

# Report zur bibliometrischen Analyse

BIH Charité (Junior) Clinician Scientist  
Programmevaluation

Nico Riedel

Dieses Werk steht unter der Creative Commons Namensnennung, nicht kommerziell, keine Bearbeitung 3.0 Deutschland Lizenz (CC-BY-NC-ND)

<https://creativecommons.org/licenses/by-nc-nd/3.0/de/>



**Projektverantwortlicher – BIH QUEST Center for Transforming Biomedical Research**

Dr. Nico Riedel

Telefon +49 (0)30 450 543 009

E-Mail: [nico.riedel@bih-charite.de](mailto:nico.riedel@bih-charite.de)

**Projektverantwortliche – BIH Biomedical Innovation Academy**

Dr. Nathalie Huber

Dr. Rüdiger Hesse

**Impressum**

**Herausgeber**

Berliner Institut für Gesundheitsforschung in der Charité/

Berlin Institute of Health at Charité (BIH)

Anna-Louisa-Karsch-Str. 2 | 10178 Berlin, Germany

[www.bihealth.org/](http://www.bihealth.org/)

**Vorsitzender BIH Direktorium | Vorstand Charité Translationsforschungsbereich**

Univ.-Prof. Dr. med. Christopher Baum

Juni 2021

# BIH Charité (Junior) Clinician Scientist Programmevaluation - Report zur bibliometrischen Analyse

Nico Riedel - QUEST Center for Transforming Biomedical Research  
Berlin Institute of Health at Charité

3.3.2021

## Inhaltsverzeichnis

<b>Kurzzusammenfassung</b>	<b>2</b>
<b>1. Einleitung – Übersicht Indikatoren</b>	<b>3</b>
Übersichtsliste Indikatoren: . . . . .	3
<b>2. Methodik</b>	<b>4</b>
2.1. Definition Jahrgangskohorten . . . . .	4
2.2. Publikationssuche . . . . .	5
2.3. Erfassen der Metriken . . . . .	5
2.4. Datenquelle CSP-Berichte . . . . .	7
<b>3. Ergebnisse - CSP + JCSP</b>	<b>9</b>
3.1. Produktivität/Übersicht . . . . .	9
3.2. Resonanz . . . . .	11
3.3 Setting: klinisch/animal/molecular & cellular . . . . .	12
3.4 Forschungsfelder . . . . .	13
3.5 Teamscience . . . . .	14
3.6 Transparenz & Open Science . . . . .	16
3.7 Alternativer Forschungsoutput . . . . .	17
<b>4. Ergebnisse - Vergleich mit Kontrollgruppe</b>	<b>18</b>
4.1. Produktivität/Übersicht . . . . .	18
4.2. Resonanz . . . . .	20
4.3 Setting: klinisch/animal/molecular & cellular . . . . .	22
4.4 Forschungsfelder . . . . .	23
4.5 Teamscience . . . . .	24
4.6 Transparenz & Open Science . . . . .	25
4.7 Alternativer Forschungsoutput . . . . .	26
<b>5. Diskussion</b>	<b>27</b>
<b>6. Zusammenfassung</b>	<b>27</b>
<b>7. Literaturverzeichnis</b>	<b>28</b>
<b>8. Abkürzungsverzeichnis</b>	<b>28</b>
<b>9. Anhang</b>	<b>28</b>
9.1 Weitere Abbildungen für alle Publikationen, inkl. Mittelautorenschaften . . . . .	28

## Kurzzusammenfassung

Der vorliegende Report untersucht den wissenschaftlichen Output des BIH Charité (Junior) Clinician Scientist Programms anhand einer breiten Vielfalt bibliometrischer Metriken. Diese Analyse ist komplementär zum zweiten Teil der Programmevaluation, welcher auf Interviews und Surveys basiert (Hendriks et al., 2021). In diesem Report werden folgende Themen abgedeckt und durch die Analyse sichtbar und messbar gemacht: Thematik der Forschungsarbeiten, Kollaborationen, Zitationen, Open Science sowie alternativer Forschungsoutput. Die Analyse ist in zwei Blöcke aufgeteilt. Im ersten Teil wird die gesamte Kohorte des Clinician Scientist Programms (2013-19, N=110) sowie des Junior Clinician Scientist Programms (2015-19, N=66) betrachtet. Im zweiten Teil wird die Untergruppe der habilitierten Clinician Scientists (N=52) mit anderen habilitierten Charité-Wissenschaftler\*innen (N=351) verglichen.

Die Analyse zeigt, dass in den Jahren ab 2017 die Clinician Scientists einen großen Anteil aller habilitierenden Forscher\*innen an der Charité ausmachten. Die Clinician Scientists decken eine große Bandbreite an Forschungsfeldern als auch verschiedene Phasen in der Translationskette ab. Hierbei ist der Fokus aber stets am stärksten auf der kliniknahen Forschung. Außerdem zeigen die Clinician Scientists ein hohes Maß an Kollaborationen auf allen Ebenen, sowohl mit anderen Clinician Scientists als auch mit nationalen und internationalen Kooperationspartnern. Für den Bereich Open Science konnten die Clinician Scientists die untersuchten Themen Open Access, Open Data und Open Code besser abdecken als die Vergleichsgruppe. Absolut gesehen gibt es bei diesen Themen jedoch noch Potential zur Verbesserung - was jedoch auf die gesamte biomedizinische Forschungsliteratur zutrifft.

# 1. Einleitung – Übersicht Indikatoren

In diesem Bericht sind die Ergebnisse der bibliometrischen Analyse des BIH Charité (Junior) Clinician Scientist Programm (CSP/JCSP) zusammengefasst. In der Analyse werden sowohl Metriken für das gesamte (J)CSP dargestellt, als auch ein Vergleich mit einer Kontrollgruppe für eine Untergruppe an habilitierten Fellows vorgenommen. Für die Analyse wurden Publikationen und weitere, alternative Veröffentlichungsformen der Programmteilnehmenden herangezogen. Auf dieser Grundlage wurden verschiedene Metriken aus unterschiedlichen Themenbereichen erfasst. Diese bibliometrische Analyse ist komplementär zum zweiten Teil der Evaluation, welcher von Hendriks et al. (2021) durchgeführt wurde und welcher Interviews und eine Umfrage unter den CSP Fellows sowie anderen wichtigen Akteuren beinhaltet.

## Übersichtsliste Indikatoren:

### Resonanz

Unter diesem Punkt sind Metriken aufgeführt, welche die Resonanz in Form von Zitationen der Publikationen messen. Es werden Metriken verwendet, die einen möglichst fairen Vergleich bemühen, indem sie die Zitationsraten mit denen ähnlich alter Artikel aus dem gleichen Fachbereich vergleichen.

- Relative Citation Ration (RCR): Zitationsrate eines Artikels im Vergleich zu anderen Artikeln aus dem gleichen Feld mit ähnlichem Alter
- NIH Perzentil: Das NIH Perzentil vergleicht den RCR einer Publikation mit denen aller NIH-geförderten Publikationen.

### Setting: Human/animal/molecular & cellular

Diese von der iCite-Datenbank des NIH zur Verfügung gestellte Metrik gibt für jede biomedizinische Publikation in Pubmed an, in welche der drei Kategorien Human/animal/molecular & cellular sich diese einordnen lässt. Dieses gibt einen Einblick darin, wo im translationalen Prozess sich die Forschungsarbeiten verorten lassen.

### Forschungsfelder

Die Forschungsfelder der Publikationen geben einen Überblick darüber, wie stark die unterschiedlichen biomedizinischen Themenbereiche durch die Publikationen des CSP abgedeckt werden.

### Teamscience

Unter dem Punkt Teamscience sind die Zusammenarbeit mit verschiedenen Arten von Kooperationspartnern aufgeführt. Neben den in den CSP-Jahresberichten erwähnten Kooperationen, werden die Kooperationen durch die Koautorenschaften in den Publikationen erfasst. Folgende Arten von Kooperationen werden erfasst:

- Kooperation mit Industriepartnern
- nationale Kooperationen
- internationale Kooperationen

### Transparenz & Open Science

Hier werden verschiedene Metriken aus dem Bereich Open Science erfasst. Hierbei geht es darum, wie transparent und offen Forschungsergebnisse geteilt werden.

- Open Access: Publikationen sind frei zugänglich für jedermann
- Open Data: zur Publikation gehörige (Roh-)Daten werden geteilt
- Open Code: zur Publikation gehöriger Analyse- oder Softwarecode wird geteilt

### Alternativer Forschungsoutput

Unter diesem Punkt werden verschiedene alternative Publikationsformate erfasst, welche durch den Fokus auf Journalpublikationen häufig nicht berücksichtigt werden. Folgende Formate werden berücksichtigt:

- Preprints
- Proceedings
- Buchkapitel
- Patente
- Konferenzposter
- Konferenzpräsentationen

Der weitere Report gliedert sich wie folgt: Zunächst wird unter Punkt 2 die verwendete Methodik detailliert beschrieben. Anschließend werden unter Punkt 3 die Ergebnisse der Analyse für das gesamte (J)CSP dargestellt während unter Punkt 4 die Ergebnisse des Vergleichs mit der Kontrollgruppe dargestellt werden.

## 2. Methodik

Als Grundlage für die bibliometrische Analyse dienen die Publikationen der Fellows des (J)CSP sowie der Vergleichskohorte von habilitierten Charité-Wissenschaftler\*innen die nicht Teil des Programms sind. Hauptquelle für die Publikationssuche und die Metriken ist die kommerzielle Datenbank Dimensions ([www.dimensions.ai](http://www.dimensions.ai)), welche sowohl Daten über wissenschaftliche Publikationen als auch unter anderem über Patente, Fördergelder, Policy-Dokumente enthält. Zusätzlich zur grafischen Oberfläche besitzt Dimensions eine Programmierschnittstelle (API), über die automatisiert Suchanfragen gestellt werden können. Dies ermöglicht, einen Großteil der Publikationsanalyse automatisiert über Skripte zu realisieren, welche eine einfache Änderung von Details der Analyse ermöglichen, sowie eine einfache Ausweitung der Analyse auf weitere Kohorten ermöglicht.

### 2.1. Definition Jahrgangskohorten

Für die bibliometrische Analyse werden verschiedene Personenkohorten betrachtet. Zunächst werden alle Personen betrachtet, die Teil des (J)CSP sind oder waren. Diese Kohorten werden jeweils getrennt nach Programm (CSP/JCSP) und Jahr des Eintritts in das Programm untersucht. Außerdem wird eine Vergleichskohorte von habilitierten Charité-Wissenschaftler\*innen außerhalb des CSP verwendet. Diese werden mit dem Subset der ebenfalls habilitierten (J)CSP-Fellows verglichen.

Um die Publikationen der Fellows zu identifizieren, wird als Grundlage eine Tabelle mit Namen der (J)CSP-Fellows sowie dem Start- und Enddatum der Zugehörigkeit zum (J)CSP genommen. Für die Vergleichskohorte wird analog eine Tabelle mit Name sowie dem Datum der Eröffnung des Habilitationsverfahrens genommen. Für jede Person wurde dann eine manuelle Suche nach der zugehörigen Dimensions researcher ID durchgeführt. Die Dimensions researcher ID ist eine von Dimensions automatisch jedem Forschenden zugeordnete ID, welche die eindeutige Zuordnung von Publikationen vereinfacht. Ein Vorteil gegenüber der ORCID ID ist, dass die Dimensions researcher ID automatisch zugeordnet wird und damit für alle Fellows verfügbar ist, ohne dass diese sich für eine ID registrieren mussten (Ausnahme: Forschende mit nur wenigen, ca. 1-3, Publikationen haben noch nicht in allen Fällen eine Dimensions ID zugewiesen bekommen; dies betrifft einige JCSP-Fellows).

Um die Dimensions researcher IDs zu identifizieren, wurde für jede Person Name und Nachname in das ‚researcher‘ Suchfeld in der grafischen Oberfläche von Dimensions ([app.dimensions.ai](http://app.dimensions.ai)) eingegeben. Wenn eine Person mit passendem Namen und Charité-Affiliation gefunden wurde, wurde die researcher ID aufgezeichnet. Wenn mehrere passende Suchergebnisse für den Namen angezeigt wurden, wurden diese auf Plausibilität geprüft (Affiliation, Publikationsjahre), und wenn mehrere passende IDs gefunden wurden, wurden alle aufgezeichnet. Dies kann in seltenen Fällen vorkommen, wenn eine Person die Affiliation gewechselt hat und die selbe Person unter zwei verschiedenen IDs geführt wird. Bei den Personen der Kontrollgruppe kann es vorkommen, dass diese aktuell nicht mehr an der Charité arbeiten und daher eine andere Affiliation angezeigt wird. In diesen Fällen werden die gefundenen IDs auf eine frühere Charité-Zugehörigkeit untersucht (durch Suche nach früheren Publikationen mit Charité-Affiliation) um die passende ID zu finden. Wenn keine passende researcher ID gefunden wurde, wurde keine ID für diesen Fellow aufgezeichnet.

## 2.2. Publikationssuche

Für die Identifikation der Publikationen die im Rahmen des (J)CSP veröffentlicht wurden, wird pro Fellow eine Publikationssuche durchgeführt. Hierfür wird eine Suche mittels der im vorigen Schritt etablierten Dimensions researcher IDs durchgeführt. Als Zeitraum der Suche wird als untere Begrenzung der Beginn der Zugehörigkeit zum (J)CSP verwendet, während als obere Grenze 3 Jahre nach Ende der Programmzugehörigkeit gewählt werden. Wenn die obere Grenze zum Zeitpunkt der Abfrage noch in der Zukunft liegt, wird stattdessen das Datum der Abfrage als obere Grenze gewählt.

