0,69	0,75	0,82	0,70	
Mathiesen 2006	401	0,80					
Nilsen 2013	505		0,65	0,47			
Nilsen 2015	77	0,75					
Posserud 2008	6229	0,82					
Sanne 2009	3166		0,65	0,55	0,75	0,61	0,63
Sherr 2014	220	0,73					0,75
		0,74					0,80
Skar 2015	141	0,73					0,75
Sveen 2013	845	0,77					
Sørensen 2012	6403		0,68				
van Roy 2008	6645	0,78	0,67	0,5	0,76	0,56	0,66
Median		0,79	0,67	0,55	0,77	0,57	0,70

TOT=Totale vansker; EMO=Emosjonelle symptomer; ATF=Atferdsproblemer; HYP=Hyperaktivitet; VEN=Problemer med venner; PRO=Prososial atferd

spesifikke problemer eller risikofaktorer.

Normer / grenseverdier for avvik

Det foreligger resultater fra fire relativt store befolkningsbaserte undersøkelser med høye svarprosent, men de har alle lokalt avgrensede studiepopulasjoner. De fleste utvalgene er også avgrenset i alder, men til sammen dekker de aldersgrupper fra 4 til 13 år. Dataene kan ikke brukes som norske nasjonale normer. Helseprofil Akershus har størst materiale med bredest aldersspekter, fra 8 til 13 år.

Ingen av de norske studiene har hatt som primært siktemål å rapportere SDQ-data fra befolkningsbaserte utvalg. Resultatene som vi har inkludert, rapporteres ofte fra underutvalg og ikke hele utvalget. Vi har derfor estimert gjennomsnittsskårer for totalutvalg når det foreligger data som gjør dette mulig, og vi har også brukt data fra kontrollgrupper (supernormale utvalg). De forskjellige underutvalgene kan være mer eller mindre representative. For eksempel rapporterer to artikler resultater ved 10-13 år fra Helseprofil Akershus (van Roy et al, 2009, van Roy et al, 2010). Artikkelen fra 2010 har et større utvalg (N=8517) enn artikkelen fra 2009 som studerer barn med sosial angst og en kontrollgruppe (Total N=7763). Vi estimerte totale vansker for totalutvalgene i de to artiklene og fant at totale vansker var 6,2 i det største utvalget versus 5,9 i den noe mindre studien. Vi antar at resultatene fra det større utvalget er mer representativt for populasjonen.

En studie av barn i tilsvarende alder fra Barn i Bergen (Hysing et al., 2009) fant mindre problemer enn i Akershus (estimert gjennomsnitt for totale vansker 5,3). Dette er bølge 2 i en longitudinell studie, og frafallet fra bølge 1 er ikke ubetydelig (6658 deltakere i bølge 1, 5781 i bølge 2), og selektivt frafall kan være av betydning for resultatet. Hoekstra og medarbeidere (2013) rapporterer resultater for 4025 barn hvor det foreligger data fra både lærere og foreldre fra begge bølger. Fokus for studien var barn med tics (N=38) For kontrollgruppen (barn uten tics, N=3 847) var gjennomsnittlig totale vansker 5,3 (SD 4,4) i bølge 1 og 4,9 (SD 4,5) i bølge 2, lavere skårer

enn andre studier fra samme materialet, både i bølge 1 (Heiervang et al., 2008) og Hysing og medarbeideres (2009) resultater fra bølge 2. Dette indikerer at resultatene fra runde 2 er mindre representative for populasjonen. Resultatene fra studien fra Helseprofil Akershus med størst utvalg er sannsynligvis de beste normdataene for aldersgruppen 10-13 år. For aldersgruppen 7-10 år er det tre store studier som rapporterer resultater fra Bergen, Akershus og Romsdal, og variasjonen er liten (totale vansker 5,9-6,2).

Kjønnforskjellene i skårer støtter annen empiri om mer problematferd hos gutter enn hos jenter, og mer prososial atferd hos jenter enn hos gutter. En studie sammenligner resultater fra Bergen og Storbritannia for 7-9 år gamle barn (Heiervang et al., 2008). Man fant lavere SDQ-T og SDQ-P-skårer for barna i Bergen, som samsvarte med lavere prevalenstall for de norske barna for ikke-emosjonelle forstyrrelser. For emosjonelle forstyrrelser derimot, hvor også de norske foreldrene og lærerne skåret barna lavere på SDQ enn de britiske, var det ikke forskjell i prevalens mellom landene. Dette kan indikere at det er forskjeller i rapporteringsstil eller terskel for rapportering av emosjonelle problemer hos foreldre i de to landene. Det var tilsvarende funn for emosjonelle forstyrrelser for lærernes SDQ-skårer. Dette viser også at krysskulturelle forskjeller i spørreskjema-resultater ikke nødvendigvis kan tolkes som reelle forskjeller i mental helse.

