

Synpunkter på RUT-utredningens bibliometriska prestationsindikator

Tore Lund

[Denna uppsats härstammar från ett PM inför Chalmers remissbehandling av RUT-utredningen. Nu känns den lite daterad - bl a märker jag att jag var taggad av utredningens i mitt tycke orättvisa behandling av den norska modellen. Vissa delar är kanske fortfarande relevanta. Jag har lagt till några hastiga kommentarer med anledning av Ulf och Erik Sandströms rapport *Resurser för citeringar* (Högskoleverket. Rapport 2008:18R), som jag tycker för diskussionen framåt på flera punkter. - Tore Lund, 1.10.2008]

Sammanfattning

Resursfördelningsutredningen (RUT2) förslår i sitt slutbetänkande *Resurser för kvalitet* (SOU 2007:81) att 20% av de statliga forskningsanslagen till högskolorna skall fördelas via en bibliometrisk prestationsindikator. Man presenterar också ett förslag till en sådan indikator. Indikatorn har tagits fram av Ulf Sandström och använder *fältnormering* för att tillåta jämförelser mellan olika ämnesområden.

Indikatorn beräknas som produkten av två delmätt. Det ena är ett välkänt och internationellt beprövat *kvalitetsmått* (citeringar per publikation, mätt i förhållande till genomsnittet inom ämnet). Det andra är ett *kvantitetsmått*, som beräknas enligt en ny och oprövad metodik och som försöker ge ett ämnesjusterat mått på hur ”stor” lärosätets vetenskapliga produktion är. Måttet blir härigenom också indirekt ett mått på *produktiviteten*. Det mesta av omfördelningseffekten uppstår (åtminstone i utredningens räkneexempel) genom produktions- och produktivitetens mått; kvalitetsmålet bidrar bara med ungefär 1/6. (Se mer under Indikatorn)

Man kan uttrycka det så, att produktionstalet ”kvalitetsjusteras” genom multiplikation med citeringsmålet, och därefter fördelas (de direkta) anslagen i proportion till vad man har åstadkommit. Eftersom produktionen bygger på de *totala* forskningsanslagen (externa + direkta) så innebär metoden att universitet med hög andel externa anslag gynnas – eller, om man så vill, belönas). (Se mer under Effekterna).

Både citeringsmålet och produktionstalet beräknas ur databasen Web of Science från Thomson Scientific (nedan av gammal vana kallad ISI). Databasens täckning varierar starkt mellan olika ämnen och är inom många områden så låg att man brukar avråda från användning för utvärderingar. För beräkningen av citeringsmålet fanns inga bättre alternativ. Mer kontroversiellt är användningen av ISI för produktionsberäkningen. Valet motiveras främst med kostnadskäl; risken är att priset får betalas i lägre legitimitet (se mer under Databasen).

Det föreslagna produktionstalet är inte ämnesneutralt på institutionsnivån. Inom Chalmers kan man uppskatta att vissa institutioners forskning skulle få minst tre gånger så hög ”vikt” som andras; inom samhällsvetenskaperna och humaniora blir skillnaderna ännu högre. Det mesta av dessa skillnader torde i dagsläget jämnas ut på lärosätetsnivån, men de öppnar dörren för taktiska manipulationer (se mer under Ämnesneutralitet).

Ett tänkbart alternativ till den föreslagna indikatorn är den kvalitetsnivågrupperade och (avsett) ämnesneutrala publikationsräkningsmodell som används i Norge, skall införas i Danmark och övervägs av Finland och Island. Utredningen diskuterar ”den norska modellen”

och jämför den med det egna förslaget. Jämförelsen är emellertid inte bara retoriskt vinklad utan bygger också på tankefel och missförstånd (se mer under Norska modellen).

Slutligen skall det erkännas att den föreslagna indikatorn *är* nyskapande och t o m elegant. Det är den första modell jag sett som kombinerar ett kvalitetsmått med ett ämnesnormerat produktions- och produktivitetmått, och som dessutom kan utnyttja det senare till att ”vikta ihop” kvalitetsmättet på ett tilltalande sätt. Min bedömning blir ändå

- att metoden [ännu] inte är tillräckligt utprovad för att användas ”skarpt” i ett så viktigt sammanhang
- att den bygger på ISI-data även inom ämnesområden där det råder en relativt bred konsensus om att ISI inte kan ge tillförlitliga resultat
- samt att utredningen inte övertygande visar att den föreslagna modellen är bättre än den ”norska”.

