Bilaga 5. Metodologiska kommentarer till metaanalyserna

Inledning

I denna bilaga redovisas hur metaanalyserna genomförts avseende effektmått och sammanvägning. Även dataextraktion och hur problem hanterats då informationen i inkluderade studier varit ofullständig, d v s då inte följt CONSORT-statement [1] respektive Extended CONSORT-statement [2].

Effektmåtten

Två typer av effektmått har använts i enlighet med etablerade konventioner inom Cochrane Collaboration [3, 4]. Vid kontinuerliga utfall har standardiserade medelskillnader (SMD) använts i form av Hedges g och vid dikotoma utfall har oddskvoter använts. Hedges g är samma sak som Cohens d men med en justering för ”small sample bias”, där m_i är medelvärdet och sd_i standardavvikelsen i respektive grupp samt s(g) är standardfelet:

$$d = \frac{m_1 - m_2}{\sqrt{\frac{(n_1 - 1)sd_1^2 + (n_2 - 1)\, sd_2^2}{N - 2}}} \quad (1)$$

$$\hat{g} = d \cdot \left(1 - \frac{3}{4N - 3{,}94}\right) \quad (2)$$

$$se\,(\hat{g}) = \sqrt{\frac{N}{n_1 n_2} + \frac{g^2}{2(N - 3{,}94)}} \quad (3)$$

Med *a* (antal drabbade personer i interventionsgruppen), *b* (antal ej drabbade personer i interventionsgruppen), *c* (antal drabbade personer i kontrollgruppen), *d* (antal ej drabbade personer i kontrollgruppen) är oddskvoten (*OR*)

$$OR = \frac{a \cdot c}{b \cdot d} \tag{4}$$

… med ett standardfel på

$$se\left\{ 1n\,(OR) \right\} = \sqrt{1/a + 1/b + 1/c + 1/d} \tag{5}$$

Metaanalyserna

Metaanalyser baseras på heterogen mängd studier av delvis olika program, populationer, subgrupper mm. Detta betyder att den visuella bilden som framträder i forest plots och den icke-kvantitativa tolkningen är mycket viktig. Finns det någon generell trend eller är resultatet helt och hållet heterogent? Finns det något program som avviker från trenden? Finns det populationer eller subgrupper som avviker från trenden? Kliniska relevanta effekter och trender är mer intressanta än om vilken sida om signifikansgränsen ett utfall ligger om resultaten ligger nära gränsen. Detta gäller såväl resultat i enskilda studier som sammanvägda effekter. Den sammanvägda effekten i metaanalyserna är den kanske minst intressanta informationen.

Ju större likhet det finns avseende programmen, kontrollvillkoren, populationerna och studiedesignen, desto mer meningsfull är den totala sammanvägda effekten och dess konfidensintervall och desto mer intressant är heterogeniteten som ett tecken på kontextens betydelse. Ju större skillnader det finns för alla dessa aspekter, desto mer intressant blir det att titta på specifika subgrupper och enskilda studier och betona icke-kvantitativa jämförelser. Standardiseringen av utfallsmått mm underlättar dock icke-kvantitativa jämförelser.

Det finns olika sätt att väga samman resultat i enskilda studier till en metaanalys. Den metod som använts här rekommenderas av Cochrane

Collaboration [3,4] kallas invers varians. Detta betyder att resultaten i de olika studierna vägs samman med en faktor w_i vilken tar hänsyn till den spridning (och därmed indirekt studiens storlek) som varje enskild studie kännetecknas av. I praktiken gäller följande: ju mindre spridning en studie har, desto mer vikt ges den i sammanvägningen.
Vid sammanvägning av standardiserad medelskillnad SMD (g_{FE}) med en ”fixed effects model” gäller att…

$$w_i = \frac{1}{se\,(\hat{g}_i)^2} \tag{6}$$

$$\hat{g}_{FE} = \frac{\sum w_i \hat{g}_i}{\sum w_i} \tag{7}$$

$$se\,(\hat{g}_{FE}) = 1/\sqrt{\sum w_i} \tag{8}$$

För oddskvoten blir motsvarande sammanvägning (OR_{FE})

$$w_i = b_i c_i \,/\, N_i \tag{7}$$

$$\hat{OR}_{FE} = \frac{\sum w_i \hat{OR}_i}{\sum w_i} \tag{8}$$

$$se\left\{ ln\,(\hat{OR}) \right\} = \sqrt{\frac{PR\,/\,R^2 + PS + QR\,/\,R \cdot S + QS\,/\,S^2}{2}} \tag{9}$$

