

# Screening för tillstånd med låg prevalens ger problem

## Göteborgsk autismstudie visar exempel



**MÅRTEN ERIKSSON**, docent, universitetslektor, avdelningen för socialt arbete och psykologi, Akademin för hälsa och arbetsliv, Högskolan i Gävle  
marten.eriksson@hig.se



**MONICA WESTERLUND**, docent, leg logoped, institutionen för kvinnors och barns hälsa, Uppsala universitet

Möjligheten att införa tidig populationsscreening för autism i samband med barns besök vid BVC har under hösten diskuterats i Läkartidningen [1, 2]. Bedömningsinstrumentet M-CHAT (Modified checklist for autism in toddlers), som bygger på 23 föräldrafrågor, prövas för närvarande av barnhälsovården i Göteborg. Flera internationella studier talar för att det är ett värdefullt instrument för tidig upptäckt av autismspektrumtillstånd.

Frågorna i M-CHAT är konkreta och torde inte vara svåra att svara på. Likväl anger Nygren et al [1] att instrumentet resulterar i för många falskt positiva fall. För att öka träffsäkerheten följer de därför upp barnen för vilka M-CHAT utfallit positivt med en strukturerad intervju av föräldrarna kring de frågor som har gett positivt utfall. Inga värden för sensitivitet ges i Läkartidningen, men en oro för lågt prediktivt värde (PPV) uttrycks på flera ställen. Prevalensen för autism anges vara 0,5–1 procent.

### Checklistor ger osäkerhet

Till detta vill vi anföra att screening medelst checklistor vid låga prevalenser är allmänt osäkert. Ur metodologisk synpunkt är det otillfredsställande att i utprovningen av ett screeninginstrument, såsom gjorts i Göteborgsstudien, bara undersöka de positiva utfallen i del 2 av screeningen (intervju). Dessutom förefaller inga scree-

ningnegativa ha examinerats vidare.

### Risk för överdiagnostik

PPV såväl som negativt prediktivt värde, NPV, är knutna inte bara till ett visst instrument utan också till prevalensen av sjukdomen i den aktuella populationen. Samma instrument ger således olika PPV vid olika prevalenser. Sensitivitet (och specificitet) är däremot oberoende av prevalensen. Termerna kan definieras utifrån Fakta 1.

Om vi antar att ett screeninginstrument har det utmärkta värdet 95 procent för sensitivitet och 95 procent för specificitet och att prevalensen är 1 procent kan man räkna ut att PPV blir 16 procent, dvs att 16 procent av dem med positivt värde på screeningen har sjukdomen. Skulle prevalensen i stället vara 0,5 procent blir PPV 9 procent.

Det betyder att trots mycket höga värden på sensitivitet och specificitet resulterar screeningen i en överdiagnostik, där endast vart 6:e respektive vart 11:e utfall är sant positivt.

### Svårt screena rätt vid låg prevalens

Sensitiviteten och specificiteten för M-CHAT är inte angivna av Nygren et al [1], men det är inte sannolikt att man kan uppnå så höga värden som 95 procent med ett instrument som baseras på checklistor och används av många olika bedömare i en så stor organisation som BVC. (Tillförlitligheten i laboratorieprov har vi ingen erfarenhet av, men skulle både sensitiviteten och specificiteten uppgå till 99 procent blir PPV 50 procent, vilket torde vara tillfredsställande i de flesta sammanhang.)

Vi menar därför att möjligheterna till



Tidig screening för autism på BVC diskuterades i Läkartidningen under hösten. Dock finns här metodologiska problem, som inte belystes närmare.

framgångsrik screening med hjälp av checklistor är små för tillstånd med låg prevalens. Det är möjligt att checklistor fortfarande kan spela en roll vid sekundärscreening, förutsatt att man kan identifiera riskgrupper på ett tillförlitligt sätt eller att, som man gör i Göteborgsstudien, använda checklistor för att identifiera en riskgrupp.

Utprovningen av screeninginstrument för sjukdomar med låg prevalens innebär också flera praktiska problem. Den största utmaningen ligger i att följa upp och undersöka ett tillräckligt stort

### SAMMANFATTAT

**Utprovning** av screeninginstrument vid låga prevalenser innebär flera metodologiska problem.

**Några av dessa** problem aktualiseras av autismspektrumstudien i Göteborg, dock utan att de diskuteras närmare där.

**Vi tänker speciellt** på dels att alla deltagare inte genomför hela screeningprogrammet, dels att inga screeningnegativa fall förefaller undersökas vidare, vilket omöjliggör upptäckt av falskt negativa fall och därmed beräkning av sensitivitet och specificitet.

antal barn med negativt screeningresultat.

Falskt negativt utfall är minst lika allvarligt som falskt positivt, men att undersöka ett stort antal individer, varav flertalet är friska, är resurskrävande och kan upplevas frustrerande av en högt belastad sjukvård.

Ofta håller man därför denna grupp liten, vilket leder till metodologiska problem.

