

Professor Marek Kwiek
 Institute for Advanced Studies in Social Sciences and Humanities, Director
 UNESCO Chair in Institutional Research and Higher Education Policy
 Adam Mickiewicz University of Poznan, Poland
kwiekm@amu.edu.pl
 ORCID: orcid.org/0000-0001-7953-1063

Dr. Wojciech Roszka
 Poznan University of Economics and Business, Poznan, Poland
wojciech.roszka@ue.poznan.pl
 ORCID: orcid.org/0000-0003-4383-3259

Academic vs. Biological Age in Research on Academic Careers: A Large-scale Study with Implications for Scientifically Developing Systems

Abstract

Biological age is an important sociodemographic factor in studies on academic careers (research productivity, scholarly impact, and collaboration patterns). It is assumed that the academic age, or the time elapsed from the first publication, is a good proxy for biological age. In this study, we analyze the limitations of the proxy in academic career studies, using as an example the entire population of Polish academic scientists visible in the last decade in global science and holding at least a PhD ($N = 20,569$). The proxy works well for science, technology, engineering, mathematics, and medicine (STEMM) disciplines; however, for non-STEMM disciplines (particularly for humanities and social sciences), it has a dramatically worse performance. This negative conclusion is particularly important for systems that have only become recently visible in global academic journals. The micro-level data suggest a delayed participation of social scientists and humanists in global science networks, with practical implications for predicting biological age from academic age. We calculate correlation coefficients, present contingency analysis of academic career stages with academic positions and age groups, and create a linear multivariate regression model. Our research suggests that in scientifically developing countries, academic age as a proxy for biological age must be used more cautiously than in advanced countries: ideally, it must be used only for STEMM disciplines.

1. Introduction

In the previous decade, research on academic careers conducted at the micro-level of the individual scientist, rather than at the aggregate country level, has increasingly relied on the category of academic age (Gingras et al., 2008; Aksnes et al., 2011; Abramo et al., 2016). The reason for this is simple: the data on biological age is generally unavailable for large-

scale studies. Age is one of the most important variables related to numerous dimensions of academic careers, such as research productivity, international mobility, or international research collaboration (Levin & Stephan, 1991; Levin & Stephan, 1989). Therefore, it has often been assumed, somewhat out of necessity, that the academic age (or the time elapsed from the first publication) is a good proxy for the biological age (Robinson-Garcia et al., 2020; Milojević, 2012; Nane et al., 2017). The date of first publication in Scopus or Web of Science can be calculated and used for research purposes for all publishing scientists at the level of institutions or their clusters, cities, regions, disciplines, journals, and countries.

Access to data on the date of first publication on a large scale (such as, for example, for a whole country) requires rather complex preparation, particularly including a list of individual identifiers of all authors, but it is technically possible. On the other hand, access to data on the biological age of researchers, given the current level of development of large administrative databases encompassing all researchers within a country and access to them for scientific purposes, in practice does not exist (with a few exceptions, including Italy and Norway) (see Rørstad & Aksnes 2015; Abramo et al., 2011). These also include Poland and a database we have created (“Polish Science Observatory”) based on the integration of an administrative register of all scientists working in public science sectors (approximately 100,000 scientists)—maintained by a national institution created for this purpose (OPI-PIB)—with the database of all Polish publications indexed in Scopus in the previous decade (approximately 380,000). Our Observatory database has been used several times in our extensive studies of academic faculty (Kwiek & Roszka 2020, 2021a). The integration of both databases was performed using probabilistic and deterministic methods and we have described its construction in detail elsewhere (see Kwiek & Roszka 2021a, pp. 4-6).

In this paper, our goal is to analyze the limitations of using academic age as a proxy for biological age in the example of the entire national science system, for which we consider both biological age and academic age for all researchers affiliated to Polish institutions of higher education, holding at least a PhD degree, and participating in global academic science via global publishing. An approximation of functioning in global science is having at least one publication indexed in the Scopus database in the decade 2009–2018.

Since we have an unambiguously defined biological age from the national registry of scientists for each scientist and we are able to determine the academic age for all of them, we can systematically show the differences between the two types of age depending on the selected key parameters.

