

Dr. Rainer Frietsch, in Zusammenarbeit mit Lutz Bornmann und Sonia Gruber

Highly-Cited Researcher - Implikationen unterschiedlicher Implementierungen

Agenda

1. Einleitung: Was sind Highly-Cited Researcher (HCR)?

- a. Was sind HCR?
- b. Die Methode von Clarivate Analytics
- c. Kritik an der Umsetzung

2. Alternative Implementierungen

3. Implikationen der Varianten

4. Zusammenfassung und Empfehlungen

Einleitung: was sind Highly-Cited-Researcher (HCR)

Was sind HCR?

Zunächst ist es ein Konstrukt von ThomsonReuters bzw. Clarivate Analytics

- Als Definition kommt prompt auch ein Verweis auf die Datenbank (bspw. Wikipedia: https://de.wikipedia.org/wiki/Highly_Cited_Researchers)
- Laut Clarivate ist das Ziel der HCR-Liste "to recognize scientists and social scientists with community-wide influence"
(Szomszor, Pendlebury, & Adams, 2020, p. 1124)

Nimmt man allerdings den Titel „Hochzitierte Wissenschaftler:innen“ ernst, dann entstehen eine Reihe von Fragen, die meist in Annahmen oder Festlegungen münden

- Was bedeutet „hochzitiert“ (in welchem Zeitraum, fachspezifisch, ...)?
- Welcher Personenkreis wird berücksichtigt?

Gegenstand unserer Analyse: Die Festlegungen führen zu sehr unterschiedlichen Ergebnissen

- Manche der Annahmen und Festlegungen sind mehr und manche weniger plausibel
- Sie haben sehr unterschiedliche Effekte auf Personen oder Organisationen
- Es gibt also nicht **DIE** Gruppe/Implementierung von HCR

Die Methode von Clarivate Analytics (ehem. ThomsonReuters)

Identifikation der 1% am höchsten zitierten Publikationen nach 21 Feldern

- aus methodischen Gründen wird jede Publikation (über die Zeitschrift) lediglich einem einzigen Feld zugewiesen
- lediglich multidisziplinäre Zeitschriften werden auf Artikel-Basis zugewiesen

Auswahl der Personen mit den meisten absoluten Zitierungen innerhalb von Feldern

- Autor:innen-Disambiguation notwendig (Clarivate dazu: "For each ESI field, author names are disambiguated through advanced clustering methods...")
- Die Quadratwurzel aus der Zahl der Autor:innen in einer nach Anzahl hochzitiertes Publikationen geordneten Liste legt den Grenzwert fest (außerdem muss auch der Grenzwert der Zitierungen im Zeitraum erreicht werden)

Seit 2018 zusätzlich Berücksichtigung von Autor:innen, die in mehreren Feldern publizieren

- Summe aus der feldspezifischen Abweichung von den Grenzwerten aller Felder einer/s Autor:in muss größer 1 sein

Kritik an der Umsetzung von Clarivate I

Selbstkritik:

- „Each measure or set of indicators, whether total citations, h-index, relative citation impact, mean percentile score, etc., accentuates different types of performance and achievement. Here we arrive at what many expect from such lists but what is unobtainable: that there is some optimal or ultimate method of measuring performance.“

Lediglich Autor:innen auf den 1% am häufigsten zitierten Publikationen werden berücksichtigt

(Aksnes und Aagaard 2021; Docampo und Cram 2019)

- Eine Autor:innen-Disambiguierung (author ID) gibt es in WoS nicht
- Händisches Kuratieren notwendig; Methode nicht eindeutig beschrieben

Keine Berücksichtigung von Rangfolgen von Autor:innen

- Papiere mit mehr als 30 Autor:innen werden ausgeschlossen (bis 2018: mehr als 30 Affiliationen)

Keine Berücksichtigung von Mehrfachaffiliationen

- was speziell bei „Ehren“-Affiliationen oder „Gekauften“ Affiliationen eine massive Verzerrung mit sich bringt
(siehe bspw. Bornmann und Bauer 2015)

Kritik an der Umsetzung von Clarivate II

Alle Publikationen und Zitierungen aus dem Zeitraum 2010-2020 (für die Liste 2021) verwendet

- keine Zitationsfenster => ältere Publikationen und Autor:innen haben bessere Chancen

