

# Bilaga 6

## Utvärdering av kritisk granskning och risk för publicerings skevhet

### Innehåll

<b>Metoder</b> .....	<b>1</b>
<b>Kategori 1.1</b> .....	<b>3</b>
Känslighetsanalys och undergruppsanalys .....	3
Publicerings skevhet.....	5
<b>Kategori 1.2 och Kategori 1.3</b> .....	<b>6</b>
<b>Kategori 1.4</b> .....	<b>6</b>
Känslighetsanalys och undergruppsanalys .....	6
Publicerings skevhet.....	7
<b>Kategori 2</b> .....	<b>8</b>
Känslighetsanalys och undergruppsanalys .....	8
Publicerings skevhet.....	10
<b>Kategori 3</b> .....	<b>11</b>
Känslighetsanalys och undergruppsanalys .....	11
Publicerings skevhet.....	13
<b>Sammanfattning och slutsatser</b> .....	<b>14</b>

### Metoder

Utvärderingen av den kritiska granskningen bygger på känslighetsanalyser och undergruppsanalyser. Vid känslighetsanalyserna har den sammanvägda medelvärdeskilnad som erhålls då samtliga studier inom en viss kategori inkluderas i en metaanalys jämförts med den sammanvägda medelvärdeskilnad som erhålls för en viss undergrupp. Vid undergruppsanalyserna har de sammanvägda medelvärdeskilnaderna för två olika undergrupper jämförts med varandra.

Indelningen i undergrupper baseras på poängsättningen i den kritiska granskningen som beskrivs i avsnitt 3.3 i huvudrapporten. I den kritiska granskningen bedöms risken för skevheter (bias) orsakade av metodologiska vägval i de enskilda studierna. Denna bedömning beskriver studiernas interna tillförlitlighet, eller *interna validitet*. Vi har också baserat några undergrupper baserat på yttre faktorer som kan påverka hur tillämpbara resultaten i en studie är på svenska förhållanden. Tillämpbarheten på svenska förhållanden beskriver studiernas externa tillförlitlighet, eller *externa validitet*. I detta fall har vi bedömt att klimatet är den viktigaste faktorn för den externa validiteten. Vi har då jämfört de viktigaste klimatzonerna i Sverige (Cfb och Dfb) med varandra och med den varmare klimatzonen Cfa, som relativt många studier har utförts i.

Vid beräkning av sammanvägda medelvärdesskillnader och jämförelser mellan två undergrupper har vi utgått från en Random Effects Model och antagit att variansen mellan studierna ( $I^2$ ) är separat för varje undergrupp (se Bilaga 3 till huvudrapporten för beräkningsmetoder). Differensen  $D$  mellan två undergrupper beräknas som

$$D = M_B - M_A$$

Där  $M_B$  är den sammanvägda medelvärdesskillnaden för undergrupp B, och  $M_A$  är den sammanvägda medelvärdesskillnaden för undergrupp A. För att avgöra om  $D$  är statistiskt signifikant görs ett Z-test där  $Z$  beräknas som

$$Z = \frac{D}{SE_D}$$

och där differensens standardfel ( $SE_D$ ) beräknas som

$$SE_D = \sqrt{V_{M_A} + V_{M_B}}$$

där  $V$  betecknar variansen för respektive medelvärdesskillnad. Vid nollhypotesen att de båda undergrupperna har samma sanna sammanvägda medelvärdesskillnad antar vi att  $Z$  följer en normalfördelning. Vid ett tvåsidigt test kan då p-värdet beräknas som

$$p = 2[1 - (\theta(|Z|))]$$

där  $\theta(|Z|)$  är den standardiserade kumulativa normalfördelningsfunktionen.

Om färre än fem jämförande observationer ingår i en viss undergrupp har ingen jämförelse med den undergruppen gjorts eftersom beräkningen av  $T^2$  då kan vara mycket osäker.

Ibland kan en sammanvägd medelvärdesskillnad som erhålls i en metaanalys vara missvisande på grund av publiceringsskevhet, det vill säga en skevhet som uppstår därför att det av olika anledningar kan vara lättare att publicera positiva och/eller signifikanta resultat. I den vetenskapliga litteraturen skulle det i så fall finnas en underrepresentation av studier med negativa eller ickesignifikanta resultat. En bedömning av risken för publikationsskevhet kan göras genom så kallade trattdiagram. I ett sådant diagram avsätts studiernas effektstorlek (i det här fallet medelvärdesskillnader) mot deras standardfel, och teoretiskt bör då datapunkterna, i avsaknad av publikationsskevhet, falla symmetriskt inom ett triangelformat område där noggranna studier med hög precision (litet standardfel) ligger nära den sammanvägda effektstorleken och studier med lägre precision (större standardfel) sprider sig i ett bredare intervall kring den sammanvägda effektstorleken. Graden av symmetri eller asymmetri bedöms något subjektivt efter en okulär besiktning av diagrammet.

## Kategori 1.1

### Känslighetsanalys och undergruppsanalys

Vid utvärderingen av den kritiska granskningen av studierna i kategori 1.1 har vi valt att inkludera de studier där kontrollgrödan inte utgjordes av en baljväxt, det vill säga de studier som ingår i jämförelse nr 2 i huvudrapporten. Med denna begränsning inkluderas totalt 113 jämförande observationer. I tabell 1 visas sammanvägda medelvärdeskilnader och konfidensintervall för samtliga dessa observationer ("Alla studier") och för olika undergrupper.

