



Vetenskapsrådet

Diarienummer  
354-2014-6411  
Dat um  
2014-04-29  
Handläggare  
Henrik Aldberg  
Carl Jacobsson

## Bibliometriskt underlag för medelsfördelning

### 1 Bakgrund

I den forskningsproposition regeringen lade fram hösten 2008, *Ett lyft för forskning och innovation* [1], ingick en bibliometrisk modell för prestationsbaserad fördelning av 5% av de direkta forskningsanslagen till landets lärosäten. Den baseras på den modell som beskrivs i Bilaga 8 i utredningen *Resurser för kvalitet* [2]. Modellen går i grova drag ut på att utifrån antalet publikationer, antalet citeringar till dessa, samt en uppskattning av det genomsnittliga antalet publikationer en forskare producerar inom olika ämnesklasser, beräknas ett bibliometriskt index för varje lärosäte (Appendix A listar de lärosäten som berörs av modellen). Detta index ligger sedan till grund för fördelningen av de prestationsbaserade medlen. Fördelningen för 2009 gjordes i forskningspropositionen och därefter har Vetenskapsrådet haft uppdraget att ta fram underlag till medelsfördelningen enligt den föreslagna modellen. Detta dokument beskriver hur beräkningen av underlaget går till: kapitel 2 beskriver innehållet i Vetenskapsrådets publikationsdatabas, kapitel 3 redogör generellt för de bibliometriska mått som ingår i fördelningsmodellen och kapitel 4 beskriver hur modellens bibliometriska index beräknas.

### 2 Databasen

Vetenskapsrådets databas bygger på samma grundmaterial som Web of Science (Thomson Reuters)<sup>1</sup>. Varje tidskrift i databasen klassificeras av Thomson Reuters till en eller flera av cirka 250 ämnesklasser och en artikel i en tidskrift ärver tidskriftens ämnesklassning.

Publikationer i de tidskrifter som av Thomson Reuters klassas som multidisciplinära - exempelvis Nature och Science - har i Vetenskapsrådets databas i stor utsträckning förts till andra ämnesklasser. Denna omklassning baseras på ämnestillhörigheten hos de publikationer den aktuella publikationen citerar och på ämnestillhörigheten hos de publikationer som citerar den aktuella publikationen.

Varje publikation är också klassad som tillhörande någon av cirka 30 olika dokumenttyper där *Article*<sup>2</sup> utgör den största andelen (65%). Appendix B listar de olika dokumenttyperna.

<sup>1</sup> Certain data included herein are derived from the *Science Citation Index Expanded*®, *Social Science Citation Index*® and *Arts and Humanities Citation Index*® prepared by Thomson Reuters®, Philadelphia, Pennsylvania, USA© Copyright Thomson Reuters® 2013. All rights reserved.

<sup>2</sup> I Vetenskapsrådets databas klassas dokumenttyperna *Letter* och *Note* som *Article*.

### 3 Underliggande mått

En av komponenterna i fördelningsmodellen är antalet publikationer för de olika lärosätena. Ett problem med att räkna antalet publikationer från olika lärosäten är att en publikation kan ha författare från flera lärosäten. Summan av antalet publikationer från de olika lärosätena blir då större än det totala antalet publikationer. Samma problem finns med att räkna antalet publikationer i olika ämnen, eftersom en tidskrift kan vara klassad som tillhörande flera ämnesklasser vilket gör att summan av antalet publikationer över de enskilda ämnena blir större än det totala antalet publikationer.