Sehr geringe Anzahl an ausgewählten Autor:innen

- Auswahl von ca. 6.000 Personen was lediglich ca. 0,1% aller Autor:innen entspricht (Ioannidis et al. 2019)

Alternative Implementierungen

Methodische Variationen als Alternativen I

Umsetzung der Clarivate-Methode auf Basis von Scopus-Daten (Variante 0)

- 3-Jahres-Zitations-Fenster, aggregierte Daten für 2010-2018, keine Selbstzitationen, Ausgangspunkt sind die 1% hochzitierten Publikationen (HCP); Papiere mit mehr als 30 Affiliationen werden ausgeschlossen

Hochzitierte Autor:innen in der gesamten Datenbank auf Basis absoluter Zitationszahlen (Variante 1)

- Anstatt auf die HCP einzuschränken, werden alle Publikationen in Scopus pro Autor:in berücksichtigt, allerdings nur Fokus auf Zitierungen (nicht auch noch Anzahl der Publikationen); 3-Jahres-Zitationsfenster; Berechnung auf Jahresbasis und anschließend Aggregation (2010-2018); keine Selbstzitationen, **top1% Autor:in pro Feld** definiert den Grenzwert
- Highly-cited ist in diesem Fall definiert als die absolut meisten Zitierungen (pro Feld) in einem vorgegebenen Zeitraum
- Vorteil: alle erhaltenen Zitierungen (in 3 Jahren) werden berücksichtigt; Nachteil: besonders produktive Autor:innen haben größere Chancen, berücksichtigt zu werden; Mehrfachzählungen von Zitierungen bei Ko-Publikationen

Methodische Variationen als Alternativen II

Fraktionierte Zählweise von Zitierungen (Variante 2)

- entspricht Variante 1, jedoch werden die Zitierungen fraktioniert gezählt

Fokus auf Korrespondenzautor:innen (Variante 3)

- entspricht Variante 1, jedoch werden die Zitierungen lediglich für den/die Korrespondenzautor:in gewertet

Nutzung der Liste von Ioannidis et al. 2020 (Variante 4) für Kritik siehe Moed 2021

- Kompositindikator aus sechs Einzelindikatoren (Zitierungen, H-Index, korrigierter H-Index (Schreiber H-Index), Anzahl der Einzelautor:in-Papiere, Zitierungen auf Einzel-Papiere, Einzel- und Erstautor:in-Papiere, Zitierungen auf Einzel- und Erst-Autor:in-Papiere) in 22 Feldern bzw. 176 Sub-Feldern; top-100.000 insgesamt oder top-2% in den Subdisziplinen; ohne Selbstzitierungen; Jahre 2010-2020

Implikationen der Varianten

Fünf Indikatoren für den empirischen Vergleich

Abdeckung der Autor:innen in der Scopus-Datenbank

- Anteil bezogen auf die komplette Anzahl von Autor:innen in der Periode 2010-2018

Vergleiche auf der institutionellen Ebene

- Anteile von HCR in den außeruniv. Forschungsorganisationen und der Unis in Deutschland und Relationen untereinander

Durchschnittliches wissenschaftliches Alter

- Clarivate behauptet, dass junge Autor:innen in ihrer Variante eine adäquate Berücksichtigung finden; wir definieren das wissenschaftliche Alter als die Differenz Jahr der ersten Veröffentlichung und dem Analysejahr (2022)

Anteil der publizierenden Nobelpreisträger:innen

- Wir haben alle Nobelpreisträger:innen seit den 1980er Jahren in Scopus identifiziert und prüfen, wie viele der publizierenden Nobelpreisträger:innen durch die jeweilige Variante in die Liste der HCR aufgenommen werden

Disruptionsindex (Bornmann et al. 2020; Leydesdorff et al. 2021)