Sett till den sammantagna interna validiteten och studiernas varaktighet tenderar medelvärdeskilnaderna (det vill säga växtföljdernas effekt) att öka med ökad validitet. Det finns dock inga statistiskt signifikanta skillnader mellan det medelvärde som erhålls då samtliga studier inkluderas och det medelvärde som erhålls i någon av undergrupperna (se tabell 2). I tabell 3 visas beräknade p-värden vid jämförelse mellan olika undergrupper. Där framgår att det finns en statistiskt signifikant skillnad när studier med 2 tillförlitlighetspoäng avseende studiens varaktighet jämförs med studier med 0 tillförlitlighetspoäng. Däremot finns inget statistiskt signifikant samband mellan effekt och studiens varaktighet vid en linjär regression där samtliga studier ingår (se figur 1). Detta kan bero på att andra faktorer som till exempel jordbearbetning och gödsling också spelar stor roll och därmed kan överskugga effekten av studiernas varaktighet. Om exempelvis enbart studier av system där ingen eller organisk gödsling sker inkluderas, vilket motsvarar jämförelse nr 7 i huvudrapporten, finns ett tydligt positivt samband vid en linjär regression ( $k=0,15 \text{ g} \cdot \text{kg}^{-1} \cdot \text{år}^{-1}$ ,  $p=0,0005$ ; visas ej).

Inom kategori 1.1 är det få studier som gjorts i klimatzon Cfb och vi kan därför inte jämföra de studierna med studier i andra klimatzoner. Men eftersom det inte finns någon statistiskt signifikant skillnad mellan Cfa och Dfb finns det ingen anledning att tro att studierna i Cfb avviker signifikant.

**Tabell 1.** Sammanvägda medelvärdeskilnader (g/kg) och konfidensintervall (95%) för alla studier (113) i kategori 1.1 och för olika undergrupper. Till undergruppen Hög sammantagen intern validitet hör studier där den sammanlagda tillförlitlighetspoängen är 5 eller högre.

Intern validitet	Alla studier	Hög	Låg
Sammantagen	0,35 [0,09 – 0,62]	0,45 [0 – 0,9]	0,24 [-0,08 – 0,56]
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	-0,02 [-0,41 – 0,37]	0,44 [0,13 – 0,76]	-0,01 [-1,84 – 1,83]
Temporala replikat	<5 studier	<5 studier	0,4 [0,13 – 0,66]
Förd. av behandling	0,37 [0,06 – 0,67]	<5 studier	0,25 [-0,1 – 0,61]
Studiens varaktighet	2,39 [0,25 – 4,52]	0,71 [0,04 – 1,38]	0,09 [-0,16 – 0,34]
Provtagningsdjup	0,36 [-0,16 – 0,87]	0,52 [0,03 – 1]	0,17 [-0,22 – 0,55]
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	0,97 [0,07 – 1,86]	<5 studier	0,24 [-0,95 – 1,44]

**Tabell 2.** Känslighetsanalys för kategori 1.1. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper och Alla studier. Skillnaden anses vara statistiskt signifikant om  $p < 0,05$ .

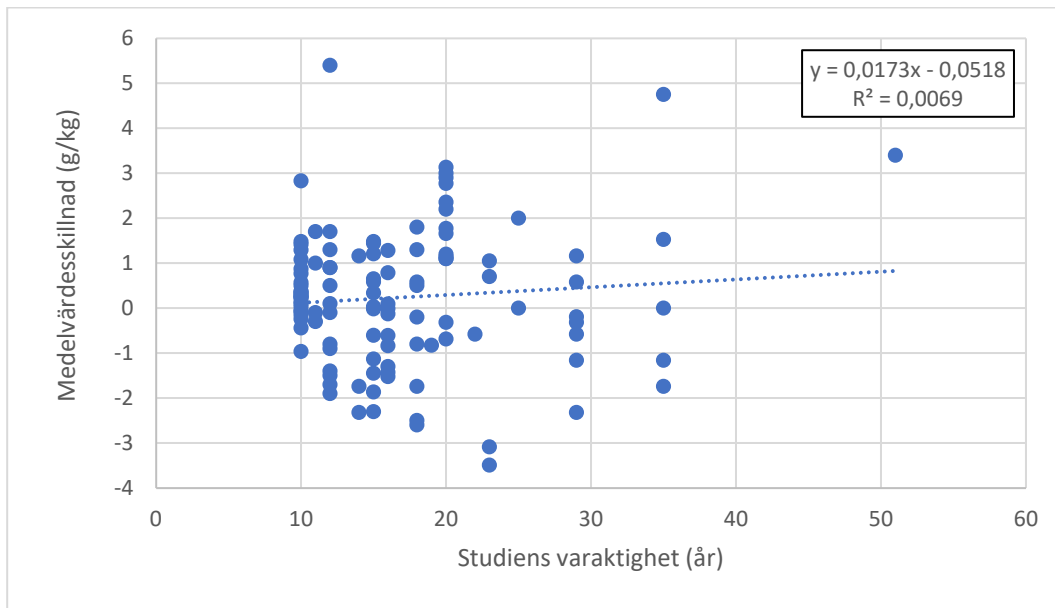
<b>Intern validitet</b>		<b>Hög</b>	<b>Låg</b>
Sammantagen		0,7066	0,6036
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	0,1214	0,6580	0,7070
Temporala replikat	a	a	0,8188
Förd. av behandling	0,9459	a	0,6661
Studiens varaktighet	0,0633	0,3310	0,1562
Provtagningsdjup	0,9866	0,5573	0,4328
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	0,1957	a	0,8642

<sup>a)</sup> Ingen jämförelse har gjorts därför att undergruppen innehåller färre än fem studier.