Indikatorn : kvalitetsmättet

Som kvalitetsmått används ”fältnormaliserade medelvärden för citeringar per publikation” (CPP/FCSm). Fältnormaliseringen innebär att man jämför artiklar inom samma ämne; indelningen följer ISI:s klassifikation (242 avdelningar); artiklarna klassas (närmast med nödvändighet) via den tidskrift de publiceras i. Måttet är internationellt välkänt, utprovat och ofta använt i utvärderingssammanhang, bland annat av Vetenskapsrådet. Citeringstalet har beräknats enligt VR:s metodik, med fraktionisering av författarna. Fältnormaliseringen har gjorts mot världsmedelvärden och inte (som vid produktivitetsberäkningen) mot nordiska medelvärden; det är därför som nästan alla har värden över 1,0.

Kvalitetsmättet varierar ganska litet mellan universiteten, från 0,96 till 1,23, med ett mv över hela gruppen på 1,16 (se tabell 1 nedan). Vilket innebär att ingen ligger mer än 6% över eller 17% under gruppmedelvärdet. Omfördelningseffekten blir därför ganska liten – om publikationsindikatorn bara hade bestått av det kvalitetsmättet hade totalt c:a 2% av den berörda tårtbiten omfördelats (eller 0,4% av det totala forskningsanslaget). För Chalmers del skulle det innebära en ökning av totala forskningsanslaget med 0,8% (tabell 3 nedan).¹

[I rapporten *Resurser för citeringar* förs en utförlig diskussion kring citeringskomponenten i den föreslagna indikatorn; bland annat diskuteras alternativa citeringsindikatorer och olika sätt att beräkna den valda indikatorn. Avsnittet är utmärkt, och de val som görs är väl motiverade. Problemet är vad som *inte* sägs. I rapporten fokuseras genomgående på denna

¹ Lite oroande är att jämföra de uppmätta siffrorna med den osäkerhetsskattning som görs på s. 430 i utredningens tekniska bilaga. Där sägs att man för aggregationsnivån ”universitet” kan använda följande gränser för när avvikelser från medelvärdet kan betraktas som signifikanta:

< 0,5	mycket under
< 0,8	under
0,8-1,2	omkring
> 1,2	över
> 1,5	mycket över

Nu ligger ju alla undersökta universitet inom intervallet 0,96-1,23 (räknat mot världen) eller 0,83-1,06 (räknat mot gruppens medelvärde). Vilket innebär att kvalitetsskillnaderna inte till någon del skulle vara statistiskt signifikanta. – De nyss anförda signifikansgränserna tycks mig ha karaktär av höftning, och dessutom tycker jag mig ha sett samma gränser anges för nivån ”forskargrupp”. Kanske en fördjupad analys skulle kunna skärpa skattningen och ge en något säkrare statistisk grund för användningen av citeringsmättet som fördelningsinstrument?

(internationellt välförankrade och praktiskt beprövade) citeringskomponent, vilket framgår redan av titeln och förstärks av rapportens kvalitets-vs-kvantitetsretorik. Vad som berörs med mycket lätt hand och knappast alls diskuteras är

- att indikatorn består av en del som bygger på citeringsräkning (kvalitet) och en som bygger på publikationsräkning (kvantitet), och
- att det är *kvantitetsdelen* som står för större delen av indikatorns omfördelningseffekt.

Den fortsatta diskussionen borde enligt min mening koncentreras på kvantitetskomponenten, på vad den mäter och hur väl den gör det, och hur den eventuellt skulle kunna förbättras. De följande avsnitten kan förhoppningsvis bidra till en sådan diskussion.]

Indikatorn: produktions- och produktivitetmättet

Man skulle kunna tänkt sig ett produktionstal som beräknades genom registrering av den totala vetenskapliga produktionen vid alla universitet. Utredningen förkastar denna tanke ("för dyrt och krångligt") och föreslår istället att man uppskattar produktionen via den del som registreras i ISI (eller noga räknat genom en *delmängd* av vad som finns i ISI, eftersom man bara kan använda de artiklar där den *korresponderande författaren* är ifrån högskolan ifråga).

Nu har ju ISI en *mycket* varierande täckning inom olika ämnesområden. För att kompensera det, och för att på samma gång kompensera för att olika ämnesområden skiljer sig i "normal produktionstakt" (ett problem vid "vanlig" publikationsräkning) föreslås i utredningen en ny och tekniskt intressant modell för "fältnormering". För vart och ett av 23 breda ämnesområden beräknas en *normal "produktionstakt-i-ISI" för nordiska forskare inom området*. Detta nordiska referensvärde används sedan för att dividera respektive högskolas antal "artiklar-i-ISI", varvid man får ett ämnesoberoende mått på institutionens produktion inom området. Produktionstalet kan också uppfattas som ett mått på institutionens "effektiva forskarstyrka" (produkten av antalet forskare och deras genomsnittliga produktivitet) inom ämnesområdet.

