

AGING AND THE PRODUCTION OF HIGH-IMPACT AND TRANSFORMATIVE SCIENCE (HITS)

Katy Börner and Robert Light (Indiana U)

Jerry Marschke and Huifeng Yu (SUNY Albany)

Joseph Staudt and Bruce Weinberg (Ohio State)

Progress report prepared for Innovation in an Aging Society meeting, Cambridge, MA, July 16, 2015. Thanks to Akina Ikudo and Kelly Chian for research assistance.

- Aim 1
 - Develop metrics and visualization tools to identify high impact and transformative science (HITScience),
 - Identify the periods in which HITScience is produced,
 - Document how these periods relate to the underlying features of fields, including researcher age.
- Aim 2
 - Document the nature of high-impact and transformative science relative to other work
 - Study the inter-relationship between high-impact and transformative work

Outline

- Highlights of last six months' progress:
 - New definition of concept of field
 - HITS metrics for all of Medline, and some new HITS
 - Relating metrics to one another
 - Relating metrics to age of field
 - A new metric display tool (for vetting with domain experts; for research)
- Next steps

Metrics

Field Level

High Impact (but not necessarily transformative)

- Works highly cited
- Large number of articles, authors
- Increase in scientist flows into field

Transformative (and high impact)

- Reduction in backward citation ages
- Durable forward citations
- Works cited broadly
- New Concepts, especially new concepts mentioned frequently in future articles
- **Obsolescence: reduction in mentions of/citations to concepts, articles, authors, important in past**

Expanded to all of Medline

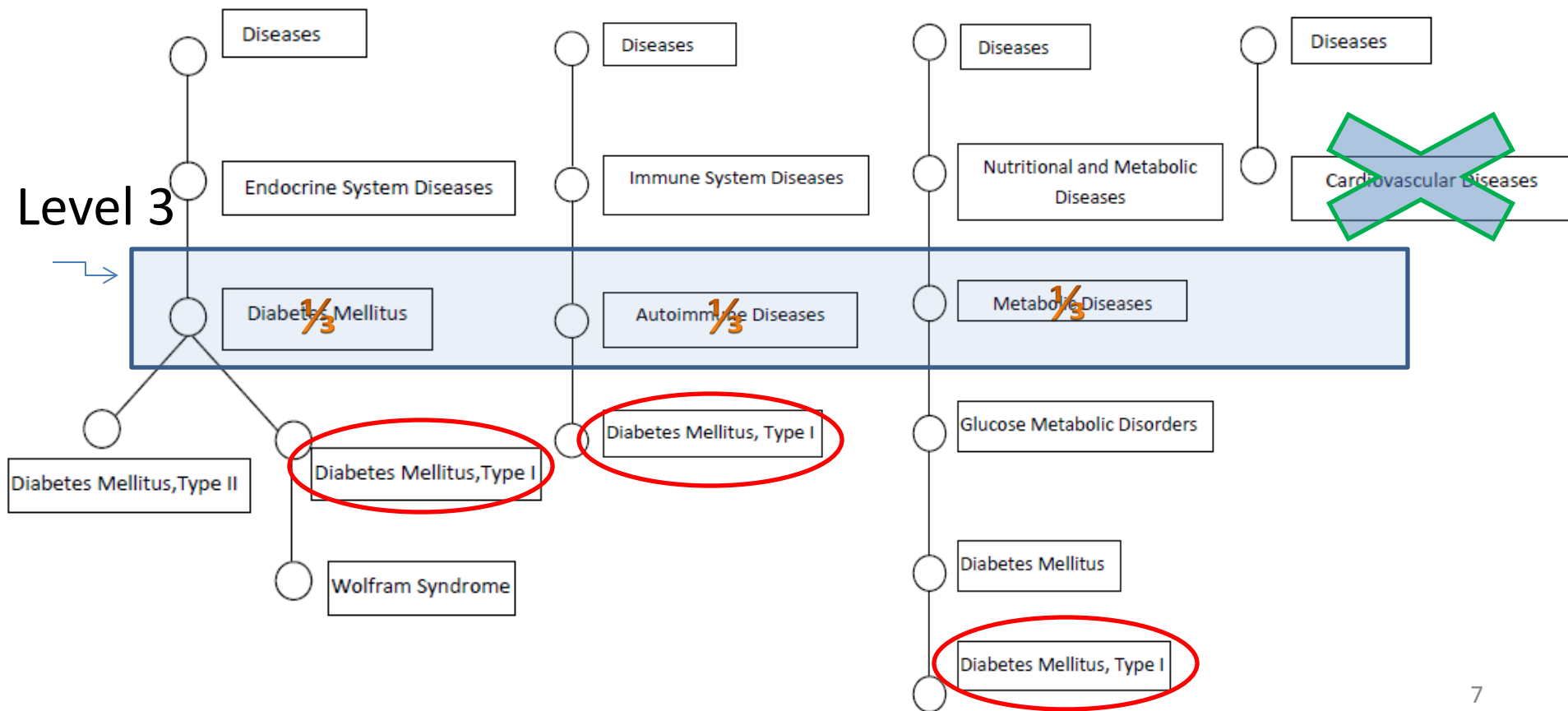
- *Last report:* As an initial step we used Neurodegenerative and Alzheimer's (NDA) as our platform for formulating measures
- *Now all of Medline:* Define fields, organize papers into fields and five year bins (1968-1972, 1973-1977, 1978-1982,...,2008-2012)
 - *Then characterize each field-year bin by HITS measures*