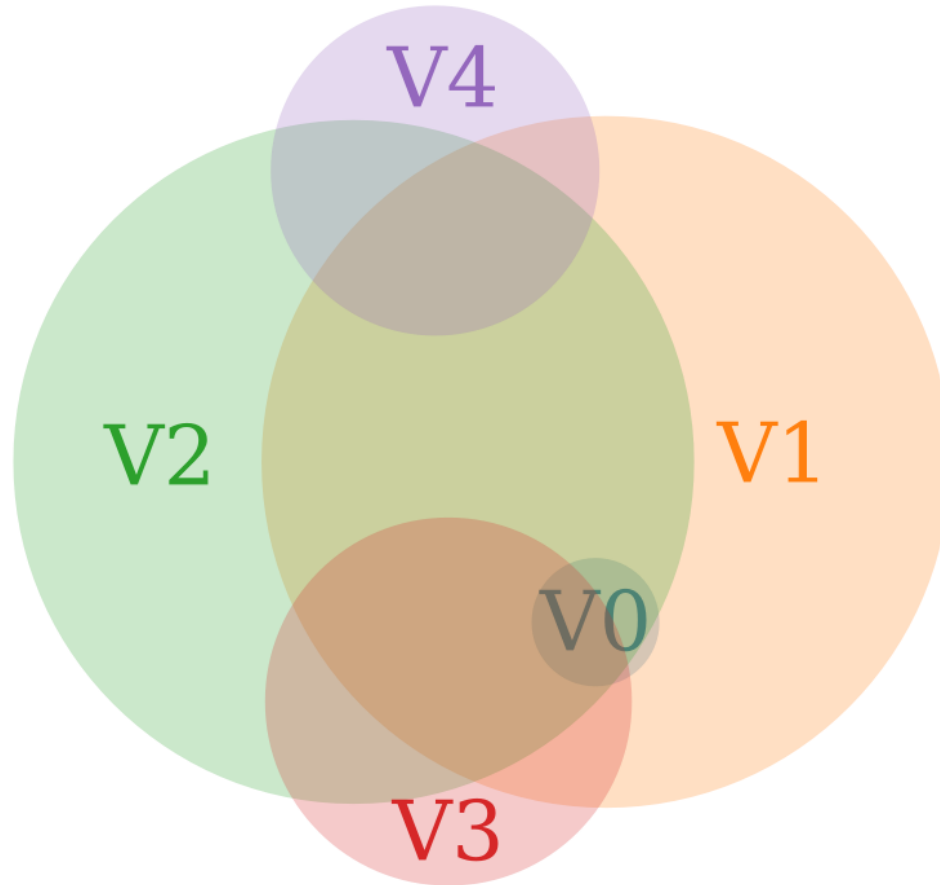
- Berücksichtigung der Originalität des Beitrags (bspw. Überblicksartikel vs. originäre Forschung); die Erwartung ist, dass eine HCR-Liste genuine Forschung besser berücksichtigt

Indikator 1: Abdeckung der Varianten

Variante	Zahl der Basis der Autor:innen	Anzahl der HCR	Anteil der HCR an allen Autor:innen
V0	446.399	14.455	0,07%
V1	20,1 Mio.	420.196	2,1% (1,5-1,6% p.a.)
V2	20,1 Mio.	410.335	2,02%
V3	5,7 Mio.	119.081	2,1% der CA und 0,6% aller
V4	20,1 Mio.	95.654	0,48%