Vi fant ingen statistisk påviste forskjeller mellom mødre- og fedrerapporterte SDQ-P-skårer. Men en visuell sammenlikning viser en tendens til at mødrerapporterte problemskårer lå noe høyere enn fedrerapporterte skårer.

Når SDQ skåres ved håndskåring eller på nettet, brukes britiske normer, og den britisk 80. og 90. persentilen benyttes som grenseverdier. Det er viktig å vite betydningen av grenseverdier. Anna og Robert Goodman (2009) har studert om SDQ er et dimensjonalt mål eller om det er terskelverdier i forhold til identifikasjon av psykiatriske lidelser. De fant gjennomgående at barn med høyere total vanske skåre hadde mer psykopatologi og

konkluderte at SDQ er et dimensjonalt mål for mental helse.

Helseprofil Akershus har angitt grenseverdier for avvik i sitt befolkningsutvalg (Van Roy 2010). I samsvar med funnene om lavere gjennomsnittsskåre i Norge versus Storbritannia, viser Tabell 6 at de norske grenseverdiene er lavere enn de britiske. For totale vansker er grenseverdi for «høy» skåre 14 i Akershus og 17 i det britiske norm-materialet, for «lett forhøyet» skåre 11 i Akershus og 14 i det britiske utvalget. En delstudie fra Barn i Bergen (Munkvold et al., 2011) har definert grenseverdier for «høy» skåre for Emosjonelle symptomer, Hyperaktivitet og Venneproblemer. Disse verdiene er lik resultatene fra Akershus, med unntak av Venneproblemer hvor Barn i Bergen setter grenseverdi 3 for høy skåre, mens Akershus bruker 4. Resultater som skåres med det britiske normgrunnlaget, må tolkes med forsiktighet.

Diagnostisk nøyaktighet

Van Roy (2010) har definert grenseverdier for avvik, basert på en gitt andel som skårer «forhøyet» eller «lett forhøyet». En annen tilnærming er å studere hvordan man ved å bruke ulike grenseverdier klarer å klassifisere korrekt de som har et reelt problem (et kriterium, ofte en diagnose). Dette er et viktig aspekt ved instrumentets validitet, kriterievaliditet. Seks studier har undersøkt diagnostisk nøyaktighet i ulike utvalg. AUC som er et overordnet mål for diagnostisk nøyaktighet, varierte noe med type utvalg og type diagnostikk. Generelt var AUC meget tilfredsstillende i alle studiene. Basert på disse analysene og resultater for sensitivitet og spesifisitet, kan man velge grenseverdier som gir best kombinasjon av sensitivitet og spesifisitet.

Sensitivitet og spesifisitet varierer med verdien som definerer avvik for testen. For eksempel vil sensitivitet øke hvis grenseverdien man bruker for å identifisere de med problemer, settes lavere. Da vil man fange opp flere med problemer, men også flere som ikke har problemer, dvs flere «falske positive». Sensitivitet og spesifisitet må ses i sammenheng, for når sensitiviteten øker, vil spesifisiteten reduseres og motsatt.

Grenseverdiene for testen må velges utfra formålet. Det kan være viktig med høy sensitivitet, slik at flest mulig med problemer fanges opp, og det er få falske negative. I andre sammenhenger er høy spesifisitet viktigst, slik at man får få falske positive. I praksis er testens prediktive evne viktig. Positiv prediktiv verdi (PPV) angir andelen med en diagnose av de som skårer positivt. Prediktiv verdi varierer med prevalens av problemer.

I studien av 4-åringer i Tidlig trygg i Trondheim fant Sveen og medarbeidere (2013) at grenseverdi for Totale vansker som ga best kombinasjon av sensitivitet og spesifisitet, var ≥ 10 . Dette er en verdi i normalområdet i det britiske norm-materialet, og ved overgang til lett forhøyet område i van Roys studie (2010) fra Helseprofil i Akershus. Selv om grenseverdien var lav, var sensitiviteten likevel lav ved denne verdien; sensitivitet/ spesifisitet 0,54/0,89 i forhold til «minst en psykiatrisk diagnose». Sensiviteten var høyere i forhold til atferdsforstyrrelser (0,79, spesifisitet 0,88) enn i forhold til emosjonelle lidelser (0,59, spesifisitet 0,88). Forfatterne understreker betydningen av en lav grenseverdi hvis SDQ skal brukes i kommunale tjenester med lav forekomst av psykiske lidelser. I sitt befolkningsutvalg finner de da at positiv prediktiv verdi (PPV) er lav i forhold til å identifisere «minst en psykiatrisk diagnose» [0,26 (95% CI 0,20-0,32)], dvs mange falske positive testresultater. Forfatterne gjør beregninger av PPV i utvalg med høyere forekomst av psykiske lidelser. Ved 10% prevalens er PPV 0,35, ved 20% prevalens 0,55.