How we define an article's field

- Options: MeSH, journal subject categories, text.... We use MeSH
- MeSH issues
 - Multiple MeSH terms per article
 - One-to-many mapping of terms to tree locations
 - Not fully hierarchical

How we define an article's field

- Example Article—MeSH terms: Diabetes Mellitus, Type 1; Cardiovascular Diseases



HITS in Medline

- First look at HITS
 - Do these measures make sense?
 - HITS measures “log residualized” —we log measures and take deviations from field and year means, because
 - Some measures are size dependent and size of subfields vary
 - End point of data affects our measures (e.g. forward citations are right censored)

Validity of our HITS metrics

- Alternative HITS metrics should be positively correlated and they mostly are
 - Forward citation-based measures are strongly positively correlated with each other
 - Forward citation-based, text-based, size-based measures are positively correlated
 - Some surprises: In field-year bins with many forward citations, citations peter out relatively quickly (forward citations negatively correlated with forward citation age)

Correlation of HITS and Age Measures

- Overall, field age-based measures negatively correlated with citations, text measures, and entry (the entry correlation is partially mechanical)
- Ages of entrants positively correlated with citations and text measures

Metric Display Tool

- We will use it to solicit expert feedback on our metrics
- Practitioners can select their field and the metrics.
- The tool will display a heat map representation of the ups and downs over the last 45 years in the field selected, according to the metrics selected.

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW ACROBAT

Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A A Wrap Text General

B I U Merge & Center \$ % .00 .00

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

AutoSum Fill Clear Sort & Filter Find & Select

	A	B	C	D	E	F	G	H	I
1		1. Select years		2. Select measures				3. Select fields on the orange tabs	
2	<input checked="" type="checkbox"/>	1968-1972		<input checked="" type="checkbox"/>	f1	Transformativeness		Anatomy & Organisms	Anatomy
3	<input checked="" type="checkbox"/>	1973-1977		<input checked="" type="checkbox"/>	f2	Impact			Organisms
4	<input checked="" type="checkbox"/>	1978-1982		<input checked="" type="checkbox"/>	fcite95	Forward Citations 95%ile		Diseases	Diseases
5	<input checked="" type="checkbox"/>	1983-1987		<input checked="" type="checkbox"/>	fcite9999	Forward Citations 99.99%ile		Chemical & Drugs	Chemicals and Drugs
6	<input checked="" type="checkbox"/>	1988-1992		<input checked="" type="checkbox"/>	fciteMean	Forward Citations Mean		Analytical	Analytical, Diagnostic and Therapeutic Techniques and Equipment
7	<input checked="" type="checkbox"/>	1993-1997		<input checked="" type="checkbox"/>	bciteAge	Backward Citations Age			Psychiatry and Psychology
8	<input checked="" type="checkbox"/>	1998-2002		<input checked="" type="checkbox"/>	mentions	Number of times New Terms are Used Subsequently			Phenomena and Processes
9	<input checked="" type="checkbox"/>	2003-2007		<input checked="" type="checkbox"/>	concepts	New Terms			Disciplines and Occupations
10	<input checked="" type="checkbox"/>	2008-2012		<input checked="" type="checkbox"/>	articles	Articles			Anthoropology, Education, Sociology and Social Pehnomena
11				<input checked="" type="checkbox"/>	uw_authors	Authors			Technology, Industry, Agriculture
12									Humanities
13							Other		Information Science
14									Named Groups
15									Health Care
16								Publication Characteristics	
17								Geographicals	
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									

Create Report

4. The results will display on the blue tabs

UserGuide Anatomy & Organisms Diseases Chemical & Drugs Analytical Other LevelByMeasure LevelByField LogResidual ...