Für die Dimensions-Suche wird für jeden (J)CSP-Fellow eine Suchanfrage aus den folgenden Filtern gebildet:

- Dimensions researcher ID
- Artikeltyp: „article“
- Veröffentlichungsdatum zwischen Start- und Enddatum der CSP-Zugehörigkeit + 3 Jahre
- Journalliste: Pubmed oder Norwegisches Register Level 1 oder 2 (dies sind kuratierte Listen, die nur Journale enthalten die gewisse Qualitätsmindeststandards erfüllen)

Zusätzlich wird für alle detektierten Publikationen geprüft ob der CSP-Fellow mit der gesuchten researcher ID die Charité in seiner Affiliation angegeben hat (und nicht eine frühere, andere Affiliation). Publikationen ohne Charité-Affiliation werden ausgeschlossen. Außerdem wird für jede Publikation aufgezeichnet, ob der gesuchte CSPler Erst- oder Letztautor auf der Publikation war. Da Koautorenschaften (Zweitautor oder vorletzter Autor, der in Publikationen typischerweise mit “equally contributing” angegeben ist) nicht in Dimensions oder anderen Publikationsdatenbanken erfasst sind, können diese nicht berücksichtigt werden. Die Suchanfrage wird für alle Wissenschaftler\*innen wiederholt und die Ergebnisse werden in einer Tabelle zusammengeführt. Für jede Publikation werden verschiedene Metadaten wie DOI, Titel, Journal, Autoren sowie einige weitere Metadaten gespeichert.

Für den Vergleich zwischen den habilitierten (J)CSP-Fellows und der Vergleichskohorte wird als Suchzeitraum ein Zeitraum von drei Jahren vor dem Datum der Eröffnung des Habilitationsverfahrens als untere Grenze sowie drei Jahren nach diesem Datum als obere Grenze verwendet um eine möglichst gute Vergleichbarkeit der Ergebnisse sicherzustellen. Außerdem werden für diesen Vergleich keine Publikationen ohne Charité-Affiliation ausgeschlossen, da die Personen aus der Vergleichskohorte teilweise bereits nicht mehr an der Charité arbeiten und daher mit anderer Affiliation publizieren.

Die finale Publikationssuche sowie die automatisierte Erfassung der Daten für die weiteren Metriken wurde am 9.2.2021 durchgeführt, so dass die Ergebnisse dieses Reports den Datenstand zu diesem Datum widerspiegeln.

## 2.3. Erfassen der Metriken

Ausgehend von der erhaltenen Publikationsliste als Grundlage (oder separat davon für manche Metriken) werden nun verschiedene Metriken erhoben.

### 2.3.1. Open Access

Um den Open Access Status der Publikationen zu bestimmen, greifen wir auf die von Dimensions angegebenen Open Access Informationen aus dem Feld ‚open\_access\_categories‘ zurück. Hierbei verwenden wir die folgenden detaillierteren Open Access Kategorien die Dimensions angibt: Es wird zwischen den Open Access-Kategorien Gold (reines OA Journal), Hybrid (Journal mit OA Option), Grün (Repositoriumsversion) und Bronze (Artikel frei verfügbar gemacht durch Journal, aber keine ausgewiesene OA-Lizenz) unterschieden.

### 2.3.2. Relative Citation Ratio

Der Relative citation ratio (Hutchins 2016) wird auch direkt von Dimensions zur Verfügung gestellt und ist für jede Publikation unter dem Feld ‚relative\_citation\_ratio‘ angegeben. Es ist zu beachten, dass nicht alle Publikationen bereits einen Wert für diese Metrik haben. Da dies eine zitationsbasierte Metrik ist, muss ausreichend Zeit vergangen sein, in dem die Arbeit zitiert werden konnte. Daher werden diese Parameter frühestens 2 Jahre nach Erscheinen der Publikation berechnet.

### 2.3.3. Fields of Research

Die Forschungsfelder die eine Publikation abdeckt werden in Dimensions im Feld ‚category\_for‘ zur Verfügung gestellt. Dimensions nimmt die Zuordnung der Forschungsfelder auf Artikellevel und nicht auf Journallevel vor. Das heißt, dass verschiedene Publikationen die im gleichen Journal erscheinen in verschiedene Forschungsfelder eingeordnet werden können. Dimensions gibt zwei Hierarchieebenen für die Fields of Research an, eine höhere Hierarchieebene die mit zwei Zahlen indiziert ist, sowie eine feinere Hierarchieebene, die mit vier Zahlen indiziert ist. Eine Publikation kann sowohl mehreren höheren Hierarchieebenen sowie auch mehreren feineren Hierarchieebenen der gleichen Oberkategorie zugeordnet sein. Für die Auswertung der Forschungsfelder des CSP werden für jede Publikation alle genannten Kategorien beider Hierarchieebenen aufgezeichnet.

### 2.3.4. Alternative Veröffentlichungsformen

Für die Suche nach anderen Veröffentlichungsformen, die nicht in klassischen wissenschaftlichen Journalen erscheinen (Preprints, Conference Proceedings, Buchkapitel), wird die oben beschriebene Publikationssuche mit Dimensions wiederholt, mit der einzigen Änderung, dass jetzt alle Artikeltypen außer dem Typ „article“ gesucht werden.

### 2.3.5. Industriekollaborationen

Um Publikationen zu identifizieren die in Kooperation mit der Industrie entstanden sind, wird eine zweistufige Suche auf Dimensions durchgeführt. Hierfür werden zunächst Organisations-Affiliationen identifiziert, die laut Dimensions als Industrie-Organisation eingestuft werden. Im zweiten Schritt wird für alle CSP-Publikationen geprüft, ob eine dieser Industrie-Organisationen in einer der Affiliationen vorkommt. Zur Identifizierung der Industrie-Organisationen wird der Dimensions Organisations-Datensatz verwendet. In Dimensions werden Organisationen analog zu den Wissenschaftler\*innen eine eigene ID zugewiesen, die dann in den Affiliationen der Publikationen angegeben werden. Zunächst werden unter allen in Dimensions erfassten Organisationen diejenigen herausgefiltert, welche den Typ ‚Company‘ zugeordnet bekommen haben. Zum Zeitpunkt der Abfrage wurden 27940 Industrie-Organisationen in Dimensions geführt. Im zweiten Schritt wird diese gefilterte Liste der Organisationen verwendet um aus allen bereits identifizierten CSP-Publikationen diejenigen herauszusuchen, bei denen in wenigstens einer der Affiliationen die ID einer der Industrie-Organisationen genannt wird. Jede dieser identifizierten Publikationen wird als Publikation mit Industriekollaboration gezählt.

### 2.3.6. Nationale & Internationale Kollaborationen

Um zu identifizieren welche Publikationen zusammen mit nationalen oder internationalen Partnern entstanden sind, werden alle in den gefundenen CSP-Publikationen angegebenen Affiliationen durchsucht. Jedem der Autor\*innen einer Publikation wird eine eigene Affiliation zugeordnet. Wenn die in der Affiliation angegebene Institution automatisch erkannt wurde, wird der Affiliation automatisch eine Organisations-ID zugeordnet und weitere Affiliations-Metadaten-Felder wie z.B. ‚Country‘ gefüllt. Wenn die Institution nicht erkannt wurde, wird nur die gesamte Affiliation angegeben, so wie sie in der Publikation angegeben ist. Zunächst werden alle Affiliationen identifiziert, welche der Charité oder dem Berlin Institute of Health zugeordnet werden können. Diese werden für die weiteren Suchschritte nicht weiter berücksichtigt. Wenn es verbleibende Affiliationen gibt, wird geprüft ob die genannten Institutionen aus Deutschland kommen. Hierfür wird das Affiliations-Feld ‚Country‘ geprüft, welches nur gefüllt ist, wenn der Affiliation eine GRID ID zugeordnet ist. Wenn das Feld leer ist wird stattdessen die gesamte angegebene Affiliation für diesen Autor nach der Nennung von ‚Germany‘ oder ‚Deutschland‘ durchsucht. Wenn es eine Nennung von Deutschland in mindestens einer der verbleibenden Affiliationen gibt, wird die Publikation als Publikation mit nationaler Kollaboration gezählt. Wenn es verbleibende Affiliationen gibt, die weder die Charité noch eine deutsche Institution nennen, wird die Publikation als Publikation mit internationaler Kollaboration gezählt. Die genaue Anzahl verschiedener Kollaborationspartner wird nicht angegeben, da für eine größere Anzahl von Affiliationen die Organisations-IDs fehlen und damit eine eindeutige und automatisierte Zuordnung und Unterscheidung der verschiedenen Institutionen die an einer Publikation beteiligt waren nur schwer möglich machen.

### 2.3.7. Kollaborationen mit anderen (J)CSP-Fellows

Um Publikationen zu identifizieren die in Kooperation mit anderen (J)CSP-Fellows entstanden sind, werden alle in den gefundenen Publikationen angegebenen Affiliationen nach researcher IDs anderer (J)CSP-Fellows durchsucht. Dafür werden für jede Publikation alle researcher IDs der anderen Autoren mit der Liste der researcher IDs aller (J)CSP-Fellows abgeglichen. Wenn wenigstens einer der anderen Autoren ebenfalls Teils des (J)CSP ist, wird diese Publikation als Kollaboration mit anderen (J)CSP-Fellows gezählt.

### 2.3.8. iCite Metriken

Zusätzlich zu den von Dimensions zur Verfügung gestellten Metriken werden verschiedene Metriken verwendet, die von der iCite-Datenbank zur Verfügung gestellt werden. iCite wird vom amerikanischen National Institutes of Health zur Verfügung gestellt und bietet verschiedene bibliometrische Metriken in den Kategorien „Influence“, „Translation“ und „Citations“ an. Die Metriken können automatisiert über eine Programmierschnittstelle (<https://icite.od.nih.gov/api/pubs>) abgerufen werden. Wir nutzen diese Schnittstelle um folgende Metriken abzufragen:

- NIH-Percentile (Influence): Das NIH Perzentil vergleicht den RCR einer Publikation mit denen aller NIH-geförderten Publikationen. Ein NIH Perzentil von 95% bedeutet, dass der RCR dieser Publikation höher ist als der RCR von 95% aller NIH-geförderten Publikationen.
- Kategorisierung nach Human/Animal/Molecular & Cellular Biology (Translation): Zur Kategorisierung der Publikationen nutzt iCite die sogenannten Medical Subject Heading (MeSH)-Begriffe, welche die inhaltliche Ausrichtung einer Publikation kategorisieren und die jeder Publikation in PubMed basierend auf ihrem Inhalt zugeteilt werden. Mittels der MeSH-Begriffe wird jede Publikation in eine der drei Kategorien Human/Animal/Molecular & Cellular Biology bzw. einer Kombination aus zwei Kategorien eingeordnet. Dadurch kann für eine Publikationsliste eine mittlere Orientierung bezüglich dieser drei Kategorien berechnet werden. Diese Orientierung kann in einem Dreieck dargestellt werden, welches die drei Kategorien den drei Ecken zuordnet. Die Orientierung ist für jede der drei Kategorien durch eine Zahl zwischen 0 und 1 angegeben, wobei die Summe der Orientierungen immer 1 ergibt. Für weitere Details zu diesem Ansatz, siehe Weber 2013.

Für die Abfrage der iCite-Metriken werden für alle identifizierten CSP-Publikationen die PubMed-IDs an die iCite-Schnittstelle gesendet um die Metriken abzufragen. Da iCite nur Daten über Publikationen mit PubMed-ID führt, kann die Abfrage nur für Publikationen durchgeführt werden, welche auch eine PubMed-ID besitzen. Da für die Publikationssuche in Dimensions jedoch die PubMed-Listung als zusätzliches Suchkriterium verwendet wurde, trifft dies auf alle identifizierten CSP-Publikationen zu.

### 2.3.9. Open Data & Open Code

Um Publikationen zu identifizieren die Rohdaten (Open Data) oder den Analysecode (Open Code) zur Verfügung stellen, verwenden wir das vom QUEST-Center entwickelte text-mining Tool ODDPub (Riedel et al. 2020; siehe auch <https://github.com/quest-bih/oddpub>). Dieses Tool durchsucht die Volltexte von Publikationen nach Sätzen die auf das Teilen von Rohdaten oder von Analysecode hinweisen. Um das Tool anwenden zu können werden zunächst die Volltexte aller identifizierten CSP-Publikationen als PDF-Version automatisiert heruntergeladen. Es können nur Volltexte heruntergeladen werden zu denen die Charité im Rahmen ihrer Lizenzvereinbarung mit den Verlagen Zugang hat. Auf die erhaltenen Publikations-Volltexte wird ODDPub angewandt, welches zurückgibt in welchen Publikationen Open Data oder Open Code identifiziert wurden.

## 2.4. Datenquelle CSP-Berichte

Zusätzlich zu den Metriken die über externe Quellen erfasst werden, werden einige Angaben aus den jährlichen Berichten der (J)CSP-Fellows verwendet. Im Gegensatz zu den über externe Quellen erfassten Metriken werden diese Angaben von den Fellows selbst gemacht.