Produktionstalet inom varje ämnesområde multipliceras så med citeringsmättet (kvalitetsmättet) för motsvarande område, och resultatet summeras till ett samlat kvalitetsjusterat produktionstal för högskolan. Vilket också innebär att kvalitetsmättet "viktas ihop" på ett sätt som inte missgynnar ämnen med liten andel artiklar i ISI.

För att beräkna referensvärdet för antalet ISI-artiklar per forskare måste man uppskatta det totala antalet aktiva nordiska forskare inom ämnet, inklusive dem som *inte* publicerat sig i ISI under undersökningsperioden. Metoden kan summeras så här. Man utgår från en empirisk lag för hur produktiviteten fördelar sig i en forskarpopulation, förutsätter att denna gäller även när man studerar produktiviteten genom ISI:s fläckiga filter, försöker ur ISI:s opålitliga efternamns+initialkombinationer plocka fram unika individuella forskare, fördelar artiklarna på dem (hur många har publicerat *en* artikel, hur många har publicerat *två* osv) och använder sedan den s k Waringfördelningens egenskaper till att extrapolera antalet forskare som har publicerat minst noll artiklar. (Det är därför som utredningens produktionstal bär Warings namn).

Metoden togs fram på 1980-talet och användes i några studier för att uppskatta totala antalet forskare inom olika länder. Stundom i god överensstämmelse med officiell statistik, stundom i sämre (varvid skulden kunde läggas på statistiken). Någon bredare användning för praktiska bibliometriska ändamål har jag inte kunnat konstatera; och jag kan inte se att den har använts för jämförelser *mellan* ämnesområden, eller i sammanhang där resultaten varit *viktiga* (t ex för fördelning av resurser). Metoden kan alltså betraktas som i sammanhanget oprövad.²

Vid diskussionen av metoden bör man nog inte fokusera alltför mycket på dess "Waringdel", dvs uppskattningen av totala antalet nordiska forskare inom olika ämnesområden. Siffrorna är kanske inte jätteexakta, men eftersom samma referensvärden används för alla kommer (som utredningen påpekar) de flesta felen att jämnas ut på högskolenivå. En större felkälla ligger nog i skattningen av högskolornas totala produktion via den i ISI synliga delen. Mer om detta under Databasen och Ämnesneutralitet.

Effekterna

På sid 414-415 (tabell 5 & 6) har indikatorns upphovsmän beräknat vilka omfördelningar den skulle ge upphov till, med utgångspunkt från vad som publicerats 2001-2004. De relevanta siffrorna har sammanförts i tabellen nedan. Produktionstalet ("Waring") finns i andra kolumnen, citeringsmättet i tredje kolumnen och produkten av dem i fjärde. I sista kolumnen har forskningsanslagen (inom 20%-tårtbiten) fördelats efter denna produkt.

Tabell 1. Sammanställning av utredningens tillämpningsexempel

	nuvarande andel %	Produktionstal (Waring)	CPP/FCSm	Waring * CPP/FCSm	beräknad andel %
LU	14,33	4401	1,15	5059	15,22
UU	13,16	4115	1,23	5049	15,19
KI	11,17	3816	1,16	4438	13,35
GU	12,53	3140	1,09	3424	10,30
SU	8,30	2560	1,22	3119	9,38
KTH	4,74	2104	1,21	2551	7,67
SLU	6,70	2022	1,23	2496	7,51
CTH	4,46	1757	1,21	2126	6,39
UMU	9,21	1641	1,14	1871	5,63
LiU	5,88	1659	1,00	1663	5,00
LTU	2,25	467	0,96	447	1,35
HHS	0,57	240	1,16	279	0,84
ORU	1,81	207	1,04	215	0,65
KAU	1,55	219	0,96	211	0,63
MIU	1,76	189	0,96	182	0,55
VXU	1,59	113	1,05	119	0,36

I nästa tabell har jag bearbetat dessa siffror. Först har jag
 - beräknat kvoten mellan beräknad och nuvarande anslagsandel ("*relativ förändring*"), och
 - fört in det *relativa kvalitetsmättet* (citeringstalet jämfört mot gruppens medelvärde 1,16).

² Jag menar alltså inte oprövad som matematisk-statistisk metod, utan som fördelningsinstrument i ett praktiskt socialt sammanhang.

Den ”relativa förändringen” är det omfördelningstal som publikationsindikatorn ger upphov till, och detta är i sin tur produkten av det *relativa kvalitetsmättet* och ett *relativt produktivitet*³. Detta produktivetsmått har synliggjorts i kolumnen längst till höger.