Ett sätt att hantera detta är att dela upp varje publikation i smådelar, fraktioner, där varje fraktion bara har *en* adress och *en* ämnesklass. Om en publikation har 3 adresser och är klassad i 4 ämnen delas den upp i 12 fraktioner som vardera har "vikten"  $1/12$  (1 dividerat med produkten av antal adresser och antal ämnen). Om vi studerar ämnenas publikationsvolym får de fyra ämnena tillgodoräkna sig 3 fraktioner, det vill säga en fjärdedel av publikationen, vardera. Om vi istället vill undersöka publikationsvolymen för olika lärosäten och två av de tre adresserna kommer från samma lärosäte finns detta lärosäte i adressen på 8 av de 12 fraktionerna, det vill säga lärosätet krediteras  $2/3$  av publikationen, medan lärosätet med den tredje adressen krediteras  $1/3$ . Om vi i tillägg skulle studera ämnesuppdelad publikationsvolym för lärosätena skulle lärosätet med två adresser få tillgodoräkna sig  $2/12$  av publikationen för vart och ett av de fyra ämnena.

När vi beräknar den fraktionerade publikationsvolymen för ett lärosäte tar vi helt enkelt summan av vikterna för alla publikationsfraktioner med en adress från lärosätet. På samma sätt blir publikationsvolymen för ett ämne summan av vikten av alla publikationsfraktioner med den ämnesklassen.

I modellen ingår även antalet citeringar. Antalet citeringar beräknas också fraktionerat, vilket innebär att citeringarna till en publikation delas upp på fraktionerna. Antalet citeringar för en publikationsfraktion är alltså antal citeringar för publikationen gånger fraktionsvikten. Anta att publikationen i exemplet ovan har 24 citeringar. Då skulle de 12 fraktionerna få 2 citeringar var. De fyra ämnena har vardera 3 fraktioner och antalet fraktionerade citeringar blir alltså 6 per ämne (antalet citat dividerat med antalet ämnen). Antalet fraktionerade citeringar för ett lärosäte med två av de tre adresserna blir på samma sätt 16 medan lärosätet med en adress får 8 citeringar från den aktuella publikationen.

När vi beräknar det fraktionerade antalet citeringar för ett lärosäte tar vi helt enkelt summan av antalet citeringar för de publikationsfraktioner som tillhör lärosätet. På samma sätt är det fraktionerade antalet citeringar för en ämnesklass summan av antalet citeringar för alla publikationsfraktioner med den ämnesklassen.

Alla citeringsvärden är fältnormerade. Den fältnormerade citeringen av en publikation med endast ett ämne är helt enkelt antalet citeringar i jämförelse med genomsnittet för ämnesklassen, det vill säga antalet citeringar till en publikation dividerat med citeringsmedelvärdet (fältreferensvärdet) för alla publikationer i databasen i samma ämnesklass (fält), från samma år och av samma publikationstyp. Om en publikation fått lika många citeringar som den genomsnittliga publikationen i samma ämnesklass, från samma år och av samma typ, får den alltså en fältnormerad medelcitering lika med 1. När en publikation har fler än ett ämne beräknas den fältnormerade citeringen för varje publikationsfraktion för sig. Den fältnormerade citeringen för en publikationsfraktion är antalet citeringar för fraktionen<sup>3</sup> dividerat med citeringsmedelvärdet för fraktionens ämnesklass.

---

<sup>3</sup> Det vill säga antalet citeringar för publikationen gånger fraktionsvikten.

Mer precist beräknas fältreferensvärdet för en ämnesklass enligt följande: Först beräknas det fraktionerade antalet citeringar och det fraktionerade antalet publikationer för ämnesklassen enligt ovan. Fältreferensvärdet är sedan det fraktionerade antalet citeringar dividerat med det fraktionerade antalet publikationer.

Den fältnormerade medelciteringen för ett lärosäte är sedan antalet fältnormerade citeringar dividerat med det fraktionerade antalet publikationer från lärosätet.

## 4 Fördelningsmodellen

I det underlag Vetenskapsrådet levererar till utbildningsdepartementet ingår

- Fraktionerad publikationsvolym
- Antal fältnormerade citeringar
- Fältnormerad medelcitering
- Bibliometriskt index.

Vi har ovan beskrivit hur de tre första av dessa indikatorer beräknas. Här beskriver vi beräkningen av det bibliometriska indexet.