One of the limitations of our study is the absence of a few academic researchers in the Scopus database—the lack of indexed publications attributed to them is related to the delayed entry of the Polish academic system into the global scientific circulation. However, this feature is common to many countries that are scientifically catching-up, in this sense, the empirical part of the paper is relevant to a large part of the world undergoing similar changes in publication patterns over time. Global science is increasingly characterized by new actors not only at the level of institutions and countries, but also at the level of individual scientists. Science is increasingly becoming a global system that comprises both advanced and less developed countries, an increasing number of institutions, and an increasing number of authors (Wagner, 2019; Wang & Barabási, 2021; Olechnicka et al., 2019). The globalization of science implies a growing number of countries and individual scientists participating in international research collaboration and the ties between countries and individual scientists

being much closer than before (Kwiek 2020a), thereby leading to decentralization (Gui et al., 2019) or pluralization (Marginson, 2018) of science. The traditional Anglo-American academic hegemony is being challenged by new entrants (Marginson and Xu, 2021), countries and individual scientists alike, in an increasing number of academic fields.

In other words, in this study, we determine the academic age for each Polish scientist based on the date of their first publication and systemically analyze the differences between academic age and their actual biological age. In addition, we also show how the individually determined academic age differs in practice from the strictly determined biological age depending on independent variables such as gender, scientific discipline, academic position, and institutional type.

Thus, using comprehensive data from an entire national system comprising 100,000 researchers and their 380,000 publications as an example, we estimate the scope of limitations of the use of academic age as a proxy for biological age depending on the selected independent variables and analyze both the practical and methodological implications of using academic age in academic career research.

As our study reveals, these limitations turn out to be greatest for the humanities and social sciences: for these disciplines, the level of correlation of both types of age turn out to be low. Therefore, in the analyzed Polish case, it is not possible to practically derive biological age from academic age in case of scientists publishing in these disciplines, as the correlation between them varies between 35% and 50% (correlation coefficient ranges from 0.35 for arts and humanities and 0.43 for economics, econometrics, and finance to 0.49 for psychology). In contrast, for science disciplines, the correlation (and determination) is dramatically higher, and using academic age as a proxy for biological age reduces the risk (correlation coefficient ranges from 0.82 for biochemistry, genetics, and molecular biology to 0.89 for chemistry and 0.90 for immunology and microbiology). Furthermore, correlations and determinations take different values for different individual and institutional parameters, which we explore in greater detail below.

Importantly, our study was conducted for a system of science that is in the pursuit of massive global science participation and has only relatively recently adopted global publication patterns. Poland's contribution to global science is significant, particularly given the relatively low investment in academic science and the inconsistent and outdated system of promotion in science. Systems of promotion strongly affect publishing patterns (Stephan, 2012) and in Poland it is based on a complicated ladder of successive academic degrees, from doctorate to habilitation to full professorship, accompanied by a parallel, complicated system of academic positions (Kwiek & Szadkowski, 2017). Poland ranks seventeenth in the Scopus database in terms of the number of scientific articles in 2020 (43,618 articles), compared to 12,868 articles in 2000 and 22,362 in 2010 (SciVal, 2021). The presence of Polish science in the global system of science is steadily increasing in terms of numbers and percentages, but it is strongly diversified from a disciplinary perspective. While the global visibility of STEMM sciences—particularly chemistry and physics—is stable and solid, the presence of humanities, social sciences and economics in global datasets of indexed journals is relatively small and has been observed only recently. In particular, the number of publications in highly prestigious journals (top 10% and top 1% of journals) has been growing more strongly in relatively recent times, which is certainly the effect of a full decade of reforms of science and higher education systems (Antonowicz et al., 2020).

