

Introducing and testing an advanced quantitative methodological approach for the evaluation of research centers: a case study on sustainability science

Omar Kassab,^{1,*} Rüdiger Mutz¹ and Hans-Dieter Daniel^{1,2}

¹Professorship for Social Psychology and Research on Higher Education, Department of Humanities, Social and Political Sciences, ETH Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland and ²Department of Psychology, University of Zurich, Binzmuehlestrasse 14, 8050 Zurich, Switzerland

*Corresponding author. Email: omar.kassab@gess.ethz.ch.

Abstract

With the growing complexity of societal and scientific problems, research centers have emerged to facilitate the conduct of research beyond disciplinary and institutional boundaries. While they have become firmly established in the global university landscape, research centers raise some critical questions for research evaluation. Existing evaluation approaches designed to assess universities, departments, projects, or individual researchers fail to capture some of the core characteristics of research centers and their participants, including the diversity of the involved researchers, at what point in time they join and leave the research center, or the intensity of their participation. In addressing these aspects, this article introduces an advanced approach for the ex post evaluation of research centers. It builds on a quasi-experimental within-group design, bibliometric analyses, and multilevel statistics to assess average and individual causal effects of research center affiliation on participants along three dimensions of research performance. The evaluation approach is tested with archival data from a center in the field of sustainability science. Against a widely held belief, we find that participation in research centers entails no disadvantages for researchers as far as their research performance is concerned. However, individual trajectories varied strongly.

Key words: research center; bibliometrics; research performance; accelerated longitudinal design; growth curve modeling; sustainability science.

1. Introduction

Research centers have evolved into indispensable organizational instruments in the university landscape (Ikenberry and Friedman 1972; Rivers and Gray 2013; Smith et al. 2016). Their strength lies in the ability to handle complex problems that could not be addressed in the traditional departmental and discipline-based context (Sabharwal and Hu 2013; Corley et al. 2017). However, research centers operate at the interface of conflicting research policy developments: On the one hand, universities are increasingly encouraged by funding entities to conduct solution-oriented research to tackle the grand societal challenges like climate change, energy supply, or urbanization. Those applied research questions require

collaboration across disciplinary boundaries and have ultimately led to an increased emergence of inter- and transdisciplinary research centers (Kueffer et al. 2012; SDSN 2017). At the same time, however, the academic ‘publish or perish’ system rewards efficiency in terms of individual research performance, which, given the coordination effort associated to inter- and transdisciplinary research, very often results in disciplinary and highly focused basic research (Talwar, Wiek and Robinson 2011; Lang et al. 2012; Wiek et al. 2014).

Despite the broad consensus on their systemic importance (Spangenberg 2011; Ziegler and Ott 2011), researchers are somewhat reluctant to participate in research centers, presumably due to concerns that this might negatively affect their careers (Stokols et al.

2008b; Su 2014). But is the skepticism empirically justified? A number of studies have investigated the question of whether and to what extent participation in research centers has an impact on the publication activities and collaboration behavior of individual researchers (Landry and Amara 1998; Wen and Kobayashi 2001; Bozeman and Rogers 2002; Gaughan and Bozeman 2002; Corley and Gaughan 2005; Lee and Bozeman 2005; Lin and Bozeman 2006; Mallon 2006; Boardman and Corley 2008; Ponomariov and Boardman 2010; Sabharwal and Hu 2013; Youtie, Kay and Melkers 2013). While most of the studies have found participation in research centers not to be disadvantageous in terms of individual research performance (usually measured in terms of publication productivity), methodological shortcomings put these findings into perspective. Against the systemic relevance of research centers, we herewith propose a comprehensive evaluation approach that further develops previous approaches theoretically and methodologically. We demonstrate and test the approach on the basis of a research center with a focus on sustainability science.

The approach we propose is comprehensive in that it entails the data collection procedure, an underlying quasi-experimental research design, and applies the latest available data analysis methods (i.e. multilevel analysis and growth curve modeling). The basic idea of the approach is to look at individual researchers and their entire publication record. The beginning of their research center participation, and of the corresponding publication activity, is thus seen as a transition from the previous publication activity. In the research design, this transition is understood as a ‘treatment’ while the time prior to the research center participation is regarded as the baseline. From an accountability perspective, it is important to evaluate not only the individual causal effects (ICEs) of research center participation on the individual researcher but also the summative average causal effect across all participants. For this purpose, advanced multilevel models serve to capture the hierarchical data structure, i.e., the publication activity over time (level 1) for different researchers (level 2) while at the same time providing ways to solve the aggregation problem. Multilevel models, thus, not only allow us to capture the average causal effect of the research center, but also make it possible to assess the effect on the individual. In contrast to conducting surveys with varying response rates, the combination of archival data with bibliometric data safeguards the objectivity of the evaluation approach as a whole.

This article is structured as follows: we start with a review of the research center evaluation literature, which we draw upon to develop the theoretical foundations of our evaluation approach. We then illustrate the shortcomings of existing approaches that we aim to resolve, before briefly describing the case that we use to test the evaluation approach. Data and methods are introduced thereafter. After presenting the results in detail, the article closes with a discussion of strengths, limitations, and policy implications of the approach.

2. Literature and theory

2.1 Research centers in the university context

Research centers are organizational entities within a university that exist chiefly to serve a research mission, are set apart from the departmental organization, and include researchers from more than one department (Bozeman and Boardman 2003: 17). Since the first research centers were founded in the USA in the 1970s, national

innovation systems around the globe have increasingly made strategic use of research centers to address problems that are too complex for a single department to manage (Geiger 1990; Stokols et al. 2008a; Mittelstrass 2011; Rivers and Gray 2013; Su 2014). Beyond their ability to facilitate inter- and transdisciplinary research, they provide various opportunities for collaboration with sectors beyond academia, for training of future generations of academic workforce, for technology transfer and dissemination activities directed to various target audiences, for building network ties, and for career changes, among others (Santoro and Gopalakrishnan 2001; Feller, Ailes and Roessner 2002; Slaughter et al. 2002; Boardman and Corley 2008; Gaughan and Ponomariov 2008; Ponomariov and Boardman 2010; The Madrillon Group Inc. 2010; Ávila-Robinson and Sengoku 2017; Corley et al. 2017).

From an organizational viewpoint, there are vast differences between research centers across a multitude of dimensions, such as the number of their participants, their institutional and disciplinary composition, collaboration and networking opportunities, their funding schemes, their strategic goals, or their operative lifespan (Rogers, Youtie and Kay 2012; Rivers and Gray 2013; Sabharwal and Hu 2013; Bishop et al. 2014; Smith et al. 2016; Corley et al. 2017). Research centers are not substitutes for university departments, but rather require and complement them. For many researchers participating in university-based research centers, the department remains their primary affiliation, while only a share of their total working time is devoted to projects at the research center (Boardman and Bozeman 2007; Kassab, Schwarzenbach and Gotsch 2018).

2.2 Understanding the dynamics at research centers and the implications for research performance

A typical research center is characterized by intricate coordination processes and inter- and transdisciplinary knowledge exchange. The dominant output orientation in evaluation practice, however, hardly does justice to this reality (Cozzens and Turpin 2000; Coryn et al. 2007). According to a study on the ‘Evaluation of Research Center and Network Programs at the National Institutes of Health’ (NIH), based on a review of 61 cases from the years 1978 to 2009, 81% of the cases focused on scientific publications as the primary output to be assessed. Moreover, the review shows that 61% of all studies relies solely on descriptive statistics (The Madrillon Group Inc. 2010).

As a remedy to this narrow focus, Bozeman, Dietz and Gaughan (2001) developed an evaluation model to delineate what they label the Scientific and Technical Human Capital (STHC), defined as ‘the sum of an individual researcher’s professional network ties, technical knowledge and skills, and resources broadly defined’ (Bozeman, Dietz and Gaughan 2001: 636). As such, their perspective focuses less on the discrete outputs but rather on the processes that enable researchers to expand their networks and improve their capabilities. Since its introduction, the STHC model has been applied in many areas of science and technology policy research, for example, to evaluate career development, research collaboration, or institutional interactions (Corley et al. 2017). It is the holistic view of the STHC model that has also made it the most prominent perspective for theorizing the dynamics in research centers, as can be seen from numerous examples in the literature (Bozeman and Corley 2004; Dietz and Bozeman 2005; Lin and Bozeman 2006; Boardman and Corley 2008; Ponomariov and Boardman 2010; Sabharwal and

Hu 2013). From this perspective, research centers are understood as ‘organizational reservoirs’ of STHC, to which all participants of the research center gain access, in particular during the research center’s lifetime but also beyond (Ponomariov and Boardman 2010: 617).

We also draw on the STHC model to describe the implications of research center participation for individual research performance.¹ To provide a comprehensive perspective, we define research performance not only on the basis of publication productivity, but instead as measured by three indicators: (1) ‘scientific productivity’ in terms of the number of publications, (2) ‘scientific impact’ in terms of the number of citations, and (3) ‘integration into the scientific community’ in terms of the number of coauthors.

The basic assumption of the STHC model is that participation in a research center expands individual capabilities and networks. With regard to the first dimension, scientific productivity, this suggests that research centers provide more financial and human resources than would be the case in a departmental setting, thus, leading to an increased scientific productivity (Corley and Gaughan 2005; Bunton and Mallon 2006; Sabharwal and Hu 2013). Due to the denser network of contacts and additional communication mechanisms provided by the research center management, it can be assumed that the scientific publications produced at the research center will have a greater visibility, which in turn will increase the citation probability. Finally, the third dimension of research performance, integration into the scientific community, is likely to be boosted by joining the research center because of increased access to a pool of potential collaborators, which in some cases is even explicitly demanded by the funding entity (Gaughan and Ponomariov 2008; Ponomariov and Boardman 2010).

While we, in line with previous studies, acknowledge that the STHC model is in principle very well-suited for investigating and explaining the dynamics in research centers, we would like to concentrate on three key characteristics of research centers and their participants that have as yet been only insufficiently taken into account in previous evaluation studies. To this end, we would like to start from the STHC perspective and its basic assumptions outlined above, take up additional aspects, and thus form the theoretical basis for our evaluation approach.