Tabell 2. *Relativ kvalitet och produktivitet*

	nuvarande andel %	beräknad andel %	Relativ förändring	Relativ kvalitet	Relativ produktivitet
LU	14,33	15,22	1,06	0,99	1,07
UU	13,16	15,19	1,15	1,06	1,09
KI	11,17	13,35	1,20	1,00	1,20
GU	12,53	10,30	0,82	0,94	0,87
SU	8,30	9,38	1,13	1,05	1,07
KTH	4,74	7,67	1,62	1,04	1,55
SLU	6,70	7,51	1,12	1,06	1,06
CTH	4,46	6,39	1,43	1,04	1,37
UMU	9,21	5,63	0,61	0,98	0,62
LiU	5,88	5,00	0,85	0,86	0,99
LTU	2,25	1,35	0,60	0,83	0,73
HHS	0,57	0,84	1,47	1,00	1,47
ORU	1,81	0,65	0,36	0,90	0,40
KAU	1,55	0,63	0,41	0,83	0,49
MIU	1,76	0,55	0,31	0,83	0,38
VXU	1,59	0,36	0,23	0,91	0,25

Vi ser att kvalitetetsmättet varierar ganska litet (mellan 0,83 och 1,06). Produktivetsmättet varierar mycket mer (från 0,25 till 1,55). Om indikatorn bara hade bestått av kvalitetetsmättet så hade totalt 2% av den aktuella tårbiten omfördelats (se tabellen nedan); men när även produktivetsmättet räknas så omfördelas sex gånger så mycket, dvs 12%.

Tabell 3. *Kvalitetetsmättets del av omfördelningseffekten*

	nuvarande andel %	beräknad andel %	skillnad %-andelar	Relativ kvalitet	beräknad andel "bara kvalitet"	skillnad %-andelar
LU	14,33	15,22	0,89	0,99	14,21	-0,12
UU	13,16	15,19	2,03	1,06	13,95	0,79
KI	11,17	13,35	2,18	1,00	11,17	0,00
GU	12,53	10,30	-2,23	0,94	11,77	-0,76
SU	8,30	9,38	1,08	1,05	8,73	0,43
KTH	4,74	7,67	2,93	1,04	4,94	0,20
SLU	6,70	7,51	0,81	1,06	7,10	0,40
CTH	4,46	6,39	1,93	1,04	4,65	0,19
UMU	9,21	5,63	-3,58	0,98	9,05	-0,16
LiU	5,88	5,00	-0,88	0,86	5,07	-0,81
LTU	2,25	1,35	-0,90	0,83	1,86	-0,39
HHS	0,57	0,84	0,27	1,00	0,57	0,00

³ Waringmättet är, som sagt, ett produktionsmått ("hur stor är högskolans vetenskapliga produktion?"). Men när det används som fördelningsindikator blir det också ett produktivetsmått ("hur stor är högskolans produktion i förhållande till dess direkta forskningsanslag?"). Skolor som har högre produktivitet än genomsnittet tilldelas därvid extra medel.

ORU	1,81	0,65	-1,16	0,90	1,62	-0,19
KAU	1,55	0,63	-0,92	0,83	1,28	-0,27
MIU	1,76	0,55	-1,21	0,83	1,46	-0,30
VXU	1,59	0,36	-1,23	0,91	1,44	-0,15
			12,12			2,02

Varför är skillnaderna i produktivetsmättet så mycket större? En del av förklaringen torde ligga i att produktivetsmättet är konstruerat så att det gynnar universitet med hög andel externa anslag. "Produktiviteten" i tabellen ovan är ju beräknad i relation till de *direkta* statliga anslagen; den som producerar mer än sin andel av dessa medel får ökade anslag. Men produktionen bygger ju på *hela* forskningsanslaget inklusive de externa medlen. Publikationsindikatorn blir alltså en Mattheusmaskin som ger mer (direkta) pengar till dem som redan har (externa). Precis som den andra tunga indikatorn, "andel externa forskningsmedel" (vilket i förbigående noteras på s. 416). Den sammanlagda effekten i riktning mot "färre och starkare aktörer" torde bli avsevärd.

Man kan illustrera denna effekt genom att istället beräkna produktiviteten i relation till de *totala forskningsmedlen* (hämtade ur Högskoleverkets statistik för 2007). Vi ser att i spännvidden i relativ produktivitet då blir mindre. De fem småuniversiteten (inkl Luleå) ligger på 0,44-0,61; Umeå på 0,81; övriga mellan 0,95 och 1,16 (utom HHS som är en outlier på 1,80). Hade man använt detta produktivitetstal istället för utredningens, i kombination med samma kvalitetstal, så hade totala omfördelningen blivit mindre än 7%. Vi kan också observera att Chalmers produktivitet enligt denna modell hamnar exakt på medelvärdet (1,00).