In this paper, academic age is defined as the time that has elapsed between the date of the first publication indexed in the Scopus database (all publication types, any role played in publication) and 2017, our reference year. It can be very roughly assumed that in the sciences, the first publication in the database most often appears during the last years of writing a PhD dissertation—that is, in the Polish case, at approximately 30 years of age. Thus, in a large simplification, an academic age of 0 may indicate 30 years of age and academic age of 40 may indicate 70 years of age. In this sense, in a very approximate sense, an academic career would encompass a maximum of 40 years of life: academic age range 0–40 years and biological age range 30–70 years. Such a view of an academic career corresponds to the Polish model, in which there is an extremely strong trend of academic inbreeding—that is, working at the same university from the time of obtaining a doctoral degree, and often graduation; remaining in the system of science for several decades; and a marginal level of admission to academic work at a later age, including new admissions, as part of the transition from the economic or public administration sector to the academic sector. The career model includes entering the workforce at around age of 30 years and working in the sector or leaving academia with no option to return. However, most importantly a long career in the academic sector does not necessarily imply a long period of publishing articles indexed in the Scopus database (particularly in non-STEMM fields), which distinguishes the Polish system from systems that have been present longer and more intensively in the global circuit of indexed science.

In the first part of the Results section, we present a detailed estimation of the scale of inconsistencies between the two types of age in different cross-sections and propose broader generalizations. In the second part of the Results section, using a linear regression model, we estimate biological age on the basis of academic age and other selected parameters directly (institutional affiliation) or indirectly available in the Scopus database (such as the ASJC discipline dominating in a scientist's 10-year publication portfolio; gender determined by first and subsequent names and possibly by external databases and special gender determination algorithms). Further, we explore the utility of the model by inquiring the extent to which we can simulate biological age based on publication metadata available in global databases.

2. Key Literature

2.1. Biological Age in Academic Career Research

Biological age has been an important sociodemographic factor in sociological studies of academic careers (especially of research productivity, scholarly impact, and collaboration patterns) for more than half a century (Lehman, 1953; Cole, 1979; Stern, 1978; Stephan & Levin, 1992; Kyvik, 1990; Kyvik & Olsen, 2008).

First, more historically, productivity studies used biological age as an independent variable: sociologists were interested when productivity peaks and significant contributions to science occurred in scientists' lifecycles, thereby revealing that the peaks occurred in the late 30s and early 40s (Lehman, 1953, pp. 3–22, based on a skewed sample of the authors of great discoveries). Performance was found to peak at mid-career and then dropped, except for inner-motivated scientists (Pelz & Andrews, 1976, pp. 174–199). Zuckerman and Merton

(1973) studied age stratification and the visibility of knowledge, dysfunctions of gerontocracy in science, which included the late age of peer reviewers, gatekeeping mechanisms in academic publishing, and rank-stratified collaboration rates. Stern (1978) studied the age and achievement of randomly selected mathematicians compared with mathematician members of the American Academy of Sciences. Levin and Stephan (1991, p. 126) found evidence that, on average, scientists become less productive as they age. In his research on a cohort of mathematicians, Cole (1979) showed that there was no decline in the quality of work produced by these mathematicians as they progressed through their careers (specifically, there was no decline for those who continued publishing throughout their careers as a “residue” that comprised the best members of their cohort).

Kyvik examined a sample of Norwegian scientists (1990, p. 38) and has shown a curvilinear relationship between age and productivity, with a productivity peak in the scientists’ 40s, followed by productivity decline. Six different hypotheses were suggested (Kyvik & Olsen, 2008, pp. 441–442) to explain reduced research performance of universities with an aging academic staff, related to either a decline in productivity (the utility maximizing hypothesis, the seniority burden hypothesis, and the cumulative disadvantage hypothesis) or to a decline in creativity of research (the age decrement hypothesis, the obsolescence hypothesis, and the intellectual deadlock hypothesis). Kyvik and Olsen used three survey data sets on Norwegian professors divided into age cohorts from three decades (1982, 1992, and 2001, with a sample of 5,367 scientists) and show that the decline in productivity with increasing age may be due to generation effects rather than aging effects.