**Tabell 3.** Undergruppsanalys för kategori 1.1. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper. Statistiskt signifikanta skillnader indikeras med kursiv text mot gul botten.

<b>Intern validitet</b>	<b>Hög vs. Låg</b>		
Sammantagen	0,4548		
<b>Tillförlitlighetspoäng</b>	<b>2 vs. 1</b>	<b>2 vs. 0</b>	<b>1 vs. 0</b>
Spatiala replikat	0,0687	0,9858	0,6368
Temporala replikat	a	a	a
Förd. av behandling	a	0,6404	a
Studiens varaktighet	0,1417	<i>0,0359</i>	0,0897
Provtagningsdjup	0,6570	0,5605	0,2648
<b>Extern validitet</b>	<b>Cfa vs. Cfb</b>	<b>Cfa vs. Dfb</b>	<b>Cfb vs. Dfb</b>
Klimatzon	a	0,3430	a

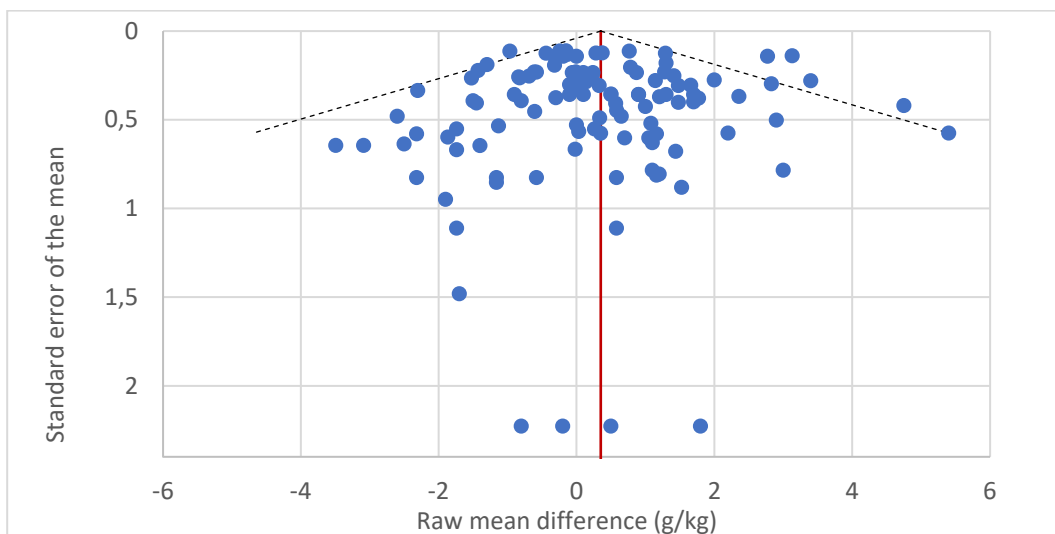
<sup>a)</sup> Ingen jämförelse har gjorts därför att minst en av undergrupperna innehåller färre än fem studier.



**Figur 1.** Medelvärdeskilnad som funktion av studiens varaktighet i kategori 1.1. Vid ett dubbelsidigt t-test för sambandet är  $p=0,3812$ .

### Publiceringskevhet

Ett trattdiagram för kategori 1.1 visas i figur 2, och vår bedömning är att datapunkterna är någorlunda symmetriskt fördelade och att det därför inte finns någon publiceringskevhet.



**Figur 2.** Trattdiagram för studier i kategori 1.1. Den lodräta röda linjen ligger på den sammanvägda medelvärdeskilnaden. Spridningen runt denna bedöms vara symmetrisk.

## Kategori 1.2 och Kategori 1.3

I kategori 1.2 och 1.3 är det för få studier för att en fullständig utvärdering ska vara möjlig. Vi kan dock konstatera att fördelningen av tillförlitlighetspoäng, och därmed den interna validiteten, inte skiljer sig i dessa studier jämfört med studierna i övriga kategorier. Det finns heller inga tecken på publiceringsskevhet bland studierna i kategori 1.2 och 1.3.

## Kategori 1.4

### Känslighetsanalys och undergruppsanalys

I kategori 1.4 ingår 23 jämförande observationer, vilket motsvarar jämförelse nr 11 i huvudrapporten. I tabell 4 visas sammanvägda medelvärdesskillnader och konfidensintervall för samtliga dessa observationer och för olika undergrupper. Av tabell 5 framgår att det inte finns några statistiskt signifikanta skillnader mellan den sammanvägda medelvärdesskillnad som erhålls då alla studier inkluderas i metaanalysen och den sammanvägda medelvärdesskillnaden i någon av undergrupperna. På grund av ett relativt litet antal studier i kategori 1.4 är det få undergrupper som har kunnat jämföras med varandra, men i de fall det har gått finns inte några statistiskt signifikanta skillnader (se tabell 6). Det finns inte heller något signifikant samband mellan medelvärdesskillnad och studiernas varaktighet vid en linjär regression ( $p=0,11$ ).

**Tabell 4.** Sammanvägda medelvärdesskillnader (g/kg) och konfidensintervall (95%) för samtliga studier i kategori 1.4 och för olika undergrupper. Till undergruppen Hög sammantagen intern validitet hör studier där den sammanlagda tillförlitlighetspoängen är 5 eller högre.