- Anzahl Patente
- Konferenzen: Vortragspräsentationen

- Konferenzen: Posterpräsentationen
- Anzahl nationale Kollaborationen
- Anzahl internationale Kollaborationen

Da diese Berichte nur für das (J)CSP vorhanden sind, können die darin enthaltenen Kennzahlen nicht im Vergleich mit der Kontrollgruppe verwendet werden.

### 3. Ergebnisse - CSP + JCSP

In diesem Abschnitt werden die Ergebnisse für das gesamte CSP + JCSP dargestellt. Die Ergebnisse für die einzelnen Metriken werden getrennt für die einzelnen Jahreshkohorten dargestellt, d.h. alle im Beobachtungszeitraum gefundenen Publikationen (oder andere Daten) aller Fellows einer Kohorte werden zusammengefasst. Wo es sinnvoll ist, werden neben dem Mittelwert über alle Fellows der Kohorte auch die Datenpunkte pro Fellow gezeigt. Die Daten pro Fellow werden durch Mittelung über alle detektieren Publikationen des Fellows berechnet. Hierbei werden alle Erst-/Letztautorenpublikationen berücksichtigt.

#### 3.1. Produktivität/Übersicht

Zunächst wird eine Übersicht über die Anzahl der (J)CSP-Fellows pro Jahrgangskohorte sowie die Anzahl der gefundenen Publikationen gegeben.

Es werden zunächst alle Jahrgangskohorten berücksichtigt, auch die Kohorten bei denen das Programm aktuell noch läuft. Hier gibt es die Einschränkung, dass der Beobachtungszeitraum noch nicht abgeschlossen ist und die Ergebnisse unvollständig und nur vorläufig sind. Um trotzdem eine möglichst gute Vergleichbarkeit mit den anderen Kohorten darzustellen, werden die einzelnen Metriken relativ zum Beobachtungszeitraum oder relativ zur Gesamtzahl der gefundenen Publikationen normiert.

Tabelle 1: Anzahl CSP-Fellows und gefundene Publikationen die für die weitere Analyse berücksichtigt werden.

Jahreskohorte	Anzahl Fellows	Anzahl Publikationen	Publikationen pro Jahr pro Fellow
2013	9	87	1.5
2014	17	128	1.2
2015	20	128	1.1
2016	17	112	1.4
2017	14	55	1.0
2018	23	83	1.3
2019	10	31	1.5

Tabelle 2: Anzahl JCSP-Fellows und gefundene Publikationen die für die weitere Analyse berücksichtigt werden.

Jahreskohorte	Anzahl Fellows	Anzahl Publikationen	Publikationen pro Jahr pro Fellow
2015	18	99	1.0
2016	20	109	1.1
2017	14	77	1.4
2018	9	23	0.9
2019	5	5	0.5

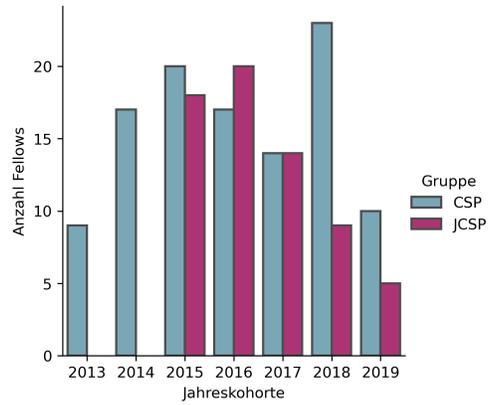


Abbildung 1: Anzahl (J)CSP-Fellows pro Jahrgangskohorte.

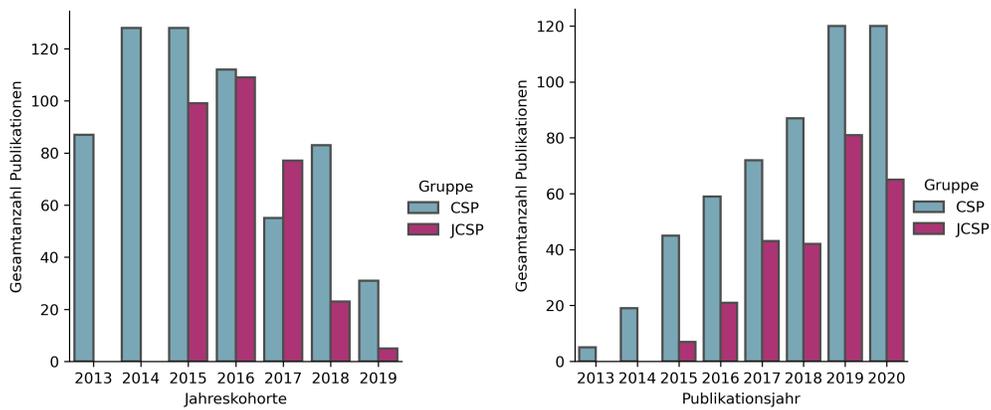


Abbildung 2: Anzahl der gefundenen Publikationen. Links: Aufgeschlüsselt nach Jahrgangskohorten. Rechts: Aufgeschlüsselt nach Publikationsjahr.

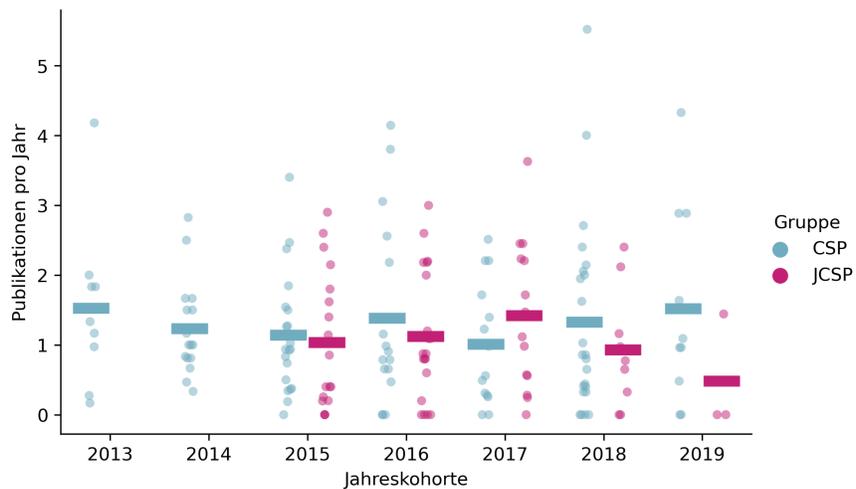


Abbildung 3: Publikationen pro Jahr und pro Fellow für das CSP- und JCSP.

### 3.2. Resonanz

Die Metrik des Relative Citation Ratio (RCR) misst wie häufig die Publikationen relativ zu ihrem Feld zitiert werden. Ein Wert über 1 bedeutet eine überdurchschnittlich häufige Zitation.

Allerdings ist der RCR erst zwei Jahre nach Erscheinen einer Publikation verfügbar. Daher fehlen für 35% (RCR) aller Publikationen im Datensatz noch die Werte für den RCR.

Das NIH Perzentil vergleicht den RCR einer Publikation mit denen aller NIH-geförderten Publikationen. Ein NIH Perzentil von 95% bedeutet, dass der RCR dieser Publikation höher ist als der RCR von 95% aller NIH-geförderten Publikationen. Da das NIH Perzentil an der RCR gekoppelt ist, gelten hier die gleichen Einschränkungen wie für nicht verfügbare Werte für den RCR.

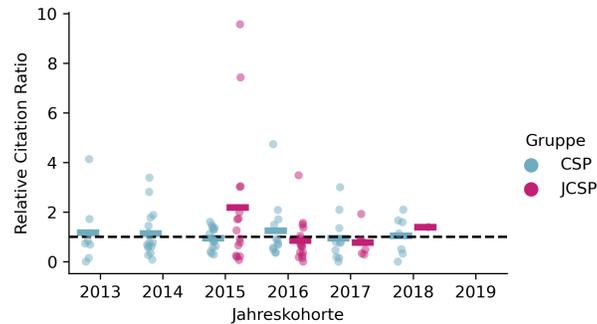


Abbildung 4: Mittlerer RCR pro Fellow für das CSP- und JCSP. Der RCR ist so normiert, dass Publikationen im Mittel einen Wert von 1 erreichen. Ein Wert über 1 bedeutet eine überdurchschnittlich häufige Zitation.

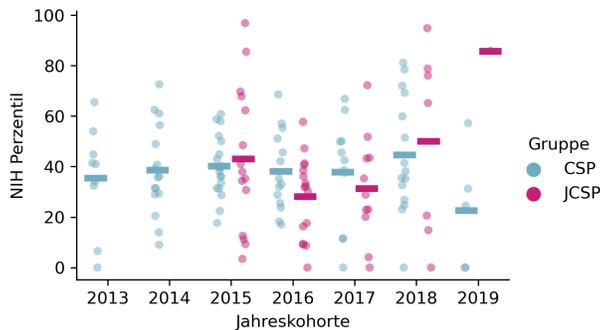


Abbildung 5: Mittleres NIH Perzentil pro Fellow für das CSP- und JCSP.

### 3.3 Setting: klinisch/animal/molecular & cellular

Über die Plattform iCite des NIH werden alle Publikationen in PubMed anhand der inhaltlichen Kategorisierung durch die MeSH-Terms in die Kategorien klinisch/animal/molecular and cellular eingeteilt. Die mittlere Ausrichtung wird pro Fellow sowie gemittelt für das CSP/JCSP gezeigt.

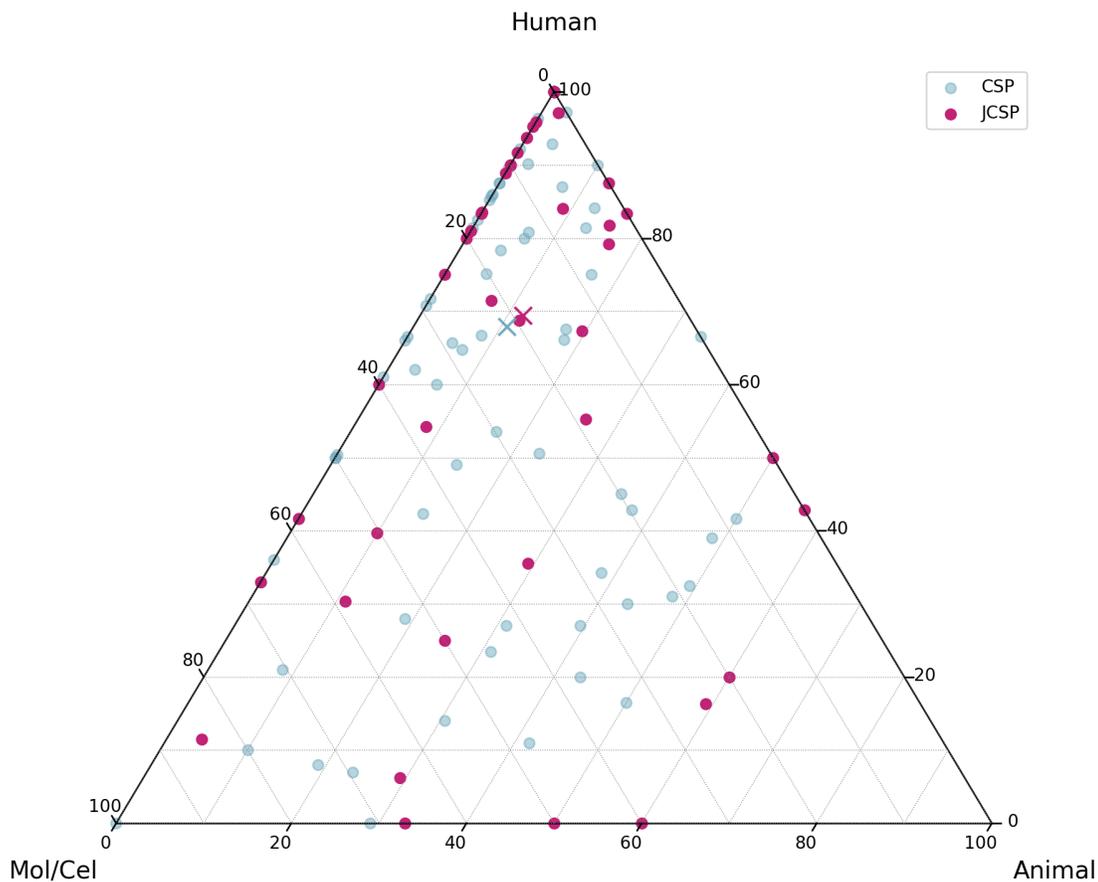


Abbildung 6: Ausrichtung der Publikationen entlang der Achsen klinisch/animal/molecular and cellular, gemittelt über alle Publikationen pro CSP-Fellow. Die Kreuze geben den Mittelwert über alle Personen einer Gruppe an.

### 3.4 Forschungsfelder

Die Forschungsfelder werden auf Dimensions für jede Publikation angegeben. Der relative Anteil der Forschungsfelder pro Kohorte wird angegeben. Hierfür werden alle Publikationen die für die jeweilige Kohorte gefunden worden berücksichtigt. Es wurde nicht über die einzelnen Fellows gemittelt. Eine Publikation kann mehreren Forschungsfeldern zugewiesen sein. Der Übersichtlichkeit halber wurden in den Abbildungen nur Fields of Research abgebildet, bei denen der relative Anteil bei wenigstens einer Gruppe über 2% lag. In den Abbildungen sind die Forschungsfelder jeweils nach der Häufigkeit in der letzten dargestellten Jahreskohorte sortiert.