### Problem med primärscreeningen

Betydligt högre PPV än 16 procent rapporteras i de studier Nygren et al refererar till [3]. Det beror på att man där begår ett klassiskt misstag, som brukar betecknas som verifikationsbias [4, 5]. Verifikationsbias innebär att man på grund av låg prevalens adderar högriskpatienter, eller redan diagnostiserade fall, till en oselektad grupp i syfte att få fler utfall. I en rent oselektad population är prevalensen så låg att endast ett fåtal fall skulle kunna upptäckas.

Genom att addera fall till undersökningspopulationen ökar man antalet sjuka i stickprovet och får därmed automatiskt ett högre PPV. Sensitiviteten överskattas därigenom och specificiteten underskattas [6].

Om ytterligare fall inkluderas måste man därför justera för detta, genom tex användning av Bayes' teorem (för ett exempel se Miniscalco Mattson [7]). Av samma skäl är det vilseledande att betrakta M-CHAT och den uppföljande intervjun av föräldrarna till utfallsbarnen som *ett* instrument, vars prediktiva värde kan beräknas. Det bör i stället beskrivas som *två* separata instrument.

Det första M-CHAT som Nygren et al [1] använder på en oselektad grupp har låg sensitivitet, specificitet och PPV. Den andra metoden, den strukturerade intervjun, används emellertid endast på

### FAKTA 1. Definition av termer

Screening	Diagnostiserad sjukdom		Totalt
	Positiv	Negativ	
Positiv	a	b	a + b
Negativ	c	d	c + d
Totalt	a + c	b + d	a + b + c + d

*Sensitivitet* är ett sannolikhetsmått på instrumentets förmåga att identifiera de sjuka eller behandlingskrävande, dvs =  $a/a+c$ .

*Specificitet* är ett sannolikhetsmått på instrumentets förmåga att korrekt identifiera de friska, dvs =  $d/b+d$ .

*Positivt prediktivt värde*, PPV, är sannolikheten att en individ med screeningutfall faktiskt är i behov av behandling/har sjukdomen, dvs =  $a/a+b$ .

*Negativt prediktivt värde*, NPV, är på motsvarande sätt sannolikheten att en individ utan screeningutfall verkligen ska visa sig vara helt frisk, dvs =  $d/c+d$ .

*Prevalensen* är andelen som har sjukdomen, dvs =  $a+c/a+b+c+d$ .

dem för vilka M-CHAT utfallit positivt, dvs på en selektad grupp. Ett sådant tillvägagångssätt är naturligtvis fullt möjligt, men det blir fel om vi betraktar dessa två led som en process med gemensamma värden, eftersom de grundar sig på *olika* populationer.

Man kan alltså inte utan vidare förbättra PPV hos ett instrument genom att införa ytterligare en procedur för utfallen. Negativa screeningutfall måste följas upp och examineras i bägge leden vid utprovningen av en screeningprocedur. Detta blir naturligtvis problematiskt för intervjudet i Göteborgsstudien, eftersom detta består i att fråga

föräldrarna om de är säkra på de svar de gett som markerar utfall. Det förefaller oss därför rimligare att detta led hör till diagnostiksidan, kanske som prediagnostik.

### Medvetenhet om metodproblem krävs

Vi har tidigare i år gått igenom ett antal screeningstudier, som alla syftat till tidig – före 3 års ålder – identifiering av språkstörning [8]. Granskningen visar att de flesta studier innehåller allvarliga begränsningar, vilket innebär att slutsatser om en eventuell screeningprocedurs möjligheter är svåra att dra.

Utöver verifikationsbias är det vanligt med oklarheter kring vilken målpopulation som avses, snedvriden rekrytering, stora och systematiska bortfall, små undersökningsgrupper och slutligen problem i samband med val av metod för diagnostik (gold standard).

För att kunna dra några slutsatser av studier som görs inom området är det därför viktigt att vara medveten om dessa metodologiska problem.

■ *Potentiella bindningar eller jävsförhållanden: Inga uppgivna.*

### REFERENSER

- Nygren G, Sandberg E, Arvidsson T, Gillberg C. BVC:s unika roll – att fånga upp autism tidigt. Erfarenheter av nya rutiner i barnhälsovården i Göteborg. 2010;107:2314-8.
- Janson S, Blennow M. Allmän screening för autism kan bli realitet vid 2,5-årsbesöket. BVC i nyckelposition för tidig diagnos – om studierna håller. 2010;107:2308-9.
- Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL, et al. The modified checklist for autism in toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *J Autism Dev Disord.* 2008; 38(5):827-39.
- Bates AS, Margolis PA, Evans AT. Verification bias in pediatric studies evaluating diagnostic tests. *J Pediatr.* 1993;122(4):585-90.
- Camp BW. Evaluating bias in validity studies of developmental/behavioural screening tests. *J Dev Behav Pediatr.* 2007;28(3):234-40.