Intern validitet	Alla studier	Hög	Låg
Sammantagen	-0,28 [-0,58 – 0,03]	-0,34 [-0,67 – -0,01]	0,45 [-0,63 – 1,53]
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	0,04 [-0,4 – 0,48]	-0,49 [-1 – 0,02]	<5 studier
Temporala replikat	<5 studier	<5 studier	-0,37 [-0,64 – -0,09]
Förd. av behandling	-0,44 [-0,72 – -0,15]	<5 studier	<5 studier
Studiens varaktighet	0,26 [-0,65 – 1,17]	-0,35 [-0,68 – -0,02]	<5 studier
Provtagningsdjup	-0,39 [-0,89 – 0,11]	<5 studier	0,04 [-0,4 – 0,48]
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	-0,35 [-0,68 – -0,02]	-0,18 [-0,82 – 0,46]	<5 studier

**Tabell 5.** Känslighetsanalys för kategori 1.4. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper och Alla studier. Skillnaden anses vara statistiskt signifikant om  $p < 0,05$ .

Intern validitet	Alla studier	Hög	Låg
Sammantagen	-	0,7726	0,2040
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	0,2482	0,4739	a
Temporala replikat	a	a	0,6634
Förd. av behandling	0,4453	a	a
Studiens varaktighet	0,2738	0,7388	a
Provtagningsdjup	0,6938	a	0,2482
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	0,7388	0,7977	a

<sup>a)</sup> Ingen jämförelse har gjorts därför att undergruppen innehåller färre än fem studier.

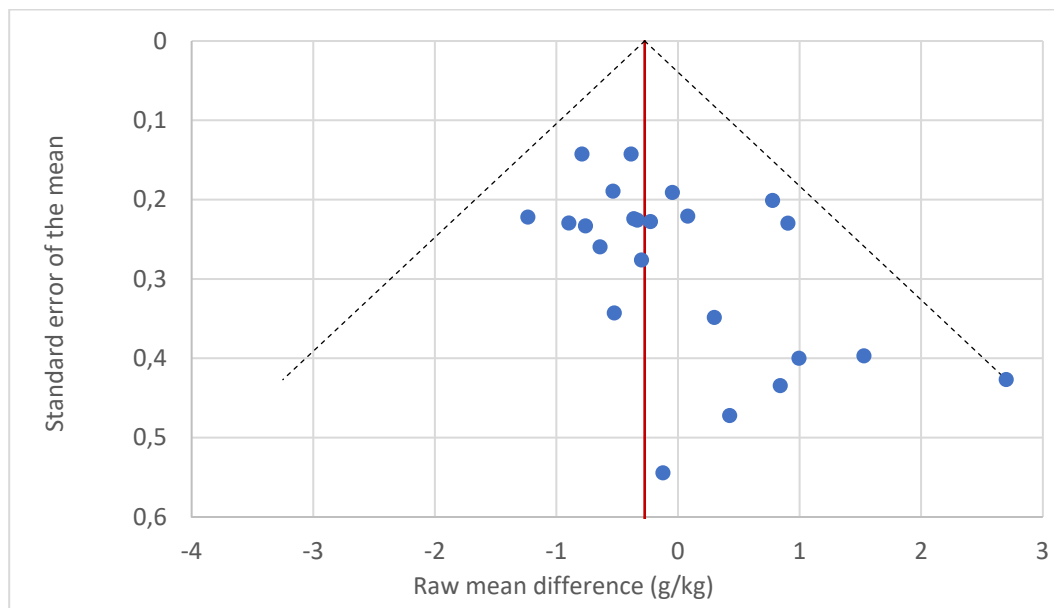
**Tabell 6.** Undergruppsanalys för kategori 1.4. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper. Inga statistiskt signifikanta skillnader ( $p < 0,05$ ) har kunnat påvisas.

Intern validitet	Hög vs. Låg		
Sammantagen	0,1683		
<b>Tillförlitlighetspoäng</b>	<b>2 vs. 1</b>	<b>2 vs. 0</b>	<b>1 vs. 0</b>
Spatiala replikat	0,1204	a	a
Temporala replikat	a	a	a
Förd. av behandling	a	a	a
Studiens varaktighet	0,2146	a	a
Provtagningsdjup	a	0,2029	a
<b>Extern validitet</b>	<b>Cfa vs. Cfb</b>	<b>Cfa vs. Dfb</b>	<b>Cfb vs. Dfb</b>
Klimatzon	0,6456	a	a

<sup>a)</sup> Ingen jämförelse har gjorts därför att minst en av undergrupperna innehåller färre än fem studier.

## Publiceringskevhet

Ett trattdiagram för kategori 1.4 visas i figur 3. En viss asymmetri i datapunkternas fördelning kan skönjas där det finns färre datapunkter till vänster om den sammanvägda medelvärdesskillnaden i diagrammets nedre del. Det kan därför finnas risk för en publiceringskevhet där det saknas studier med negativt resultat och stora osäkerheter. Å andra sidan verkar det finnas en viss övervikt av studier till vänster om den sammanvägda medelvärdesskillnaden i diagrammets övre del. Den samlade bedömningen är att den sammanvägda medelvärdesskillnaden inte är missvisande på grund av publiceringskevhet.



**Figur 3.** Trattdiagram för studier i kategori 1.4. Den lodräta röda linjen ligger på den sammanvägda medelvärdesskillnaden. Spridningen runt denna bedöms vara något asymmetrisk, vilket kan indikera en viss publiceringsskevhet.