Til forskjell fra denne studien, fant Lehmann og medarbeidere (2014) at grenseverdi 13 for Totale vansker var optimalt. Totale vansker 13 er i øvre del av normalområdet i det britiske normmaterialet, men i øvre del av lett forhøyet område i van Roys data (van Roy, 2010) fra Helseprofil Akershus. Hvis man ønsker å finne barn som bør følges opp videre (dvs unngå falske negative testresultater), anbefaler forfatterne at man identifiserer barn som enten har totale vansker ≥ 13 eller «Impact» ≥ 2 .

Det er flere forskjeller mellom disse to studiene som kan være av betydning for at man fant ulike

optimale grenseverdier. Lehmanns studie brukte et mer strukturert intervju (DAWBA), barna i fosterhjem hadde høy forekomst av psykiatriske lidelser og mer heterogen alder (6-12 år). Sveen og medarbeideres studie brukte et semistrukturert intervju (PAPA) i et befolkningsbasert utvalg med yngre barn som alle var 4 år.

Ullebø og medarbeideres studie (2011) av 7-9-åringene med ADHD fant at ved grenseverdi 5 for Hyperaktivitets-skalaen (i normalområdet basert på britisk normering, i overgangen til lett forhøyet område basert på van Roys data fra Helseprofil Akershus) var sensitiviteten 0,79 spesifisiteten 0,86 og PPV 0,24 (ved prevalens: 5 %).

Björgås og medarbeideres studie (2013) av barn med CP var liten (N=47), men det var høy forekomst av psykiatriske lidelser, primært ADHD/ADD. I motsetning til Ullebøs studie av barn med ADHD, fant Björgås og kolleger at Hyperaktivitetsskalaen var lite sensitiv i forhold til å identifisere barna med ADHD (sensitivitet 0,13, spesifisitet 0,87). Sensitivitet/spesifisitet for totale vansker i forhold til «minst en psykiatrisk diagnose» var 0.85/0,55 med PPV 0,71 og NPV 0,73. 41 av 47 barn med CP hadde venneproblemer. Det er derfor ikke overraskende at Venneproblemer i denne studien hadde høy sensitivitet (1,00), men lav spesifisitet (0,25) i forhold til «minst en psykiatrisk diagnose».

For undergrupper av 14-åringene med lav fødselsvekt hadde 90. persentilen for Totale vansker utmerket evne til å skille mellom individer med og uten minst én psykiatrisk diagnose, mens denne evnen var mindre pålitelig for en jevngammel kontrollgruppe. Dette var en liten studie, og forekomsten av problemer var lav, ikke minst i kontrollgruppen.

De ulike studiene har litt forskjellige konklusjoner vedrørende betydning av «Impact»-skåren for den diagnostiske nøyaktighet. Sveen og medarbeidere (2013) fant at den diagnostiske nøyaktighet ikke ble bedret ved å inkludere «Impact»-skåre, mens Lehmann og medarbeidere (2014) anbefaler å bruke både symptomer og «Impact». Hysing og medarbeidere (2007) brukte en algoritme som

kombinerer symptomer og «Impact»-skåren og fant at spesifisiteten ble lav for symptomer alene.

Andre psykometriske egenskaper

Konfirmatoriske faktoranalyser og forventete gruppeforskjeller fra svært mange studier støtter opp under instrumentets begrepsvaliditet. Tre studier har undersøkt korrelasjoner mellom SDQ og skalaer fra andre instrumenter, men bare én av studiene (Indredavik et al., 2005) sammenlignet med et annet spørreskjema som måler generelle psykiske helseproblemer (CBCL). De fant at samsvaret mellom CBCL og SDQ-P var «godt». Mathiassen og kolleger (2012) fant, som forventet, signifikante korrelasjoner mellom SDQ-P (med unntak av Emosjonelle symptomer) og en kliniker vurdert skåre, HoNOSCA Total skåre, mens de for CGAS bare fant signifikant korrelasjon for Atferdsproblemer. Det var ingen korrelasjon mellom SDQ-P skalaer og IQ, hvilket støtter diskriminant validitet. Barn i Bergen-forskerne fant at SDQ-P-skalaene korrelerte, som forventet, med subskalaene i Autism Spectrum Screening Questionnaire (ASSQ), med høyest korrelasjon mellom ASSQ Social Difficulties og SDQ Venneproblemer ($r=0,66$).

Mål på indre konsistens var rapportert fra mange studier, men med stor variasjon i utvalgsstørrelse. Flere skalaer hadde inadekvat median indre konsistens (Emosjonelle symptomer, Atferdsproblemer og Problemer med venner), mens de øvrige skalaene hadde tilfredsstillende verdier. Resultatene var dårligere enn for SDQ-T-skalaen (Kornør & Heyerdahl, 2013), men bedre enn den norske selvrapporteringsversjonen av SDQ (SDQ-S), hvor kun én av skalaene hadde akseptabel indre konsistens (Kornør & Heyerdahl, 2013). Resultatene er i tråd med funn fra en nasjonal britisk studie (Goodman, 2001).