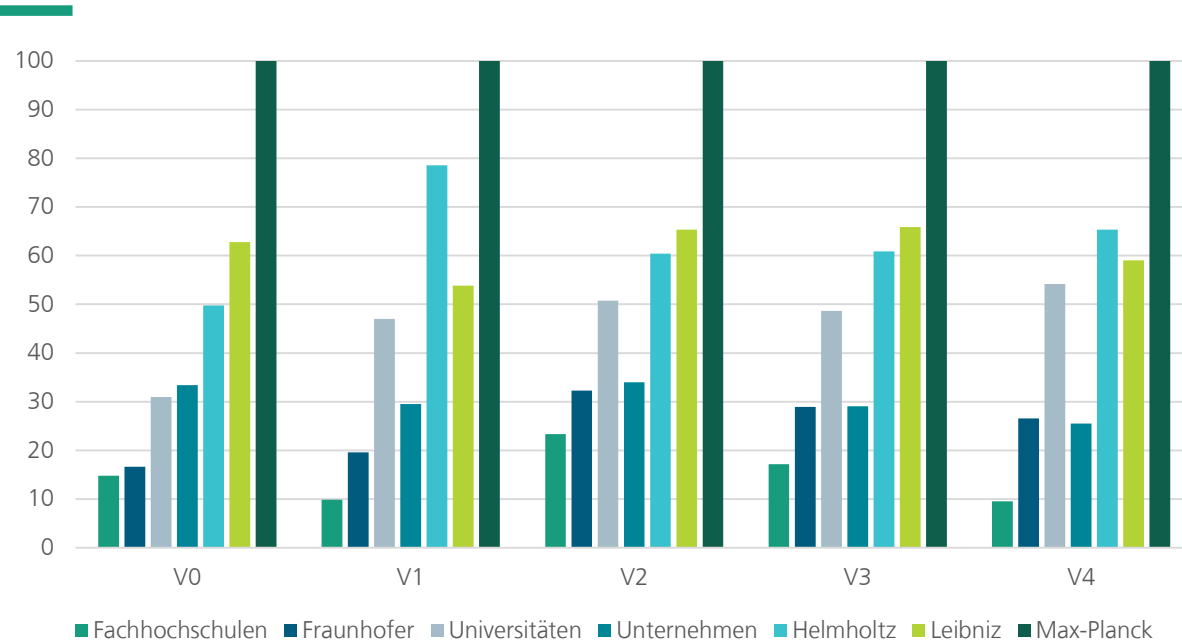
Überschneidung der Varianten

Varianten Überschneidung



	V0	V1	V2	V3	V4
V0		3,4%	3,3%	7,8%	6,6%
V1	99,8%		54,8%	64,7%	48,3%
V2	94,5%	53,5%		73,2%	60,5%
V3	64,4%	18,3%	21,3%		28,9%
V4	42,9%	10,9%	13,9%	22,9%	

Indikator 2: AUF/Unis und deren Relationen untereinander



	V0	V1	V2	V3	V4