2.2.1 Diversity of participants (*‘diversity’*)

It is in the nature of a research center that participants differ from each other in many respects and, by definition, have diverse disciplinary backgrounds. Leveraging this diversity effectively is one of the greatest strengths of research centers, because it makes the conduct and success of inter- and transdisciplinary research possible in the first place (Clark 2007: 1737; Lang et al. 2012). From the STHC perspective, the individual ‘internal resources’, understood as cognitive abilities or technical knowledge, are aggregated for the duration of the researcher’s affiliation with the research center, thus, making them accessible to all other participants (Bozeman, Dietz and Gaughan 2001; Ponomariov and Boardman 2010). With regard to the research performance and scientific progress in general, the resources that a research center can bring together add up to a whole that is ‘greater than the sum of its parts’. While this understanding of diversity is largely based on the disciplinary aspect, other characteristics such as the role of the participants in the research center as well as their academic age, gender, or institutional culture have shown to play a crucial role as well (Bishop et al. 2014; Corley et al. 2017).

Those who take on a management role, for example, not only have full access to the aggregated resources of the research center,

but at the same time have an opportunity to develop leadership skills and thus an increased level of STHC (Elkins and Keller 2003; Gray 2008). A further strength of the STHC model is the ‘recognition of the evolution of the scientist throughout his or her productive life cycle’ (Bozeman, Dietz and Gaughan 2001). This is particularly important in view of the fact that previous studies have identified a generational ‘cohort effect’ on the impact of a research center when it comes to research performance (Sabharwal and Hu 2013).

While the STHC model in its original form does not make any gender-specific distinctions, a further development of the model brings in a cultural dimension, defined as ‘the sum of an individual scientist’s experiences that are gained while interacting with people from diverse cultural backgrounds’ (Corley et al. 2017), one of which is gender. As women engage in inter- and transdisciplinary research centers at least as often as men (Corley and Gaughan 2005), participants in research centers are typically in contact with colleagues of different sexes, which, according to Corley et al. (2017), ultimately increases their overall level of STHC.

2.2.2 Transition in and out of research centers (*‘transition’*)

Another characteristic of research centers and their participants that has not yet been sufficiently taken into account is related to the fact that ‘[o]ver time, individuals, groups, and firms encounter acute events that involve transitioning from one state or role to another’ (Bornmann, Mutz and Daniel 2009; Bliese, Adler and Flynn 2017). In the concrete context of research centers, those transitions can take place when marking the starting or end points of the affiliation, or during temporary commitments to projects.

The STHC model indeed assumes that the individual STHC is constantly changing. Theory says that ‘the individual may “load” at a different level on the dimension[s] at any particular point in time’ (Bozeman, Dietz and Gaughan 2001). If the individual STHC changes over time and by means of interaction, then in the context of a research center, this indicates that the moment of transition and period of affiliation must be taken into account. Previous studies have focused on incorporating affiliation versus nonaffiliation on an annual basis with a binary coding regime (Boardman and Corley 2008; Ponomariov and Boardman 2010; Sabharwal and Hu 2013; Bishop et al. 2014). However, this is not fully satisfactory for two reasons: First, because it has to be assumed that participants may be involved in more than one project at the research center, consecutively or simultaneously, which in turn implies a greater STHC development potential and impact on research performance. Second, the research center routine not only includes activities on the level of the project but also networking activities on the level of the center as a whole. Essentially, if one intends to assess the impact of participation in the research center based on individual research performance, one should consider the aspect of transition in all its complexity as conceptualized by the STHC model, both on the project level and on the organizational level.

2.2.3 Intensity of participation (*‘intensity’*)

The extent to which participation in a research center ultimately affects individual research performance is also a matter of exposure. As in classical experimental settings, the effect depends on the ‘intensity’ of the treatment (West, Cham and Liu 2014). As the participants in the vast majority of cases have further obligations in addition to their research center affiliation, it must also be assumed that the effect on their research performance varies accordingly. When spending only a share of the total working time at the

research center, the individual researcher not only has limited access to the aggregated STHC resources, but he or she also has fewer opportunities to develop their own STHC than would be possible in case of full-time affiliation. Boardman and Corley (2008) took an important step in this direction by asking the research center participants in their sample how much time they spent working alone and how much of their work involved other groups, sectors, or countries. However, similar to the approach by Ponomarev and Boardman (2010), which takes ‘core institution affiliation’ into account, both studies only integrate a ‘binary’ research center affiliation indicator. In other words, the intensity is not measured.

In the preceding sections, we have discussed three aspects that have not yet been sufficiently taken into account in previous quantitative evaluations of research centers. With this article, we introduce an advanced methodological approach for the ex post evaluation of research centers. In the chapter that follows, the evaluation approach and the remedies it brings are described in more detail.

3. Case description: Competence Center Environment and Sustainability

The case used to demonstrate the evaluation approach is the Competence Center Environment and Sustainability (CCES), a research center in Switzerland that operated for 10 years between 2006 and 2016 with a focus on sustainability science (Kassab, Schwarzenbach and Gotsch 2018). CCES is one of the four inter- and transdisciplinary research centers that were established to promote research, education, and societal outreach activities within and between the six institutions that constitute the ETH Domain. The ETH Domain comprises the two Federal Institutes of Technology in Zurich (ETH Zurich) and Lausanne (EPFL), as well as four independent research institutions: the Paul Scherrer Institute (PSI), the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), the Swiss Federal Laboratories for Materials Science and Technology (Empa), and the Swiss Federal Institute of Aquatic Science and Technology (Eawag). While the six institutions differ greatly in terms of their research cultures—ETH Zurich and EPFL being rather oriented toward basic research, while the other four are more application oriented—they also share thematic research priorities, which the ETH Board, the ETH Domain’s management body, intended to consolidate through the foundation of the four research centers. As can be seen in Table 1, 170 senior researchers from the six institutions came together in CCES to work on a total of 26 inter- and transdisciplinary projects, covering five thematic priority areas of sustainability science: (1) Climate and Environmental Change, (2) Sustainable Land Use, (3) Food, Environment, and Health, (4) Natural Resources, and (5) Natural Hazards and Risks (i.e. ‘diversity’).

CCES was designed to operate in two phases, the first running from 2006 to 2010 and the second from 2011 to 2016. Of the 26 projects, 18 were conducted in the first phase and eight in the second phase. During the startup of the research center, review processes and administrative arrangements caused substantial delays to the beginnings of the projects. As a matter of fact, CCES affiliation did not take effect for all participants in the same year (i.e. 2006), but rather in a staggered manner. Depending on their project involvement, researchers also had varying exposure to the research center context. In some cases, the researchers’ affiliation did not extend over the entire project duration, but ended along the way, opening

Table 1. Sample description (researchers and publications)

Researchers (N = 198)			
		Absolute	Percent
Gender	Male	171	86
	Female	27	14
Cohort	Phase 1 (2006–2010)	102	52
	Phase 2 (2011–2016)	23	12
	Both phases (2006–2016)	45	23
	Comparison group	28	14
Institution	ETH Zurich	96	48
	EPFL	26	13
	Eawag	23	12
	WSL	27	14
	PSI	11	6
	Empa	5	3
	Other	10	5
Scientific background ^a	Social scientists	17	9
	Other disciplines	181	91
Year of PhD (cohorts)	Before 1990	66	33
	Between 1990 and 2000	87	44
	After 2000	21	11
	Missing information	24	12
Role	Professor	110	56
	Not professor	88	44
Publications (publication years x researchers; N = 3,250), annual			
	M	SD	Min/Max
Number of publications	4.37	3.98	1/31
Number of citations	46.52	80.52	0/1,503
Number of coauthors	24.10	121.58	0/5,405

^aClassification based on Frascati Manual (OECD 2007).

opportunities for new participants to join at a time when the project and the research center were already in operation (i.e. ‘transition’).

The participants of CCES were also involved in their respective projects to varying degrees. Very few were engaged in a project full time; the majority of the researchers participated in CCES on a part-time basis, suggesting other research activity beyond the research center. Moreover, their exposure differed not only with regard to the temporality (full time vs. part time), but also with regard to participation in more than one project over the course of the research center’s operation, either at the same time or consecutively (i.e. in terms of ‘intensity’).

4. Data and methods

4.1 Data

The main data basis consists of archival data in the form of 99 annual project reports of 26 projects over the course of 10 years, kindly provided by the CCES management.² From the reports, we retrieved data on the (1) individual researchers, (2) bibliometric data, and (3) institutional data on the research center. This data collection effort resulted in a longitudinal dataset that included observations of the same individuals over the course of their academic careers, thereby encompassing their affiliation with CCES.

4.1.1 Researcher data

We started by compiling a list of all 170 participants that were affiliated with CCES as principal investigators and project partners.³ We

incorporated into the dataset an additional 28 researchers who had submitted project proposals for CCES but were rejected after review. We do not regard these researchers as a randomized control group in the classical experimental sense, but as a comparison group of individuals who were formally qualified for CCES affiliation, but were not selected. The group is, therefore, not matched to the personal profiles of CCES participants.

Following the research evaluation literature and to capture the multiple aspects of diversity (see Section 2.2.1), we coded researcher-specific information derived from the annual project reports such as gender, the scientific background, year of PhD (indicating the beginning of the academic career), role in project, and academic title (Corley and Gaughan 2005; OECD 2007; Cañibano and Bozeman 2009; Sabharwal and Hu 2013). Where necessary, complementing and confirming information was retrieved from personal websites.

To capture the aspect of transition (see Section 2.2.2), we coded the actual project affiliations of the individual researchers, including both starting and ending time points of their affiliation and stating whether they were affiliated with multiple CCES projects, either in parallel or subsequently (Cafri, Hedeker and Aarons 2015). Among the 170 participating researchers, affiliations ranged from single project affiliation during one phase of CCES to up to four project affiliations over the course of both phases.

Addressing the aspect of intensity (see Section 2.2.3), we retrieved information on the time commitment associated with individual CCES projects, as documented in the annual project reports in full-time equivalent (FTE) per researcher and year. Table 2 gives an idea of how the data were structured and coded (exemplary).

4.1.2 Bibliometric data

In a second step, we downloaded the full publication histories of all 198 individual researchers, using the *Clarivate Analytics Web of Science* database. As the research center was still operating at the time that we conducted this study, the cutoff date for publications was the end of 2014. In total, we collected bibliometric data on 13,578 peer-reviewed journal articles. As the first publication dated back to the year 1980, the study covers a timeframe of 35 years.

4.1.3 Institutional data

Using unique identifiers, each of the 13,578 articles published in peer-reviewed journals was assigned to one or more of the 198

researchers. Publications produced within the context of CCES were indicated accordingly, with reference to the project and the researcher(s). All remaining publications (before, during, or after CCES) were coded with reference to the researcher(s) as well.

