

## Bilaga 4. Mall för kvalitetsgranskning av diagnostiska studier (QUADAS) [1,2]

Författare: \_\_\_\_\_ År: \_\_\_\_\_ Artikelnummer: \_\_\_\_\_

Mallen består av 11 enskilda kriterier [2]. Hur olika typer av bias kan påverka resultat visas i Tabell 7.2 i SBU:s handbok och i förklaring/kommentarer.

	Ja	Nej	Oklart
1. Var sammansättningen av patientgruppen (spektrum) representativ för de patienter som kommer att få testet i praktiken? <i>Undvikande av spektrumbias</i>			
2. Är det troligt att referenstestet korrekt klassificerar det sökta tillståndet? <i>Undvikande av felklassifikationsbias</i>			
3. Var tidsintervallet mellan referenstest och indextest så kort att det studerade tillståndet inte kunnat förändras mellan de båda testen? (Acceptabel fördröjning mellan testerna) <i>Undvikande av sjukdomsprogressionsbias</i>			
4. Genomgick samtliga patienter eller ett slumpmässigt urval av patienter det avsedda referenstestet? <i>Undvikande av partiell verifikationsbias</i>			
5. Användes ett och samma referenstest oberoende av vilket resultat som erhöles på indextestet? <i>Undvikande av differentiell verifikationsbias</i>			
6. Var referenstestet oberoende av indextestet (dvs indextestet ingick inte som en del av referenstestet)? <i>Undvikande av inkorporationsbias</i>			
7. Tolkades resultaten från referenstestet utan kännedom om resultaten från indextestet? (Indextestresultat blindade) <i>Undvikande av informationsbias</i>			
8. Tolkades resultaten från indextestet utan kännedom om resultaten från referenstestet? (Referenstestresultat blindade) <i>Undvikande av informationsbias</i>			
9. Fanns samma kliniska data tillgängliga då testresultaten tolkades som skulle vara tillgängliga då testen används i praktiken? (Relevant klinisk information)			
10. Rapporteras ej tolkningsbara/intermediära testresultat?			
11. Förklarades bortfall av patienter från studien?			

## **Förklaring/kommentarer till enskilda kriterier och instruktioner till hur enskilda frågor i mallen ska bedömas och koda<sup>1</sup>**

### **1. Var sammansättningen av patientgruppen (spektrum) representativ för de patienter som kommer att få testet i praktiken?**

Det finns två aspekter på frågan:

- Var den *rekryterade patientgruppen rätt* för att besvara den aktuella frågeställningen?
- Var *metoden för att rekrytera patienter* adekvat så att ett representativt urval erhöles?

*Rätt patientgrupp.* Det är viktigt att de patienter som ingår i undersökningen har en adekvat sammansättning (patientspektrum), eftersom skillnader i demografiska och kliniska karakteristika mellan populationer kan ge varierande resultat för diagnostisk tillförlitlighet. Om sammansättningen av de patienter som ingått i studien inte motsvarar dem som kommer att bli föremål för testet i klinisk praxis, saknar resultaten relevans. Observera att en studie kan ha god intern validitet men sakna relevans för den aktuella frågeställningen. ”Spektrum” avser inte bara allvarlighetsgraden hos det underliggande sökta tillståndet, utan också hur patienterna valts ut avseende demografiska förhållanden, differentialdiagnos och samsjuklighet. För att bedöma om sammansättningen av patienterna är relevant, kan klinisk information om patienterna, till exempel symtom och eventuella föregående tester också vara väsentlig. Sammanfattningsvis är det alltså viktigt att det finns en tydlig beskrivning av den undersökta populationen och tydliga kriterier för inklusion respektive exklusion av patienter. Uppgifter om detta bör man kunna få via rapporterade inklusions- och exklusionskriterier och/eller tabeller avseende patientkarakteristika.