## Kategori 2

### Känslighetsanalys och undergruppsanalys

För att utvärdera studiernas validitet inom kategori 2 har vi valt att analysera de studier som inte innehåller några flerårsväxter, det vill säga motsvarande jämförelse nr 13 i huvudrapporten. Det innebär att totalt 22 studier ingår. Vi har begränsat analysen till dessa studier eftersom förekomsten av flerårsväxter har stor påverkan på kolinlagringen och effekten av dessa riskerar därför att överskugga en eventuell effekt av skillnader i studiernas validitet.

I tabell 7 visas sammanvägda medelvärdesskillnader och konfidensintervall för Alla studier och för olika undergrupper. Av tabell 8 framgår att det inte finns några statistiskt signifikanta skillnader mellan den sammanvägda medelvärdesskillnaden som erhålls då alla studier inkluderas i metaanalysen och den sammanvägda medelvärdesskillnaden i någon av undergrupperna. Det finns dock en tendens att den sammanvägda medelvärdesskillnaden ökar med ökad tillförlitlighetspoäng med avseende på studiernas varaktighet, och i tabell 9 framgår det att det finns en statistiskt signifikant skillnad mellan studier som fått 2 poäng för varaktighet och studier som fått 1 poäng för varaktighet. Det är dock få (1–3) studielokaler som ingår i varje undergrupp och det går därför inte att dra några generella slutsatser. Vid en linjär regression där alla 22 observationer ingår finns inget statistiskt signifikant samband mellan effekt och studiernas varaktighet (se figur 4).



Det finns en statistiskt signifikant skillnad mellan studierna i klimatzon Cfb och Cfa, med en högre effekt i den i Sverige förekommande Cfb. Även här är det dock få studielokaler som ingår i varje undergrupp och det går därför inte att dra några generella slutsatser om klimatets betydelse.

**Tabell 7.** Sammanvägda medelvärdeskillnader (g/kg) och konfidensintervall (95%) för alla studier (22) i kategori 2 och för olika undergrupper. Till undergruppen Hög sammantagen intern validitet hör studier där den sammanlagda tillförlitlighetspoängen är 5 eller högre.

Intern validitet	Alla studier	Hög	Låg
Sammantagen	0,61 [0,39 – 0,84]	0,61 [0,43 – 0,78]	1,17 [-0,17 – 2,51]
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	<5 studier	0,61 [0,45 – 0,78]	<5 studier
Temporala replikat	<5 studier	<5 studier	0,65 [0,43 – 0,88]
Förd. av behandling	0,61 [0,39 – 0,84]	<5 studier	<5 studier
Studiens varaktighet	0,85 [0,65 – 1,04]	0,36 [0,14 – 0,58]	0,16 [-1,09 – 1,41]
Provtagningsdjup	<5 studier	<5 studier	1,17 [-0,17 – 2,51]
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	0,36 [0,14 – 0,58]	0,85 [0,65 – 1,04]	<5 studier

**Tabell 8.** Känslighetsanalys för kategori 2. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper och Alla studier. Skillnaden anses vara statistiskt signifikant om  $p < 0,05$ .

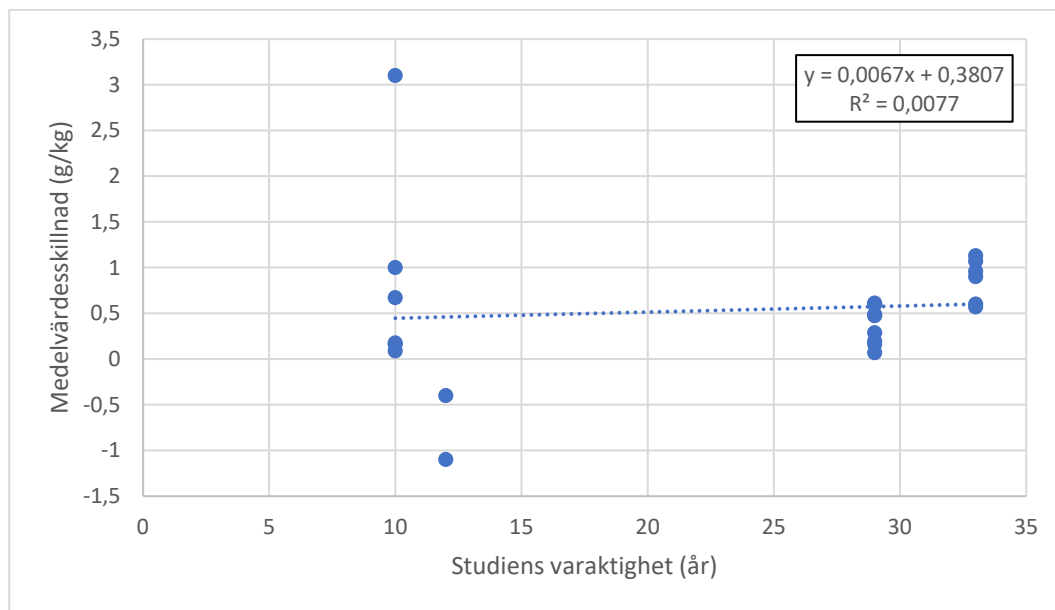
Intern validitet	Alla studier	Hög	Låg
Sammantagen	-	0,9726	0,4230
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	a	1,0000	a
Temporala replikat	a	a	0,7957
Förd. av behandling	1,0000	a	a
Studiens varaktighet	0,1195	0,1143	0,4889
Provtagningsdjup	a	a	0,4230
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	0,1143	0,1195	a

<sup>a)</sup> Ingen jämförelse har gjorts därför att undergruppen innehåller färre än fem studier.

**Tabell 9.** Undergruppsanalys för kategori 2. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper. Statistiskt signifikanta skillnader indikeras med kursiv text mot gul botten.