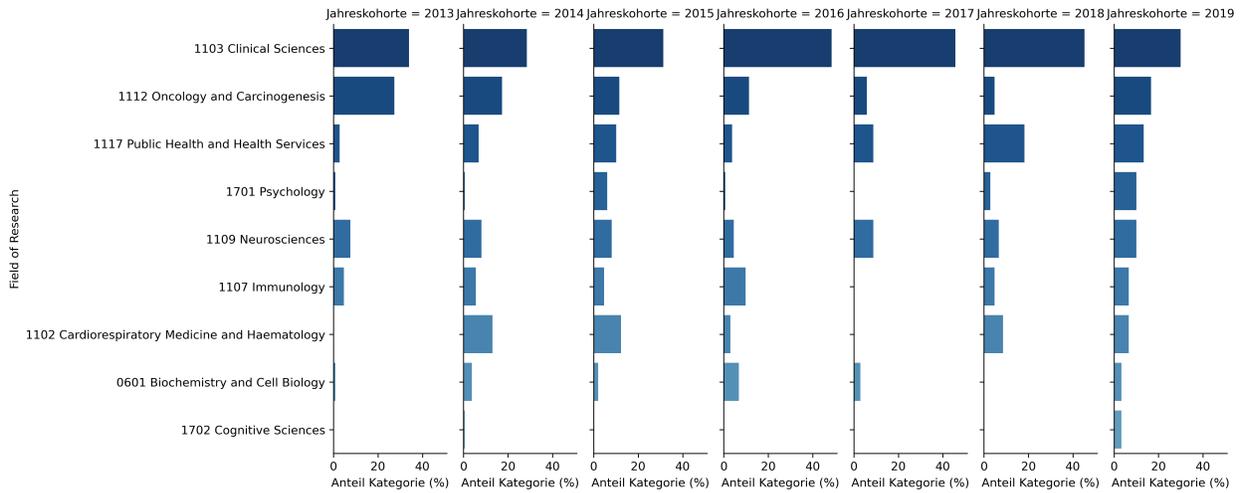


Abbildung 7: Anteil der Forschungsfelder für die Publikationen der CSP-Fellows.

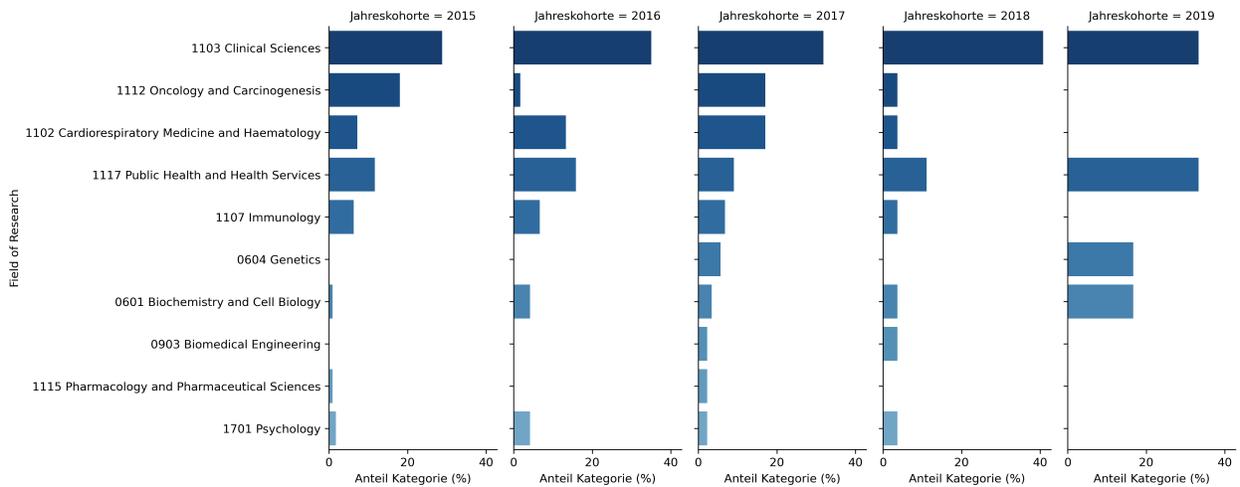


Abbildung 8: Anteil der Forschungsfelder für die Publikationen der JCSP-Fellows.

### 3.5 Teamscience

In dieser Kategorie werden Kollaborationen in verschiedenen Bereichen erfasst. Es werden nationale, internationale und Industriekollaborationen sowie Kollaborationen mit anderen CSP-Fellows berücksichtigt. Die nationalen und internationalen Kollaborationen wurden sowohl durch die Angaben in den CSP-Jahresberichten als auch durch die Affiliationen von Koautoren auf Publikationen erfasst.

#### Berichtete Kollaborationspartner aus Jahresberichten

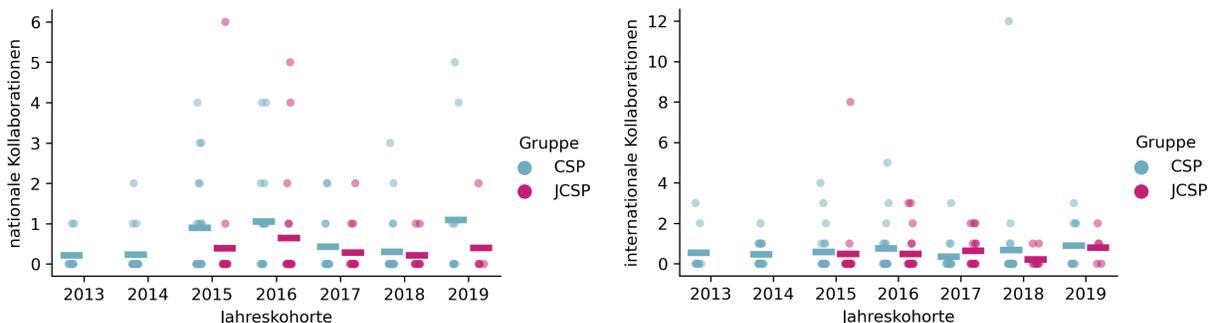


Abbildung 9: Anzahl der in den Jahresberichten angegebenen nationalen/internationalen Kollaborationen. Es wird die Gesamtzahl der angegebenen Kollaborationspartner pro Fellow über die gesamte Förderperiode hinweg angegeben. Wird der gleiche Kollaborationspartner in verschiedenen Reports eines Fellows mehrfach angegeben, wird dieser nicht doppelt gezählt.

#### Kollaborationspartner aus Publikationen

Tabelle 3: Für das CSP: Anzahl der Publikationen die in Kooperation mit Industriepartnern, anderen CSP-Fellows oder mit nationalen oder internationalen Kollaborationspartnern entstanden sind.

Jahreskohorte	Anzahl Publikationen	mit Industriepartner	mit nationalen Kollaborationen	mit internationalen Kollaborationen	mit anderen (J)CSP-Fellows
2013	87	3	46	32	8
2014	128	7	58	35	14
2015	128	2	68	34	21
2016	112	0	47	22	22
2017	55	2	29	12	3
2018	83	1	47	30	6
2019	31	2	14	7	4

Tabelle 4: Für das JCSP: Anzahl der Publikationen die in Kooperation mit Industriepartnern, anderen CSP-Fellows oder mit nationalen oder internationalen Kollaborationspartnern entstanden sind.

Jahreskohorte	Anzahl Publikationen	mit Industriepartner	mit nationalen Kollaborationen	mit internationalen Kollaborationen	mit anderen (J)CSP-Fellows
2015	99	2	50	26	7
2016	109	3	46	18	11
2017	77	3	35	32	14
2018	23	0	14	7	3
2019	5	0	2	0	0

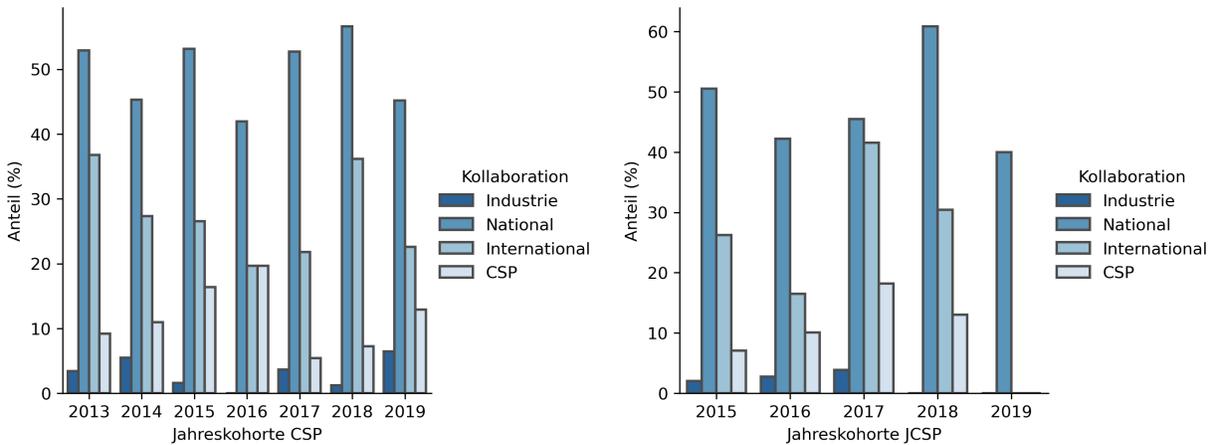


Abbildung 10: Anteil der Publikationen pro Jahreskohorte mit mindestens einem nationalen, internationalen oder Industriepartner bzw. einem anderen (J)CSP-Fellow, welcher in den Affiliationen genannt wird. Links: CSP. Rechts: JCSP.

### 3.6 Transparenz & Open Science

In dieser Kategorie werden verschiedene Open Science Metriken erfasst. Wir erfassen den Anteil sowie die Art der Open Access-Publikationen. Außerdem erfassen wir die Nennung von data & code sharing Statements in den Publikationsvolltexten mithilfe des von QUEST entwickelten text-mining Tools ODDPub (Tool: <https://github.com/quest-bih/oddpub>, Publikation: <http://doi.org/10.5334/dsj-2020-042>). Da es durch das automatisierte Screening zu falsch positiven Ergebnissen kommen kann, wurden die detektierten Fälle zusätzlich manuell überprüft. Nur die manuell verifizierten Fälle von Open Data und Open Code wurden hier dargestellt.

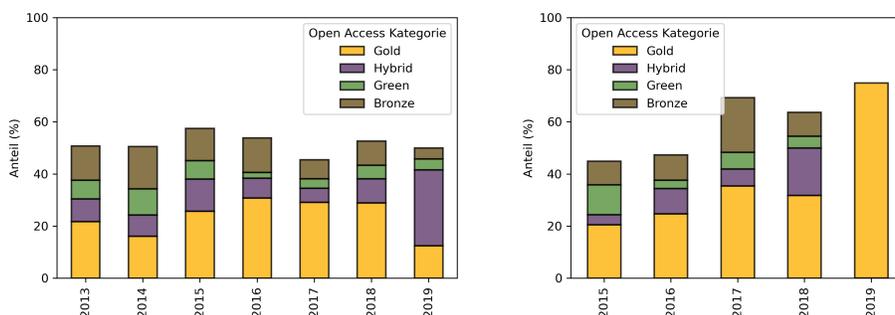


Abbildung 11: Anteil der Open Access-Publikationen pro Jahreskohorte. Es wird zwischen den Open Access-Kategorien Gold (reines OA Journal), Hybrid (Journal mit OA Option), Grün (Repositoriumsversion) und Bronze (Artikel frei verfügbar gemacht durch Journal, aber keine ausgewiesene OA-Lizenz) unterschieden. Links: CSP. Rechts: JCSP.

Tabelle 5: Für das CSP: Anzahl der Publikationen mit Open Data und Open Code.

Jahreskohorte	Anzahl Publikationen	mit Open Data	mit Open Code
2013	87	5	1
2014	128	4	2
2015	128	5	0
2016	112	7	5
2017	55	0	0
2018	83	2	2
2019	31	0	1

Tabelle 6: Für das JCSP: Anzahl der Publikationen mit Open Data und Open Code.

Jahreskohorte	Anzahl Publikationen	mit Open Data	mit Open Code
2015	99	2	0
2016	109	1	4
2017	77	4	3
2018	23	0	0
2019	5	0	0

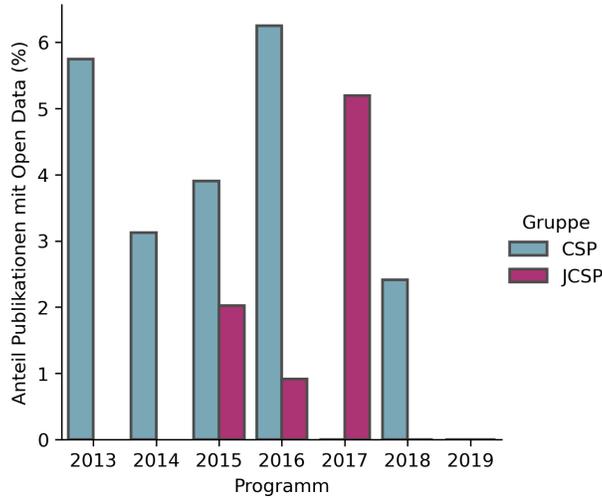


Abbildung 12: Anteil der Publikationen für die Open Data detektiert wurde. Da so wenige Fälle von Open Code gefunden wurden, wurde auf eine grafische Darstellung der Open Code-Anteile verzichtet.