Teoretiskt är sensitivitet och specificitet oberoende av sjukdomsprevalensen i populationen. I praktiken påverkas dock både sensitivitet och specificitet av patientkarakteristika, det vill säga det patientspektrum som testet appliceras på. Det innebär att sensitivitet och specificitet för ett test på patienter som remitterats till en specialistklinik inte kan förväntas vara desamma som för patienter i allmänpraxis. Den förra utgör en selekterad population, ofta med symtom på (ännu odiagnostiserad) sjukdom, medan den senare primärt är mer oselektad och med större andel patienter utan sjukdom. Sensitivitet och specificitet påverkas därmed också av sjukdomsprevalensen, som är lägre i en oselektad population [3]. I regel är sensitiviteten lägre och specificiteten högre i en oselektad population [3–5].

<sup>1</sup> Efter Whiting [1] och Reitsma [2].

*Metod för att rekrytera patienter.* Det är också viktigt att patienterna rekryterades på ett adekvat sätt. Den bästa studiedesignen för att få ett representativt urval är ett prospektivt och konsekutivt urval av patienter som uppfyller valda kriterier för inklusion. Så kallade fall–kontrollstudier där man jämför en grupp sjuka patienter med en grupp friska kontrollindivider övervärderar den diagnostiska tillförlitligheten.

*Relevans.* Kriteriet är alltid relevant och ska alltid ingå i kvalitetsbedömningen.

*Specificera.* Granskarna bör i möjligaste mån på förhand specificera vilka patientgrupper som är relevanta utifrån den aktuella frågeställningen (inklusionskriterier). Notera vilka faktorer som kan påverka den diagnostiska tillförlitligheten såsom den kliniska situationen (t ex primärvård, specialistvård, sjukhusvård), allvarlighet hos sjukdomen, sjukdomsprevalens och eventuella tester som föregår det aktuella testet. Om man accepterar en mindre andel ej relevanta patienter, bör storleken på den andelen anges. Här bör också anges om man accepterar studier som innehåller en kontrollgrupp med friska individer.

*Hur resultaten ska bedömas.* Bedömningen ska baseras på både karakteristika hos de inkluderade patienterna och på metoden som använts för att rekrytera patienter. Frågan besvaras med ”Ja” om man anser att den sammansättning (spektrum) av patienter som ingår i studien är representativ för dem som ska få testet i klinisk praxis, och metoden för rekrytering är adekvat. Om den redovisade populationen/metod för rekrytering inte uppfyller givna villkor bör svaret vara ”Nej”. Studier som rekryterar en grupp friska kontrollpersoner och en grupp där man vet att patienterna har det sökta tillståndet ska under nästan alla förhållanden bedömas med ”Nej”. Om information om patientspektrum och/eller rekryteringsmetod är otillräcklig, besvaras frågan med ”Oklart”.

## **2. Är det troligt att referenstestet (referensstandard, gold standard) korrekt klassificerar det sökta tillståndet?**

Referenstestet används för att bestämma närvaro eller frånvaro av det sökta tillståndet. Skattningar av indextestens diagnostiska förmåga bygger på antagandet att det jämförts med en referenstest som är 100 procent sensitivt och 100 procent specifikt. Om man finner skillnader i resultat mellan indextest och referenstest drar man slutsatsen att indextestet har brister. Valet av referenstest är således av stor betydelse. Tyvärr är perfekta referenstest ovanliga. Ett imperfekt referenstest kan ge upphov till bias avseende den diagnostiska tillförlitligheten hos indextestet. Man bör därför i regel begränsa sin inklusion till studier som baseras på en eller flera acceptabla referenstester.

Om det finns allvarliga betänkligheter, till exempel att ett indextest kanske är bättre än tillgängligt referenstest, är sedvanlig beräkning av diagnostisk tillförlitlighet inte längre tillämplig. Under sådana förhållanden bör beräkning av diagnostisk tillförlitlighet inte

göras utan att först noga överväga om det finns alternativa metodologiska metoder som är bättre [5,6].

*Relevans.* Kriteriet är alltid relevant och ska alltid ingå i kvalitetsbedömningen.

*Specificera.* Vad som är acceptabel referensstandard måste definieras (inklusionskriterier). Inom vissa områden är referensstandarderna bestämda genom konsensus. Ibland används en blandning av referensstandarder, och man kan då behöva överväga om alla är acceptabla.