Intern validitet	Hög vs. Låg		
Sammantagen	0,4163		
Tillförlitlighetspoäng	2 vs. 1	2 vs. 0	1 vs. 0
Spatiala replikat	a	a	a
Temporala replikat	a	a	a
Förd. av behandling	a	a	a
Studiens varaktighet	<i>0,0010</i>	0,2881	0,7606
Provtagningsdjup	a	a	a
Extern validitet	Cfa vs. Cfb	Cfa vs. Dfb	Cfb vs. Dfb
Klimatzon	<i>0,0010</i>	a	a

a) Ingen jämförelse har gjorts därför att minst en av undergrupperna innehåller färre än fem studier.

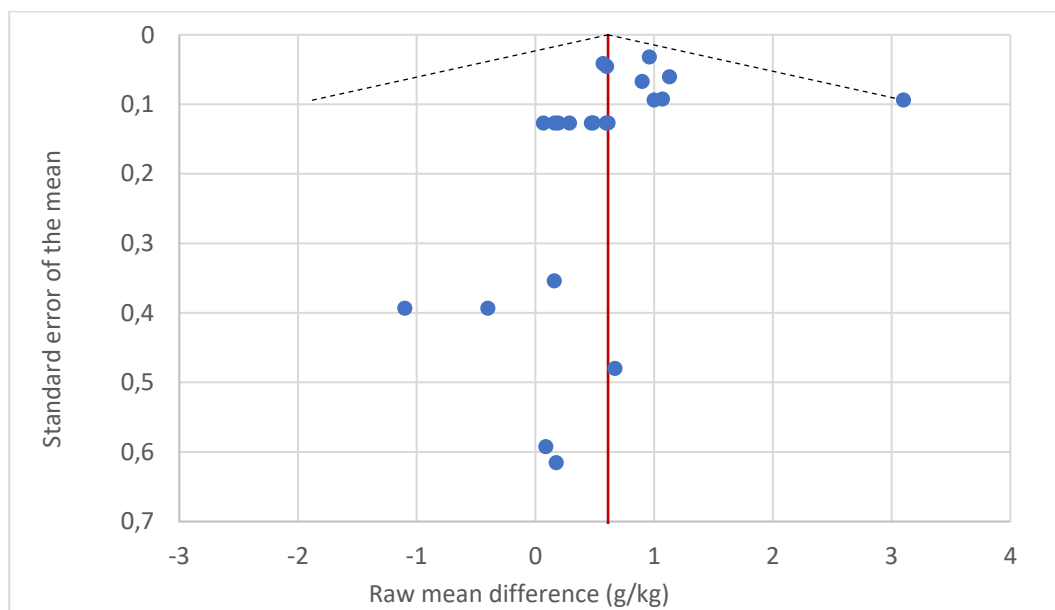


**Figur 4.** Medelvärdeskillnad som funktion av studiens varaktighet i kategori 2. Vid ett dubbelsidigt t-test för sambandet är  $p=0,698$ .

### Publiceringskevhet

Ett trattendigram för kategori 2 visas i figur 5. Datapunkternas fördelning är asymmetrisk. Det finns färre datapunkter till höger om den sammanvägda medelvärdeskillnaden i diagrammets nedre del.

Det kan därför finnas risk för en publiceringsskevhet där det saknas studier med positivt resultat och stora osäkerheter. Om så är fallet kan den effekt som rapporteras i den systematiska översikten för kategori 2 vara underskattad snarare än överskattad. Det stärker slutsatsen om att baljväxter har en positiv effekt på kolinlagringen i jorden inom varierade växtföljder.



**Figur 5.** Trattdiagram för studier i kategori 2. Den lodräta röda linjen ligger på den sammanvägda medelvärdeskilnaden. Spridningen runt denna bedöms vara något asymmetrisk, vilket kan indikera en viss publiceringsskevhet.

### Kategori 3

#### Känslighetsanalys och undergruppsanalys

I kategori 3 ingår 39 jämförande observationer. Det är dock ganska stor skillnad i effekt mellan kortvariga och långvariga flerårsväxter, och för att lättare upptäcka eventuella skillnader som beror på variationer i studiernas validitet väljer vi här att enbart analysera långvariga flerårsväxter, det vill säga växtföljder som innehåller en flerårsväxt som får växa utan förstörande jordbearbetning i minst tre år i rad. Detta motsvarar jämförelse nr 16 i huvudrapporten och inkluderar 33 studier eller observationer. I tabell 10 visas sammanvägda medelvärdeskilnader och konfidensintervall för samtliga dessa observationer och för olika undergrupper. Av tabell 11 framgår att studier där den sammantagna interna validiteten bedömts som låg har en sammanvägd medelvärdeskilnad som signifikant skiljer sig (är lägre) än den sammanvägda medelvärdeskilnad som erhålls när alla 33 observationer ingår i metaanalysen. Det finns också en signifikant skillnad mellan studier där den sammantagna interna validiteten bedömts som hög och studier där densamma bedömts som låg (se

tabell 12). Däremot finns ingen signifikant skillnad mellan studier med en hög sammantagen intern validitet och alla 33 studier.