Olika forskningsområden har olika publicerings- och citeringstraditioner. De olika citeringstraditioner kompenseras genom fältnormering av antalet citeringar. För att kunna jämföra forskningsområden med olika publiceringstakt används i fördelningsmodellen en uppskattning av det genomsnittliga antalet artiklar en forskare producerar inom olika områden. De områden som används i fördelningsmodellen baseras på en uppdelning av tidskrifterna i Web of Science i 34 så kallade makroklasser. Dessa makroklasser är skapade utifrån en klustring som presenterades i dåvarande Högskoleverkets rapport *Resurser för citeringar* [3, Tabell 10, sid. 64f.].<sup>4</sup> För varje område introduceras en fältfaktor som är en uppskattning av det genomsnittliga antalet artiklar en forskare producerar under en fyraårsperiod inom var och en av de 34 makroklasserna. Tabell 1 nedan listar de 34 makroklasserna och deras fältfaktorer.

Makroklass	Fältfaktor	Makroklass	Fältfaktor
Agriculture	1.17	Health Studies	1.26
Biology	1.16	Humanities	0.16
Biomolecular	1.51	Immunology	1.23
Blood	1.36	Information Science	0.82
Chemistry	2.22	Materials Science	1.52
Computer Science	0.88	Mathematics	1.40
Dentistry	1.16	Mechanics	1.35
Ecology	1.25	Medicine External	1.09
Economics	0.72	Medicine Internal	1.59
Education	0.32	Neuroscience	1.48
Engineering	0.96	Oncology	1.53
Engineering Mathematics	1.11	Pharmacology	1.30
Environmental Health	1.40	Physics	1.46
Environmental Studies	0.43	Psychology	1.06
Ergonomics	0.58	Social Science	0.48
Geoscience	0.93	Statistics	0.95
Health	0.78	Surgery	1.19

**Tabell 1:** De olika makroklasserna och deras fältfaktorer.

De fyra indikatorerna i fördelningsmodellen beräknas, för varje lärosäte, inom varje makroklass.

<sup>4</sup> Eftersom den klustringsmetodik som används i [3] inte dokumenterats på ett sådant sätt att den kunnat återskapas av Vetenskapsrådet makroklassas varje ny tidskrift till den makroklass som är den mest citerade i tidskriften.

Låter vi  $F_m$  beteckna fältfaktorn för makroklass  $m$  (Tabell 1) och  $N_m$  beteckna antalet fältnormerade citeringar för ett visst lärosäte i makroklass  $m$  beräknas lärosätets bibliometriska index i den aktuella makroklassen,  $B_m$ , enligt

$$B_m = \frac{N_m}{F_m}.$$

Lärosätets bibliometriska index  $B$  är summan av de bibliometriska indexen för alla makroklasser, det vill säga  $B = B_1 + \dots + B_{34}$ . I beräkningen av fördelningsmodellen ingår bara publikationer som har dokumenttyperna *Article* och *Review*<sup>5</sup>. Antalet citeringar är räknade under ett öppet citeringsfönster vilket betyder att samtliga citeringar som gjorts till en publikation räknas, dock inte självciteringar. Underlaget baseras på publikationer från en fyraårsperiod. I det underlag som levererades 2013 inför fördelningen av forskningsanslag 2014 ingick publikationer från 2009-2012.