4.2 Variables

As introduced above, we understand research performance along three dimensions, including ‘scientific productivity’ in terms of the number of publications, ‘scientific impact’ in terms of the number of citations, and ‘integration into the scientific community’ in terms of the number of coauthors. The three corresponding dependent variables to be used in the analysis are count variables. As Table 1 shows, we counted the number of publications, the number of citations with a citation window of 5 years, and the number of coauthors, per researcher and publication year ($N = 3,250$). Furthermore, with the exception of the number of publications, we used count rates (Fleiss, Levin and Paik 2003). For count rates, we did not analyze the annual citations, but instead looked at the annual number of citations per publication (annual number of citations divided by annual number of publications), that is, how many citations a researcher receives for a publication per year on average. The citations were counted on the same citation window, and the citations were not field normalized, because the vast majority of the papers were published in natural and life science journals.

We distinguish two types of covariates or factors: covariates ‘between individuals’, which describe the researchers, and covariates or factors ‘within individuals’, which characterize the time course. Specifically, our approach includes the following covariates:

1. *Between individuals*: Researchers had different characteristics (gender, age and year of PhD, role in project, academic title, and scientific background) and belonged not only to different age cohorts but also to four different person clusters. One comparison group of researchers did not participate in CCES (see Section 4.1.1), a second group participated only in phase 1 (2006–10), a third group participated in phase 2 (2011–6) only, and a fourth group participated in both phases (2006–16).
2. *Within individuals*: Researchers published their articles in the time range from 1980 to 2014 (publication year). They participated in CCES projects at different points in time and run as a

Table 2. Example of coding scheme on the researcher level indicating unique researcher_id (researcher identification), year (publication year), project_id_1 (identification of project 1), project_id_2 (identification of project 2), FTE_id_1 (full-time equivalent in project 1), not_FTE_id_1 (full-time equivalent outside of project 1), FTE_id_2 (full-time equivalent in project 2), not_FTE_id_2 (full-time equivalent outside of project 2), pub_cces (number of publications in CCES), and pub_non_cces (number of publications not in CCES)

researcher_id	year	project_id_1	project_id_2	FTE_id_1	not_FTE_id_1	FTE_id_2	not_FTE_id_2	pub_cces	pub_non_cces
57	2002	0	0	0	1	0	1	0	5
57	2003	0	0	0	1	0	1	0	6
57	2004	0	0	0	1	0	1	0	4
57	2005	0	0	0	1	0	1	0	5
57	2006	111	0	0.2	0.8	0	1	0	5
57	2007	111	0	0.2	0.8	0	1	2	6
57	2008	111	321	0.2	0.8	0.3	0.7	3	7
57	2009	111	321	0.2	0.8	0.3	0.7	5	8
57	2010	0	321	0	1	0.3	0.7	0	10
57	2011	0	321	0	1	0.3	0.7	0	11
57	2012	0	0	0	1	0	1	0	10
57	2013	0	0	0	1	0	1	0	12
57	2014	0	0	0	1	0	1	0	11

rule through different phases: no CCES, phase 1 (2006–10), phase 2 (2011–6).

4.3 Research design

The basic research design underlying the evaluation approach we propose is a quasi-experimental within-group design that models the full publication history of a group of researchers over time (Shadish, Cook and Campbell 2002). Given the theoretical considerations and characteristics discussed above (see Section 2.2), we found the *longitudinal interrupted accelerated design* (McDowall et al. 1983; Willett, Singer and Martin 1998; Galbraith, Bowden and Mander 2017), a more sophisticated version of the basic research design, to be most suitable.

Alleviating the challenge of diversity (Section 2.2.1), the design makes possible the examination of individual researchers of different age cohorts and different stages of their career with respect to their individual trajectory of bibliometric indicators in a longitudinal perspective (*longitudinal accelerated design*). Their participation is captured as a ‘treatment’ over time (binary: participation or no participation), which has causal effects on their bibliometric indicators. We assume that the individual times series of publication trajectories is ‘interrupted’ (*interrupted time series*) by the affiliation with CCES (Wagner et al. 2002). The bibliometric indicators (such as number of publications) change due to participation in CCES in a way that could not be predicted based on expectations arising from the previous course of the bibliometric indicators.

The research design considers the research performance on two levels: *micro* impact and *macro* impact. To assess the *micro* impact, the effect of each project on the bibliometric indicators is examined, weighted according to the time commitment in the project in FTE per researcher per year. Although this procedure concisely addresses the challenge of intensity (see Section 2.2.3), it comes with the disadvantage that information about the effects of a project is limited to its duration. But as scientific output is frequently published after the project completion (i.e. publication lag), *macro* impacts are additionally examined, that is, effects attributed to affiliation with the research center as a whole.

The research design we propose makes the evaluation approach relatively robust against many of the common threats to internal validity (Shadish, Cook and Campbell 2002: 55). The most typical ones in this context are instrumentation, maturation, and history. The threat of instrumentation is alleviated through the objectivity of the bibliometric data as retrieved from the standardized *Web of Science* database retrospectively. The threat of maturation is mitigated through the statistical modeling of a baseline, as will be described in more detail below (see Section 4.4). In the opportunity-driven context of research funding, the most severe threat to internal validity is history. It is not unlikely that concurrent research center affiliations or other events could cause the observed effect on the participating researchers. In the research design applied here, the treatment that research center participants receive is not a single-shot treatment but rather a continuous exposure. Furthermore, as neatly documented in the annual reports and the corresponding data, that exposure is different for every researcher. The treatment is operationalized in two ways: in a binary way (participation or no participation), as has been done in previous studies, and by capturing the participation intensity.

The evaluation approach is also robust against many of the threats to external validity (Shadish, Cook and Campbell 2002: 86).

The research design and the statistical approach are sufficiently broad to be tailored to individual researchers, regardless of their disciplinary background or other characteristics relevant to the evaluation of their research performance. The approach can also be applied in different settings, provided the objectivity of the data is ensured (Christensen and Waraczynski 1988; Ferguson 2004). Relying on archival data is a strong safeguard against this threat. While this article demonstrates the evaluation approach using a concrete case of a research center, it can be used to study other cases as well, making it generalizable in the methodological sense. Finally, it is particularly robust, as it assumes a natural setting without the effects that could intervene and influence the effect under scrutiny in a laboratory setting (Shadish, Cook and Campbell 2002: 83).

4.4 Statistical approach

This section describes in detail the statistical approach as a central element of the evaluation approach. While the approach could be presented in general terms as well, the case of CCES is used as an example to increase transparency and to demonstrate the applicability of the approach. We propose an univariate multilevel approach (Goldstein 2011; Hox, Moerbeek and van de Schoot 2018), consisting of the following five elements:

- (1) *Measurement model*: Bibliometric indicators such as ‘number of publications’ are ordinary Poisson distributed count variables, positive integer values including zero (Cameron 1998; Hilbe 2014; Mutz, Wolbring and Daniel 2016). In the case of stronger overdispersion, where the variance does not equal the mean of the variables, a negative binomial regression model is applied. The criterion for overdispersion is the ratio of Pearson χ^2 and degrees of freedom, which according to the model estimation should not be much greater than 1.0 (Hilbe 2014: 82). The problem of zero-inflation with ‘number of citations’ (disproportional number of noncited publications) is considered to be a problem of overdispersion and handled with a negative binomial distribution. Rates are represented by an ‘offset’ in the regression model. As the logarithm of rates $\ln(n_p/n)$ equals the difference $\ln(n_p) - \ln(n)$, the corresponding regression model can be complemented simply by an additional variable, $\ln(n)$, that has a regression coefficient of 1.0, so that, again, $\ln(n_p)$ can be modeled as an outcome (Fleiss, Levin and Paik 2003: 347; SAS Institute Inc. 2014: 3144).
- (2) *Impact of CCES publications*: As briefly described above, two types of publications are differentiated. One type consists of articles that were published in the context of a CCES project, as precisely documented in the annual reports of the CCES projects. The second type consists of publications that were not produced in the context of CCES (non-CCES). These are all publications in the dataset that were published prior to the establishment of the research center in 2006, as well as all publications since 2006 in which the 170 researchers were involved but that were not specified in the annual reports as CCES publications. The two types of publications are defined as variables as follows: One variable represents all publications (cumulative) of a researcher across all years. Another variable represents the publications that were not produced within CCES (non-CCES). Accordingly, two records (data rows) per year are produced for every researcher. In turn, the difference between the two records is the number of publications that a researcher published in the context of CCES. For the logarithmic transformed bibliometric

indicator y_{jic} of researcher i in publication year j in cohort c , the following first model component can be defined, whereby x_{jic} identifies the two different variables ($x_{jic} = 0$: all without CCES publications, $x_{jic} = 1$: all publications):

$$\ln(y_{jic}) = \beta_{00} + \beta_{01}x_{jic}, \quad (1)$$

where β_{00} as fixed effect denotes the mean value of the bibliometric indicator regarding all non-CCES publications, and β_{01} as fixed effect denotes the mean value of the bibliometric indicators of CCES publications. Eventually, the overall model estimation was based on the total number of publications (CCES/non-CCES), because model estimation and testing is more efficient for large sample sizes than for small ones.

- (3) *Growth curve model*: With regard to the natural log link for the dependent variables, a linear trend is assumed, which if necessary can be extended to a nonlinear trend (polynomial time trend) as further model assumption. The growth model represents the development of a researcher irrespective of any effects in the sense of 'maturation', which might result from participating in the research center (see Section 4.3). As discussed above, it may be assumed that researchers in their individual trajectories of bibliometric indicators could deviate from this general trend more or less (interindividual differences in intraindividual changes). It may also be assumed that there are different growth trends in the different age cohorts: A researcher who started publishing in 1980 will most likely have a different trajectory than a researcher who began publishing in 2002 (Way et al. 2017). Growth curve models can be estimated either as a multivariate structural equation model (SEM; Duncan, Duncan and Strycker 2013) or as a univariate multilevel model (Raudenbush and Bryk 2002: 160; Goldstein 2011: 147; Hox, Moerbeek and van de Schoot 2018: 79). As the time span from 1980 to 2014 is rather long and a SEM would need a variable for each year, a multilevel model is preferred. Also, in a multilevel model, no missing values occur, as no fixed measurement occasions must be assumed (Hox, Moerbeek and van de Schoot 2018: 106). Based on all of these considerations, we chose the following three-level growth curve model for the bibliometric indicator y_{jic} , where individual trajectories (level 1) of researchers (level 2) are nested within age cohorts (level 3):

$$\begin{aligned} \ln(y_{jic}) &= \beta_{00ic} + \beta_{1ic}(t_i - 2006) \\ \beta_{00ic} &= \beta_{00c} + u_{00ic} \\ \beta_{1ic} &= \beta_{1c} + u_{1ic} \quad u_{00ic}, u_{1ic} \sim N\left(\begin{pmatrix} \sigma_{u00}^2 & \sigma_{u001} \\ \sigma_{u001} & \sigma_{u1}^2 \end{pmatrix}\right) \\ \beta_{00c} &= \beta_{00} + v_{00c} \\ \beta_{1c} &= \beta_{1} + v_{1c} \quad v_{00c}, v_{1c} \sim N\left(\begin{pmatrix} \sigma_{v00}^2 & \sigma_{v001} \\ \sigma_{v001} & \sigma_{v1}^2 \end{pmatrix}\right), \end{aligned} \quad (2)$$

where the publication year t is centered at the year 2006. The year 2006 is favored against the year 1980 (the first year of a publication in the sample) due to the fact that the starting point of CCES is 2006, and therefore, the year effect vanishes in 2006 (for $t_i = 2006$: $\beta_0(t_i - 2006) = 0$). In this way, other 'treatment' effects can be more easily identified (Galbraith, Bowden and Mander 2017: 5; Hox, Moerbeek and van de Schoot 2018: 110f). Due to more general time trends in growth of science (Bornmann and Mutz 2015), the overall timeline from 1980 to 2014 was of primary interest in our study and was given preference over the individual timelines starting from the first publications of a researcher, which would require centering on the publication year of the first publication of each researcher.