Av tabell 10 och tabell 11 framgår också att studier som tilldelats 2 poäng för studiens varaktighet har en sammanvägd medelvärdeskillnad som är signifikant högre än den sammanvägda medelvärdeskillnad som erhålls när alla 33 observationer ingår i metaanalysen. Det finns även en signifikant skillnad mellan studier som tilldelats 2 poäng för studiens varaktighet och studier som tilldelats 0 poäng för studiens varaktighet (tabell 12). Vid en linjär regression där samtliga studier ingår finns ett statistiskt signifikant samband mellan effekt (sammanvägd medelvärdeskillnad) och studiens varaktighet, där effekten ökar med ökad varaktighet ( $k=0,18 \text{ g}\cdot\text{kg}^{-1}\cdot\text{år}^{-1}$ ; se figur 6).

Avseende den externa validiteten finns det inte någon signifikant skillnad mellan studierna som utförts i klimatzon Cfb och de som utförts i klimatzon Cfa.

**Tabell 10.** Sammanvägda medelvärdeskillnader (g/kg) och konfidensintervall (95%) för samtliga studier (33) i kategori 3.2 och för olika undergrupper. Till undergruppen Hög sammantagen intern validitet hör studier där den sammanlagda tillförlitlighetspoängen är 5 eller högre.

Intern validitet	Alla studier	Hög	Låg
Sammantagen	3,73 [2,73 – 4,73]	4,14 [2,92 – 5,36]	1,19 [-0,21 – 2,58]
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	<5 studier	3,63 [2,61 – 4,64]	<5 studier
Temporala replikat	<5 studier	<5 studier	3,12 [2,09 – 4,16]
Förd. av behandling	3 [1,94 – 4,05]	<5 studier	<5 studier
Studiens varaktighet	7,07 [5,97 – 8,17]	<5 studier	2,71 [1,67 – 3,74]
Provtagningsdjup	3,82 [2,62 – 5,01]	<5 studier	1,72 [-0,2 – 3,65]
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	2,34 [-0,13 – 4,81]	3,38 [1,63 – 5,12]	<5 studier

**Tabell 11.** Känslighetsanalys för kategori 3.2. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper och Alla studier. Statistiskt signifikanta skillnader indikeras med kursiv text mot gul botten.

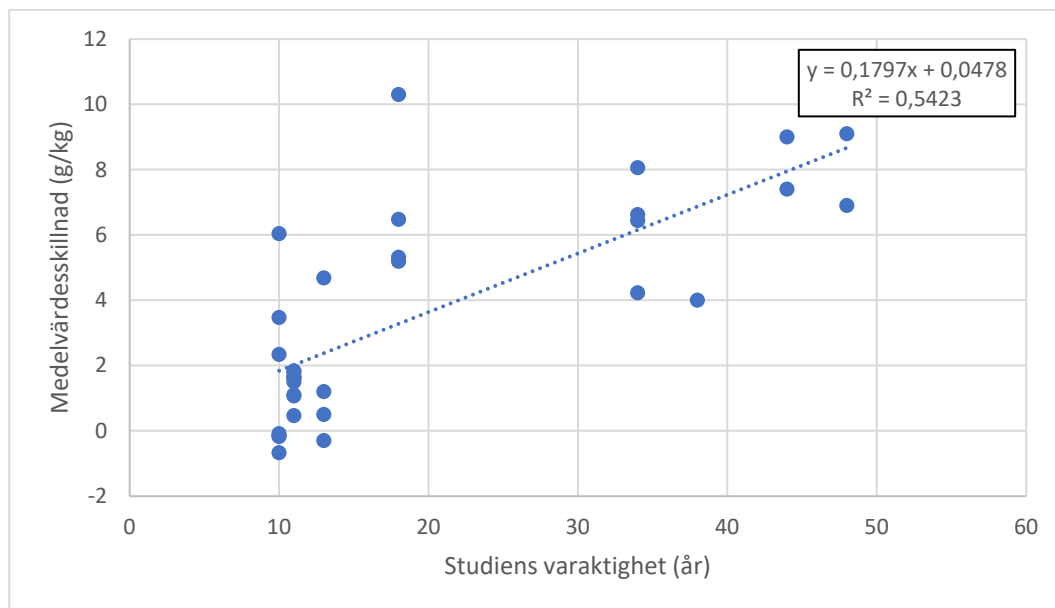
Intern validitet	Alla studier	Hög	Låg
Sammantagen	-	0,6081	<i>0,0037</i>
<b>Tillförlitlighetspoäng</b>	<b>2</b>	<b>1</b>	<b>0</b>
Spatiala replikat	a	0,8928	a
Temporala replikat	a	a	0,4123
Förd. av behandling	0,3240	a	a
Studiens varaktighet	<i>&lt;0,0001</i>	a	0,1657
Provtagningsdjup	0,9117	a	0,0701
<b>Extern validitet</b>	<b>Cfa</b>	<b>Cfb</b>	<b>Dfb</b>
Klimatzon	0,3084	0,7316	a

<sup>a)</sup> Ingen jämförelse har gjorts därför att undergruppen innehåller färre än fem studier.

**Tabell 12.** Undergruppsanalys för kategori 3.2. Tabellen visar beräknade p-värden vid jämförelse mellan olika undergrupper. Statistiskt signifikanta skillnader indikeras med kursiv text mot gul botten.

Intern validitet	Hög vs. Låg		
Sammantagen	<i>0,0018</i>		
Tillförlitlighetspoäng	2 vs. 1	2 vs. 0	1 vs. 0
Spatiala replikat	a	a	a
Temporala replikat	a	a	a
Förd. av behandling	a	a	a
Studiens varaktighet	a	<i>&lt;0,0001</i>	a
Provtagningsdjup	a	0,0699	a
Extern validitet	Cfa vs. Cfb	Cfa vs. Dfb	Cfb vs. Dfb
Klimatzon	0,5028	a	a

a) Ingen jämförelse har gjorts därför att minst en av undergrupperna innehåller färre än fem studier.