Tabell 4. Om produktiviteten mättes mot totala forskningsanslaget

	Forskn. intäkter	Produktionstal	Prod per miljon	rel produktivitet	rel kvalitet	P*K	nuv andel %	ber andel %	diff
LU	3240	4401	1,36	1,09	0,99	1,08	14,33	15,45	1,12
UU	2868	4115	1,43	1,15	1,06	1,22	13,16	16,03	2,87
KI	3200	3816	1,19	0,95	1,00	0,95	11,17	10,67	-0,50
GU	2465	3140	1,27	1,02	0,94	0,96	12,53	12,01	-0,52
SU	1773	2560	1,44	1,16	1,05	1,22	8,30	10,09	1,79
KTH	1719	2104	1,22	0,98	1,04	1,02	4,74	4,85	0,11
SLU	1663	2022	1,22	0,97	1,06	1,03	6,70	6,92	0,22
CTH	1411	1757	1,25	1,00	1,04	1,04	4,46	4,64	0,18
UMU	1620	1641	1,01	0,81	0,98	0,80	9,21	7,34	-1,87
LiU	1261	1659	1,32	1,05	0,86	0,91	5,88	5,34	-0,54
LTU	612	467	0,76	0,61	0,83	0,51	2,25	1,14	-1,11
HHS	107	240	2,24	1,80	1,00	1,80	0,57	1,02	0,45
ORU	260	207	0,80	0,64	0,90	0,57	1,81	1,03	-0,78
KAU	263	219	0,83	0,67	0,83	0,55	1,55	0,86	-0,69
MIU	275	189	0,69	0,55	0,83	0,46	1,76	0,80	-0,96
VXU	207	113	0,55	0,44	0,91	0,40	1,59	0,63	-0,96
	22944	28650	1,25						6,75

Databasen

Den föreslagna indikatorn vilar tungt på ISI, både för kvalitetsmättet och för det kvantitativa mättet (produktions- och produktivetsberäkningen). Problemet är att databasen bara täcker en del av den vetenskapliga litteraturen. Dels är den begränsad till tidskriftsartiklar, som ju

har olika stor betydelse inom olika ämnesområden. Inom biomedicin, fysik, kemi och matematik dominerar de. Konferenspublikationerna dominerar däremot inom data/elektronik och spelar en stor roll inom andra teknikområden. Inom humaniora och samhällsvetenskaper finner man en bred mix av olika publikationstyper. Utredningen är medveten om detta men menar att tidskriftsdelen av produktionen kan fungera som en rättvisande delmängd. Stundom (t ex s. 242) tycks man antyda att det bara är genom detta medium man verkligen når en internationell publik och blir citerad (vilket, som Google Scholar-användaren vet, inte riktigt stämmer).

Vidare täcker ISI bara en del av tidskriftsproduktionen, och en del som varierar starkt mellan olika ämnesområden. Utredningen tycks utgå från att ISI täcker ”internationella tidskrifter” på ”vetenskapens språk” (engelska) inom alla ämnen (se t ex s. 389), men så är det inte. För Chalmers som helhet är ISI:s täckning av (internationella engelskspråkiga) tidskriftsartiklar över 85%, men man behöver bara gå till ett ”samhällsvetenskapligt” ämne som Teknikens ekonomi och organisation för att vara nere i 35%. Inom humaniora är det tveksamt om man alls kan tala om någon seriös täckning (ISI:s konkurrent Scopus har helt undvikit detta område pga svårigheterna att åstadkomma vettiga resultat).

För Chalmers gäller att ISI täcker ungefär 50% av vår vetenskapliga produktion (räknat som summan av fackgranskade tidskriftsartiklar, d:o konferensbidrag, böcker och bokkapitel), men med stora ämnesskillnader (från <5% för Arkitektur och 15 för TeknEkon och Produkt- och Produktionsutveckling via 25-30 för Data&IT till 50 i många teknikämnen och 80+ i fysik och kemi). Ungefär samma totalsiffra gäller (enligt utredningen *Kvalitet och Förnyelse*, 2007) för ett Uppsala universitet, men med lika stora skillnader mellan fakulteterna. – Till det kommer att RUT:s kvantitetsmått (den fördelningsmässigt viktiga delen av indikatorn) pga brister i ISI bara bygger på *en del* av databasens artiklar, nämligen på de där *korresponderande författaren* kan knytas till högskolan (i vårt fall 60% av alla artiklar med Chalmersadress, eller 30% av vår totala vetenskapliga produktion).