The individual trajectories of bibliometric indicators of a researcher can be represented by an individual random intercept, u_{00ic} , and a random slope of the year trend, u_{1ic} , and their corresponding variance-covariance matrix. The same is true for the cohort effects with random intercept and slope v_{00c} and v_{1c} for each cohort c . This model makes it possible to model not only the average linear time trends (fixed effects: β_{00} , β_{1}), but also the individual trajectories of researchers and cohorts, represented by the random effects model aspect. We can, thus, speak of a *cohort-sequential model* (Klaiber, Seeling and Mutz 2002; Hox, Moerbeek and van de Schoot 2018: 109). In addition, covariates can be added to the model that can explain the interindividual differences in intraindividual changes over time (e.g. age and year of PhD). These could essentially be represented as interactions. We also tested whether in addition to the linear and exponential trends there were also quadratic and cubic time trends.

- (4) *Micro impact—multiple membership*: To estimate the *micro impact*, i.e. the intensity of participation in a project (see Section 2.2.3), we chose a multiple membership model (Goldstein 2011: 255f), in which for each publication year and researcher, using dummy variables (0/1), we coded the projects in which the researcher participated in (see Table 2). In addition, a zero project was coded for the FTE of the researcher's work and publication activity outside of the research center. To include the FTEs for each project, the FTEs were entered into the design matrix $D[D_{kijc}]$ instead of the ones (dummy variable). From this, the following model components resulted for $k = 1$ to K CCES projects (Cafri, Hedeker and Aarons 2015: 409f):

$$\ln(y_{jic}) = \beta_{00} + \sum_{k=1}^K D_{kijc}u_{2k} \quad u_{21}, u_{22}, \dots, u_{2K} \sim N(0, \sigma_{u2}^2), \quad (3)$$

where u_{21}, \dots, u_{2k} are the project effects as random effects, σ_{u2}^2 the corresponding variance component, and D_{kijc} is the corresponding design matrix with the FTEs for each project per publication year and researcher.

- (5) *Macro impact—segmented regression*: To test whether there is a *macro impact* of participation in CCES on the bibliometric indicators, we computed a segmented regression, which is a commonly used statistical approach for analysis of interrupted time series (Sauter, Mutz and Munro 1999; Wagner et al. 2002; Ramsay et al. 2003). Three situations are differentiated (no CCES, CCES phase 1, and CCES phase 2) and coded using dummy variables (0/1). Phases 1 and 2 account for an average and an individual change over time (interrupted time series). These effects can be interpreted causally, because according to the potential outcome concept (Rubin 2005), both the expected value under control (before CCES) and the expected value under treatment (phases 1 and 2) are available for each researcher (Mutz and Daniel 2012a, 2012b). From the differences between these expected values, individual and average causal effects of CCES participation can be calculated while controlling for all individual factors. The model components can be formulated as follows (Wagner et al. 2002: 302f):

$$\begin{aligned} \ln(y_{jic}) &= \beta_{00} + \beta_3(t_{jic} > 2005) + \beta_4(t_{jic} > 2010) \\ &+ u_{3i}(t_{jic} > 2005) \quad u_{3i} \sim N(0, \sigma_{u3}^2), \end{aligned} \quad (4)$$

where β_3 is the average causal effect of phases 1 and 2 (>2005) and β_4 is the additional average causal effect of phase 2

compared with phase 1. The random effect u_{3i} denotes the ICE of the two CCES phases for researcher i with the corresponding variance component σ^2_{u3} . Additional ‘time after treatment effects’ can be considered by including $t_{jic-2005}$ or $t_{jic-2010}$ in the model (Wagner et al. 2002: 302).

From the five statistical model components, a hierarchically nested sequence of increasingly complex models can be generated that represent different model assumptions (e.g. cohort effect, kind of polynomial trend, and effects of covariates), whereby the model components ‘segmented regression’ (macro impact) and ‘multiple membership’ (micro impact), as different models of the treatment effect, are not combined.

The individual models are then compared using the Bayesian information criterion (BIC). The smaller the BIC, the better the model fits the data. Models and the associated model components are rejected and discarded if the model components do not improve the BIC. The statistical analyses were carried out with a procedure in SAS software (PROC GLIMMIX) using maximum likelihood estimation/Laplace approximation (SAS Institute Inc. 2014: 3052f).

In econometrics and sociology, longitudinal data are usually modeled by fixed effects regression focusing on average effects (Allison 2009). To consider this alternative modeling strategy, additionally, a *fixed effects segmented regression model* was estimated (assuming residuals are not auto correlated), which consisted of five components: the effect of being a CCES publication or not, a quadratic trend model, the effects of the two phases, a time-lagged outcome variable ($y_{(j-1)ic}$), and the overall fixed effect for each researcher a_i , which may correlate with the predictor variables (a major difference to the growth curve model).

5. Results

5.1 Model comparison

The different model assumptions were formulated as statistical models for the three dependent variables of research performance, which could then be estimated using the data. The models are hierarchically nested, that is, derived from other models that are shown in Table 3 in the column labeled ‘base’ (e.g. M_3 from M_2). Instead of showing the model parameters of each model, the previous models are evaluated comparatively. We used the BIC as a relative measure for the model comparison. The model comparison allowed us to identify the crucial models and thus to rule out more complex model assumptions (e.g. cohort effects and time after treatment effect) at this stage.

The effect sizes were expressed in absolute units (e.g. number of publications). Effect size in terms of proportion of explained variance is only relevant for models that include predictors. For count regression data, several R-squared measures have been proposed (Cameron and Windmeijer 1996; Heinzl and Mittlböck 2003).

In addition to the null model (M_0), which contains only one random intercept for each researcher (i.e. u_{00ic} , Eq. 2), we tested whether the impact on the bibliometric indicators differed between CCES publications and non-CCES publications (M_1). Growth curve models are a class of models that describe the individual development of researchers over time, depicted in an individual linear regression with time as a predictor (M_2). A cohort effect model also includes the possible effect of age cohorts (M_3). With a polynomial time trend (M_4), the linear time trend is abandoned in favor of a quadratic polynomial ($y = b_0 + b_1x + b_2x^2$). With the next model

component, a causal impact model (M_5), we tested whether there were individual effects of the projects with inclusion of the FTEs on the bibliometric indicators (micro impact). Of central importance are the models M_6 (macro impact) and M_7 (time after treatment), in which the average causal effects were estimated using segmented regression. Model M_8 tested whether there were different ICEs for each researcher. Models M_9 – M_{15} provided indications concerning the effect of external variables on the individual growth process.

Regarding the variable ‘number of publications’ (scientific productivity), we found that in addition to the differentiation between CCES publications and non-CCES publications (M_1), the inclusion of the growth curve models (M_2 , M_3 , and M_4) led to great improvement of the BIC. For the variables ‘number of citations’ (scientific impact) and ‘number of coauthors’ (integration into the scientific community) the inclusion of further model components also led to the improvement of the BIC, but the improvement was comparatively small. As expected, across all three variables, the variability of the researchers in their individual trajectories played an important role. In addition, we found a causal effect of research center participation (M_5 , M_6 , and M_8). We also found a micro effect when including not only the FTEs, but also and especially an average causal effect of the research center (M_5) as well as ICEs (M_8). In contrast, in all three cases, the covariates did not lead to any appreciable improvement of the model. This also means that the person cluster (M_{10}) had no effect. The person cluster primarily differentiates between the 170 participating researchers and the 28 researchers (comparison group) who did not participate in CCES.

The growth curve model outperformed the fixed effects segmented regression (M_{16}) with respect to all outcome variables, also because the fixed effects regression considerably increases the number of estimated parameters (e.g. for each researcher), which in turn increases the BIC. For all outcome variables, statistically significant treatment effects can be found.

In sum, the model comparison shows that participation in the research center had a positive effect on all three dimensions of research performance—both overall (average causal effect) and regarding the individual development of a researcher (ICEs).

5.2 Model interpretation

In the following, we present the results of the parameter estimation for the models that were selected on the basis of the model comparison (M_0 , M_8). This is done in comparison with a basic or null model that allows only the intercept of the otherwise fixed polynomial regression model varying across researchers (Eq. 2). Overall, the models fit the data well. The Pearson χ^2/df was close to 1.00. The Poisson distribution assumption was not violated. Each of the selected models represents one of the three dimensions of research performance.