### 3.7 Alternativer Forschungsoutput

Zusätzlich zu den klassischen Journalartikeln wurden Veröffentlichungen in anderen Formaten erfasst. Über Dimensions werden zusätzlich Preprints, Proceedings sowie Buchkapitel erfasst. Aus den Jahresberichten werden Patente und Konferenzposter und -präsentationen erfasst.

Tabelle 7: Für das CSP: Anzahl der Beiträge für die verschiedenen Kategorien des alternativen Forschungsoutputs.

Jahreskohorte	Anzahl Fellows	Preprints	Proceedings	Buchkapitel	Patente	Konferenzposter	Konferenzpräsentationen
2013	9	0	1	0	3	8	53
2014	17	2	8	7	1	34	48
2015	20	5	3	8	0	32	52
2016	17	0	7	10	5	41	90
2017	14	0	2	0	0	13	21
2018	23	1	2	1	0	18	59
2019	10	1	0	0	1	10	15

Tabelle 8: Für das JCSP: Anzahl der Beiträge für die verschiedenen Kategorien des alternativen Forschungsoutputs.

Jahreskohorte	Anzahl Fellows	Preprints	Proceedings	Buchkapitel	Patente	Konferenzposter	Konferenzpräsentationen
2015	18	5	2	1	0	10	24
2016	20	2	10	0	1	19	27
2017	14	4	8	0	0	17	26
2018	9	0	2	0	0	7	21
2019	5	0	0	0	1	2	2

## 4. Ergebnisse - Vergleich mit Kontrollgruppe

Um einen direkten Vergleich mit einer ähnlich zusammengesetzten Kontrollgruppe zu haben, werden alle bisher habilitierten (J)CSP-Fellows/Alumni (N rund 50) mit anderen habilitierten Wissenschaftler\*innen an der Charité (seit 2012; N rund 300) verglichen.

Da in dieser Untergruppe der habilitierten (J)CSP-Fellows/Alumni deutlich weniger Personen und Publikationen pro Jahrgangskohorte gab und es vor allem in den früheren Jahren sehr wenig Personen pro Jahr gab, ist eine Aufschlüsselung der Ergebnisse nach Jahrgangskohorte kaum möglich. Stattdessen werden die Ergebnisse für alle Jahrgangskohorten zusammengefasst. Ein Problem hierbei ist die ungleiche Verteilung der Kohorten auf die Jahre im Vergleich zwischen (J)CSP und Kontrollgruppe. Das kann eine Verzerrung des Vergleichs bedeuten für die Metriken die sich über die Jahre deutlich geändert haben, wie z.B. Open Data.

Es werden alle publikationsbasierten Metriken für die Erst- und Letztautorenpublikationen beider Gruppen dargestellt. Eine Berücksichtigung aller Publikationen, unabhängig von der Autorenposition führt bei allen Metriken zu sehr ähnlichen Ergebnissen (siehe Anhang).

### 4.1. Produktivität/Übersicht

Zunächst wird eine Übersicht über die Anzahl der berücksichtigten habilitierten Personen pro Jahrgangskohorte für CSP und Kontrollgruppe sowie die Anzahl der gefundenen Publikationen gegeben.

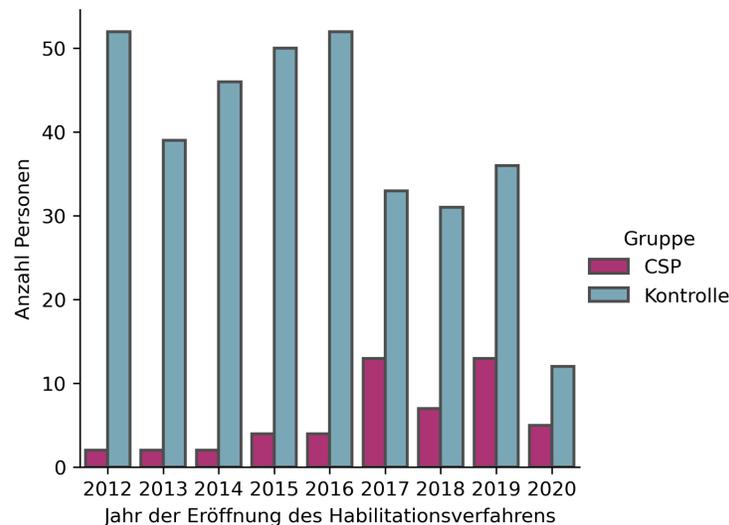


Abbildung 13: Anzahl der berücksichtigten habilitierten Personen pro Jahrgangskohorte für CSP und Kontrollgruppe. Die Personenanzahl der habilitierten (J)CSP-Fellows ist ungleich über die Jahre verteilt, mit der größten Anzahl in den letzten Jahren. Dies zu einer Verzerrung einzelner Metriken führen, welche sich über die Jahre entwickelt haben, wie z.B. Open Data. Besonders kritisch ist dieser Punkt, da der Vergleich mit der Kontrollgruppe aufgrund der kleinen Anzahl Fellows nicht in einzelne Jahrgänge aufgespalten werden kann.

Insgesamt sind die habilitierten Personen für beide untersuchten Gruppen sehr unterschiedlich über die Jahre verteilt. Dies wird vor allem daran liegen, dass es in den Anfangsjahren des CSP noch wenig habilitierte Personen aus dieser Gruppe gab, gerade wenn es um die jüngeren JCSPler geht, und erst über die Jahre mehr Habilitationen aus dieser Gruppe dazukamen. Für die gesamte Charité gibt es diesen Effekt jedoch nicht, so dass es hier eine gleichmäßigere Verteilung über die Jahre gibt. Bemerkenswert ist in diesem Zusammenhang, dass ab dem Jahr 2017 ein beträchtlicher Anteil der habilitierten Personen (ca. 20-30%) aus dem CSP stammten.

Die unterschiedliche Verteilung der beiden Personengruppen schlägt sich auch auf die Verteilung der Publika-

tionen in Abb. 14 nieder. Dies kann eine Auswirkung auf den Vergleich der beiden Gruppen bezüglich der Metriken haben. Insbesondere bei Metriken die sich über die Jahre verändert haben, wie z.Bsp. Open Data, kann es einen Vorteil für die CSP-Gruppe geben, da diese einen höheren Anteil an Publikationen aus den jüngeren Jahren haben. Daher sollte dies beim Vergleich der Gruppen mitberücksichtigt werden.

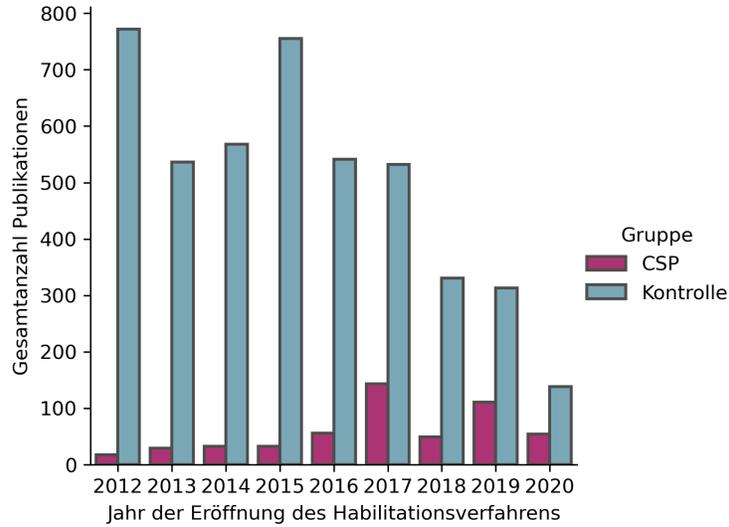


Abbildung 14: Anzahl der gefundenen Publikationen pro Jahrgangskohorte für CSP und Kontrollgruppe (nur Erst-/Letztautorenpublikationen).

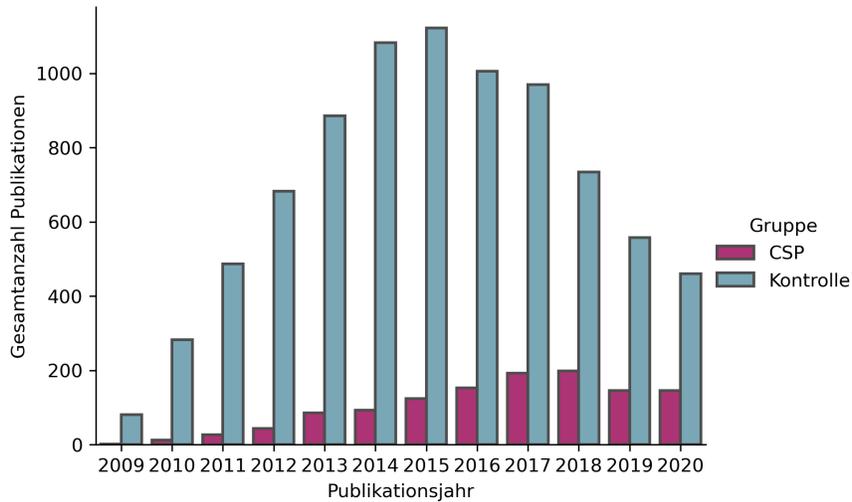


Abbildung 15: Anzahl der gefundenen Publikationen sortiert nach Publikationsjahr für CSP und Kontrollgruppe (nur Erst-/Letztautorenpublikationen).

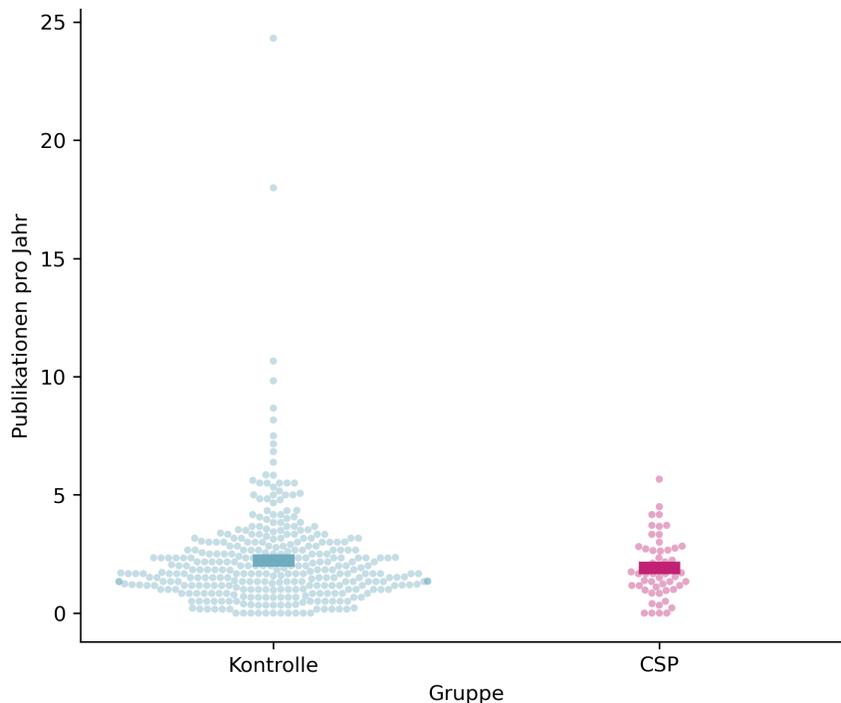


Abbildung 16: Erst-/Letztautorenpublikationen pro Jahr und pro Fellow für CSP und Kontrollgruppe.

## 4.2. Resonanz

Die Metrik des Relative Citation Ratio (RCR) misst wie häufig die Publikationen relativ zu ihrem Feld zitiert werden. Ein Wert über 1 bedeutet eine überdurchschnittlich häufige Zitation.

Allerdings ist der RCR erst zwei Jahre nach Erscheinen einer Publikation verfügbar. Daher fehlen für für die Kontrollgruppe für 9% aller Publikationen im Datensatz noch die Werte für den RCR. Für die habilitierten CSP-Fellows fehlen die Werte sogar für 16% aller Publikationen im Datensatz, da die Fellows vor allem in den letzten Jahren habilitiert haben.

Das NIH Perzentil vergleicht den RCR einer Publikation mit denen aller NIH-geförderten Publikationen. Ein NIH Perzentil von 95% bedeutet, dass der RCR dieser Publikation höher ist als der RCR von 95% aller NIH-geförderten Publikationen. Da das NIH Perzentil an der RCR gekoppelt ist, gelten hier die gleichen Einschränkungen wie für nicht verfügbare Werte für den RCR.