$$R = \sum a_i d_i \,/\, N_i \tag{10}$$

$$S = \sum b_i c_i \,/\, N_i \tag{11}$$

$$PR = \sum (a_i + d_i)\, a_i\, d_i / N_i^2 \tag{12}$$

$$PS = \sum (a_i + d_i)\, b_i\, c_i / N_i^2 \tag{13}$$

$$QR = \sum (b_i + c_i)\, a_i\, d_i / N_i^2 \tag{14}$$

$$QR = \sum (b_i + c_i)\, b_i\, c_i / N_i^2 \tag{15}$$

$$se(\hat{O}R_{FE}) = 1\sqrt{\sum w_i} \tag{16}$$

Heterogeniteten räknas fram genom att summera vägda avvikelser för enskilda studiers effekter, vilket ger Q som har en χ^2-fördelning med k-1 frihetsgrader där k är antalet studier…

$$Q = \sum w_i\, (g_i - g_{FE})^2 \tag{17}$$

… och

$$Q = \sum w_i\, (\hat{O}R_i - \hat{O}R_{FE})^2 \tag{18}$$

Konventionen enligt *Cochrane Collaboration* [3] är att använda random effects model och inte fixed effects model, då signifikant heterogenitet påvisats. Fixed effects model innebär att heterogeniteten inkluderas i modellen med en faktor kallad τ.

$$r^2 = \frac{Q - (k-1)}{\sum w_i - \frac{\sum (w_i^2)}{\sum w_i}} \tag{19}$$

Med hjälp av t tas en ny vägningsfaktor w_i' fram

$$w'_i = \frac{1}{se\,(\hat{g}_i)^2 + r^2} \tag{20}$$

… respektive

$$w'_i = \frac{1}{se\,(\hat{OR}_i)^2 + r^2} \tag{21}$$

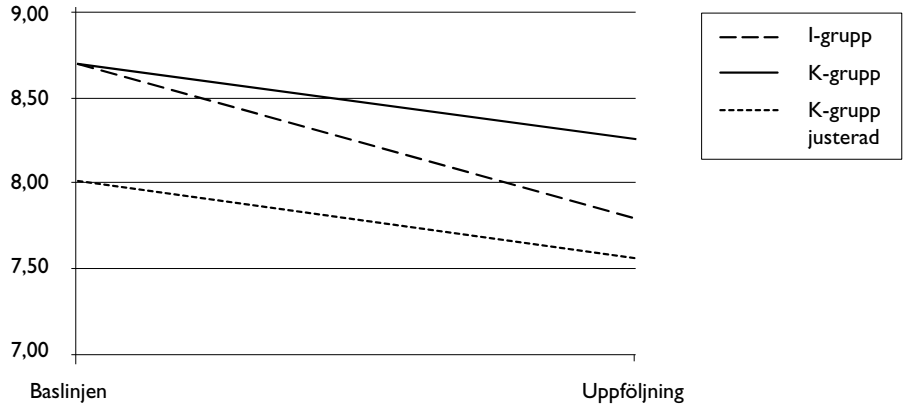
… med bl a ett större standardfel som följd.

I denna rapport har dock fixed effects model används konsekvent. Skälet är att random effects model kan ge en illusorisk precision i fall där man inte bör ge för stor vikt till den sammanvägda effekten då resultatet är heterogent utan istället bör fokusera subgruppsanalyser.

Dataextraktion och omräkningar

Justering för skillnader vid baslinjen

Justeringar av skillnader vid baslinjen har gjorts eftersom sådana skillnader kan medföra bias vid klusterrandomiseringar [3] där det har varit möjligt. I följande fall har det inte varit möjligt pga av brist på information i studierna: Barrera m fl 2002, Gross 2009, Lochman m fl 2004 och Wolchik m fl 2000. I Figur 1 illustreras vad justeringarna kan innebär:



Figur 1 *Exempel - Spencer 2005, BDI, hela gruppen, justeras för BL-skillnader avseende medelvärden.*

Simulering av m och sd då detta inte redovisats i inkluderade studier

Metaanalyserna har gjorts med Cochrane Collaborations programvara *Review Manager 5.0*. För att analyserna ska vara möjliga krävs att man har tillgång till m, sd och n för interventionsgrupp respektive kontrollgrupp vid varje aktuellt mättillfälle. I flera studier saknas detta. Om n och Cohens d finns kan artificiella m och sd simuleras och programmet kan då användas. Dessa m och sd får dock inte användas till något annat eftersom de i princip är godtyckliga. I följande studier har detta gjorts: Barrera m fl 2002, Gross m fl 2009, Trembley 1991 och Wolchik m fl 2000.