Further, a few cohorts may be more research productive than others due to different competition levels in hiring in their early years and also due to different research funding opportunities. Scientists hired under different conditions may stay on in academia for decades. As Stephan and Levin (1992, p. 117) noted, “the relative attractiveness of careers in science changes over time,” bringing young people with different talents, motivations, and abilities to academia (Levin and Stephan 1991, p. 126). Academic cohorts may be more or less productive from the moment they have entered the academic profession and a few cohorts may have had the same low level of productivity during their entire careers (Kyvik, 1990, p. 51). Thus, the academic careers of scientists are affected by events that occur at the time of obtaining a PhD: in other words, “cohort matters” (Stephan, 2012, p. 175), with “opportunity structures” shifting over time (Stephan & Levin, 1992, p. 117).

Second, more recently, productivity studies continued with age and, occasionally, gender as independent variables: a large-scale study of 11,500 Norwegian scientists indicated that productivity increases with age, reaching a peak late in the career and declining thereafter, thereby suggesting that scientists over 60 are significantly less cited than their younger colleagues (Aksnes et al., 2011, p. 42–44). Moreover, a study of 6,388 university professors in Quebec (Gingras et al., 2008, pp. 6–7) indicated that productivity first increases sharply with age and then increases at a slower pace at about 40 years of age, reaching its peak at about 50 years of age. However, the average number of papers in high impact journals and among highly cited papers rises continuously until retirement, with older professors more likely to be the leaders of their own teams. Another Quebec-focused study with age data reveals (Larivière et al., 2011) that after they have crossed the age of 38 years, women receive, on average, less funding for research than men, are generally less productive, and are slightly less cited. Finally, a general change has reportedly taken place in

the norms for academic publishing and the norms for appropriate academic behaviors, thereby not only affecting newer generations of scientists but also older generations pressed for more publications (Kyvik & Aksnes, 2015, p. 1451).

In a study of all Italian full professors, Abramo et al. (2016, p. 318) concluded that, in general, productivity declines significantly with age. However, professors appointed at a young age were more likely to maintain and increase their productivity than colleagues promoted at a later age. A negative monotonic relationship between age and research performance was found to accompany a positive relationship between seniority in rank and performance. In addition, the age-related productivity of scientists was also compared to the productivity of top performers. In their bibliometric study of Spanish National Research Council scientists, based on a class-based approach (top, medium, and low performance), Costas et al. (2010) concluded that the productivity of top- and medium-performing scientists increases or remains stable with age and the productivity of low-performing researchers tends to decrease with age (Costas et al., 2010, p. 1578). Furthermore, in a study of age and productivity of Italian National Research Council scientists, Bonaccorsi and Daraio (2003, p. 75) concluded that productivity declines with age and that the average age of researchers is increasing, with severe policy implications for national science systems (“the graying of the academic profession”). Guns, Eykens, and Engels (2019) examined different publication patterns of successive age cohorts at a social sciences and humanities department in Flanders, Belgium, and showed that the oldest cohort appears to maintain the traditional publication pattern focused on book publications, with a useful distinction between intracohort change and cohort succession. Finally, Rørstad and Aksnes (2015) in their study of publication rate by age, gender, and academic position on a large sample of 12,400 Norwegian university professors have shown that academic position is more important for productivity than age and gender.

Third, biological age was also incidentally used in research collaboration studies: there are only few studies of the impact of age on collaboration because few data sets combine biographical, publication, and citation data at the individual level. Collaboration, with data on age, can be studied at the meso-level of institutions; however, studies at the macro-level of countries are rare because of limited data availability (examples include Abramo et al. (2011) and Abramo et al. (2016) on Italian scientists; Norwegian studies based on a comprehensive national database Current Research Information System in Norway (CRISTIN), such as Aksnes et al. (2011) and Rørstad & Aksnes (2015)).

Biological age has also figured prominently in studies focused on bibliometric indicators at the micro-level of individual scientists and teams (Costas & Bordons, 2005; Costas & Bordons, 2007; Costas et al., 2010; Sugimoto et al. 2016). For example, Costas and Bordons (2010) proposed a bibliometric classificatory approach to research performance and, referring to biological age, presented evidence that top researchers are younger than the other two research productivity classes (low and medium classes). Finally, Wang and Barabási (2021, p. 39–50) examined the question of when scientists do their greatest work and what determines the timing of peaks, thereby revealing that the peak age of Nobel laureates has increased over time by approximately six years. The micro-level has also been applied in our own studies in which we used specifically constructed “individual publication portfolios” based on bibliometric metadata from a decade of publications indexed in Scopus (Kwiek & Roszka, 2020; Kwiek & Roszka, 2021a).