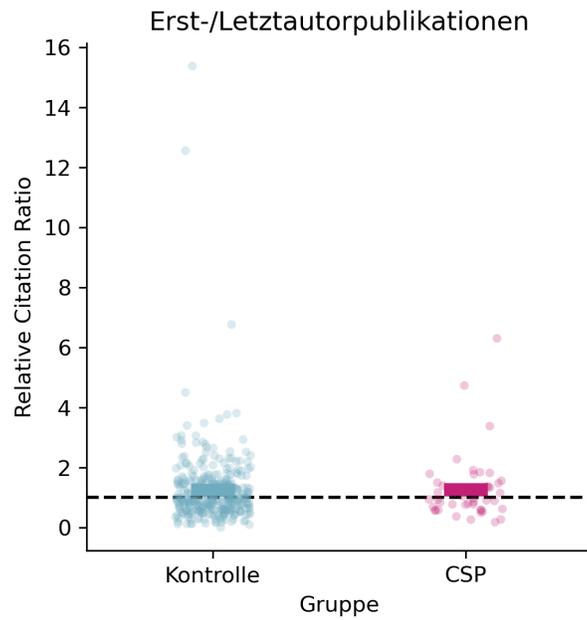


Abbildung 17: Mittlerer RCR pro Fellow im Vergleich zwischen CSP und Kontrollgruppe (nur Erst-/Letztautorenpublikationen). Der RCR ist so normiert, dass Publikationen im Mittel einen Wert von 1 erreichen. Ein Wert über 1 bedeutet eine überdurchschnittlich häufige Zitation.

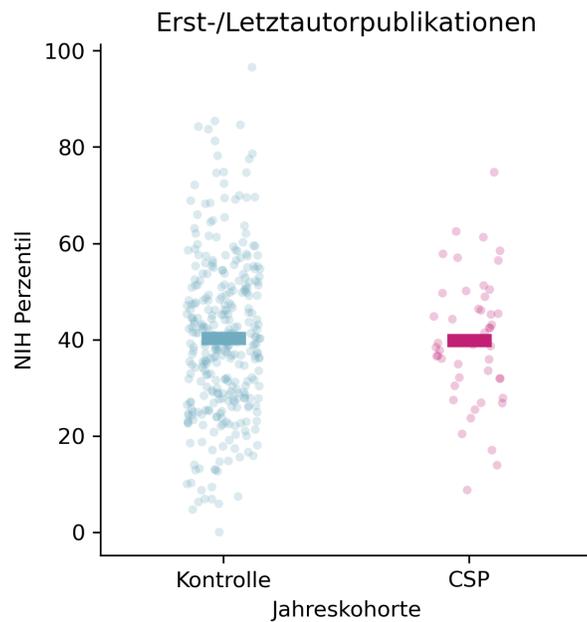


Abbildung 18: Mittleres NIH Percentil pro Fellow im Vergleich zwischen CSP und Kontrollgruppe (nur Erst-/Letztautorenpublikationen).

### 4.3 Setting: klinisch/animal/molecular & cellular

Über die Plattform iCite des NIH werden alle Publikationen in PubMed anhand der inhaltlichen Kategorisierung durch die MeSH-Terms in die Kategorien klinisch/animal/molecular and cellular eingeteilt. Die mittlere Ausrichtung pro Wissenschaftler wird für alle sowie für jeden Jahrgang separat gezeigt.

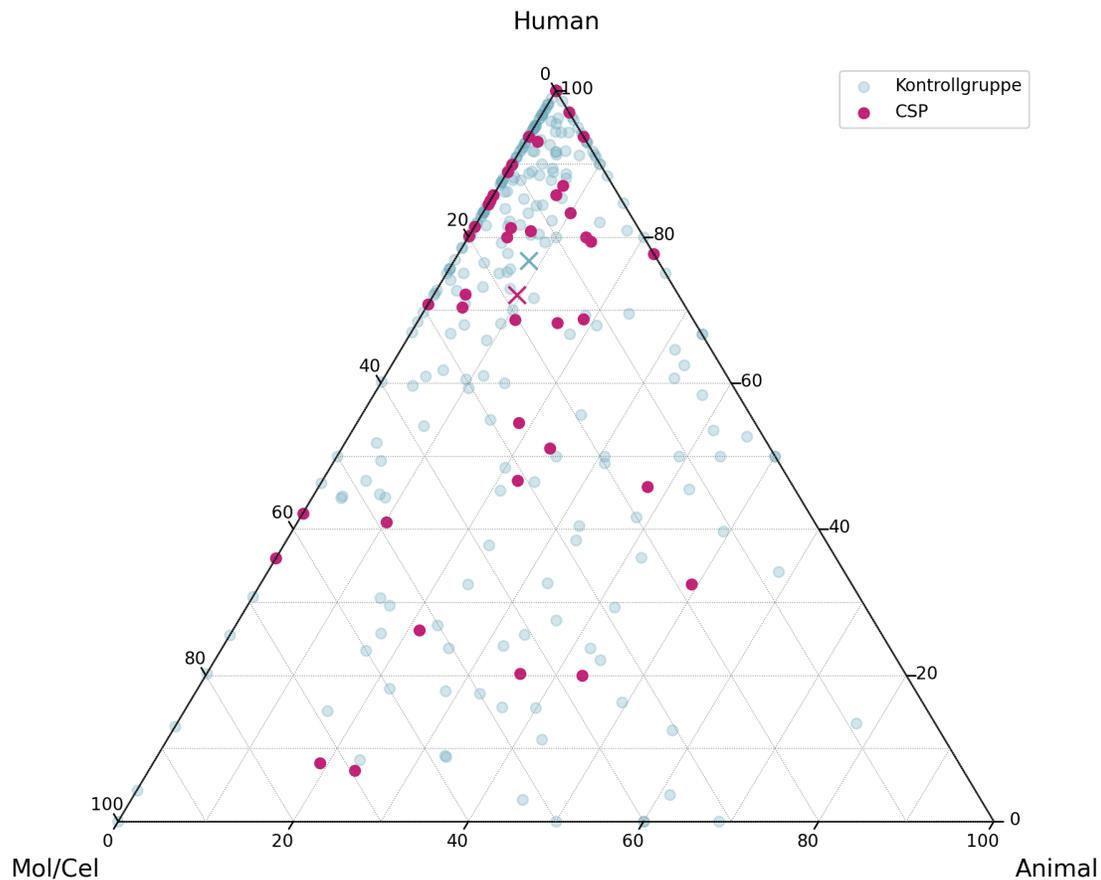


Abbildung 19: Ausrichtung der Publikationen entlang der Achsen klinisch/animal/molecular and cellular, gemittelt über alle Publikationen pro Person. Die Kreuze geben den Mittelwert über alle Personen einer Gruppe an. In diesem Diagramm wurden nur Erst-/Letztautorenpublikationen berücksichtigt.

## 4.4 Forschungsfelder

Die Forschungsfelder werden auf Dimensions für jede Publikation angegeben. Der relative Anteil der Forschungsfelder pro Kohorte wird angegeben. Hierfür werden alle Publikationen die für die jeweilige Kohorte gefunden worden berücksichtigt. Es wurde nicht über die einzelnen Fellows gemittelt. Eine Publikation kann mehreren Forschungsfeldern zugewiesen sein. Der Übersichtlichkeit halber wurden in den Abbildungen nur Fields of Research abgebildet, bei denen der relative Anteil bei wenigstens einer Gruppe über 1% lag.

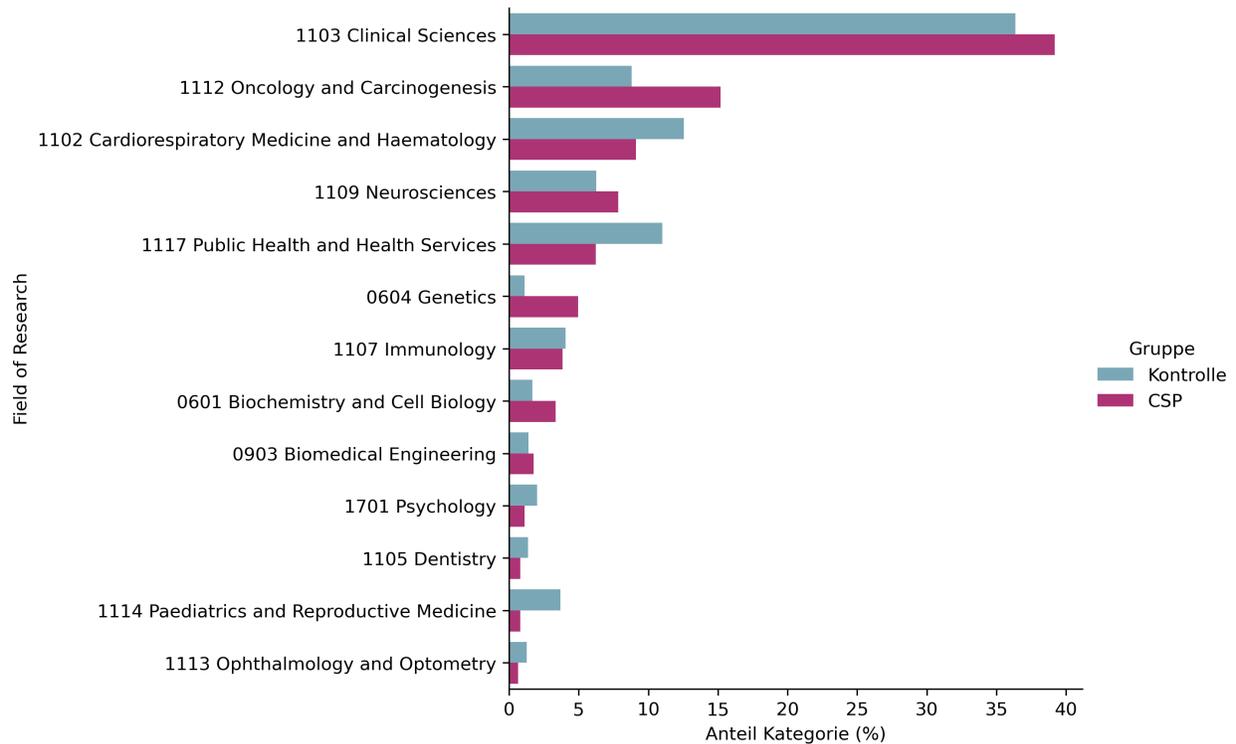


Abbildung 20: Vergleich der Anteile der Forschungsfelder zwischen CSP und Kontrollgruppe für die gefundenen Erst-/Letztautor-Publikationen.

## 4.5 Teamscience

In dieser Kategorie werden Kollaborationen in verschiedenen Bereichen erfasst. Es werden nationale, internationale sowie Industriekollaborationen berücksichtigt. Anders als im Abschnitt mit den Ergebnissen für das Gesamt-CSP werden hier keine Angaben aus den CSP-Jahresberichten verwendet, da diese nicht für die Kontrollgruppe vorliegen.

### Kollaborationspartner aus Publikationen

Tabelle 9: Anzahl der Publikationen die in Kooperation mit Industriepartnern, anderen CSP-Fellows oder mit nationalen oder internationalen Kollaborationspartnern entstanden sind. Nur die Erst-/Letztautorenpublikationen wurden hier berücksichtigt.

Gruppe	Anzahl Publikationen	mit Industriepartner	mit nationalen Kollaborationen	mit internationalen Kollaborationen	mit anderen (J)CSP-Fellows
CSP	526	18	245	146	57
Kontrolle	4486	174	1756	1231	137

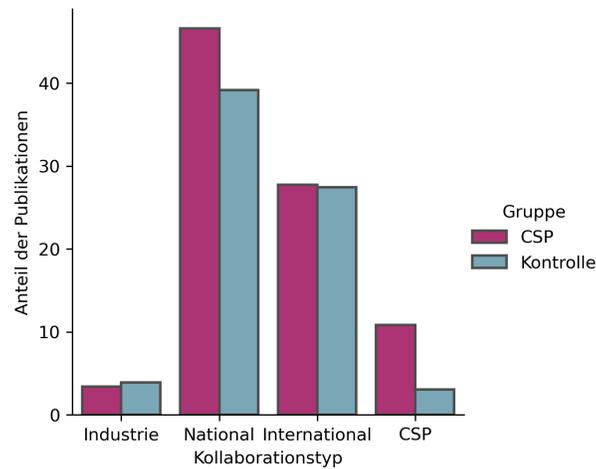


Abbildung 21: Anteil der Publikationen pro Jahreskohorte mit mindestens einem nationalen, internationalen oder Industriepartner bzw. einem anderen (J)CSP-Fellow, welcher in den Affiliationen genannt wird (nur Erst-/Letztautorenpublikationen).

## 4.6 Transparenz & Open Science

In dieser Kategorie werden verschiedene Open Science Metriken erfasst. Wir erfassen den Anteil sowie die Art der Open Access-Publikationen. Außerdem erfassen wir die Nennung von data & code sharing Statements in den Publikationsvolltexten mithilfe des von QUEST entwickelten text-mining Tools ODDPub (Tool: <https://github.com/quest-bih/oddpub>, Publikation: <http://doi.org/10.5334/dsj-2020-042>). Da es durch das automatisierte Screening zu falsch positiven Ergebnissen kommen kann, wurden die detektierten Fälle zusätzlich manuell überprüft. Nur die manuell verifizierten Fälle von Open Data und Open Code wurden hier dargestellt.

Open Data und Open Code ist ein eher neues Phänomen und tritt vor allem in neueren Publikationen auf. Da es in der Kontrollgruppe anteilig mehr Publikationen aus früheren Jahren gibt, wurden für Open Data und Open Code nur die Publikationen aus den Jahren 2017 - 2020 berücksichtigt um die Kohorten vergleichbarer zu machen.