Detta är alltså möjligt om man känner till d och n (se ekvationerna 1–3 ovan). Även om enskilda *m* och *sd* är godtyckliga (men relationen mellan dem är det inte) har värden nära baslinjevärden använts. Det man gör är att sätta in d och n i formlerna och då får man fram hur relationen mellan m och sd måste se ut för att få sd och m som fungerar i RevMan. I några studier finns inte heller Cohens d i följande studier, utan istället F-värden från ANOVA, eftersom Cohens d är [5]

$$d = \sqrt{F_{between} \frac{n_1 + n_2}{n_1 \cdot n_2}} \quad (22)$$

Genom att sätta in *F* och *n* i formeln får man d och därefter, med stöd av (1), (2) och (3) kan lämpliga m och sd tas fram i RevMan. Detta har gjort för Trembley m fl 1991.

I en studie fanns endast ett ”ANCOVA *f* effect size measure” för två grupper. I detta fall kan man få fram Cohens d genom att multiplicera f med 2 [6]: Barrera m fl 2002. Därefter kan (1), (2) och (3) användas för att få fram nödvändig information. I ytterligare en studie har all nödvändig information saknats för att kunna beräkna sd eller Cohens *d* för uppföljningen. Då har *s*d från baslinjen använts: Trembley m fl 1991.

Extrahering av data för Sandler et al 2003 gjordes på följande sätt. Cohens d räknades fram från t med hjälp av följande formel. Denna formel har artikelförfattarna använt och hänvisar till Little, R. C., Milliken, G. A., Stroup, W. W., & Wofinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS.

$$d = t \frac{n_1 + n_2}{\sqrt{df}\sqrt{n_1 n_2}} \quad (23)$$

Detta ger nästa identiska resultat jämfört med den som oftast brukar användas [5].

$$d = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (24)$$

Efter (1) sattes *d, n_1* och *n_2* in i en ekvation med vars hjälp fiktiva *m* och *sd* kunde tas fram, så att RevMan kunde räkna fram aktuella resultat i Forest plot. Det fanns inga *n* för uppföljningen vid 11 månader. Därför extrapolerades dessa på följande sätt:

Vid baslinjen var n_1=135 och n_2=109. Det totala bortfallet vid 11 månader innebar n_1=117 och n_2=99 med följande proportioner n_1=54,2% och n_2=0,45,8%. Eftersom frihetsgraderna vid t-testet inte innebar en total på 216 utan 198, tyder detta på att bortfallet för det aktuella utfallet (CBCL-T externalisering) på ett större bortfall. Genom att dela upp 198

i enlighet med proportionerna från 11 månader blir skattningen n_1=107 och n_2=91.

Det bör noteras att *n* för Clarke 2001 är extrapolering från generellt bortfall på 9 som fördelats på följande sätt: intervention 45-4=41 och 49-5=44. Samtliga kontrollgrupper verkar vara ”usual care”. För Clarke 1995 och 2001 samt Garber 2009 är utfallet kumulativ frekvens av personer med diagnos någon gång under hela perioden (MDD för Clarke och CDRS för Garber). För Dadds 1999 är utfallet andel personer med diagnos (anxiety) vid mättillfället.

Olika n i tabellerna och metaanalyserna

I vissa fall är *n* inte samma i tabellerna som i metaanalyserna. Detta beror oftast på att n i tabellerna bygger på generell information om storlek på interventions- och kontrollgruppen vid olika skeden under utvärderingsprocessen, t ex n vid första urval, n vid baslinjen, n vid posttest, n vid uppföljning osv.

Redovisat n (totalt) vid uppföljningen är inte alltid samma som n för ett specifikt utfallsmått vid uppföljningen p g a internbortfall. Ibland framgår detta av redovisade frihetsgrader vid *F*-test för ANOVA eller vid direkt redovisning av antal personer som fanns med. I dessa fall kan *n* i metaanalyserna vara lägre än *n* i tabellerna.

Ibland kan det vara så att beräkningarna av effektstorlekar vid uppföljning även inkluderar personer som fallit bort (t ex med hjälp av värden från baslinjen för bortfallet eller andra skattningar). I dessa fall kan n vara större än redovisat *n* i tabellerna.