*Hur resultaten ska bedömas.* Det är inte alltid självklart hur man ska bedöma referenstestets validitet. I regel behövs klinisk erfarenhet i ämnet för att bedöma om ett test eller en kombination av tester utgör en adekvat referensstandard. Om man bedömer att referenstestet är acceptabelt är svaret ”Ja”. Om man bedömer att referenstestet sannolikt inte korrekt klassificerar det sökta tillståndet bör man svara ”Nej”. Om det saknas tillräcklig information för att bedöma detta, svarar man ”Oklart”.

### **3. Var tidsintervallet mellan referenstest och indextest så kort att det studerade tillståndet inte kunnat förändras mellan de båda testen?**

Bäst är om index- och referenstest genomförs samtidigt på samma patienter. Om så inte var fallet, kan fel uppstå genom att patienterna antingen försämrats eller förbättrats (spontant eller genom behandling) under tidsperioden mellan de båda testen. Betydelsen av tidsintervallets längd varierar beroende på det studerade tillståndet. En fördröjning på några dagar är till exempel sannolikt inget problem för kroniska tillstånd, men kan vara oacceptabelt vid akuta infektioner.

*Relevans.* Kriteriet är relevant i de flesta situationer.

*Specificera.* Bestäm vad som ska betraktas som acceptabelt intervall mellan indextest och referenstest. Bestäm också om det är acceptabelt att en viss andel (ange storlek) av patienterna ligger utanför intervallet.

*Hur resultaten ska bedömas.* Det gäller att ta ställning till hur stor risken är för felklassificering. Det är vanligt att tidsintervallet mellan testen varierar mellan ingående patienter. Man bör då utgå från det längsta tidsintervall som förekommit mellan index- och referenstest. Om detta bedöms vara tillräckligt kort bedöms kriteriet vara uppfyllt (”Ja”). Om inte är svaret ”Nej”. Om risken för felklassificering inte går att bedöma, är svaret ”Oklart”.

#### **4. Genomgick samtliga patienter eller ett slumpmässigt urval av patienter det avsedda referenstestet?**

På engelska använder man bland annat termerna partial verification bias, work-up bias, primary selection bias eller sequential ordering bias för att beteckna situationer där inte alla som genomgick indextestet också genomgick referenstestet. Om resultat från indextestet påverkar beslutet att genomföra referenstestet riskerar man att få snedvridna resultat.

Om patienterna randomiseras till att genomgå/inte genomgå referenstestet påverkas inte resultaten. Det händer dock ofta att urvalet av patienter som genomgår referenstest inte sker slumpmässigt.

*Relevans.* Partiell verifikationsbias förekommer generellt sett bara i prospektiva kohortstudier, där patienterna genomgår indextestet före referenstestet. I situationer där referenstestet bedöms före indextestet får möjligheten till bias bedömas utifrån den aktuella situationen.

*Specifcera.* Eventuellt kan det vara relevant att bestämma hur stor andel som inte verifieras som kan accepteras.

*Hur resultaten ska bedömas.* Om det står klart att samtliga patienter eller ett slumpmässigt urval av dem som fått indextestet även verifierades enligt referenstestet, är svaret ”Ja”. Om vissa av de patienter som fick indextestet inte fick sitt sanna tillstånd verifierat och urvalet av patienter som fick referenstestet inte var slumpmässigt, är svaret ”Nej”. Om uppgift saknas kodas ”Oklart”.

#### **5. Användes ett och samma referenstest oberoende av vilket resultat som erhöles på indextestet?**

Differentiell verifikation uppstår när patienternas indextester valideras mot olika referenstester. Om dessa referenstester definierar det sökta tillståndet på olika sätt, finns risk för differentiell verifikationsbias. Detta inträffar ofta när patienter som får positivt resultat på indextestet blir föremål för ett mer avancerat, inte sällan invasivt referenstest, jämfört med dem som har ett negativt resultat på indextestet. En sådan situation inträffar till exempel när det bedöms som oetiskt att använda ett invasivt referenstest hos individer som fått ett negativt resultat på indextestet. Om negativa testresultat hos ett indextest verifieras med ett mindre korrekt referenstest, kommer detta att påverka den diagnostiska tillförlitligheten. En extrem form av differentiell verifikation är när en del av de negativa indextestresultaten inte verifieras alls. Detta leder till övervärdering av både sensitivitet och specificitet.

Differentiell verifikation kan också uppstå när olika centra använder olika referenstest.