Data Display Prototype v3.xlsx - Excel

Marschke, Gerald R

FILE **HOME** **INSERT** **PAGE LAYOUT** **FORMULAS** **DATA** **REVIEW** **VIEW** **ACROBAT**

Clipboard: Paste, Cut, Copy, Format Painter
 Font: Calibri, 11, Bold, Italic, Underline, Color, Background Color
 Alignment: Wrap Text, Merge & Center
 Number: General, Currency, Percentage, Decimals
 Styles: Conditional Formatting, Format as Table, Cell Styles
 Cells: Insert, Delete, Format
 Editing: AutoSum, Fill, Clear, Sort & Filter, Find & Select

H30

	A	B	C	D	E	F	G	H	I
1		1. Select years		2. Select measures				3. Select fields on the orange tabs	
2	<input checked="" type="checkbox"/>	1968-1972		<input checked="" type="checkbox"/>	f1	Transformativeness		Anatomy & Organisms	Anatomy Organisms
3	<input checked="" type="checkbox"/>	1973-1977		<input checked="" type="checkbox"/>	f2	Impact		Diseases	Diseases
4	<input checked="" type="checkbox"/>	1978-1982		<input checked="" type="checkbox"/>	fcite95	Forward Citations 95%ile		Chemical & Drugs	Chemicals and Drugs
5	<input checked="" type="checkbox"/>	1983-1987		<input checked="" type="checkbox"/>	fcite9999	Forward Citations 99.99%ile		Analytical	Analytical, Diagnostic and Therapeutic Techniques and Equipment Psychiatry and Psychology
6	<input checked="" type="checkbox"/>	1988-1992		<input checked="" type="checkbox"/>	fciteMean	Forward Citations Mean			Phenomena and Processes
7	<input checked="" type="checkbox"/>	1993-1997		<input type="checkbox"/>	bciteAge	Backward Citations Age		Other	Disciplines and Occupations
8	<input checked="" type="checkbox"/>	1998-2002		<input checked="" type="checkbox"/>	mentions	Number of times New Terms are Used Subsequently			Anthropology, Education, Sociology and Social Phenomena
9	<input checked="" type="checkbox"/>	2003-2007		<input type="checkbox"/>	concepts	New Terms			Technology, Industry, Agriculture
10	<input checked="" type="checkbox"/>	2008-2012		<input type="checkbox"/>	articles	Articles			Humanities
11				<input checked="" type="checkbox"/>	uw_authors	Authors			Information Science
12									Named Groups
13								Health Care	
14								Publication Characteristics	
15								Geographicals	
16									
17									
18									
19									
20								Create Report	
21									
22								4. The results will display on the blue tabs	
23									
24									
25									
26									
27									
28									
29									
30									
31									

UserGuide | Anatomy & Organisms | Diseases | Chemical & Drugs | Analytical | Other | LevelByMeasure | LevelByField | LogResidualE ...

READY 100%

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW ACROBAT

Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A A Wrap Text

B I U Merge & Center \$ % .00 .00

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

AutoSum Fill Clear Sort & Filter Find & Select

E27 : Coinfection

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1281		Cardiovascular Diseases																		
1359		Hemic and Lymphatic Diseases																		
1393		Congenital, Hereditary, and Neonatal Diseases and Abnormalities																		
1504		Skin and Connective Tissue Diseases																		
1577		Nutritional and Metabolic Diseases																		
1603		Endocrine System Diseases																		
1604		Adrenal Gland Diseases																		
1610		Bone Diseases, Endocrine																		
1611		Diabetes Mellitus																		
1612					<input type="checkbox"/>	Diabetes Complications														
1613					<input type="checkbox"/>	Diabetes, Gestational														
1614					<input type="checkbox"/>	Diabetes Mellitus, Experimental														
1615					<input checked="" type="checkbox"/>	Diabetes Mellitus, Type 1														
1616					<input checked="" type="checkbox"/>	Diabetes Mellitus, Type 2														
1617					<input type="checkbox"/>	Donohue Syndrome														
1618					<input type="checkbox"/>	Prediabetic State														
1619		Dwarfism																		
1623		Endocrine Gland Neoplasms																		
1633		Gonadal Disorders																		
1641		Parathyroid Diseases																		
1645		Pituitary Diseases																		
1653		Polyendocrinopathies, Autoimmune																		
1654		Thyroid Diseases																		
1664		Tuberculosis, Endocrine																		
1665		Immune System Diseases																		