5.2.1 Average and individual causal effects on ‘number of publications’ (scientific productivity)

The estimates for the segmented regression component in model M_8 indicate the average causal effect per researcher per year that participation in the different phases of the research center had on the researchers’ number of publications (Table 4). For phase 1, the effect was $\beta_5 = 0.15$ and for phase 2, it was $\beta_6 = 0.1$, which means that the two phases had a comparable effect. To obtain the effect of both phases together, we added the two for a combined value of 0.25. As considerably more non-CCES publications were available than

Table 3. Model comparison for the three dependent variables

		Number of publications/NB		Total number of citations/NB		Number of coauthors/NB	
No	Model description	Base	BIC	Base	BIC	Base	BIC
Basic models							
M ₀	Intercept—only	—	17,687.9	—	30,438.9		22,281.1
M ₁	CCES/non-CCES	M ₀	17,096.0	M ₀	30,422.2	M ₀	22,190.3
Growth curve models							
M ₂	Linear growth curve model	M ₁	15,091.2	M ₁	30,227.6	M ₀	21,668.3
M ₃	Cohort effect	M ₂	15,043.3	M ₂	30,136.5	M ₂	21,665.0
M ₄	Polynomial time trend	M ₃	14,972.5	M ₃	29,680.3	M ₂	21,662.6
Causal impact models							
M ₅	Micro impact	M ₄	14,927.8	M ₄	29,591.0	M ₄	21,612.8
M ₆	Macro impact: ACEs	M ₄	14,969.0	M ₄	29,629.6	M ₄	21,667.6
M ₇	Time after treatment	M ₆	14,973.7	M ₆	29,514.8	M ₆	21,674.2
M ₈	Macro impact: ICEs	M ₆	14,897.5	M ₇	29,505.5	M ₆	21,633.9
Individual CCES trajectories							
M ₉	Individual CCES trajectories	M ₈	14,905.1	M ₈	29,508.7	M ₈	21,618.9
Predictors of growth							
M ₁₀	Person cluster	M ₈	14,931.6	M ₈	29,546.4	M ₈	21,685.5
M ₁₁	Gender	M ₈	^a	M ₈	29,497.6	M ₈	21,646.0
M ₁₂	Institution budget	M ₈	14,905.0	M ₈	29,521.0	M ₈	21,655.3
M ₁₃	Institution: ETH or not	M ₈	14,909.0	M ₈	29,524.0	M ₈	21,658.7
M ₁₄	Number of researchers (FTE)	M ₈	14,910.4	M ₈	29,492.0	M ₈	21,647.6
M ₁₅	Covariates (e.g. year of PhD)	M ₈	14,913.5	M ₈	29,531.4	M ₈	21,693.9
M ₁₆	Fixed effects segmented regression	M ₆	15,366.1	M ₆	124,160.8	M ₆	37,644.59

Notes: ICEs, individual causal effects; ACEs, average causal effects.

^aModel estimation was not plausible (missing parameter values, zero standard errors, ...).

CESS publications, the overall model estimation was based on the total number of publications because model estimation and testing is more efficient for large sample sizes than for small ones. Therefore, the specific effect ‘CCES versus non-CCES’ publications were not tested directly, but all publications (CCES + non-CCES) were compared non-CCES publications. Expressed in the form of publications per year, for $e^{\beta_0 + \beta_1 + \beta_5 + \beta_6} = e^{1.28 + 0.26 + 0.25}$ we had a value of 5.99 publications (CCES and non-CCES publications) compared with the phase before CCES participation, where the number of publications was $e^{\beta_0 + \beta_1} e^{1.28 + 0.26} = 4.66$. This means that CCES participation had an annual effect per researcher of approximately 1 1/3 more publications, when holding all other factors (e.g. time course) constant.

Likewise telling is the growth curve model that described the individual trajectory of a researcher. With the parameters (β_0 , β_2 , β_3 , and β_4) there was nonlinear weakened growth with negative quadratic (β_3) and cubic components (β_4) in addition to the linear component (β_2 ; Figure 1). With the CCES publication effect (CCES-Pub, β_1) we are able to compare the scientific productivity in the context of CCES to the scientific productivity beyond CCES: whereas on average $e^{\beta_0 + \beta_1} - e^{\beta_0} = e^{1.28 + 0.26} - e^{1.28} = 1.06$ annual publications were generated per researcher in the context of CCES, 3.60 ($e^{\beta_0} = e^{1.28}$) papers were published outside of CCES (non-CCES publications). Somewhat less than one-fourth of all annual publications of a researcher were thus published in the context of CCES.

In the random effects model, the individual trajectory of a researcher's publication activity, irrespective of any effect from participation in CCES, can be seen clearly in different cohorts (Table 4). The time course is cubic overall (Figure 1). Only the linear component of the trajectory, which is made up of an intercept and a slope of publication year (‘pubyear’), varied across individuals, as

well as the slope of phase 1, which represents the individual bibliometric impact of CCES. To interpret that trajectory, we can use the variance and covariance components (e.g. $\sigma_{001(2)}$, $\sigma_{201(2)}$) and correlation coefficients (e.g. $\rho_{011(2)}$) that correspond with the ‘random effects’: There were differences in the intercepts and slope of ‘pub-year’, which means that researchers’ publication careers began in very different ways, with different increases over time (slope). It is interesting that there is a high positive correlation between the individual intercept and the individual slope of a researcher, $\rho_{011(2)} = 0.70$, that is, a high number of publications at the start of CCES in 2006 (and, eventually, the start of his or her career in general) is associated with a strong increase in the number of publications in the following years, and vice versa. However, this is modified when looking at the cohorts for, which a negative relationship between intercept and slope was found ($\rho_{012} = -0.88$). In other words, the higher the average number of publications at the start of CCES in an age cohort (or the start of the age cohort in general, e.g. in the year 1999), the less steep the growth curve of this cohort and vice versa. This is a ‘ceiling effect’: For a cohort with a high level in scientific productivity in 2006, there is not much room left to increase their publication level in comparison to a cohort with a low level of scientific productivity in 2006.

Of particular importance is the statistically significant variance component of phase 1, $\sigma_{221(2)}$ of 0.20, which indicated that participation in CCES also had ICEs on a researcher's publication activity. In other words, 95% of the ICEs lie within a confidence interval of $\pm 1.96 \sqrt{0.20} = \pm 0.877$ around the average causal effect of CCES participation, $\beta_5 = 0.15$, in phase 1. Expressed in units of publications, the ICEs for researchers varies between ($e^{\beta_0 + \beta_1 + \beta_5 - 0.877} - e^{\beta_0 + \beta_1} =$) -2.43 and ($e^{\beta_0 + \beta_1 + \beta_5 + 0.877} - e^{\beta_0 + \beta_1} =$) 8.36 publications. In other words, participation in CCES (despite the positive average

Table 4. Results for selected models for ‘number of publications’ (logarithmic transformed)

Predictor	Parm	Models			
		M ₀		M ₈	
		Estim	SE	Estim	SE
Fixed effects					
Intercept	β_0	1.29*	0.04	1.28*	0.09
CCES-Pub (=yes)	β_1			0.26*	0.02
Pubyear	β_2			0.41*	0.08
Pubyear**2	β_3			−0.25*	0.04
Pubyear**3	β_4			−0.05*	0.02
Segmented regression					
Phase 1 (=yes)	β_5			0.15*	0.06
Phase 2 (=yes)	β_6			0.10*	0.04
Random effects					
Level 1: Researcher					
Intercept	$\sigma^2_{001(2)}$	0.31**	0.03	0.25**	0.04
Intercept-pubyear	$\sigma_{011(2)}$			0.15**	0.03
	$\rho_{011(2)}$			0.70	
Pubyear	$\sigma^2_{111(2)}$			0.18**	0.04
Phase 1-Intercept	$\sigma_{201(2)}$			−0.09**	0.03
	$\rho_{201(2)}$			−0.40	
Phase 1-pubyear	$\sigma_{211(2)}$			−0.08**	0.03
	$\rho_{211(2)}$			−0.40	
Phase 1	$\sigma^2_{221(2)}$			0.20**	0.04
Level 2: Cohort					
Intercept	σ^2_{002}			0.15**	0.06
Intercept-pubyear	σ_{012}			−0.09**	0.03
	ρ_{012}			−0.88	
Pubyear	σ^2_{112}			0.06**	0.03
Pearson χ^2/df		1.79		0.85	
−2LogLikelihood		17,677.4		14,841.6	
BIC		17,687.9		14,897.5	

*P < 0.05 (t-value, $df_0 = 3,413$, $df_8 = 3,408$, $df_{16} = 3,403$),

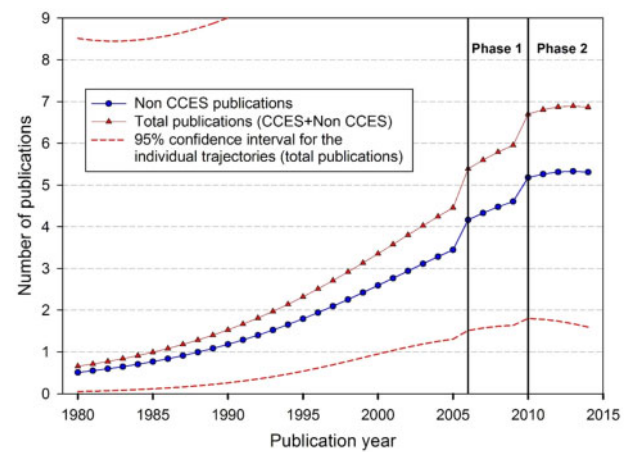
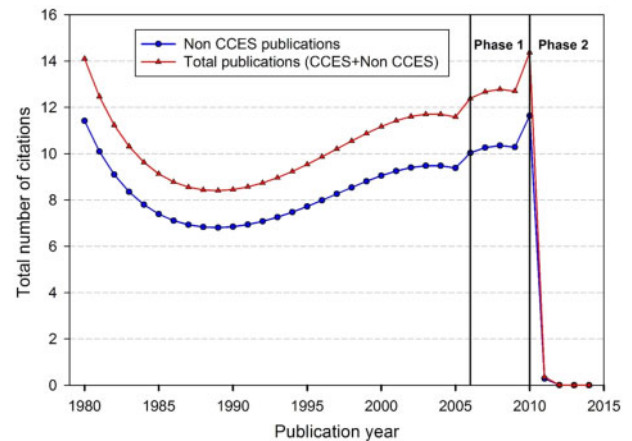
**P < 0.05 (z-test).

effect) can also have had, individually, a negative effect on the number of publications.

As described above, the evaluation approach allows us to examine not only the macro impact but also the micro impact. The micro impact is the effect of the individual CCES project on a researcher's publication activity compared with the researcher's publication activity outside CCES (non-CCES project). Here we took into account the aspect of intensity, assessed in FTEs. This finds expression in model M₅ (micro impact), which also did well in the model comparison. Instead of a complete overview of the parameter estimates, however, we report only the crucial variance component, σ_p^2 that described the variability of these project effects: This amounted to $\sigma_p^2 = 0.12$. Expressed as micro impacts, the project effects varied in number of publications (CCES and non-CCES) per year and researcher from ($e^{\beta_0 + \beta_1 - 1.96 \cdot 0.12} - e^{\beta_0 + \beta_1} = -2.12$ publications to ($e^{\beta_0 + \beta_1 + 1.96 \cdot 0.12} - e^{\beta_0 + \beta_1} = 4.19$ publications).