«Impact»-skåren kan være et nyttig supplement til problemskåren, men resultatene er ikke konsistente i forhold til hvilken betydning det har å inkludere «Impact»-skåren.

Konklusjon

Det foreligger store datasett for SDQ-P i befolkningsutvalg, men ikke nasjonale normer. Resultatene viser lavere skåre for

problemskalaene enn de britiske normene. Flere gode studier av diagnostisk nøyaktighet viser god kriterievaliditet og at grenseverdiene for avvik må være lavere enn de britiske grenseverdiene for å få optimal sensitivitet/spesifisitet. Det er behov for mere forskning for å få klarlagt hva som er mest optimal grenseverdi for å identifisere barn med problemer. I praksis må grenseverdi velges utfra problemstillingen, hvor viktig det er å identifisere flest mulig av de som har reelle problemer, eller om det er viktig å unngå «falske positive». Det er behov for bedre normdata, ikke minst for eldre

barn og ungdommer, og også studier som bruker SDQ for både foreldre, lærere og ungdom.

Den norske versjonen av SDQ-P kan være egnet til å screene for psykiatriske diagnoser i selekterte utvalg med risiko for psykiske helseproblemer. Det er ikke grunnlag for å anbefale SDQ-P for universell screening. Man må være oppmerksom på at den prediktive verdi kan være lav slik at det kan bli relativt mange falske positive og/ eller negative hvis instrumentet brukes til screening i normalbefolkning.