Data Display Prototype v3.xlsm - Excel

Marschke, Gerald R

FILE **HOME** **INSERT** **PAGE LAYOUT** **FORMULAS** **DATA** **REVIEW** **VIEW** **ACROBAT**

Clipboard: Paste, Cut, Copy, Format Painter

Font: Calibri, 11, Bold, Italic, Underline, Color, Background Color

Alignment: Wrap Text, Merge & Center

Number: General, Currency, Percentage, Decimals

Styles: Conditional Formatting, Format as Table, Cell Styles

Cells: Insert, Delete, Format

Editing: AutoSum, Fill, Clear, Sort & Filter, Find & Select

H30	A	B	C	D	E	F	G	H	I
1	1. Select years		2. Select measures					3. Select fields on the orange tabs	
2	<input checked="" type="checkbox"/>	1968-1972	<input checked="" type="checkbox"/>	f1	Transformativeness			Anatomy & Organisms	Anatomy
3	<input checked="" type="checkbox"/>	1973-1977	<input checked="" type="checkbox"/>	f2	Impact				Organisms
4	<input checked="" type="checkbox"/>	1978-1982	<input checked="" type="checkbox"/>	fcite95	Forward Citations 95%ile			Diseases	Diseases
5	<input checked="" type="checkbox"/>	1983-1987	<input checked="" type="checkbox"/>	fcite9999	Forward Citations 99.99%ile			Chemical & Drugs	Chemicals and Drugs
6	<input checked="" type="checkbox"/>	1988-1992	<input checked="" type="checkbox"/>	fciteMean	Forward Citations Mean			Analytical	Analytical, Diagnostic and Therapeutic Techniques and Equipment
7	<input checked="" type="checkbox"/>	1993-1997	<input type="checkbox"/>	bciteAge	Backward Citations Age				Psychiatry and Psychology
8	<input checked="" type="checkbox"/>	1998-2002	<input checked="" type="checkbox"/>	mentions	Number of times New Terms are Used Subsequently			Other	Phenomena and Processes
9	<input checked="" type="checkbox"/>	2003-2007	<input type="checkbox"/>	concepts	New Terms				Disciplines and Occupations
10	<input checked="" type="checkbox"/>	2008-2012	<input type="checkbox"/>	articles	Articles				Anthoropology, Education, Sociology and Social Pehnomena
11			<input checked="" type="checkbox"/>	uw_authors	Authors				Technology, Industry, Agriculture
12									Humanities
13									Information Science
14								Named Groups	
15								Health Care	
16								Publication Characteristics	
17								Geographicals	
18									
19									
20								Create Report	
21									
22									
23								4. The results will display on the blue tabs	
24									
25									
26									
27									
28									
29									
30									
31									

READY
UserGuide Anatomy & Organisms Diseases Chemical & Drugs Analytical Other LevelByMeasure LevelByField LogResidualt ...
100%

Data Display Prototype v3 [Repaired] - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Acrobat