Empiriska studier har visat att differentiell verifikation är en viktig källa till bias [7,8]. För att uppskatta risken för allvarlig bias, är det viktigt att förstå varför olika individer verifierades med olika referenstest och skillnaden i kvalitet mellan de olika referenstesterna. Om valet är relaterat till resultat av indextestet eller till sannolikheten för sjukdom (eller tillståndet ifråga), är bias en reell möjlighet.

*Relevans.* Risk för differentiell verifikationsbias föreligger alltid i studier av diagnostiska test.

*Specifcera.* I regel behövs inga detaljer.

*Hur resultaten ska bedömas.* Om referenstestet genomgående varit detsamma kodas ”Ja”. Om patienterna genomgått alternativa referenstest blir svaret på frågan ”Nej”. När uppgift saknas kodas ”Oklart”.

## **6. Var referenstestet oberoende av indextestet (dvs indextestet ingick inte som en del av referenstestet)?**

Ibland bestäms referensstandarderna med hjälp av flera komponenter eller baseras på information som samlats in under en längre period (t ex en diagnos hos en patient som skrivs ut från sjukhus). När resultatet från indextestet också är inkorporerat i underlaget för att fastställa diagnos (referensstandard), kommer detta att överskatta värdet av indextestet (engelska: incorporation bias). Ett exempel är en studie där man undersökte den diagnostiska tillförlitligheten hos MRI (magnetic resonance imaging) för att diagnostisera multipel skleros. Referensstandarderna, den slutliga diagnosen, baserades på all tillgänglig information inkluderande resultat från MRI, analys av cerebrospinalvätska (CFS) och klinisk uppföljning av patienten.

*Relevans.* Kriteriet är endast av betydelse när referenstestet utgörs av flera komponenter. Då är det viktigt att en fullständig definition lämnas av hur tillståndet fastställts och med vilka test detta skett. I studier där referenstestet utgörs av en enda undersökning saknar kriteriet relevans och ska antingen kodas ”Ja” eller utelämnas från kvalitetsbedömningen.

*Specifcera.* I regel behövs inga detaljer här.

*Hur resultaten ska bedömas.* Om det förefaller som om indextestet utgjorde en del av referenstestet kodas ”Nej”, annars kodas ”Ja”.

## **7 och 8. Tolkades resultaten från referenstestet utan kännedom om resultaten från indextestet? Tolkades resultaten från indextestet utan kännedom om resultaten från referenstestet?**

Denna fråga motsvarar ”blindning” i behandlingsstudier. Tolkning av testresultat kan påverkas av kännedom om resultatet från det alternativa testet. Detta kallas på engelska för test review bias och kan leda till att värdet av testet överskattas. I vilken mån denna bias påverkar resultaten beror framför allt på graden av subjektivitet i tolkningen. Ju mer utrymme för subjektivitet desto större risk för review bias. Det är därför angeläget att bedöma i vilken mån kännedom om resultatet från det ena testet kunnat påverka tolkningen av det andra testet. Huruvida blindning använts eller inte redovisas inte alltid explicit. Vid några tillfällen, till exempel när laborietester skickas till ett oberoende laboratorium, kan man anta att testet tolkas oberoende av referenstestet. Konfirmering om blindning från författarna är dock alltid önskvärd.

*Relevans.* Kriteriet är alltid relevant och ska alltid ingå i kvalitetsbedömningen. I de fall testresultaten är fullständigt objektiva (t ex mätvärden) eller utvärderingen görs på ett oberoende laboratorium är risken för review bias liten.

*Specificera.* I regel behövs inga detaljer här.

*Hur resultaten ska bedömas.* Om det tydligt framgår att testresultaten tolkats blint blir svaret på frågan ”Ja”. Om det klart framgår att testresultat tolkades med kännedom om det alternativa testet är svaret ”Nej”. Om det är oklart om blindning tillämpades svarar man ”Oklart”.