5.2.2 Average and individual causal effects on ‘number of citations’ (scientific impact)

The variable ‘number of citations’ per researcher and year showed a striking cubic curve over time (Figure 2). On average, the citations decreased in the 1990s, which can also be attributed to different starting time points of publication activity, and then rose again up

**Figure 1.** Predicted mean growth curve for number of publications (cumulative).**Figure 2.** Predicted mean growth curve for number of citations (5-year citation window and cumulative).

to 2010, with a dramatic decline after 2010, which is reflected in the negative sign of the regression coefficient (β_2 , β_3 , and β_4 ; Table 5). This decline occurs due to the citation window of 5 years. More recent publications simply have a lower probability of being cited than older publications.

Regarding the model estimations (M₈), we found an average effect per researcher and year that participation in the different phases of the research center had on the number of received citations (Table 5). For phase 1, the effect was 0.09 (β_5) and for phase 2, it was 0.15 (β_6). Expressed in the form of citations per year, for $e^{\beta_0 + \beta_1 + \beta_5 + \beta_6} = e^{2.22 + 0.21 + 0.09 + 0.15}$ there was a value of 14.44 citations compared with the phase before CCES with a number of citations of $e^{2.22 + 0.21} = 11.35$; this means that CCES had an annual effect per researcher of approximately 3.09 more citations, when holding all other factors constant.

Due to the problem of the citation windows, the ‘time after effects’ ($\beta_7 = 0.57$, $\beta_8 = -3.66$, $\beta_{16} = -1.20$, and $\beta_{17} = 0.70$) are not interpreted.

Also regarding this second dimension of research performance, we found individual trajectories, represented by the random effects ‘intercept’ and ‘pubyear’ and the corresponding variance components ($\sigma^2_{001(2)}$ and $\sigma^2_{111(2)}$). Individual trajectories varied strongly,

Table 5. Results for selected models for 'number of citations' (logarithmic transformed)

Predictor	Parm	Models			
		M ₀		M ₈	
		Estim	SE	Estim	SE
Fixed effects					
Intercept	β_0	2.12*	0.04	2.22*	0.08
CCES-Pub (=yes)	β_1			0.21*	0.04
Pubyear	β_2			-0.28	0.27
Pubyear**2	β_3			-0.63*	0.26
Pubyear**3	β_4			-0.21*	0.08
Segmented regression					
Phase 1 (=yes)	β_5			0.09	0.07
Phase 2 (=yes)	β_6			0.15*	0.06
Time after-phase 1	β_7			0.57	0.41
Time after-phase 2	β_8			-3.66*	0.40
Random effects					
Level 1: Researcher					
Intercept	$\sigma^2_{001(2)}$	0.34**	0.04	0.30**	0.04
Pubyear	$\sigma^2_{111(2)}$			0.28**	0.07
Phase 1	$\sigma^2_{221(2)}$			0.10**	0.03
Level 2: Cohort					
Intercept	σ^2_{002}			0.01	0.02
Pubyear	σ^2_{112}			0.08	0.08
Scale	α	0.68**	0.02	0.45**	0.01
Pearson χ^2/df		1.1		1.09	
-2LogLikelihood		30,423.1		29,453.0	
BIC		30,438.9		29,505.5	

*P < 0.05 (t-value, df₀ = 3,471, df₈ = 3,463, df₁₆ = 3,454),

**P < 0.05 (z-test).

also within the cohorts (σ^2_{002} and σ^2_{112}), which are not shown in Figure 2. Of particular interest were the ICEs of researchers, which were described with the variance component of phase 1, $\sigma^2_{221(2)} = 0.10$. The ICEs of CCES participation (compared with the time before CCES) were thus in an interval (with a probability of 0.95) from ($e^{\beta_0 + \beta_1 + \beta_3 - 1.96 \sqrt{0.10}} - e^{\beta_0 + \beta_1} = -4.67$ to ($e^{\beta_0 + \beta_1 + \beta_3 + 1.96 \sqrt{0.10}} - e^{\beta_0 + \beta_1} = 11.74$ citations per researcher and year.

A scale parameter of $\alpha = 0.45$ indicated that a negative binomial distribution, which came from overdispersion in the count data, fit the data better than a Poisson model with α restricted to 0.

For the additional model for the effects of CCES projects (M₅; micro impact), we found a variance component parameter for the projects of $\sigma^2_p = 1.90$. Expressed as number of citations for all publications (CCES and non-CCES) per researcher and year, the project effects varied from ($e^{\beta_0 + \beta_1 - 1.96 \sqrt{1.9}} - e^{\beta_0 + \beta_1} = -10.49$ citations to ($e^{\beta_0 + \beta_1 + 1.96 \sqrt{1.9}} - e^{\beta_0 + \beta_1} = 156.80$ citations.

5.2.3 Average and individual causal effects on 'number of coauthors' (integration into the scientific community)

Regarding the model estimations (M₈), we found an average effect per researcher and year that participation in the different phases of the research center had on number of coauthors (Table 6). For phase 1, the effect was 0.08 (β_5) and for phase 2, it was 0.07 (β_6). Expressed in the form of number of coauthors per researcher and per year, for $e^{1.26 + 0.12 + 0.08 + 0.07}$ there was a value of 4.6 coauthors compared with the phase before CCES with a number of publications (CCES and

Table 6. Results for selected models for 'number of coauthors' (logarithmic-transformed)

Predictor	Models				
	M ₀			M ₈	
	Parm	Estim	SE	Estim	SE
Fixed effects					
Intercept	β_0	1.35*	0.03	1.26*	0.04
CCES-Pub (=yes)	β_1			0.12*	0.03
Pubyear	β_2			0.21*	0.05
Pubyear**2	β_3			−0.15*	0.04
Pubyear**3	β_4			−0.03	0.02
Segmented regression					
Phase 1 (=yes)	β_5			0.08	0.05
Phase 2 (=yes)	β_6			0.07*	0.03
Random effects					
Level 1: Researcher					
Intercept	σ^2_{00}	0.13**	0.02	0.20**	0.03
Intercept-pubyear	σ_{01}			0.06**	0.02
	ρ_{01}			0.43	
Pubyear	σ^2_{11}			0.10**	0.02
Phase 1-Intercept	σ_{20}			−0.10**	0.03
	ρ_{20}			−0.57	
Phase 1-pubyear	σ_{21}			−0.07**	0.03
	ρ_{21}			−0.61	
Phase 1	σ^2_{22}			0.14**	0.03
Scale	α	0.18**	0.006	0.12**	0.005
Pearson χ^2/df		1.2	1.1		
−2LogLikelihood		22,265.3		21,560.2	
BIC		22,281.1		21,633.9	

*P < 0.05 (t-value, df₀ = 3, 414, df₈ = 3,465, df₁₆ = 3,458),

**P < 0.05 (z-test).

non-CCES publications) of $e^{1.26 + 0.12} = 3.97$, when holding all other factors constant. The time course of the number of coauthors was similar to that of the variable 'number of publications' (see Figure 3).

We again found strong individual differences between the researchers, which were also expressed in the variance/covariance components (σ^2_{00} , σ^2_{11} , and σ_{01}). Regarding the number of coauthors, the ICEs of CCES (compared with the time before CCES) were in an interval (with a probability of 0.95) from ($e^{\beta_0 + \beta_1 + \beta_3 - 1.96 \sqrt{0.14}} - e^{\beta_0 + \beta_1} = -1.91$ to ($e^{\beta_0 + \beta_1 + \beta_3 + 1.96 \sqrt{0.14}} - e^{\beta_0 + \beta_1} = 4.99$ coauthors per researcher per year ($\sigma^2_{22} = 0.14$).

For the additional model for the effects of CCES projects (M₅; micro impact), we found a variance component parameter for the projects of $\sigma^2_p = 0.08$, but it was not statistically significant ($z = 1.03$, $P > 0.05$). For this reason, single project effects are not interpreted.

6. Discussion

The global emergence of research centers has challenged traditional evaluation approaches as they are widely used to assess universities, departments, or individual researchers. Building on existing approaches, we introduced with this study a theoretically and methodologically refined approach for the ex post evaluation of research centers. The demonstration of the approach highlighted not only its major strengths but also a few limitations. Beyond the theoretical

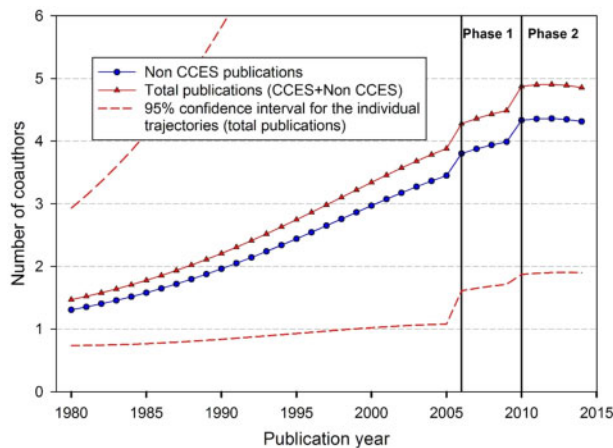


Figure 3. Predicted mean growth curve for number of coauthors (cumulative).

and methodological contributions, the concrete results of the evaluation have implications for research policy.

6.1 Strengths of the evaluation approach

One strength of the approach is its theoretical foundation, with the STHC model providing the central line of argumentation. From there, three characteristics of research centers and their participants were identified as major challenges to existing evaluation approaches: (1) the diversity of the participants ('diversity'), (2) at what moment in time the participants join and leave the research center ('transition'), and (3) the intensity of their participation ('intensity'). The evaluation approach introduced with this article addresses the three aspects and provides remedies by means of fine-grained data, the underlying research design, and an advanced statistical approach.

The data capture the 'diversity' of the participants through various covariates, including gender, scientific background, and academic age (year of PhD). Another data-related issue that the evaluation approach accounts for is the information on the researcher's affiliation with projects and phases of CCES, as retrieved from the archival data, thereby addressing the challenge of transition. The intensity of the researcher's participation is captured by the data on the FTE they spent at the research center per year. Another data-related strength of the evaluation approach is the reliance on archival and retrospectively collected bibliometric data, which safeguards the objectivity of the evaluation approach.

The quasi-experimental research design (*longitudinal interrupted accelerated design*) is central to the evaluation approach and primarily addresses the challenge of transition. It assumes that the affiliation with the research center interrupts the individual time series of publication trajectories in a way that could not be predicted based on the previous development of the bibliometric indicators, which is interpreted as a 'treatment' effect. The research design, then, is quite robust, as it withstands the major threats to internal and external validity, as described above (see Section 4.3). As a quasi-experimental within-group design, moreover, it does not require a randomized control group in the classical experimental sense.