Max-Planck	1,2%	5,1%	21,5%	8,0%	5,3%
Leibniz	0,8%	2,8%	14,1%	5,3%	3,1%
Helmholtz	0,6%	4,0%	13,0%	4,9%	3,5%
Unternehmen	0,4%	1,5%	7,3%	2,3%	1,4%
Universitäten	0,4%	2,4%	10,9%	3,9%	2,9%
Fraunhofer	0,2%	1,0%	7,0%	2,3%	1,4%
Fachhochschulen	0,2%	0,5%	5,0%	1,4%	0,5%

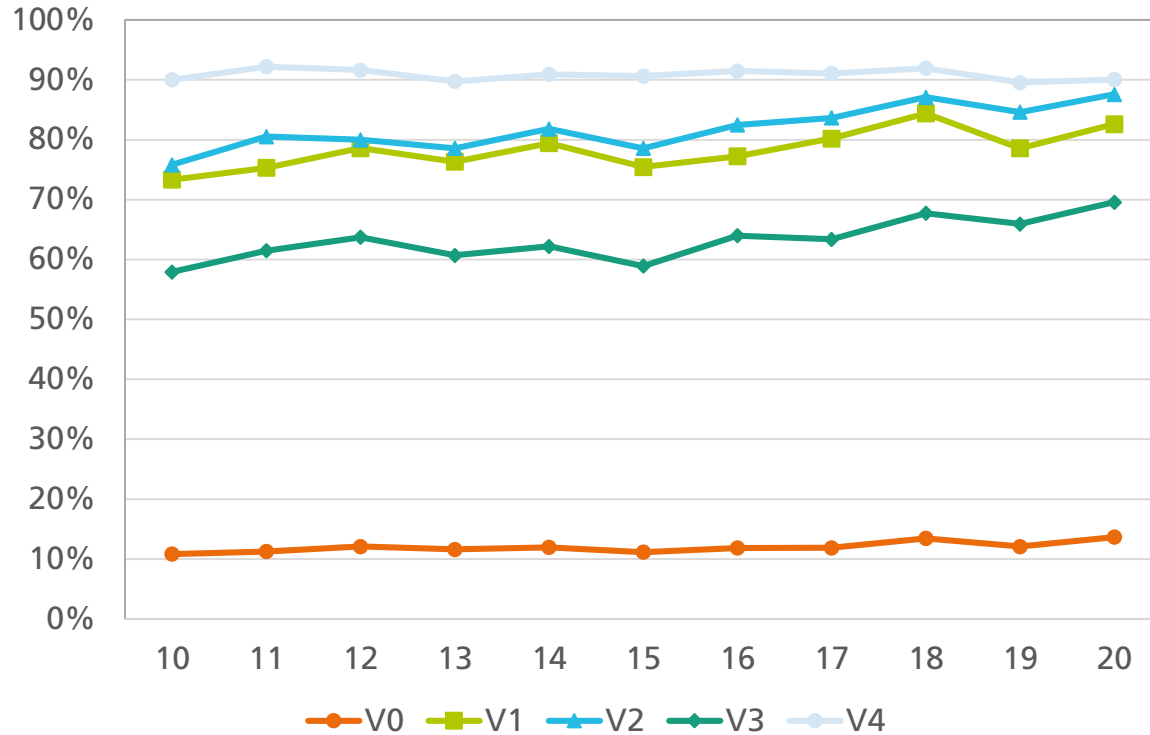
Index (Max-Planck = 100) Index der HCR, 2018