## **9. Fanns samma kliniska data tillgängliga då testresultaten tolkades som skulle vara tillgängliga då testen används i praktiken?**

För vissa indextest kan tillgång till anamnestiska/kliniska data (t ex ålder, närvaro och allvarlighet av kliniska symtom eller andra testresultat) påverka resultatet, särskilt om indextestet förutsätter tolkning. Ett exempel är tolkning av bilder, som kan påverkas av kännedom om förekomst, karaktär och lokalisering av symtom. Om sådana kliniska data finns tillgängliga när indextestet bedöms, bör de också finnas tillgängliga i klinisk praxis (och vice versa). Det kan vara svårt att separera det diagnostiska värdet av existerande klinisk information innan testet görs från det adderade värdet av indextestet. Hur detta ska hanteras får bedömas utifrån den aktuella frågeställningen. Tillgång till klinisk information hos den som bedömer ett test (gäller i huvudsak röntgenbilder) ökar den diagnostiska sensitiviteten, medan specificiteten inte försämras [3,9].

Om testet avses ersätta andra kliniska test, bör resultat från dessa inte finnas tillgängliga vid tolkningen.

*Relevans.* Om indextestet avser objektiva mätningar (t ex biokemiska analyser), som inte förändras på grund av extern information, är risken för bias liten, och kriteriet saknar relevans.

*Specificera.* Ange vilka kliniska data som normalt sett är tillgängliga i klinisk praxis när testet görs och tolkas, eller alternativt att ingen information vanligtvis är tillgänglig.

*Hur resultaten ska bedömas.* Om kliniska data som vanligen finns tillgängliga när testresultaten ska tolkas och liknande data fanns tillgängliga i studien svaras ”Ja”. Om tillgängliga kliniska data har undanhållits, eller om mer information än som vanligen finns tillgänglig är svaret ”Nej”. Om information om tillgängliga kliniska data inte rapporteras är svaret ”Oklart”.

## **10. Rapporterades ej tolkningsbara/intermediära testresultat?**

Ett diagnostiskt test kan vara ofullständigt eller ge resultat som inte kan tolkas av olika skäl. Detta kan förekomma i varierande omfattning beroende på testets egenskaper. Problem av denna karaktär rapporteras sällan i studier av diagnostiska test. De utesluts helt enkelt från analysen. Detta kan leda till felaktiga bedömningar. I vilken mån detta inträffar beror på korrelationen mellan ej tolkningsbara resultat och förekomst av positiva resultat enligt referenstestet. Oavsett orsakerna till varför vissa resultat inte kunde tolkas, är det viktigt att förekomsten av sådana problem redovisas i kvantitativ form.

*Relevans.* Kriteriet är alltid relevant och ska alltid ingå i kvalitetsbedömningen.

*Specificera.* I regel behövs inga detaljer här.

*Hur resultaten ska bedömas.* Om man bedömer att det finns ett bortfall av ej tolkningsbara resultat, och att bortfallet kan vara korrelerat med sanna positiva resultat enligt referenstestet, blir svaret på frågan ”Nej”.

## **11. Förklarades bortfall av patienter från studien?**

Bortfall av patienter kan förekomma för ett eller båda testerna. Om bortfallet skiljer sig systematiskt mellan dem som finns kvar och dem som förloras (oavsett anledning), kan utfallet av testresultaten bli snedvridna.

*Relevans.* Kriteriet är alltid relevant och ska alltid ingå i kvalitetsbedömningen.

*Specificera.* I regel behövs inga detaljer här.



*Hur resultaten ska bedömas.* Om det framgår hur stort bortfallet varit för de använda testen, till exempel i form av ett flödesdiagram, svaras ”Ja”. Om det framgår att några av de individer som primärt inkluderades inte genomgick både indextest och referenstest, och dessa individer inte har beaktats i analysen, svaras ”Nej”. Om det är oklart hur bortfallet hanterades, svaras ”Oklart”.

## Referenser

1. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
2. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration, 2009. <http://srdta.cochrane.org/>
3. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189-202.
4. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* 2009;62:5-12.
5. Knottnerus JA. Diagnostic prediction rules: principles, requirements and pitfalls. *Prim Care* 1995;22:341-363.
6. Glasziou P, Irwig L, Deeks JJ. When should a new test become the current reference standard? *Ann Intern Med* 2008;149:816-22.
7. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
8. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469-76.
9. Loy CT, Irwig L. Accuracy of diagnostic tests read with and without clinical information: a systematic review. *JAMA* 2004; 292:1602-9.