## 5 Anmärkning

Fältfaktorerna i Tabell 1 är en uppskattning av produktiviteten för en genomsnittlig forskare inom de olika makroklasserna under en fyraårsperiod. För att göra denna uppskattning behövs dels en uppskattning av hur många forskare det finns i Sverige inom de olika makroklasserna, dels en uppskattning av hur många publikationer dessa forskare producerar inom respektive klass. Det senare kan man få en uppfattning om genom att räkna antalet publikationer med svenska adresser i Web of Science (detta blir just en uppskattning eftersom databasen bara täcker en del av den svenska forskningen). Att uppskatta antalet forskare är svårare eftersom bara de forskare som publicerar sig i tidskrifter som indexeras av Web of Science finns med i databasen. I fördelningsmodellen används den så kallade waringmetoden för att uppskatta antalet forskare. När Vetenskapsrådet i januari 2009 fick i uppdrag av regeringen att ta fram nödvändigt underlag för att tillämpa fördelningsmodellen upptäcktes brister i waringmetodens tillförlitlighet när den används på dataunderlag av den storlek som förekommer i modellen. Waringmetoden bör användas på väsentligt större dataunderlag för att ge tillförlitliga resultat. Modellen och dessa brister finns beskrivna i [4] och [5].

## Referenser

- [1] *Ett lyft för forskning och innovation* (2008), Prop. 2008/09:50.
- [2] *Resurser för kvalitet* (2007), SOU 2007:81.
- [3] *Resurser för citeringar* (2008), Högskoleverket 2008:18 R.
- [4] *Bibliometrisk indikator som underlag för medelsfördelning* (2009), Vetenskapsrådet.
- [5] Fröberg, J., Gunnarsson, M., Jonsson, A., Karlsson, S. (2010) *Kan man använda Waringmetoden för att uppskatta antalet forskare?*, Vetenskapsrådets lilla rapportserie 5:2010.

---

<sup>5</sup> Dokumenttyperna *Letter* och *Note* ingår också eftersom dessa klassas som *Article* i Vetenskapsrådets databas.

## 6 Appendix

### A Berörda lärosäten

Dataunderlag tas fram för följande lärosäten<sup>6</sup>:

Blekinge tekniska högskola	Kungliga tekniska högskolan
Chalmers tekniska högskola	Linköpings universitet
Gymnastik- och idrottshögskolan	Linnéuniversitetet
Göteborgs universitet	Luleå tekniska universitet
Högskolan Dalarna	Lunds universitet
Högskolan i Borås	Malmö högskola
Högskolan i Gävle	Mittuniversitetet
Högskolan i Halmstad	Mälardalens högskola
Högskolan i Jönköping	Stockholms universitet
Högskolan i Skövde	Sveriges lantbruksuniversitet
Högskolan Kristianstad	Södertörns högskola
Högskolan på Gotland	Umeå universitet
Högskolan Väst	Uppsala universitet
Karlstads universitet	Örebro universitet
Karolinska institutet	

### B Dokumenttyper

Thomson Reuters använder följande dokumenttyper i sin databas:

Dokumenttyp	Andel (%)	Dokumenttyp	Andel (%)
Article	64,97	Fiction, Creative Prose	0,10
Meeting Abstract	11,58	Discussion	0,09
Book Review	6,67	Theater Review	0,06
Editorial Material	4,59	Dance Performance Review	0,05
Letter	3,52	Reprint	0,04
Review	2,76	Software Review	0,04
Note	2,00	Music Score Review	0,04
News Item (new in 1996)	1,15	Bibliography	0,03
Correction, Addition	0,63	TV-, Radio-, Videocassette-Review	0,03
Poetry	0,52	Excerpt	0,02
Biographical-Item	0,42	Hardware Review	0,01
Art Exhibit Review	0,22	Script	0,01
Record Review	0,15	Database Review	0,00
Music Performance Review	0,14	Chronology	0,00
Film Review	0,14	Music Score	0,00

1996 tog Thomson bort tre dokumenttyper. Publikationer som tidigare klassats som *Note* och *Chronology* klassas numera som *Article* och publikationer av typen *Discussion* klassas som *Editorial*.

Vetenskapsrådet klassar om dokumenttyperna *Letter* och *Note* som *Article*. I beräkningen av fördelningsmodellen ingår bara publikationer av dokumenttyperna *Article* och *Review*.

<sup>6</sup> Den 1 januari 2010 gick Växjö universitet och Högskolan i Kalmar ihop och bildade Linnéuniversitetet.