Referanser

- Aasheim, R. (2012). *Psykisk helse blant barnevernsbarn. En studie av psykiske vansker målt med kartleggingsverktøyet Strengths and Difficulties Questionnaire*. Master thesis. Universitetet i Bergen, Bergen.
- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, Vt.: Dept. of Psychiatry, University of Vermont.
- Achenbach, T. M., Becker, A., Dopfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C. et al. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 49(3), 251-275. doi: <http://dx.doi.org/10.1111/j.1469-7610.2007.01867.x>
- Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Alfstad, K. A., Clench-Aas, J., Van Roy, B., Mowinckel, P., Gjerstad, L. & Lossius, M. I. (2011). Psychiatric symptoms in Norwegian children with epilepsy aged 8-13 years: Effects of age and gender? *Epilepsia*, 52(7), 1231-1238. doi: <http://dx.doi.org/10.1111/j.1528-1167.2011.03042.x>
- Alvik, A., Aalen, O. O. & Lindemann, R. (2013). Early fetal binge alcohol exposure predicts high behavioral symptom scores in 5.5-year-old children. *Alcoholism: Clinical and Experimental Research*, 37(11), 1954-1962. doi: <http://dx.doi.org/10.1111/acer.12182>
- Bjørngaas, H. M., Elgen, I., Bøe, T. & Hysing, M. (2013). Mental health in children with cerebral palsy: does screening capture the complexity? *TheScientificWorldJournal*, 2013, 468402.
- Brandlistuen, R. E., Stene-Larsen, K., Holmstrom, H., Landolt, M. A., Eskedal, L. T. & Vollrath, M. E. (2011). Occurrence and predictors of developmental impairments in 3-year-old children with congenital heart defects. *Journal of Developmental and Behavioral Pediatrics*, 32(7), 526-532. doi: <http://dx.doi.org/10.1097/DBP.0b013e318222eb5a>
- Clucas, C., Skar, A.-M. S., Sherr, L. & Tetzchner, S. v. (2014). Mothers and fathers attending the international child development programme in Norway. *The Family Journal*, 22(4), 409-418. doi: <http://dx.doi.org/10.1177/1066480714533640>
- Diseth, T. H., Tangeraas, T., Reinfjell, T. & Bjerre, A. (2011). Kidney transplantation in childhood: Mental health and quality of life of children and caregivers. *Pediatric Nephrology*, 26(10), 1881-1892. doi: <http://dx.doi.org/10.1007/s00467-011-1887-9>
- Egger, H. L., Erkanli, A., Keeler, G., Potts, E., Walter, B. K. & Angold, A. (2006). Test-Retest Reliability of the Preschool Age Psychiatric Assessment (PAPA). *J Am Acad Child Adolesc Psychiatry*, 45(5), 538-549. doi: 10.1097/01.chi.0000205705.71194.b8
- Ehlers, S., Gillberg, C. & Wing, L. (1999). A Screening Questionnaire for Asperger Syndrome and Other High-Functioning Autism Spectrum Disorders in School Age Children. *Journal of Autism and Developmental Disorders*, 29(2), 129-141. doi: 10.1023/A:1023040610384
- Eilertsen, M. E., Rannestad, T., Indredavik, M. S. & Vik, T. (2011). Psychosocial health in children and adolescents surviving cancer. *Scand J Caring Sci*, 25(4), 725-734. doi: 10.1111/j.1471-6712.2011.00883.x
- Eknorås, 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>
- Elgen, S. K., Leversen, K. T., Grundt, J. H., Hurum, J., Sundby, A. B., Elgen, I. B. et al. (2012). Mental health at 5 years among children born extremely preterm: A national population-based study. *European Child and Adolescent Psychiatry*, 21(10), 583-589. doi: <http://dx.doi.org/10.1007/s00787-012-0298-1>
- 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.
- Feragen, K. B. & Stock, N. M. (2014). When there is more than a cleft: Psychological adjustment when a cleft is associated with an additional condition. *Cleft Palate-Craniofacial Journal*, 51(1), 5-14. doi: <http://dx.doi.org/10.1597/12-328>
- Fevang, S. K. E., Hysing, M., Markestad, T. & Sommerfelt, K. (2016). Mental Health in Children Born Extremely Preterm Without Severe Neurodevelopmental Disabilities. *Pediatrics*. doi: 10.1542/peds.2015-3002
- Fjermestad, K. W., Stokke, S., Wergeland, G. J. H., Anticich, S., Haugland, B. S. M., Havik, O. E. et al. (2015). Socio-emotional problems in boys with sex chromosome aneuploidies compared to a clinical sample. *Children's Health Care*, 44(1), 40-53. doi: <http://dx.doi.org/10.1080/02739615.2013.876538>
- 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
- Freuchen, A., Kjelsberg, E., Lundervold, A. J. & Groholt, B. (2012). Differences between children and adolescents who commit suicide and their peers: A psychological autopsy of suicide victims compared to accident victims and a community sample. *Child and Adolescent Psychiatry and Mental Health Vol 6 Jan 2012, ArtID 1, 6*. doi: <http://dx.doi.org/10.1186/1753-2000-6-1>
- Gjertsen, K. & Sanne, E. (2010). *Synshemming - En faktor som bidrar til å påvirke psykososial tilpassing hos grunnskoleelever i Aust-Agder, Vest-Agder og Rogaland?* Mastergrad. Norges Teknisk-Naturvitenskapelige Universitet, Trondheim.
- Goodman, A. & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J Am Acad Child Adolesc Psychiatry*, 48(4), 400-403. doi: 10.1097/CHI.0b013e3181985068
- Goodman, A., Lamping, D. & Ploubidis, G. (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. (1999). The Extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. *Journal of Child Psychology and Psychiatry*, 40(5), 791-799. doi: 10.1111/1469-7610.00494
- Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345. doi: 10.1097/00004583-200111000-00015
- 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. *J Child Psychol Psychiatry*, 41(5), 645-655.
- Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, 177(6), 534-539. doi: 10.1192/bjp.177.6.534
- 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. *Eur Child Adolesc Psychiatry*, 9(2), 129-134.
- Gowers, S. G., Harrington, R. C., Whitton, A., Lelliott, P., Beevor, A., Wing, J. et al. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for children and Adolescents (HoNOSCA). *The British Journal of Psychiatry*, 174(5), 413-416. doi: 10.1192/bjp.174.5.413
- Hanssen-Bauer, K., Heyerdahl, S. & Eriksson, A.-S. (2007). Mental health problems in children and adolescents referred to a national epilepsy center. *Epilepsy & Behavior*, 10(2), 255-262. doi: 10.1016/j.yebeh.2006.11.011