Last, the statistical approach addresses all three aspects, by including growth curve modeling, a cohort-sequential model, a multiple membership model, and two ways of treatment operationalization.

The statistical approach is quite comprehensive, as it not only allows the average causal effects to be assessed but also accounts for the ICEs, cohort effects, micro and macro effects of research center participation, as well as whether the effect on the research performance of the participant is restricted to the research center context or beyond. In particular, a great deal of value is added to the evaluation approach by the ability to identify the ICEs, as fixed effects models, conventionally applied, would fail to detect these.

6.2 Limitations of the evaluation approach

What is true for all longitudinal research designs is that the time horizon considered must cover a significant length. In the context of the evaluation approach proposed in this article, this implies that the assessment of the effect on research performance is constrained to more senior researchers with a 'long enough' academic career. Future research should indeed focus more on the career development of junior researchers to assess the capacity-building effect of research center participation (Corley, et al. 2017). Another crucial aspect for the evaluation approach is the availability of data. The data collection process required to apply the evaluation approach was rather time-consuming, as it entailed the coding from comprehensive archival data to a relational database to qualify for statistical analyses. Another, more critical limitation arises from the potential lags between the work on a publication and actual publication date as given in the annual reports. One solution could be to require reporting schemes to make such a differentiation. Overall, this article is conceived as giving an indication of how future reporting guidelines could be designed to facilitate the quantitative evaluation of research centers.

Another possible limitation of the study is the validity of the annual reports on which the study is based. It can be argued that the numbers, e.g. share of the total working time at the research center, respond more to bureaucratic rules than reflecting any realities of time allocation. However, this limitation does not necessarily apply to all data taken from the annual reports. The annual project reports had been prepared very meticulously as a base for the annual achievement report of the whole research center. For example, publications listed in the annual project reports were cross-checked by the research center management to avoid multiple mentions, thereby increasing the quality of the data.

Last, we acknowledge that some authors call for a differentiated use of the bibliometric method for evaluative purposes. We would like to highlight that the evaluation approach we propose is only suitable for assessing the research performance of a research center. However, and needless to say, other alternative evaluation approaches would be required to capture societal impacts, economic impacts, or educational or capacity-building impacts (Lin and Bozeman 2006; Corley 2007; Youtie and Corley 2011; Bornmann 2013; Rivers and Gray 2013; Hicks et al. 2015; Husbands Fealing et al. 2018; Kassab, Schwarzenbach and Gotsch 2018; Kassab 2019).

6.3 Implications

As outlined in the introduction, researchers are somewhat critical of research centers (and inter- and transdisciplinary research, for that matter) in the face of a supposed career-relevant conflict of interest. The results of this study, however, provide evidence that this skepticism is unfounded. Quite strikingly, on average, participation in research centers entails no disadvantages for researchers as far their

overall research performance is concerned, as measured in scientific productivity, the citation impact of their output, and their integration within the scientific community. These findings confirm the results of several previous studies, and yet the results presented here can be traced back to a distinctly more accurate methodological basis. The implications of this study are good news for intrinsically motivated researchers as well as for research policymakers, and finally, they are also invaluable in helping to improve the image of research centers and of inter- and transdisciplinary research in general.

Notes

1. As has been described above, research centers pursue a variety of goals. In this article, we focus on the research aspect, which we understand in terms of research performance.
2. Between 2013 and 2015, OK worked as an executive assistant to the CCES management. Afterwards he joined the Professorship for Social Psychology and Research on Higher Education at ETH Zurich, where he conducted this study in collaboration with the co-authors R.M. and H.D.D.
3. Project partners are those researchers whose names were on the project proposals, and who headed a subunit of the project.

Acknowledgements

We would like to thank the management team of the Competence Center Environment and Sustainability (CCES) of the ETH Domain, in particular, Dr Nikolaus Gotsch and Prof. em. Dr René Schwarzenbach, for their support in conducting this study. In addition to the two anonymous reviewers, we also thank Adam B. Jaffe, Brandeis University, for his valuable inputs, as well as the participants of the Behavioral Studies Colloquium at ETH Zurich.

References

- Allison, P. D. (2009) *Fixed Effects Regression Models*. London: Sage.
- Ávila-Robinson, A., and Sengoku, S. (2017) 'Multilevel Exploration of the Realities of Interdisciplinary Research Centers for the Management of Knowledge Integration', *Technovation*, 62–63: 22–41.
- Bishop, P. R., Huck, S. W., Ownley, B. H., Richards, J. K., and Skolits, G. J. (2014) 'Impacts of an Interdisciplinary Research Center on Participant Publication and Collaboration Patterns: A Case Study of the National Institute for Mathematical and Biological Synthesis', *Research Evaluation*, 23: 327–40.
- Bliese, P. D., Adler, A. B., and Flynn, P. J. (2017) 'Transition Processes: A Review and Synthesis Integrating Methods and Theory', *Annual Review of Organizational Psychology and Organizational Behavior*, 4: 263–86.
- Boardman, C., and Bozeman, B. (2007) 'Role Strain in University Research Centers', *The Journal of Higher Education*, 78: 430–63.
- Boardman, C., and Corley, E. A. (2008) 'University Research Centers and the Composition of Research Collaborations', *Research Policy*, 37: 900–13.
- Bornmann, L. (2013) 'What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey', *Journal of the American Society for Information Science and Technology*, 64: 217–33.
- Bornmann, L., and Mutz, R. (2015) 'Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References', *Journal of the Association for Information Science and Technology*, 66: 2215–22.
- Bornmann, L., Mutz, R., and Daniel, H.-D. (2009) 'The Influence of the Applicants' Gender on the Modeling of a Peer Review Process by Using Latent Markov Models', *Scientometrics*, 81: 407–11.
- Bozeman, B., and Boardman, C. (2003) *Managing the New Multipurpose, Multidiscipline University Research Centers: Institutional Innovation in the Academic Community*, *Transforming Organizations Series*. Washington, DC: IBM Endowment for the Business of Government.
- Bozeman, B., and Corley, E. A. (2004) 'Scientists' Collaboration Strategies: Implications for Scientific and Technical Human Capital', *Research Policy*, 33: 599–616.
- Bozeman, B., and Rogers, J. D. (2002) 'A Churn Model of Scientific Knowledge Value: Internet Researchers as a Knowledge Value Collective', *Research Policy*, 31: 769–94.
- Bozeman, B., Dietz, J., and Gaughan, M. (2001) 'Scientific and Technical Human Capital: An Alternative Model for Research Evaluation', *International Journal of Technology Management*, 22: 716–40.
- Bunton, S. A., and Mallon, W. T. (2006) 'The Impact of Centers and Institutes on Faculty Life: Findings from a Study of Basic Science and Internal Medicine Faculty at Research-Intensive Medical Schools', *Academic Medicine*, 81: 734–43.
- Cafri, G., Hedeker, D., and Aarons, G. A. (2015) 'An Introduction and Integration of Cross-Classified, Multiple Membership, and Dynamic Group Random-Effects Models', *Psychological Methods*, 20: 407.
- Cameron, A. C. (1998) *Regression Analysis of Count Data*. Cambridge: CUP.
- Cameron, A. C., and Windmeijer, F. A. G. (1996) 'R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization', *Journal of Business & Economic Statistics*, 14: 209–20.
- Cañibano, C., and Bozeman, B. (2009) 'Curriculum Vitae Method in Science Policy and Research Evaluation: The State-of-the-Art', *Research Evaluation*, 18: 86–94.
- Christensen, L. B., and Waraczynski, M. A. (1988) *Experimental Methodology*. Boston: Allyn and Bacon.
- Clark, W. C. (2007) 'Sustainability Science: A Room of Its Own', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 1737–8.
- Corley, E. A. (2007) 'A Use-and-Transformation Model for Evaluating Public R&D: Illustrations from Polycystic Ovarian Syndrome (PCOS) Research', *Evaluation and Program Planning*, 30: 21–35.
- Corley, E. A., and Gaughan, M. (2005) 'Scientists' Participation in University Research Centers: What Are the Gender Differences?', *The Journal of Technology Transfer*, 30: 371–81.
- Corley, E. A., Bozeman, B., Zhang, X., and Tsai, C.-C. (2017) 'The Expanded Scientific and Technical Human Capital Model: The Addition of a Cultural Dimension', *The Journal of Technology Transfer*, 44: 681–99.
- Coryn, C. L. S., Hattie, J. A., Scriven, M., and Hartmann, D. J. (2007) 'Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison', *American Journal of Evaluation*, 28: 437–57.
- Cozzens, S. E., and Turpin, T. (2000) 'Processes and Mechanisms for Evaluating and Monitoring Research Outcomes from Higher Education: International Comparisons', *Research Evaluation*, 9: 3–4.
- Dietz, J. S., and Bozeman, B. (2005) 'Academic Careers, Patents, and Productivity: Industry Experience as Scientific and Technical Human Capital', *Research Policy*, 34: 349–67.
- Duncan, T. E., Duncan, S. C., and Strycker, L. A. (2013) *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Application*. New York: Taylor & Francis Group.
- Elkins, T., and Keller, R. T. (2003) 'Leadership in Research and Development Organizations: A Literature Review and Conceptual Framework', *The Leadership Quarterly*, 14: 587–606.
- Feller, I., Ailes, C. P., and Roessner, J. D. (2002) 'Impacts of Research Universities on Technological Innovation in Industry: Evidence from Engineering Research Centers', *Research Policy*, 31: 457–74.
- Ferguson, L. (2004) 'External Validity, Generalizability, and Knowledge Utilization', *Journal of Nursing Scholarship*, 36: 16–22.
- Fleiss, J. L., Levin, B., and Paik, M. C. (2003) *Statistical Methods for Rates and Proportions*, 3rd edn. Hoboken, New Jersey: John Wiley & Sons.
- Galbraith, S., Bowden, J., and Mander, A. (2017) 'Accelerated Longitudinal Designs: An Overview of Modelling, Power, Costs and Handling Missing Data', *Statistical Methods in Medical Research*, 26: 374–98.
- Gaughan, M., and Bozeman, B. (2002) 'Using Curriculum Vitae to Compare Some Impacts of NSF Research Grants with Research Center Funding', *Research Evaluation*, 11: 17–26.
- Gaughan, M., and Ponomarev, B. (2008) 'Faculty Publication Productivity, Collaboration, and Grants Velocity: Using Curricula Vitae to Compare Center-Affiliated and Unaffiliated Scientists', *Research Evaluation*, 17: 103–10.