Normal Page Layout Page Break Preview Custom Views Full Screen Workbook Views

Ruler Formula Bar Gridlines Headings Message Bar Show/Hide

Zoom 100% Zoom to Selection

New Window Arrange All Freeze Panes Unhide

Split Hide View Side by Side Synchronous Scrolling Reset Window Position Window

Save Workspace Switch Windows Macros

K18

	A	B	C	D	E	F	G	H	I	J	K
1	Diabetes Mellitus, Type 1	1968-1972	1973-1977	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	2008-2012	
2	f1	-1.10725	-0.36084	-0.74788	0.717132	0.567536	0.250604	-0.46511	-0.16394	-0.20632	
3	f2	2.989908	2.475632	1.851335	-0.97202	-1.43197	-0.2091	0.898719	0.20178	0.073313	
4	fcite95	0.133955	0.365711	0.064599	-0.10409	-0.01001	-0.06864	-0.10702	0.07886	0.079409	
5	fcite9999	-0.90094	-1.45887	-1.36386	-0.7719	-0.33445	0.046859	-0.19717	1.05471	0.148656	
6	fciteMean	0.286956	0.228229	0.038785	-0.07553	-0.10931	-0.04545	-0.0455	0.060458	0.088735	
7	mentions	-0.37858	-0.17042	-0.12238	0.388676	0.31059	0.170016	-0.01608	-0.19195	-0.28779	
8	uw_authors	-1.44427	-1.13483	-0.72996	0.214549	0.332003	0.265272	0.03797	-0.12007	-0.19324	
9											
10	Diabetes Mellitus, Type 2	1968-1972	1973-1977	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	2008-2012	
11	f1		-9.57391	-11.596	-2.08143	-0.83095	-1.01831	-0.5906	0.202235	0.669828	
12	f2		2.877427	-0.57936	-1.28772	-0.49657	-0.34534	1.196931	0.779376	-0.68823	
13	fcite95		-1.04509	-0.88817	-0.29106	0.005422	-0.11579	0.080469	0.06332	-0.01764	
14	fcite9999		-4.60472	-4.17048	0.48253	1.085294	-0.53311	-0.17822	0.385315	-0.23783	
15	fciteMean		-0.82386	-1.15679	-0.15189	0.113772	-0.09862	0.029558	0.044777	-0.01793	
16	mentions		-2.62443	-3.58498	-0.96976	-0.72126	-0.39525	-0.20941	0.126127	0.294134	
17	uw_authors		-8.57771	-5.35164	-1.02833	-0.60971	-0.29344	-0.13119	0.130275	0.226455	
18											

Analytical Other LevelByMeasure LevelByField LogResidualByMeasure LogResidualByField

Ready 150%

Home Insert Page Layout Formulas Data Review View Acrobat

Normal Page Layout Page Break Preview Custom Views Full Screen Workbook Views

Ruler Formula Bar Gridlines Headings Message Bar Show/Hide

Zoom 100% Zoom to Selection

New Window Arrange All Freeze Panes Unhide

Split Hide Synchronous Scrolling Reset Window Position Window

View Side by Side Save Workspace Switch Windows

Macros

C6 fx 0.2282291

	A	B	C	D	E	F	G	H	I	J	K
1	Diabetes Mellitus, Type 1	1968-1972	1973-1977	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	2008-2012	
2	f1	-1.10725	-0.36084	-0.74788	0.717132	0.567536	0.250604	-0.46511	-0.16394	-0.20632	
3	f2	2.989908	2.475632	1.851335	-0.97202	-1.43197	-0.2091	0.898719	0.20178	0.073313	
4	fcite95	0.133955	0.365711	0.064599	-0.10409	-0.01001	-0.06864	-0.10702	0.07886	0.079409	
5	fcite9999	-0.90094	-1.45887	0.26286	0.7710	0.32445	0.046850	0.10717	1.05471	0.148656	
6	fciteMean	0.286956	0.228229	0	suscept gene; regulatori t						
7	mentions	-0.37858	-0.17042	Title: ISLET-CELL ANTIBODIES IN DIABETES-MELLITUS WITH AUTOIMMUNE POLY-ENDOCRINE							
8	uw_authors	-1.44427	-1.13483	Authors: BOTTAZZO, GF; FLORINCH.A; DONIACH, D							
9				Journal: LANCET							
10	Diabetes Mellitus, Type 2	1968-1972	1973-1977	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	2008-2012	
11	f1		-9.57391	Title:							
12	f2		2.877427	Authors: KUZUYA, H; BLIX, PM; ...; RUBENSTEIN, AH							
13	fcite95		-1.04509	Journal:							
14	fcite9999		-4.60472	Year: 1977							
15	fciteMean		-0.82386	Title: HL-A ANTIGENS AND DIABETES-MELLITUS							
16	mentions		-2.62443	Authors: NERUP, J; PLATZ, P; ...; SVEJGAAR.A							
17	uw_authors		-8.57771	Journal: LANCET							
18				Year: 1974							

Next steps

- Vetting ideas for measures (Survey Monkey-WoS emails; Conferences; NIH Study Section scores)
- We are now wrapping up article-level measures of scientific impact/transformativeness and age (team age, age of oldest collaborator, etc)
- Publications:
 - We're close to having the field-level measures we said we would produce
 - Submit 1-2 articles to general science journals
 - Article-based measures for article based analysis