Nu är ju både citeringsmättet och produktionssmättet i RUT:s indikator fältnormerade, dvs man försöker kompensera för skillnader i ISI-täckning genom att jämföra ”lika-med-lika” inom olika ämnesfält. Men frågan är ändå: hur säkra blir kvalitets- och kvantitetsmått som bara beräknats på en liten del av produktionen? Tillräckligt säkra för att fungera som underlag för omfördelning av anslag i stor skala?

I Henk Moeds standardverk *Citation Analysis in Research Evaluation* (2005) dras klara gränser för när ISI-databasen bör användas för forskningsutvärdering. ”When the overall ISI coverage percentage in a field is substantially lower than 50%, as it is in sociology, political science, education and anthropology, and in humanities and arts [...], it is suggested that citation analysis based on the ISI sources plays a limited role or no role at all in a research assessment study.”⁴

Linda Butler på *Research Evaluation and Policy Project* i Australien menar att man *inte bara* skall bygga en utvärdering på ISI-data inom ämnen där täckningsgraden är <50%, och *inte alls* där den är <25%. Till den sistnämnda gruppen hör (i Australien) bl a samhällsvetenskap, statskunskap, management, data & IT, juridik, arkitektur och humaniora.⁵ RUT-förslaget har tagit viss hänsyn till dålig täckning genom att utesluta humaniora från *citeringsberäkningen* (de får istället ”medelvärde”). Däremot används ISI-data för beräkningen av det (viktigare)

⁴ Henk Moed, *Citation analysis in research evaluation*, Berlin: Springer, 2005, s. 142.

⁵ Linda Butler, ICT assessment: moving beyond journal outputs, *Scientometrics* 74(2008) s. 39-55.

kvantitetsmättet även inom humanvetenskaperna. Inom övriga ämnen används ISI-data både för citerings- och kvantitetsberäkning.

Man kan tillägga att Moed och Butler talar om vid vilken täckningsgrad man bör sluta använda ISI för *citeringsmätning*. RUT-modellen innebär att man använder ISI både för citeringsmätning och kvantitetsmätning och sedan *multiplicerar* måtten, dvs adderar osäkerheterna.

Utredningens förslag bygger alltså på en användning av ISI även inom områden där databasen inte kan väntas ge upphov till säkra resultat. Man kan hävda att det väsentliga är en tillräckligt hög täckning på aggregationsnivån "lärosäte" och att felaktigheter på lägre nivå kommer att jämnas ut varandra. Men den låga täckningsnivån inom många ämnen kan medföra att deras bidrag till totalsumman domineras av andra effekter än verkliga kvalitets- och produktivitetsskillnader; och risken är att detta skadar metodens legitimitet.

[I *Resurser för citeringar* markeras (s. 8 och 40-41) en tydligare medvetenhet om ISI:s begränsningar inom humanistisk forskning, och man för fram förhoppningar om att data från ERIH skall kunna användas i kompletterande syfte.]

Ämnesneutralitet

Den föreslagna indikatorn är "fältnormaliserad", dvs den strävar efter att utjämna skillnader i publicerings- och citeringsmönster och ISI-täckningsgrad mellan olika ämnen genom att bara jämföra "lika med lika". Vid beräkningen av citeringsmättet använder man ISI:s indelning med 242 olika ämnesgrupper, vilket torde ge ett någorlunda hyggligt resultat på de i sammanhanget intressanta aggregationsnivåerna.

Vid beräkningen av produktiviteten använder man istället 23 breda "makroområden". Utredningen hävdar att dessa områden inom sig har likartade publiceringsmönster. Men vad som krävs är att ämnena inom gruppen har "likartad-produktivitet-avspeglad-i-ISI". För detta krävs

- (a) att man inom ämnena har ungefär samma "normalproduktivitet" i antal vetenskapliga publikationer per år
- (b) att fördelningen mellan tidskriftsartiklar och andra publikationstyper (t ex konferensbidrag) är ungefär densamma
- (c) att tidskriftsartiklarna täcks i ungefär samma grad av ISI

Jag har utgått från att den första premissen stämmer och har testat de två återstående, dels på de makroområden som är väsentligast för Chalmers, dels för makroområdet SOC (dvs ett område där ISI är svagt). Som underlag för Chalmers del används en studie av ISI:s täckning (<http://www.lib.chalmers.se/bibliometri/Tackningsrapport.pdf>) kompletterad med data ur CPL. För SOC-området har jag använt statistik ur Uppsala universitets utvärderingsrapport *Kvalitet och förnyelse*.

För att ämnen (på aggregationsnivån "institution") skall behandlas "neutralt" vid normeringen krävs att ISI:s täckning av den totala vetenskapliga produktionen skall vara någorlunda lika inom makroområdet. Inom vissa områden (som PHY och CHEM) gäller detta, inom andra inte. Inom MATSCI, COMPMATH, ENG och ENV hittar man en spännvidd på ungefär 1:3 mellan mest och minst "gynnad" institution. Går vi till SOC-området med sin ännu lägre ISI-

täckning finner vi en spännvidd på 1:10 mellan olika institutioner (och således också i deras relativa ”vikt” vid beräkning av produktivitetstalet).