- Heiervang, E., Goodman, A. & Goodman, R. (2008). The Nordic advantage in child mental health: Separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 49(6), 678-685. doi: <http://dx.doi.org/10.1111/j.1469-7610.2008.01882.x>
- Heiervang, E., Stormark, K. M., Lundervold, A. J., Heimann, M., Goodman, R., Posserud, M. B. et al. (2007). Psychiatric disorders in Norwegian 8- to 10-year-olds: An epidemiological survey of prevalence, risk factors, and service use. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(4), 438-447. doi: <http://dx.doi.org/10.1097/chi.0b013e31803062bf>
- Helgeland, H., Van Roy, B., Sandvik, L., Markestad, T. & Kristensen, H. (2011). Paediatric functional abdominal pain: Significance of child and maternal health a prospective study. *Acta Paediatrica*, 100(11), 1461-1467. doi: <http://dx.doi.org/10.1111/j.1651-2227.2011.02349.x>
- Helland, W. A., Helland, T. & Heimann, M. (2014). Language Profiles and Mental Health Problems in Children With Specific Language Impairment and Children With ADHD. *Journal of Attention Disorders*, 18(3), 226-235. doi: <http://dx.doi.org/10.1177/1087054712441705>
- Helland, W. A., Lundervold, A. J., Heimann, M. & Posserud, M. B. (2014). Stable associations between behavioral problems and language impairments across childhood - The importance of pragmatic language problems. *Research in Developmental Disabilities*, 35(5), 943-951. doi: <http://dx.doi.org/10.1016/j.ridd.2014.02.016>
- Hoekstra, P. J., Lundervold, A. J., Lie, S. A., Gillberg, C. & Plessen, K. J. (2013). Emotional development in children with tics: A longitudinal population-based study. *European Child and Adolescent Psychiatry*, 22(3), 185-192. doi: <http://dx.doi.org/10.1007/s00787-012-0337-y>
- Holen, S., Lervåg, A., Waaktaar, T. & Ystgaard, M. (2012). Exploring the associations between coping patterns for everyday stressors and mental health in young schoolchildren. *Journal of School Psychology*, 50(2), 167-193. doi: <http://dx.doi.org/10.1016/j.jsp.2011.10.006>
- Holen, S., Waaktaar, T., Lervåg, A. & Ystgaard, M. (2012). The effectiveness of a universal school-based programme on coping and mental health: A randomised, controlled study of Zippy's Friends. *Educational Psychology*, 32(5), 657-677. doi: <http://dx.doi.org/10.1080/01443410.2012.686152>
- 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>
- Hysing, M., Sivertsen, B., Stormark, K. M., Elgen, I. & Lundervold, A. J. (2009). Sleep in children with chronic illness, and the relation to emotional and behavioral problems-A population-based study. *Journal of Pediatric Psychology*, 34(6), 665-670. doi: <http://dx.doi.org/10.1093/jpepsy/jsn095>
- Indredavik, M. S., Vik, T., Heyerdahl, S., Kulseng, S. & Brubakk, A.-M. (2005). Psychiatric symptoms in low birth weight adolescents, assessed by screening questionnaires. *European Child & Adolescent Psychiatry*, 14(4), 226-236. doi: 10.1007/s00787-005-0459-6
- Iversen, A. C., Jakobsen, R., Havik, T., Hysing, M. & Stormark, K. M. (2007). Mental health problems among child welfare clients living at home. *Child Care in Practice*, 13(4), 387-399. doi: <http://dx.doi.org/10.1080/13575270701488790>
- Javo, C., Ronning, J. A., Handegard, B. H. & Rudmin, F. W. (2009). Cross-informant correlations on social competence and behavioral problems in Sami and Norwegian

- preadolescents. *European Child and Adolescent Psychiatry*, 18(3), 154-163. doi: <http://dx.doi.org/10.1007/s00787-008-0714-8>
- Kjeldsen, A., Janson, H., Stoolmiller, M., Torgersen, L. & Mathiesen, K. S. (2014). Externalising behaviour from infancy to mid-adolescence: Latent profiles and early predictors. *Journal of Applied Developmental Psychology*, 35(1), 25-34. doi: <http://dx.doi.org/10.1016/j.appdev.2013.11.003>
- Kleven, L. S. (2008). *Språk- og sosioemosjonelle vansker: en studie av sammenhenger mellom språk og sosioemosjonelle faktorer hos norske 4-åringer*. Universitet i Oslo, Oslo.
- Kokkvoll, A., Grimsgaard, S., Steinsbekk, S., Flaegstad, T. & Njolstad, I. (2014). Health in overweight children: 2-year follow-up of Finnmark Activity School-a randomised trial. *Arch Dis Child*. doi: 10.1136/archdischild-2014-307107
- Kornør, H. & Heyerdahl, S. (2013). Måleegenskaper ved den norske versjonen av Strengths and Difficulties Questionnaire, selvrappport (SDQ-S). *PsykTestBarn*(2), 13.
- Lehmann, S., Breivik, K., Heiervang, E. R., Havik, T. & Havik, O. E. (2015). Reactive attachment disorder and disinhibited social engagement disorder in school-aged foster children - a confirmatory approach to dimensional measures. *Journal of Abnormal Child Psychology*, No Pagination Specified. doi: <http://dx.doi.org/10.1007/s10802-015-0045-4>
- 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>
- Lundervold, A., Heimann, M. & Manger, T. (2008). Behaviour-emotional characteristics of primary-school children rated as having language problems. *British Journal of Educational Psychology*, 78(4), 567-580. doi: <http://dx.doi.org/10.1348/000709908X320480>
- Mathiassen, B., Brondbo, P. H., Waterloo, K., Martinussen, M., Eriksen, M., Hanssen-Bauer, K. et al. (2012). IQ as a predictor of clinician-rated mental health problems in children and adolescents. *The British journal of clinical psychology / the British Psychological Society*, 51(2), 185-196. doi: <http://dx.doi.org/10.1111/j.2044-8260.2011.02023.x>
- Mathiesen, K. S. & Prior, M. (2006). The impact of temperament factors and family functioning on resilience processes from infancy to school age. *European Journal of Developmental Psychology*, 3(4), 357-387. doi: <http://dx.doi.org/10.1080/17405620600557797>
- Munkvold, L., Lundervold, A., Lie, S. A. & Manger, T. (2009). Should there be separate parent and teacherbased categories of ODD? Evidence from a general population. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 50(10), 1264-1272. doi: <http://dx.doi.org/10.1111/j.1469-7610.2009.02091.x>
- Munkvold, L. H., Lundervold, A. J. & Manger, T. (2011). Oppositional defiant disorder-gender differences in co-occurring symptoms of mental health problems in a general population of children. *Journal of Abnormal Child Psychology*, 39(4), 577-587. doi: <http://dx.doi.org/10.1007/s10802-011-9486-6>
- Nilsen, T. S., Handegard, B. H., Eisemann, M. & Kvernmo, S. (2015). Evaluating change in symptomatic and functional level of children and youth with emotional disorders: a naturalistic observation study. *Eur Child Adolesc Psychiatry*. doi: 10.1007/s00787-014-0671-3
- Nilsen, W., Gustavson, K., Roysamb, E., Kjeldsen, A. & Karevold, E. (2013). Pathways from maternal distress and child problem behavior to adolescent depressive symptoms: A prospective examination from early childhood to adolescence. *Journal of Developmental and Behavioral Pediatrics*, 34(5), 303-313. doi: <http://dx.doi.org/10.1097/DBP.0b013e318293ab05>
- Nordhov, S., Rønning, J. A., Ulvund, S. E., Dahl, L. B. & Kaaresen, P. I. (2012). Early intervention improves behavioral outcomes for preterm infants: Randomized controlled trial. *Pediatrics*, 129(1), e9-e16. doi: <http://dx.doi.org/10.1542/peds.2011-0248>