Resultat endast redovisade för subgrupper

I några studier har författarna endast redovisat resultat för subgrupper och inte hela interventionsgruppen respektive hela kontrollgruppen: Bodenman m fl 2008 (pojkar respektive flickor), Gillham 2007 (skola A+B respektive skola C), Pössell 2005 (grupp med hög self-efficacy respektive grupp med låg self-efficacy) samt Pössell m fl 2008 (pojkar respektive flickor).

Eftersom det finns såväl interventions- som kontrollgrupp för varje respektive subgrupp har dessa resultat poolat som om subgrupperna utgjorde enskilda studier. Det finns tveksamheter rörande detta eftersom subgrupperna knappast kan betraktas som oberoende studier, men det är mer transparent att göra på detta sätt än att ta fram något genomsnitt för båda respektive subgrupp. Försiktig i tolkningarna av de sammanvägda effektstorlekarna bör därför iakttas.

Klusterrandomiseringar och ”effective sample size”

Klusterrandomiseringar kan medföra att man underskattar standardfelet. Detta medför att konfidensintervallens längd kan underskattas samt att skattade p-värden är för små [3] pga så kallade ”units of analysis error”. Skattningarna av effektens storlek bör dock inte vara biased.

För att hantera risken med underskattade standardfel kan man försöka skatta ”the effective sample size” *ESS* [3, 7]. Detta kan göras enligt följande ekvation där ”the design effect” DE räknas ut enligt följande:

$$DE = 1 + (M - 1) \cdot ICC \tag{23}$$

M är den genomsnittliga klusterstorleken och *ICC* är en korrelationskoefficient avseende variationen inom relativt variationen mellan kluster, ju mindre variation inom klustren (S_W^2) och ju högre variation mellan klustren (S_B^2), desto högre korrelation (ekvation 6)

$$ICC = \frac{S_B^2}{S_B^2 + S_W^2} \tag{24}$$

Det är ovanligt att tillräcklig information finns i aktuella studier för att *ICC* ska kunnas räknas ut. Korrelationen brukar vara lägre än 0.05 [3] och ofta [7] brukar den ligga mellan 0.01 och 0.02. Totala antalet personer (n) i interventionsgruppen (n_1) respektive kontrollgruppen (n_2) divideras med designeffekten (*DE*). På så sätt får man fram DE för dessa båda grupper och kan därefter beräkna storleken på respektive *ESS* [3].

$$ESS = \frac{n}{DE} \quad (25)$$

Det är uppenbart att även en mycket svag korrelation kan ha stora konsekvenser för ESS om man betraktar ekvationerna (5), (6) och (7). Vad detta skulle kunna betyda kan illustreras med data från Aune m fl 2009. Denna studie byggde på 2 kluster med k_1=801 och k_2= 638, vilket innebär M=719.5. Med en hypotetisk *ICC* på 0.01 blir ESS för n_1=98 och *ESS* för n_2=78. Att effektiva ”sample size” blir så få beror på att antalet kluster är så litet. Konsekvenserna av *DE* i detta fall är omfattande eftersom antalet kluster är så få. Konfidensintervallets längd ökar från –0.31↔–0.10 till –0.50↔0.09, vilket ju innebär att en statistiskt signifikant effekt inte längre är signifikant. Emellertid om samma population hade fördelats över 20 kluster hade resultatet med en *ICC*=0.01 blir *ESS* för n_1=469 och *ESS* för n_2=373 och konsekvenserna hade då inte blivit så dramatiska.

Eftersom det saknas tillräcklig information i de redovisade studier som bygger på klusterrandomisering för att beräkna ICC har några justeringar för detta inte gjorts. Läsaren bör därför vara medveten om att konfidensintervallen kan vara längre än vad som framgår i redovisade forrest plots (men de måste inte vara längre).

Referenser

1. Altman DG, DSc; Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T, for the CONSORT Group. (2001). The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. Ann Intern Med 2001;134:663-94.

2. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D, for the CONSORT and Pragmatic Trials in Healthcare (Practihc) groups. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337: a2390 doi: 10.1136/bmj.a2390 1-8.

3. Higgins JPT, Green S, (eds.) (2008). Cochrane Handbook for Systematic Reviews of Interventions [updated September 2008]. The Cochrane Library. John Wiley & Sons, Ltd.

4. RevMan 4.2 User' Guide (2004). Cochrane Collaboration, (www.cochrane.org).

5. Lipsey MW, Wilson DB (2001), Practical Meta-analysis. Thousand Oaks, London, New Delhi: Sage Publications, Applied Research Methods Series, vol. 49.

6. Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale New Jersey: Lawrence Erlbaum Associates, Publishers.

7. Killip S, Mahfoud Z, Perace K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med 2004;2:204-8.