Tabell. De för Chalmers viktigaste makroområdena

Makroområde	Motsvarande institutioner	% tidskr av prod	% ISI av tidskr	% ISI totalt
CHEM	Kemi och bioteknik (delar)	94	98	92
COMPMATH	Data & IT	35	77	27
	Matematiska vetenskaper	89	86	77
ENG	Tillämpad mekanik	49	89	44
	Produkt- o produktionsutveckling	26	65	17
	Bygg och miljö	49	81	40
	Energi och miljö (delar)	62	86	53
	Kemi- och bioteknik (delar)	94	98	92
ENV	Radio och rymd	78	100	78
	Sjöfart och marin teknik	52	53	28
	Bygg och miljö (delar)	49	81	40
	Energi och miljö (delar)	62	86	53
MATSCI	Teknisk Fysik (delar)	93	96	89
	Mikroteknologi o nanovetenskap	55	98	54
	Material och tillverkning	63	87	55
	Signaler och system	30	97	29
PHY	Teknisk Fysik (delar)	93	96	89
	Fundamental Fysik	64	100	64
	Kemi och bioteknik (delar)	94	98	92

Tabell. Institutioner vid Uppsala universitet inom makroområdet SOC

Institutioner/Centra	fakultet	Antal tidskriftsartiklar			Antal vet publikationer	
		i ISI	alla	% i ISI	alla	% i ISI
Genusvetenskap	HUM	17	37	46	77	22
Multietnisk forskning	HUM	1	12	8	43	2
Kulturanropologi/etnologi	HUM	6	94	6	151	4
Filosofi	HUM	10	29	34	67	15
Företagsekonomi	SAM	21	70	30	316	7
Pedagogik	SAM	6	22	27	102	6
Eurasiska studier	SAM	9	139	6	211	4
Statsvetenskap	SAM	27	75	36	180	15
Juridik	JUR	5	164	3	307	2
Freds- o konfliktforskning	SAM	12	34	35	156	8
Kulturgeografi	SAM	12	23	52	58	21
Sociologi	SAM	15	54	28	170	9
Bostads- och urbanforskning	SAM	31	74	42	242	13

Bland de Chalmersinstitutioner som ”råkar illa ut” i utredningens produktivitetsberäknings-system är Signaler och system, Data & IT, Produkt- och produktionsutveckling och Sjöfart & Marin teknik. I de tre första fallen pga stort konferensberoende, i det fjärde pga låg täckning i ISI. Man kan misstänka att även Teknikens ekonomi och organisation samt Arkitektur skulle ”drabbas” – Ulf Sandström har nämligen testat sin modell på KTH och då funnit att deras

motsvarande institution för Industriell ekonomi hade en "produktivitet" på bara 25% av KTH:s genomsnitt, medan arkitekternas produktivitet uppmättes till 0 (källa: presentation vid bibliometriskt seminarium på SU 29 november 2007).

Den bristande ämnesneutraliteten på institutionsnivån innebär att kvantitetsmättet är känsligt för variationer i ämnesnischning mellan olika lärosäten (kanske mest på de mindre, som inte kan förväntas ha ett fullt spektrum). Den öppnar också vägen för förbättringar av produktivetsmättet genom taktiska omfördelningar mellan ämnen. Utredningen påpekar med rätta att det föreslagna kvalitetsmättet är svårt att manipulera. Men detsamma gäller nog *inte* produktivetsmättet.

Det här påpekade problemet skulle antagligen kunna reduceras genom en normering av kvantitetsmättet mot fler och mindre ämnesgrupper. Om detta ställer till problem med avseende på referensgrupperna (de måste ha tillräcklig storlek) kanske man kunde överväga ett större tidsfönster vid beräkning av referensdata (t ex 1999-2006 istället för 2001-2004) och/eller normering mot EU-området istället för mot Norden?

[I *Resurser för citeringar* medges att indelningen i 23 ämnesområden "kan vara alltför grov" (s. 39) och att man bör pröva "en annan metod för klustringen". Möjligheterna till en förfinad ämnesindelning begränsas dock av varje klass måste innehålla tillräckling många nordiska-artiklar-i-ISI.]