Indikator 3: durchschnittliches wissenschaftliches Alter

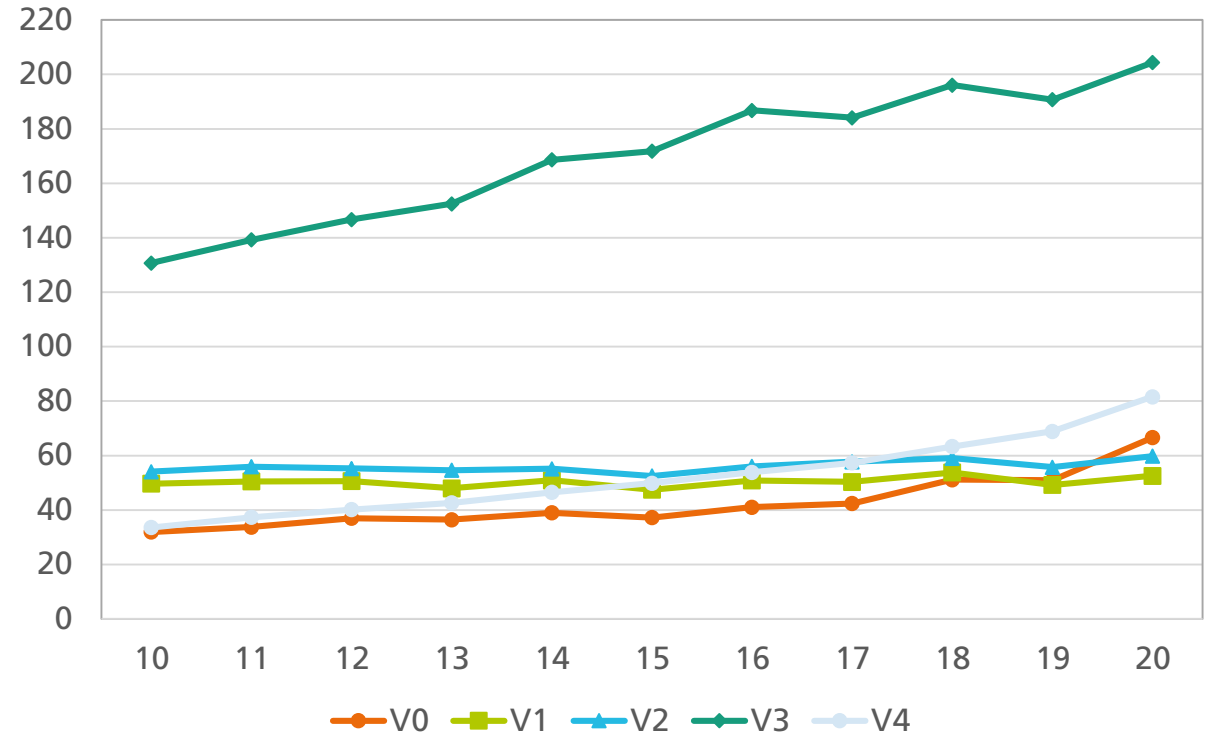
	Total	V0	V1	V2	V3	V4
Total	9,1	19,4	15,8	16,4	18,1	16,2
Mathematics	12,2	20,4	16,3	16,5	17,0	25,1
Computer and information sciences	11,1	20,4	15,7	15,7	16,3	25,1
Physical Sciences	11,3	19,3	16,2	16,8	18,2	25,3
Chemical Sciences	10,8	19,4	15,1	15,9	19,0	25,3
Earth and related envir. sciences	11,4	20,7	17,7	18,1	17,9	25,2
Biological sciences	10,8	19,9	16,7	17,3	20,2	25,4
Civil engineering	10,6	19,4	15,3	15,3	16,2	24,7
Electr. electr., information engineering	10,9	18,4	15,8	16,1	16,9	24,9
Mechanical engineering	10,6	18,4	15,1	15,4	17,7	25,1
Chemical engineering	11,0	18,2	15,2	15,8	18,4	25,1
Materials engineering	10,4	18,4	14,9	15,8	18,5	25,2
Environmental engineering	11,1	19,5	16,0	16,0	17,5	25,2
Environmental biotechnology	11,9	18,6	15,3	15,7	18,3	25,0
Other engineering and technologies	11,0	18,8	15,0	15,3	17,6	25,1
Basic medicine	10,8	20,7	15,9	18,0	20,4	25,4
Clinical medicine	10,2	21,6	17,9	18,3	20,3	25,5
Health sciences	11,2	21,6	17,2	17,2	18,2	25,4
Agriculture, forestry, and fisheries	12,0	21,7	17,7	17,7	18,7	25,4
Animal and dairy science	12,2	21,9	18,2	18,4	19,1	25,5
Veterinary science	11,8	22,0	17,6	18,3	19,5	25,5
Psychology	11,8	20,9	17,2	17,4	17,3	25,2
Economics and business	10,7	19,8	15,7	15,7	16,3	24,8
Educational sciences	11,1	21,0	16,0	15,4	15,9	25,1
Sociology	11,5	20,6	17,2	16,5	17,0	25,0
Law	11,2	19,9	15,6	15,7	16,2	24,9
Political science	11,8	20,4	16,3	16,2	17,0	24,9
Social and economic geography	11,4	20,2	16,1	15,7	16,1	24,8
Media and communications	11,6	19,5	14,8	14,7	15,2	24,7
History and archaeology	11,9	21,9	17,9	16,9	17,8	25,2
Language and literature	10,9	20,7	16,7	16,6	16,6	24,9
Philisophy, ethics and religion	11,2	20,7	15,6	15,7	16,1	25,0
Art	12,1	21,2	16,6	16,3	16,9	25,1
Other humanities	11,8	20,0	16,2	15,6	16,2	24,8
Multidisciplinary	13,1	20,5	16,0	16,2	20,2	25,3