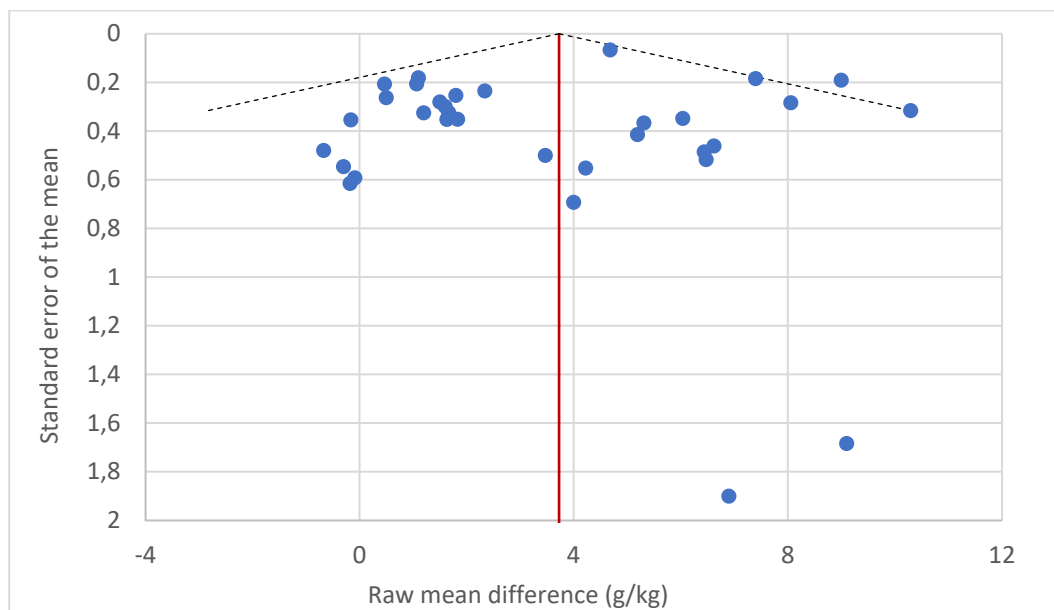


**Figur 6.** Medelvärdeskillnad som funktion av studiens varaktighet i kategori 3.2. Vid ett dubbelsidigt t-test för sambandet är  $p < 0,0001$ .

### Publiceringskevhet

Ett trattendigram för kategori 3.2 visas i figur 7. Datapunkterna är, med undantag för två punkter till höger i nedre delen av diagrammet, symmetriskt fördelade runt den sammanvägda

medelvärdesskillnaden. De två observationer som representeras av de nämnda punkterna har dock en mycket liten påverkan den sammanvägda medelvärdesskillnaden (deras relativt stora standardfel gör att deras viktning i metaanalysen blir låg). Det betyder att även om det kan finnas en liten publicerings skevhet så påverkar inte det den beräknade sammanvägda medelvärdesskillnaden nämnvärt mycket.



**Figur 7.** Trattdiagram för studier i kategori 3.2. Den lodräta röda linjen ligger på den sammanvägda medelvärdes skillnaden. Spridningen runt denna bedöms i stort sett, med undantag för två punkter i diagrammets nedre del, vara symmetrisk.

## Sammanfattning och slutsatser

I den här bilagan har vi undersökt hur variationer i studiernas interna och externa validitet kan ha påverkat den sammanvägda medelvärdes skillnaden för olika växtföljds kategorier, det vill säga växtföljdernas sammanvägda effekt på kolinlagringen i jorden. Vi har också undersökt om eventuella publicerings skevheter kan ha lett till att de sammanvägda medelvärdes skillnaderna är missvisande.

Överlag har vi endast i ett fåtal fall i känslighetsanalyserna kunnat identifiera några signifikanta skillnader mellan den sammanvägda medelvärdes skillnad som erhålls då alla studier inkluderas i metaanalysen, och den sammanvägda medelvärdes skillnaden i någon av undergrupperna. På samma sätt har vi i undergruppsanalyserna kunnat identifiera endast ett fåtal signifikanta skillnader mellan olika undergrupper. De tydligaste skillnaderna finns för undergrupper som baseras på antal tillförlitlighetspoäng avseende studiernas varaktighet, där den sammanvägda effekten ökar med ökad tillförlitlighetspoäng. Även linjära regressionsanalyser visar i vissa fall ett statistiskt signifikant samband mellan sammanvägd effekt och studiens varaktighet.

Också inom andra validitetsaspekter finns en tendens att den sammanvägda effekten ökar med ökad tillförlitlighetspoäng, även om det sällan finns några statistiskt säkerställda skillnader.

Vår samlade bedömning är därför att de sammanvägda effekter som vi rapporterar om i huvudrapporten i vart fall inte är överskattade på grund av att vi har inkluderat studier med låg validitet. Om något kan de tvärtom vara underskattade. Det faktum att effekten i flera fall ökar med ökad varaktighet talar också för att det finns en verklig mekanism bakom de effekter vi rapporterar om.

De rapporterade sammanvägda effekterna verkar inte heller vara missvisande på grund av publiceringsskevheter. Det allvarligaste tecknet på publiceringsskevheter som vi har identifierat gäller studier i kategori 2. Där ser det dock ut som att det finns en underrepresentation av studier där resultaten pekar på en effekt som är större än den sammanvägda effekten. Även i detta fall skulle den sammanvägda effekten alltså vara underskattad snarare än överskattad.