Den norska modellen

Utredningens föreslagna modell för resurstilldelning jämförs genomgående med den modell som införts i Norge (och även valts av Danmark). "Den norska modellen" befinner dock "ofullkomlig" och underlägsen den egna (s. 238-243, 394-395). Jämförelsen är dock vinklad, man letar efter förtjänster i den egna modellen och fel i den norska, som dessutom presenteras på ett delvis missvisande sätt (vilket också påtalats av både norska och danska instanser).⁶ Det tyngsta underlaget för kritiken är en studie från Vetenskapsrådet (bilaga 7 i utredningen) som sägs visa att den norska modellen missgynnar medicin och gynnar humaniora. Tyvärr bygger studien på ett feltänk och är värdelös för sitt ändamål.

[Formuleringen ovan är orättvist hård. Men jag menar ändå att studiens resultat inte håller, i synnerhet när de som i utredningens huvuddel (s. 240) används utan studiens nyanseringar (s. 387). Utredningen menar att en doktorsavhandling i medicin bara ger ungefär 1/3 så många poäng som en humsamavhandling i den norska modellen. Men studiens jämförelsemetodik ger både för höga siffror för humsamavhandlingarna (inte alltid utgivna på poänggivande förlag) och för låga för medicinavhandlingarna (avhandlingsarbetet genererar fler "pek" än de som tillfaller respondenten). TL 13.10.2008]

[I *Resurser för citeringar* har hänvisningarna till denna studie försvunnit, och behandlingen av den "norska modellen" är genomgående mer korrekt.⁷]

Utredningens förslag och den norska modellen har båda inslag såväl av kvalitetsmätning som av produktivetsmätning (produktionsräkning). I den norska räknar man *hela* den vetenskap-

⁶ Se http://www.vr.se/download/18.5ec4ab911161ebf14dc80004245/svensk_bibliometri.pdf och http://www.lib.gu.se/konferenser/2007/bibliomet/Bibliometrisem_GS_nov07.pdf

⁷ Fast man tycks ännu inte ha förstått normmännens sätt att behandla konferenspublikationer (s. 34). De ignoreras inte, utan klassas som antingen seriella publikationer (löpande konferensserier med ISSN) eller monografiska ("enstaka" proceedings med bara ISBN).

liga produktionen oavsett publikationstyp eller ämne; i RUT-modellen räknar man hälften (kvalitet) eller en tredjedel (produktion), med mycket stora ämnesmässiga variationer, och försöker sedan kompensera dataluckorna med statistiska estimat. Den norska modellens kvalitetsdel bygger på att man får extrapoäng för publicering i vissa ”högkvalitativa” tidskrifter, konferensserier etc, medan utredningens bygger på citeringsanalys. För både produktions- och kvalitetsmättet gäller att RUT-modellen torde fungera bättre än den norska inom biomedicin-naturvetenskap och sämre inom andra områden; kanske är området med ”fördel Norge” något större för produktionsmättet. Till det kommer de ”ämnespolitiska” fördelarna med att den norska modellen inte utgår från en metodik och en datakälla som i grunden är byggda kring naturvetenskapernas behov. – RUT-modellens omfördelningseffekt berodde som vi sett huvudsakligen på produktivitetmättet; om samma gäller den norska vet jag inte. – I utredningen sägs att den norska modellen omfördelar för lite pengar, men några siffror ges inte, och inte heller står det klart om detta är satt i relation till den fördelningsbara tårtbitens storlek (som är fyra gånger så stor i RUT-förslaget som i den norska modellen).⁸

[Av räkneexemplen i *Resurser för citeringar* tycks framgå att båda modellerna har ungefär lika stor omfördelningspotential; nu poängteras istället deras skilda incitamentsstrukturer.]

Den norska modellens starkaste sidor är transparens, förankring och legitimitet. RUT-förslaget är däremot, som utredningen också poängterar, billigare. Den norska modellen förutsätter att man bygger upp en nationell, kvalitetssäkrad och heltäckande publikationsdatabas.

[De befintliga lokala publikationsdatabaserna håller, som en studie i *Resurser för citeringar* visar, inte måttet.]

Å andra sidan skulle en sådan databas vara värdefull oavsett hur det aktuella resursfördelningssystemet ser ut. De vetenskapliga publikationerna är en av forskningens viktigaste utprodukter – är det inte ganska självklart att denna produktion bör registreras, på ett mellan institutionerna och inom landet (eller Norden) jämförbart sätt?

Tore Lund
30.1.2008
rev 1.10.2008
rev 13.10.2008

⁸ Norge: 60% basanslag, 25% konkurrensutsatt utbildningsanslag, 15% konkurrensutsatt forskningsanslag, av dessa fördelas 12% (1,8% av totala anslaget) via publikationsindikatorn. Om utbildning och forskning förhåller sig som 25/15 även inom basanslaget så fördelas 4,5% av forskningsanslaget via publikationsindikatorn mot 20% i RUT-modellen.