- Obel, C., Heiervang, E., Rodriguez, A., Heyerdahl, S., Smedje, H., Sourander, A. et al. (2004). The strengths and difficulties questionnaire in the nordic countries. *European Child and Adolescent Psychiatry, Supplement*, 13(2), II/32-II/39. doi: <http://dx.doi.org/10.1007/s00787-004-2006-2>
- Paap, M. C., Haraldsen, I. R., Breivik, K., Butcher, P. R., Hellem, F. M. & Stormark, K. M. (2013). The Link between Peer Relations, Prosocial Behavior, and ODD/ADHD Symptoms in 7-9-Year-Old Children. *Psychiatry Journal Print*, 2013, 319874. doi: <http://dx.doi.org/10.1155/2013/319874>
- Posserud, B., Lundervold, A. J., Steijnen, M. C., Verhoeven, S., Stormark, K. M. & Gillberg, C. (2008). Factor analysis of the Autism Spectrum Screening Questionnaire. *Autism*, 12(1), 99-112. doi: <http://dx.doi.org/10.1177/1362361307085268>
- Ramstad, K., Jahnsen, R., Skjeldal, O. H. & Diseth, T. H. (2012). Mental health, health related quality of life and recurrent musculoskeletal pain in children with cerebral palsy 8-18 years old. *Disability and rehabilitation*, 34(19), 1589-1595. doi: <http://dx.doi.org/10.3109/09638288.2012.656794>
- Sandvik, V. (2007). Bruk av Strengths and Difficulties Questionnaire i kartlegging av psykiske vansker hos barn med utviklingshemning. *Tidsskrift for Norsk psykologforening*, 44(6), 750-754 : ill., port.
- Sanne, B., Torsheim, T., Heiervang, E. & Stormark, K. M. (2009). The Strengths and Difficulties Questionnaire in the Bergen Child Study: A Conceptually and Methodically Motivated Structural Analysis. *Psychological Assessment*, 21(3), 352-364. doi: <http://dx.doi.org/10.1037/a0016317>
- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H. et al. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, 40(11).
- Sherr, L., Skar, A.-M. S., Clucas, C., von Tetzchner, S. & Hundeide, K. (2014). Evaluation of the International Child Development Programme (ICDP) as a community-wide parenting programme. *European Journal of Developmental Psychology*, 11(1), 1-17. doi: <http://dx.doi.org/10.1080/17405629.2013.793597>
- Sivertsen, B., Hysing, M., Elgen, I., Stormark, K. M. & Lundervold, A. J. (2009). Chronicity of sleep problems in children with chronic illness: A longitudinal population-based study. *Child and Adolescent Psychiatry and Mental Health Vol 3 Aug 2009, ArtID 22*, 3. doi: <http://dx.doi.org/10.1186/1753-2000-3-22>
- Sivertsen, B., Posserud, M.-B., Gillberg, C., Lundervold, A. J. & Hysing, M. (2012). Sleep problems in children with autism spectrum problems: A longitudinal population-based study. *Autism*, 16(2), 139-150. doi: <http://dx.doi.org/10.1177/1362361311404255>
- Skar, A.-M. S., von Tetzchner, S., Clucas, C. & Sherr, L. (2015). The long-term effectiveness of the International Child Development Programme (ICDP) implemented as a community-wide parenting programme. *European Journal of Developmental Psychology*, 12(1), 54-68. doi: <http://dx.doi.org/10.1080/17405629.2014.950219>
- Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A. & Janssens, J. M. A. M. (2010). Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review. *Clinical Child and Family Psychology Review*, 13(3), 254-274. doi: 10.1007/s10567-010-0071-2
- Sveen, T. H., Berg-Nielsen, T. S., Lydersen, S. & Wichstrom, L. (2013). Detecting psychiatric disorders in preschoolers: Screening with the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, .52(7), pp. doi: 10.1016/j.jaac.2013.04.010 23800486
- Sørensen, L., Hugdahl, K. & Lundervold, A. J. (2008). Emotional symptoms in inattentive primary school children: A population-based study. *Journal of Attention Disorders*, 11(5), 580-587. doi: <http://dx.doi.org/10.1177/1087054707308491>
- Sørensen, L., Plessen, K. J. & Lundervold, A. J. (2012). The Impact of Inattention and Emotional Problems on Cognitive Control in Primary School Children. *Journal of Attention Disorders*, 16(7), 589-599. doi: <http://dx.doi.org/10.1177/1087054711417394>