2.2. Biological Age and Academic Generations in Survey-based Academic Career Studies

Small-scale research includes survey-based studies of the academic profession that routinely use biological age, often aggregated to academic generations, as the independent variable. For example, Kwiek used age in his research on top performers in 11 European systems as well as in Poland (Kwiek, 2016, 2018a); and he used age in his research on internationalists and locals in 11 European systems and in Poland (Kwiek 2015a, 2018b, 2020b). As a recent study revealed, collaboration in Norwegian research universities increased from 58% in 1992 to 66% in 2001, and to 71% in 2013. Not only are younger generations more internationalized, but almost all generations become increasingly involved in international research collaboration as they age (Kyvik & Aksnes, 2015, pp. 1448–1449). As Kwiek's (2019) cross-generational European comparison indicated, the oldest generations account for the highest share of scientists collaborating with international research partners. In the 11 countries studied, the youngest academic cohort never represented the highest share of internationally collaborating scientists. Just as some generations become more productive as they age, some generations are more likely to collaborate as they age; this is possible to test when the data on age are available, as in the case of survey-based research. This is clearly linked to changing job market conditions over time, as competition for university jobs waxes and wanes. Kyvik and Aksnes (2015, p. 1448) clearly demonstrated for scientists who were the youngest age cohort in 1989–1991 that certain generations excel in international collaboration over time and as they age.

In the Polish case, career opportunities and academic norms in Poland differed significantly for those entering the academic labor force prior to 1989 and for those who came after (Kwiek, 2015b). Polish scientists in general resisted pressure to publish internationally until the recent wave of reforms in 2018–2020. These scientists' careers were clearly affected by events occurring at the time that their cohorts graduated. In other words, in Poland as elsewhere, “success in science depends, in part, on things outside of the control of the individual scientist” (Stephan & Levin, 1992, p. 4). In the specific Polish case, the academic survivors from the cohort of young scientists in the mid-1990s are predominantly internationalists today. Even though they had to cope with unprecedented challenges in research internationalization while working in the 1990s and 2000s, with scarce research funding and with Polish universities linked to their overwhelming teaching-focus, these scholars are more engaged in international research collaboration than those in the younger cohorts (Kwiek 2020b).

2.3. Academic Age as a Proxy of Biological Age Thus Far

Although biological age is an important independent sociodemographic variable in examinations of academic careers, academic age as its proxy has not been used often in research studies thus far. The major reason for this has certainly been a limited access to first publication data. Except for a relatively limited number of studies briefly summarized above where biological age was actually available, biological age has been studied through two major proxies: (1) academic age related to the date of first publication and (2) academic age related to the date of receiving a PhD.

First, in studies using academic age, the date of the first publication was applied approximately a dozen times (see, e.g., Lee and Bozeman, 2005; Milojević, 2012; Radicchi & Castellano 2013; Nane et al., 2017; Robinson-Garcia et al., 2020; Liao, 2017; Costas et al., 2015; Aref et al., 2019; Chan & Torgler, 2020; Simoes & Crespo, 2020; Petersen, 2015; Wildgaard, 2015). The samples in these studies were of different sizes, with resultant problems of generalizations from small samples to large national populations in several cases. The small samples used in these studies ranged from 137 scholars in information systems (Liao, 2017) to 472 top economists (Simoes & Crespo, 2020), 473 collaboration profiles of scientists in biology and physics (Petersen, 2015), 512 European researchers from 4 disciplines (Wildgaard, 2015), to 3,574 (Nane et al., 2017) and 3,596 (Costas et al., 2015) scientists from Quebec. Larger samples included 21,562 scientists from up to 10 core journals in 5 disciplines (Milojević, 2012); 35,136 scientists with profiles in Google Scholar Citations database (Radicchi & Castellano, 2013); 94,000 scientists across 43 countries (Chan & Torgler, 2020); and as many as 222,925 distinct authors with at least 5 publications (Robinson-Garcia et al. 2020) and 1.7 million authorship records in Web of Science (Aref et al., 2019). However, academic age was used for different purposes and occasionally played only a marginal role in the above research.