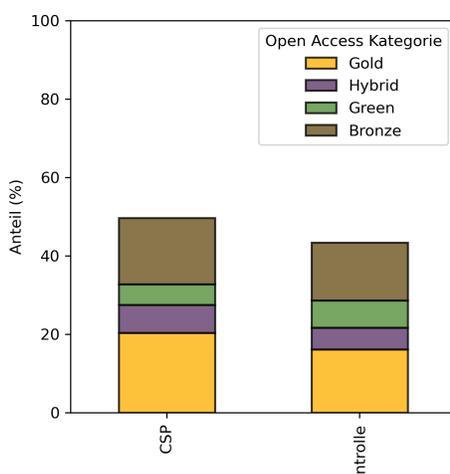


Abbildung 22: Anteil der Open Access-Publikationen für CSP und Kontrollgruppe (nur Erst-/Letztautorenpublikationen). Es wird zwischen den Open Access-Kategorien Gold (reines OA Journal), Hybrid (Journal mit OA Option), Grün (Repositoriumsversion) und Bronze (Artikel frei verfügbar gemacht durch Journal, aber keine ausgewiesene OA-Lizenz) unterschieden.

Tabelle 10: Anzahl der Publikationen mit Open Data und Open Code. Zur besseren Vergleichbarkeit der Kohorten wurden nur Publikationen aus den Jahren 2017 - 2020 berücksichtigt. Außerdem wurden nur Erst-/Letztautorenpublikationen berücksichtigt.

Gruppe	Anzahl Publikationen	mit Open Data	mit Open Code	Anteil Open Data	Anteil Open Code
CSP	247	13	6	5.3	2.4
Kontrolle	1082	14	2	1.3	0.2

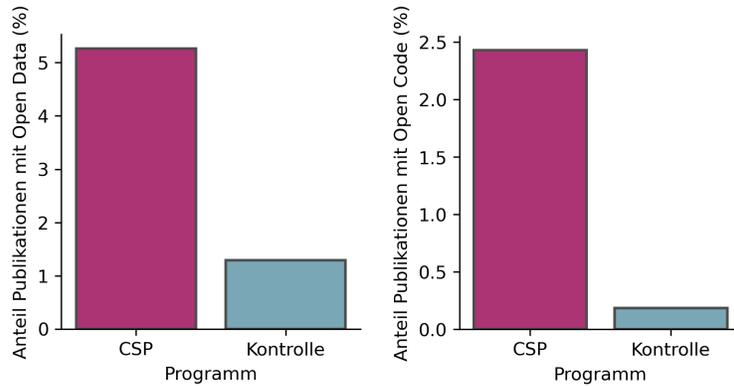


Abbildung 23: Anzahl der Publikationen mit Open Data und Open Code. Da Open Data und Open Code vor allem in neueren Publikationen auftauchen, wurde für diese Diagramme nur Publikationen der Jahre 2017 - 2020 berücksichtigt, um die unterschiedliche Verteilung der Publikationen über die Jahre für die beiden Gruppen auszugleichen. Nur Erst-/Letztautorenpublikationen wurden berücksichtigt.

#### 4.7 Alternativer Forschungsoutput

Zusätzlich zu den klassischen Journalartikeln wurden Veröffentlichungen in anderen Formaten erfasst. Über Dimensions werden zusätzlich Preprints, Proceedings sowie Buchkapitel erfasst. Anders als im Abschnitt mit den Ergebnissen für das Gesamt-CSP werden hier keine Angaben aus den CSP-Jahresberichten für die Kategorien Patente sowie Konferenzposter und -präsentationen verwendet, da diese nicht für die Kontrollgruppe vorliegen.

Tabelle 11: Anzahl der Beiträge für die verschiedenen Kategorien des alternativen Forschungsoutputs. Nur Erst-/Letztautorenpublikationen wurden berücksichtigt.

Jahreskohorte	Anzahl Fellows	Preprints	Proceedings	Buchkapitel
CSP	52	11	21	23
Kontrolle	351	25	108	191

## 5. Diskussion

Die in diesem Bericht beschriebene bibliometrische Analyse soll einen Überblick über verschiedene bibliometrische Metriken geben, welche verschiedenste Aspekte der Publikationstätigkeit beleuchten. Unter Punkt 3 wurden die Ergebnisse für das gesamte (J)CSP erfasst, welche sowohl klassische als auch alternative Publikationsformen, Forschungsbereiche, Kollaborationen als auch Transparenzmetriken umfassen.

Bezüglich der thematischen Ausrichtung der Forschungsarbeiten der (J)CSP-Fellows lässt sich ein klarer Fokus auf der klinischen Seite feststellen. Hierbei gibt es jedoch eine große Varianz, wobei die Forschungsarbeiten einiger Personen deutlich den Bereichen ‘animal’ oder ‘molecular & cellular’ zuzuordnen sind. Darüber hinaus werden verschiedenste Forschungsfelder abgedeckt wie Onkologie, Neurowissenschaften oder Kardiologie. Für den Themenbereich Teamscience konnte ein hoher Anteil an nationalen und internationalen Kollaborationen, aber auch Kollaborationen zwischen den (J)CSP-Fellows festgestellt werden.

Im Vergleich der habilitierten CSP-Fellows zur Kontrollgruppe von anderen habilitierten Wissenschaftler\*innen an der Charité unter Punkt 4 lassen sich verschiedene Beobachtungen machen. Zunächst einmal verteilen sich die Personen aus den beiden Gruppen unterschiedlich auf die verschiedenen Jahre. Dies wird vor allem daran liegen, dass das (J)CSP in den ersten betrachteten Jahren noch sehr jung war und noch nicht viele habilitierte Fellows hervorgebracht hatte. Für die letzten Jahre seit 2017 lässt sich aber beobachten, dass ein beträchtlicher Anteil aller habilitierten Personen an der Charité (ca. 20-30%) aus dem (J)CSP stammten.

Für viele der erfassten Metriken konnte kein merklicher Unterschied festgestellt werden. Dies gilt für die Anzahl der Publikationen, der Resonanz (RCR, NIH Perzentil), sowie das Setting (klinisch/animal/molecular & cellular). Für die Forschungsfelder gibt es leichte Unterschiede für manche Bereiche (u.a. mehr Onkologie und Genetik, weniger Public Health beim CSP), in Betracht der geringen Personenzahl in der CSP-Gruppe sind Schwankungen hier aber zu erwarten und die Verteilungen sind insgesamt sehr ähnlich. Im Bereich Teamscience sind in der CSP-Gruppe etwas mehr nationale Kollaborationen sowie deutlich mehr Kollaborationen mit anderen (J)CSP-Fellows zu beobachten. Dies kann so interpretiert werden, dass die Vernetzung der CSP-Fellows untereinander gut ist, so dass sie deutlich mehr Kooperationen untereinander eingehen.

Im Bereich Open Science wurden für die Metriken nur die Jahre 2017 - 2020 verglichen um die ungleiche Verteilung der Publikationen auf die Jahre zwischen den Gruppen auszugleichen und so einen faireren Vergleich zu ermöglichen. Dies ist nötig, da für die Open Science Metriken bekannt ist, dass sie gerade erst in den letzten Jahren wichtiger und bekannter geworden sind, so dass sie eher von aktuelleren Publikationen erfüllt werden. Der Vergleich ergab einen leicht erhöhten Anteil an Open Access Publikationen bei der CSP-Gruppe, wobei es hier vor allem mehr Gold Open Access Publikationen (Publikationen in reinen Open Access Journalen) gab. Ein deutlich größerer Unterschied war für die Metriken Open Data und Open Code zu sehen. Hier lag die CSP-Gruppe deutlich vorne. Insgesamt kommen Open Data & Code jedoch auch in der CSP-Gruppe selten vor, was sich mit generellen Zahlen für den biomedizinischen Bereich deckt.

## 6. Zusammenfassung

Die in diesem Bericht beschriebene bibliometrische Analyse soll einen Überblick über verschiedene Aspekte der Publikationstätigkeit für das Clinician Scientist Programm beleuchten. Hierbei wurden sowohl klassische als auch alternative Publikationsformen, die fachliche Ausrichtung der Arbeiten, Kollaborationen auf verschiedenen Ebenen sowie Metriken zur Transparenz betrachtet. Für die Analyse wurden Daten aus der Publikationsdatenbank Dimensions, aus der iCite-Datenbank des NIH sowie Eigenangaben der (J)CSP-Fellows aus den Jahresberichten verwendet. Die Daten wurden auf Ebene der einzelnen Jahreskohorten ausgewertet. Neben der Erfassung der Metriken für das gesamte (J)CSP wurde eine Untergruppe der habilitierten CSP-Fellows mit einer Kontrollgruppe bestehend aus anderen habilitierten Personen an der Charité verglichen. Insgesamt zeigt die vergleichende Analyse, dass sich beide Kohorten für viele Metriken sehr ähnlich sind. Bemerkenswert ist aber, dass in den letzten Jahren ein hoher Anteil aller habilitierten Personen an der Charité (ca. 20-30%) aus dem CSP stammten. Außerdem konnte für die CSP-Gruppe etwas mehr nationale Kollaborationen sowie mehr Kooperationen mit anderen (J)CSP-Fellows beobachtet werden. Schlussendlich wurde in der Kategorie Open Science für die CSP-Gruppe ein höherer Anteil an Publikationen mit Open Access, Open Data und Open Code beobachtet.

## 7. Literaturverzeichnis

Hendriks B.; Schendzielorz, C.; Heger, C.; Reinhart, M. (2021): Kritische Bestandsaufnahme des BIH Charité (Junior) Clinician Scientist Programms: Untersuchungen einer integrierten Forschungs- und Facharztweiterbildung in der Universitätsmedizin. Ergebnisse der Programmevaluation 2019/20. Herausgeber: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW).

Hutchins B.I.; Yuan X.; Anderson J.M.; Santangelo G.M. (2016) Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level. PLoS Biol., 14(9):e1002541. doi: 10.1371/journal.pbio.1002541.

Riedel, N.; Kip, M. and Bobrov, E. (2020). ODDPub – a Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications. Data Science Journal, 19(1), p.42. DOI: <http://doi.org/10.5334/dsj-2020-042>

Weber GM. (2013) Identifying translational science within the triangle of biomedicine. J Transl Med., 11:126. DOI: 10.1186/1479-5876-11-126.

## 8. Abkürzungsverzeichnis

API – Application Programming Interface  
BIH – Berlin Institute of Health  
CSP – Clinician Scientist Program  
DOI – Digital Object Identifier  
FCR – Field Citation Ratio  
GRID ID – Global Research Identifier Database  
ID – Identifier  
JCSP – Junior Clinician Scientist Program  
MeSH – Medical Subject Heading  
Mol/Cel – Molecular/Cellular  
NIH – National Institutes of Health  
OA – Open Access  
ODDPub – Open Data Detection in Publications  
ORCID ID – Open Researcher and Contributor ID  
RCR – Relative Citation Ratio

## 9. Anhang

### 9.1 Weitere Abbildungen für alle Publikationen, inkl. Mittelautorenschaften

#### 9.1.1 Ergebnisse - Vergleich mit Kontrollgruppe - alle Autorenpositionen

Tabelle 12: Anzahl der Publikationen die in Kooperation mit Industriepartnern, anderen CSP-Fellows oder mit nationalen oder internationalen Kollaborationspartnern entstanden sind. Alle Publikationen wurden hier berücksichtigt.

Gruppe	Anzahl Publikationen	mit Industriepartner	mit nationalen Kollaborationen	mit internationalen Kollaborationen	mit anderen (J)CSP-Fellows
CSP	1502	84	710	467	198
Kontrolle	10595	496	4829	3367	450

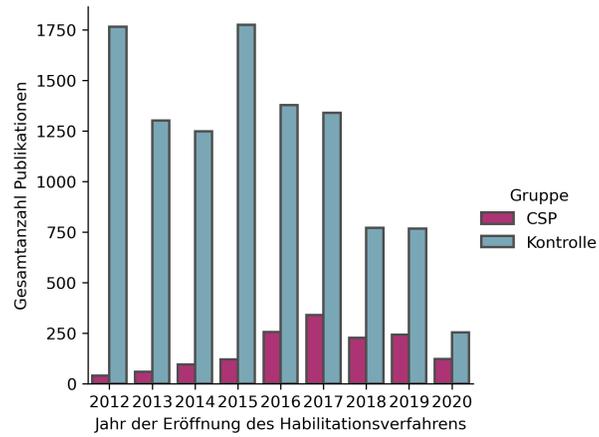


Abbildung 24: Anzahl der gefundenen Publikationen pro Jahrgangskohorte für CSP und Kontrollgruppe (alle Autorenpositionen).