	Total	V0	V1	V2	V3	V4
DE	10,9	19,0	14,6	14,5	16,3	25,3
US	10,9	18,8	14,5	14,5	17,4	25,4
CN	7,1	13,8	9,5	9,6	13,6	24,7
GB	11,0	19,5	14,8	14,7	16,5	25,3

Indikator 4: Nobelpreisträger:innen

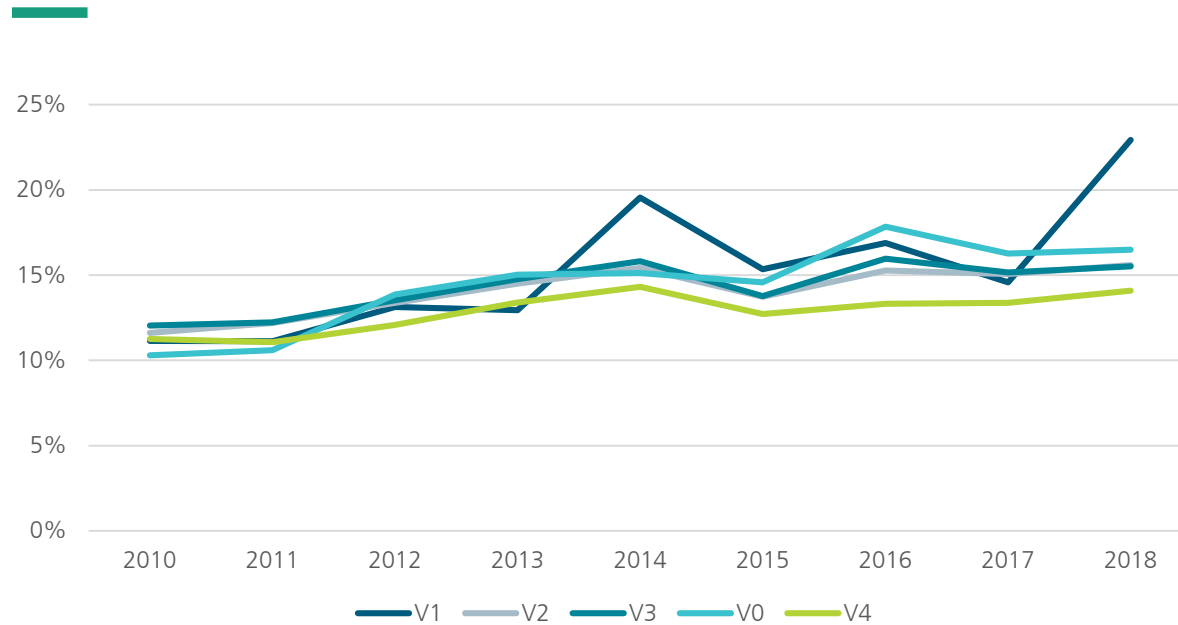


Anteile der im jeweiligen Jahr publizierenden Nobelpreisträger:innen

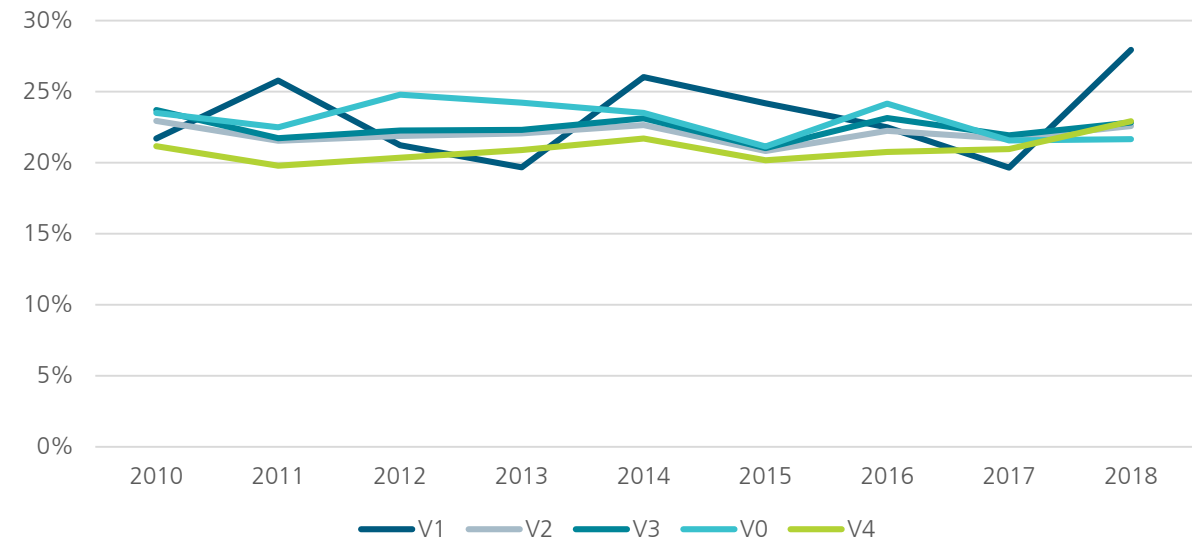


Wahrsch. der Erfassung von publizierenden Nobelpreisträger:innen

Indikator 4: Nobelpreisträger:innen



Anteile der 10% “disruptivsten” Autor:innen



Anteile der 25% “disruptivsten” Autor:innen

Zusammenfassung und (wissenschafts-)politische Implikationen

Zusammenfassung

Es gibt keine eindeutige Definition von HCR

Je nach Implementierung kommen bei Verwendung der HCR sehr unterschiedliche Ergebnisse zutage

Die Umsetzung von Clarivate scheint nach den hier vorgelegten Vergleichsrechnungen die restriktivste und gleichzeitig am meisten verzerrte Variante zu sein

Bei der Entscheidung für die ein oder andere Implementierung sollte die zu untersuchende Fragestellung bereits berücksichtigt werden

Vielen Dank für Ihre Aufmerksamkeit