- Tromop, C. (2011). *Akuttarbeid med barn og unge i det psykiske helsevern fra et kompleksitetsperspektiv*. Masteroppgave. Høgskolen i Oslo og Akershus Oslo.
- 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 & Adolescent Psychiatry*, 20(9), 451-458. doi: <http://dx.doi.org/10.1007/s00787-011-0198-9>
- Vaage, A. B., Thomsen, P. H., Rousseau, C., Wentzel-Larsen, T., Ta, T. V. & Hauff, E. (2011). Paternal predictors of the mental health of children of Vietnamese refugees. *Child and Adolescent Psychiatry and Mental Health*, 5(2). doi: <http://dx.doi.org/10.1186/1753-2000-5-2>
- Vaage, A. B., Tingvold, L., Hauff, E., Van Ta, T., Wentzel-Larsen, T., Clench-Aas, J. et al. (2009). Better mental health in children of Vietnamese refugees compared with their Norwegian peers - A matter of cultural difference? *Child and Adolescent Psychiatry and Mental Health*, 3, 34. doi: <http://dx.doi.org/10.1186/1753-2000-3-34>
- Van Roy, B., Groholt, B., Heyerdahl, S. & Clench-Aas, J. (2010). Understanding discrepancies in parent-child reporting of emotional and behavioural problems: Effects of relational and socio-demographic factors. *BMC Psychiatry Vol 10 Jul 2010, ArtID 56, 10*. doi: <http://dx.doi.org/10.1186/1471-244X-10-56>
- Van Roy, B., Kristensen, H., Groholt, B. & Clench-Aas, J. (2009). Prevalence and characteristics of significant social anxiety in children aged 8-13 years: A Norwegian cross-sectional population study. *Social Psychiatry and Psychiatric Epidemiology*, 44(5), 407-415. doi: <http://dx.doi.org/10.1007/s00127-008-0445-7>
- Van Roy, B., Veenstra, M. & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, 49(12), 1304-1312. doi: <http://dx.doi.org/10.1111/j.1469-7610.2008.01942.x>
- Van Roy, E. (2010). *Mental health problems and their impact in a cross-sectional study of Norwegian children and adolescents aged 8-19 years: a study based on results using the Strengths and Difficulties Questionnaire as a measure of psychological adjustment* (Vol. no. 924). Oslo: Unipub.
- Walhovd, K. B., Tamnes, C. K., Ostby, Y., Due-Tønnessen, P. & Fjell, A. M. (2012). Normal variation in behavioral adjustment relates to regional differences in cortical thickness in children. *European Child and Adolescent Psychiatry*, 21(3), 133-140. doi: <http://dx.doi.org/10.1007/s00787-012-0241-5>
- Wechsler, D. (1992). *Wechsler Intelligence Scale for Children* (3. utg.). London: The Psychological Corporation.
- Wichstrom, L., Berg-Nielsen, T. S., Angold, A., Egger, H. L., Solheim, E. & Sveen, T. H. (2012). Prevalence of psychiatric disorders in preschoolers. *Journal of child psychology and psychiatry, and allied disciplines*, 53(6), 695-705. doi: <http://dx.doi.org/10.1111/j.1469-7610.2011.02514.x>