- Geiger, R. L. (1990) 'Organized Research Units—Their Role in the Development of University Research', *The Journal of Higher Education*, 61: 1–19.
- Goldstein, H. (2011) *Multilevel Statistical Models*, 4th edn. Hoboken, New Jersey: John Wiley & Sons.
- Gray, B. (2008) 'Enhancing Transdisciplinary Research through Collaborative Leadership', *American Journal of Preventive Medicine*, 35 (Supplement): S124–32.
- Heinzl, H., and Mittlböck, M. (2003) 'Pseudo R-Squared Measures for Poisson Regression Models with Over-or Underdispersion', *Computational Statistics & Data Analysis*, 44: 253–71.
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., and Rafols, I. (2015) 'The Leiden Manifesto for Research Metrics', *Nature*, 520: 429.
- Hilbe, J. M. (2014) *Modeling Count Data*. Cambridge: CUP.
- Hox, J. J., Moerbeek, M., and van de Schoot, R. (2018) *Multilevel Analysis: Techniques and Applications*, 3rd edn. New York: Routledge.
- Husbands Fealing, K., Lane, J. I., King, J. L., and Johnson, S. R. (2018) *Measuring the Economic Value of Research: The Case of Food Safety*. Cambridge: CUP.
- Ikenberry, S. O., and Friedman, R. C. (1972) *Beyond Academic Departments: The Story of Institutes and Centers*. London: Jossey-Bass.
- Kassab, O. (2019) 'Does Public Outreach Impede Research Performance? Exploring the "Researcher's Dilemma" in a Sustainability Research Center', *Science and Public Policy*, 46: 710–20.
- Kassab, O., Schwarzenbach, R. P., and Gotsch, N. (2018) 'Assessing Ten Years of Inter- and Transdisciplinary Research, Education, and Outreach: The Competence Center Environment and Sustainability (CCES) of the ETH Domain', *GAIA - Ecological Perspectives for Science and Society*, 27: 226–34.
- Klaiber, V., Seeling, U., and Mutz, R. (2002) 'Veränderung der Holzfeuchte im Rundholz von Fichten (*Picea abies* (L.) Karst.) in Abhängigkeit von Art und Dauer der Lagerung—Abschätzung Mittels Einer Kohorten-Sequenzanalyse [Reduction of Moisture Content in Logs of Norway Spruce (*Picea abies* (L.) Karst.) as a Function of Type and Duration of the Storage—Estimation with a Cohort Sequential Analysis]', *Allgemeine Forst- und Jagdzeitung*, 173: 162–72.
- Kueffer, C. et al. (2012) 'Enabling Effective Problem-Oriented Research for Sustainable Development', *Ecology and Society*, 17: 8.
- Landry, R., and Amara, N. (1998) 'The Impact of Transaction Costs on the Institutional Structuration of Collaborative Academic Research', *Research Policy*, 27: 901–13.
- Lang, D. J. et al. (2012) 'Transdisciplinary Research in Sustainability Science: Practice, Principles, and Challenges', *Sustainability Science*, 7: 25–43.
- Lee, S., and Bozeman, B. (2005) 'The Impact of Research Collaboration on Scientific Productivity', *Social Studies of Science*, 35: 673–702.
- Lin, M. W., and Bozeman, B. (2006) 'Researchers' Industry Experience and Productivity in University-Industry Research Centers: A "Scientific and Technical Human Capital" Explanation', *The Journal of Technology Transfer*, 31: 269–90.
- Mallon, W. T. (2006) 'The Benefits and Challenges of Research Centers and Institutes in Academic Medicine: Findings from Six Universities and Their Medical Schools', *Academic Medicine*, 81: 502–12.
- McDowall, D., McCleary, R., Meidinger, E. E., and Hay, R. A. Jr. (1983) *Interrupted Time Series*. London: Sage.
- Mittelstrass, J. (2011) 'On Transdisciplinarity', *Trames*, 15: 329–38.
- Mutz, R., and Daniel, H.-D. (2012a) 'The Generalized Propensity Score Methodology for Estimating Unbiased Journal Impact Factors', *Scientometrics*, 92: 377–90.
- Mutz, R., and Daniel, H.-D. (2012b) 'Skewed Citation Distributions and Bias Factors: Solutions to Two Core Problems with the Journal Impact Factor', *Journal of Informetrics*, 6: 169–76.
- Mutz, R., Wolbring, T., and Daniel, H.-D. (2016) 'The Effect of the "Very Important Paper" (VIP) Designation in Angewandte Chemie International Edition on Citation Impact: A Propensity Score Matching Analysis', *Journal of the Association for Information Science and Technology*, 68: 2139–53.
- OECD (2007) *Revised Field of Science and Technology (FOS) Classification in the Frascati Manual, Working Party of National Experts on Science and Technology Indicators*. Paris: OECD.
- Ponomarev, B., and Boardman, C. (2010) 'Influencing Scientists' Collaboration and Productivity Patterns through New Institutions: University Research Centers and Scientific and Technical Human Capital', *Research Policy*, 39: 613–24.
- Ramsay, C. R., Matowe, L., Grilli, R., Grimshaw, J. M., and Thomas, R. E. (2003) 'Interrupted Time Series Designs in Health Technology Assessment: Lessons from Two Systematic Reviews of Behavior Change Strategies', *International Journal of Technology Assessment in Health Care*, 19: 613–23.
- Raudenbush, S. W., and Bryk, A. S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage.
- Rivers, D., and Gray, D. O. (2013) 'Evaluating Cooperative Research Centers: A Strategy for Assessing Proximal and Distal Outcomes and Associated Economic Impacts', in Link A. N. and Vonortas N. (eds) *Handbook on the Theory and Practice of Program Evaluation*, pp. 291–322. Cheltenham: Edward Elgar Publishing.
- Rogers, J. D., Youtie, J., and Kay, L. (2012) 'Program-Level Assessment of Research Centers: Contribution of Nanoscale Science and Engineering Centers to US Nanotechnology National Initiative Goals', *Research Evaluation*, 21: 368–80.
- Rubin, D. B. (2005) 'Causal Inference Using Potential Outcomes: Design, Modeling, Decisions', *Journal of the American Statistical Association*, 100: 322–31.
- Sabharwal, M., and Hu, Q. (2013) 'Participation in University-Based Research Centers: Is It Helping or Hurting Researchers?', *Research Policy*, 42: 1301–11.
- Santoro, M. D., and Gopalakrishnan, S. (2001) 'Relationship Dynamics between University Research Centers and Industrial Firms: Their Impact on Technology Transfer Activities', *The Journal of Technology Transfer*, 26: 163–71.
- SAS Institute Inc. (2014) *SAS/STAT® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.
- Sauter, U. H., Mutz, R., and Munro, B. D. (1999) 'Determining Juvenile-Mature Wood Transition in Scots Pine Using Latewood Density', *Wood and Fiber Science*, 31: 416–25.
- SDSN (2017), 'Getting Started with the SDGs in Universities: A Guide for Universities, Higher Education Institutions, and the Academic Sector. Australia, New Zealand and Pacific Edition', Sustainable Development Solutions Network—Australia/Pacific. Melbourne: SDSN. <http://ap-unsdsn.org/wp-content/uploads/University-SDG-Guide_web.pdf>
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Slaughter, S., Campbell, T., Holleman, M., and Morgan, E. (2002) 'The "Traffic" in Graduate Students: Graduate Students as Tokens of Exchange between Academe and Industry', *Science, Technology & Human Values*, 27: 282–312.
- Smith, A. M., Lai, S. Y., Bea-Taylor, J., Hill, R. B. M., and Kleinhenz, N. (2016) 'Collaboration and Change in the Research Networks of Five Energy Frontier Research Centers', *Research Evaluation*, 25: 472–85.
- Spangenberg, J. H. (2011) 'Sustainability Science: A Review, an Analysis and Some Empirical Lessons', *Environmental Conservation*, 38: 275–87.
- Stokols, D., Hall, K. L., Taylor, B. K., and Moser, R. P. (2008a) 'The Science of Team Science: Overview of the Field and Introduction to the Supplement', *American Journal of Preventive Medicine*, 35: 77–89.
- Stokols, D. et al. (2008b) 'The Ecology of Team Science: Understanding Contextual Influences on Transdisciplinary Collaboration', *American Journal of Preventive Medicine*, 35 (Supplement): S96–S115.
- Su, X. (2014) 'Academic Scientists' Affiliation with University Research Centers: Selection Dynamics', *Research Policy*, 43: 382–90.
- Talwar, S., Wiek, A., and Robinson, J. (2011) 'User Engagement in Sustainability Research', *Science and Public Policy*, 38: 379–90.

- The Madrillon Group Inc. (2010) *Evaluation of Research Center and Network Programs at the National Institutes of Health—A Review of Evaluation Practice, 1978-2009*. Bethesda: The Madrillon Group Inc.
- Wagner, A. K., Soumerai, S. B., Zhang, F., and Ross-Degnan, D. (2002) 'Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research', *Journal of Clinical Pharmacy and Therapeutics*, 27: 299–309.
- Way, S. F. et al. (2017) 'The Misleading Narrative of the Canonical Faculty Productivity Trajectory', *Proceedings of the National Academy of Sciences of the United States of America*, 114: E9216–23.
- Wen, J., and Kobayashi, S. (2001) 'Exploring Collaborative R&D Network: Some New Evidence in Japan', *Research Policy*, 30: 1309–19.
- West, S. G., Cham, H., and Liu, Y. (2014) 'Causal Inference and Generalization in Field Settings: Experimental and Quasi-Experimental Designs', in H. T. Reis and C. M. Judd (eds) *Handbook of Research Methods in Social and Personality Psychology*, pp. 49–80. Cambridge: CUP.
- Wiek, A. et al. (2014) 'Toward a Methodological Scheme for Capturing Societal Effects of Participatory Sustainability Research', *Research Evaluation*, 23: 117–32.
- Willett, J. B., Singer, J. D., and Martin, N. C. (1998) 'The Design and Analysis of Longitudinal Studies of Development and Psychopathology in Context: Statistical Models and Methodological Recommendations', *Development and Psychopathology*, 10: 395–426.
- Youtie, J., and Corley, E. A. (2011) 'Federally Sponsored Multidisciplinary Research Centers: Learning, Evaluation, and Vicious Circles', *Evaluation and Program Planning*, 34: 13–20.
- Youtie, J., Kay, L., and Melkers, J. (2013) 'Bibliographic Coupling and Network Analysis to Assess Knowledge Coalescence in a Research Center Environment', *Research Evaluation*, 22: 145–56.
- Ziegler, R., and Ott, K. (2011) 'The Quality of Sustainability Science: A Philosophical Perspective', *Sustainability: Science, Practice, & Policy*, 7: 31–44.