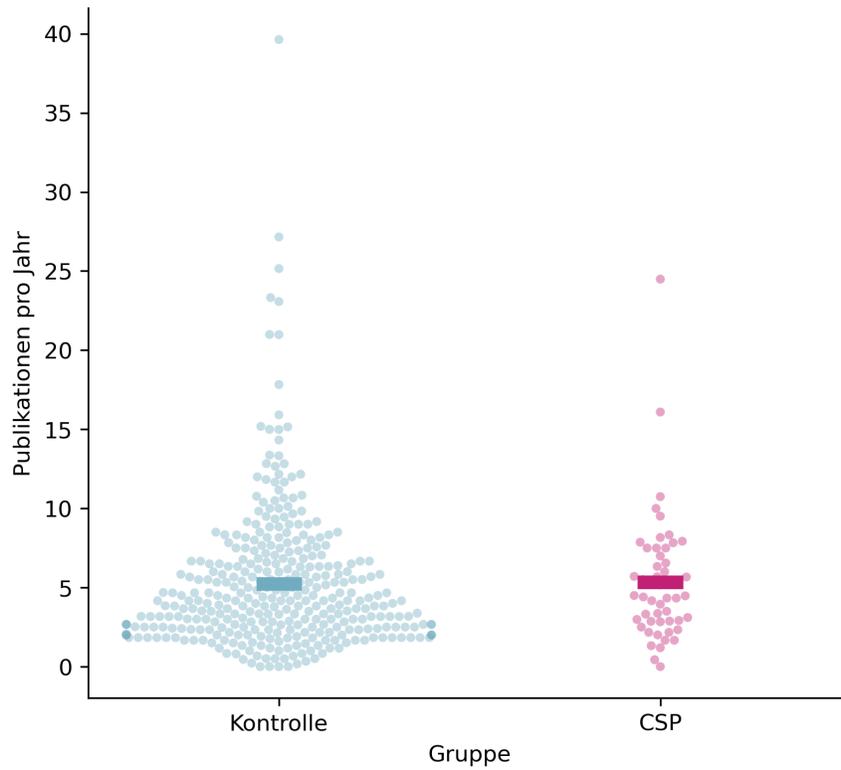


Abbildung 25: Publikationen pro Jahr und pro Fellow für CSP und Kontrollgruppe.

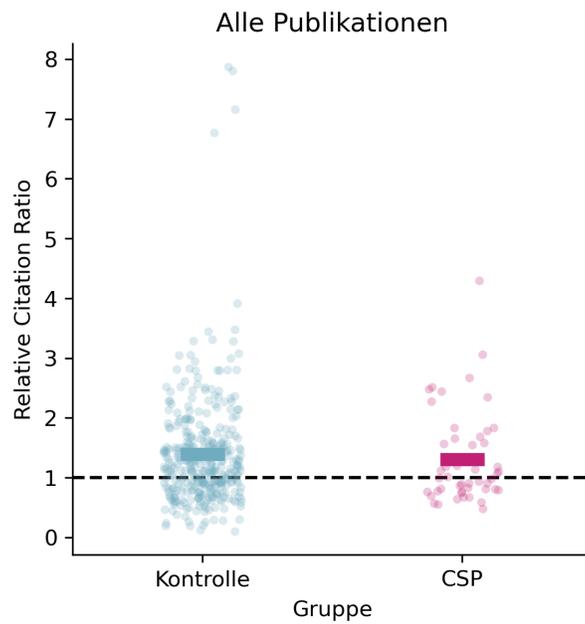


Abbildung 26: Mittlerer RCR pro Fellow im Vergleich zwischen CSP und Kontrollgruppe (alle Autorenpositionen). Der RCR ist so normiert, dass Publikationen im Mittel einen Wert von 1 erreichen. Ein Wert über 1 bedeutet eine überdurchschnittlich häufige Zitation.

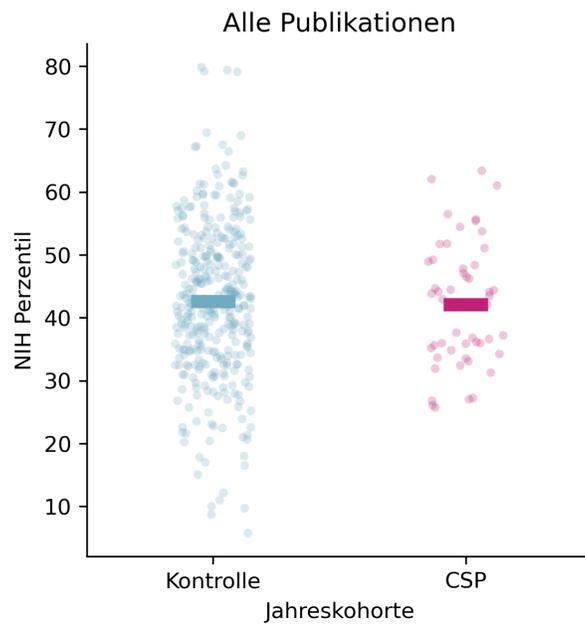


Abbildung 27: Mittleres NIH Percentil pro Fellow im Vergleich zwischen CSP und Kontrollgruppe (alle Autorenpositionen).

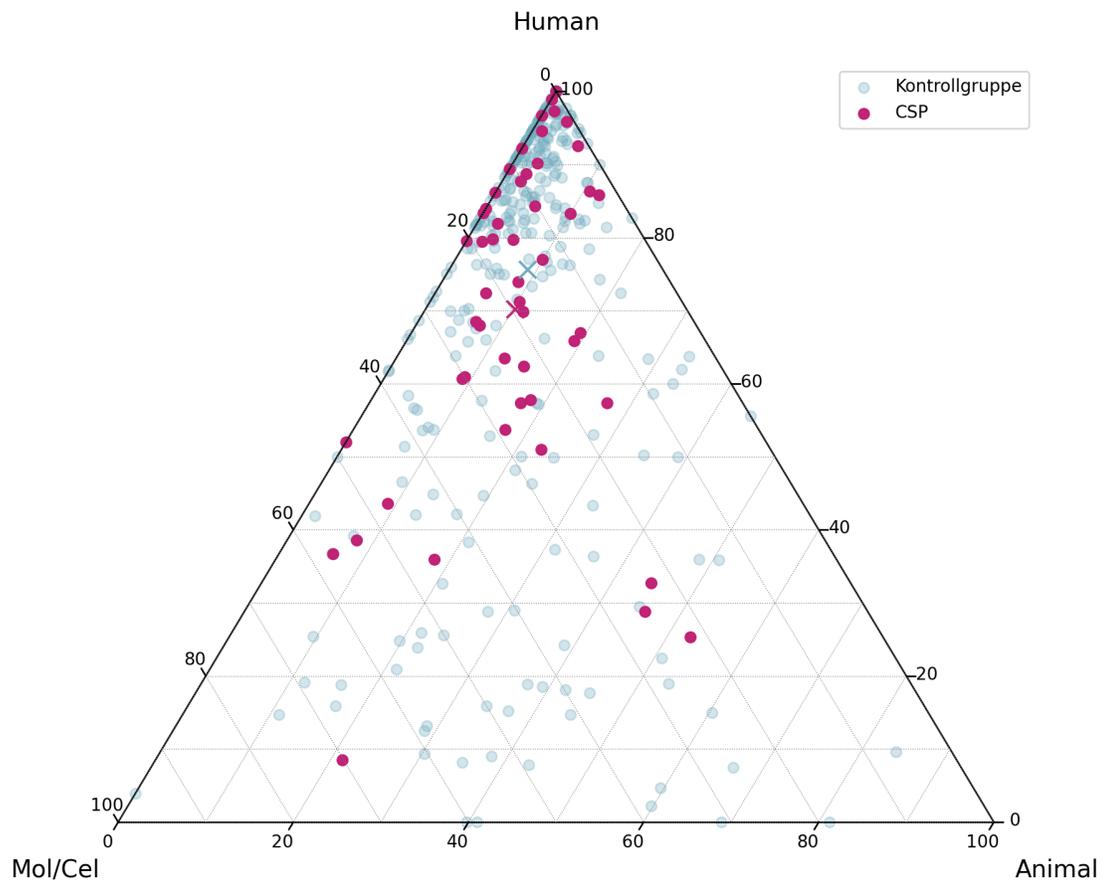


Abbildung 28: Ausrichtung der Publikationen entlang der Achsen klinisch/animal/molecular and cellular, gemittelt über alle Publikationen pro Person. Die Kreuze geben den Mittelwert über alle Personen einer Gruppe an.

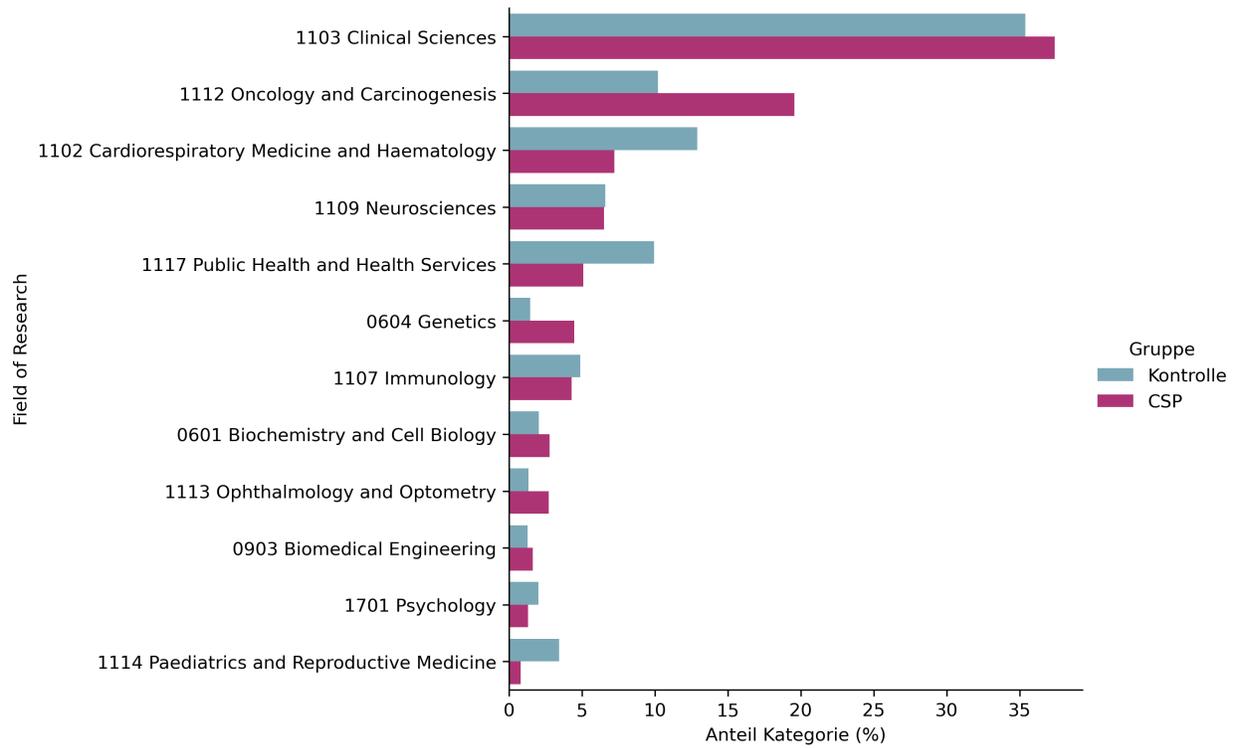


Abbildung 29: Vergleich der Anteile der Forschungsfelder zwischen CSP und Kontrollgruppe für alle gefundenen Publikationen.

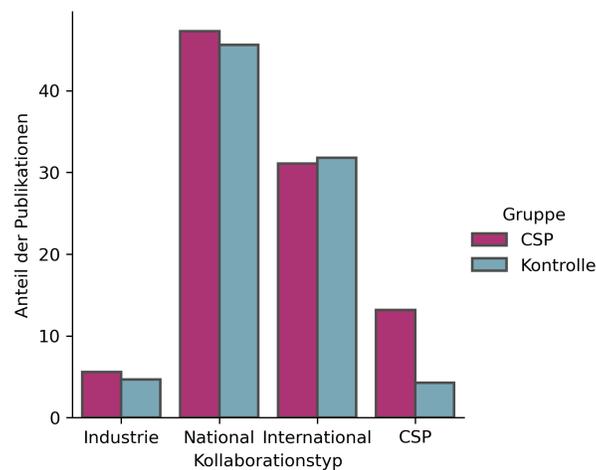


Abbildung 30: Anteil der Publikationen pro Jahreskohorte mit mindestens einem nationalen, internationalen oder Industriepartner, welcher in den Affiliationen genannt wird (alle Autorenpositionen).

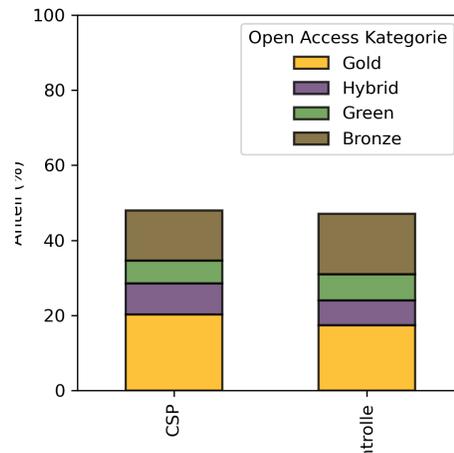


Abbildung 31: Anteil der Open Access-Publikationen für CSP und Kontrollgruppe (alle Autorenpositionen). Es wird zwischen den Open Access-Kategorien Gold (reines OA Journal), Hybrid (Journal mit OA Option), Grün (Repositoriumsversion) und Bronze (Artikel frei verfügbar gemacht durch Journal, aber keine ausgewiesene OA-Lizenz) unterschieden.

Tabelle 13: Anzahl der Publikationen mit Open Data und Open Code. Alle Publikationen wurden berücksichtigt.

Gruppe	Anzahl Publikationen	mit Open Data	mit Open Code	Anteil Open Data	Anteil Open Code
CSP	1502	68	19	4.5	1.3
Kontrolle	10595	184	33	1.7	0.3

Tabelle 14: Anzahl der Beiträge für die verschiedenen Kategorien des alternativen Forschungsoutputs. Alle Publikationen wurden berücksichtigt.

Jahreskohorte	Anzahl Fellows	Preprints	Proceedings	Buchkapitel
CSP	52	34	74	41
Kontrolle	351	93	347	295