Second, the date of receiving a PhD as a means to determine academic age, usually derived from CVs, was used even less intensively and referred to generally smaller samples. Badar et al. (2013) studied 239 faculty members in Pakistan; Perianes-Rodriguez (2014) examined 2,530 economists working in the 81 top world economics departments; Sugimoto et al. (2016) studied 1,002 scientists working in top 10 US programs in sociology, political science, and economics; van den Besselaar & Sandström (2016) examined 243 early-career researchers in economics, behavior and education, and psychology who applied to an early career grant program in the Netherlands; and, finally, Coomes et al. (2013) studied 369 scientists from 17 US and Canadian geography departments.

Third, and in more detail, academic age was used in various data contexts to support different streams of research:

(1) Milojević (2012) examined the citing behavior of scientists in relation to age, productivity, and collaboration and was interested in cohorts of scientists who are at the same stage of their academic careers and, thus, have the same academic age. She argued that the operationalization of the academic age through publications is a good one because it “represents the length of an author’s active engagement within a scientific community” (2012, p. 3).

(2) Radicchi and Castellano (2013) analyzed bibliometric indicators at the level of individual scientists, explored their population in terms of academic age, and calculated the *h*-index and other parameters of academic publishing for academic age.

(3) Among the 17 author-level bibliometric indicators, Wildgaard (2015, p. 5) distinguishes between three categories—publication-based indicators, citation-based indicators, and hybrid indicators—and locates academic age in publication-based ones. She defines it simply as the “number of years since first publication by the researcher recorded in the database.”

(4) Aref et al., (2019) studied highly mobile researchers or “super-movers”, contributing to the “brain circulation” (rather than brain gain/drain) literature and offering,

for the first time, a snapshot of their key features by academic age. They show the most common countries of academic destination for major age brackets of super-movers (early career, intermediate, and senior super-movers) or mobile scientists who have had main affiliations in at least three distinct countries. No limitations regarding the use of academic age are discussed, as opposed to other types of limitations such as the limitations of bibliometric data in general (Aref et al., 2019, p. 59).

(5) Simoes and Crespo (2020, p. 336) applied academic age to their model of measuring author-level research productivity, assuming that any author-level performance evaluation requires the comparison of scientists in different stages of their careers. Measuring the accumulated performance, favoring authors with longer careers, was contrasted with analyzing performance data per unit of time.

(6) Chan and Torgler (2020) used academic age in the context of gender differences in performance of science elites (top cited scientists) across 43 countries and in 21 academic fields; male top-cited scientists were found to be 3.2 years older than female top-cited scientists.

(7) Robinson-Garcia et al. (2020) used academic age in their recent research in task specialization across research careers; they used a data set of over 70,000 publications from PLoS journals (with approximately 350,000 distinct authors) to predict the contributions of over 220,000 authors in over 6 million publications. Their major variable was the four academic career stages. The four career archetypes (junior, early-career, mid-career, and late-career) were used to profile scientists and to examine career trajectories, productivity, and citation impact. The authors estimated researchers' age based on the year of first publication and built the four career stages based on such year (Robinson-Garcia et al. 2020, p. 18).

Finally, and most importantly for our line of research, Nane et al. (2017) analyzed the prediction of the age of researchers using bibliometric data and focused on the predictive power of various regression models. They assessed how reliable the estimation of the real ages of scholars was, defined as their biological age, based on models that exclusively rely on the various bibliometric indicators in the empirical context of researchers from Quebec for whom biological age was available. The authors believe that their sample is representative for researchers in general (Nane et al., 2017, p. 714); moreover, the sample represents core global science producing systems in the same manner in which we believe that our sample represents (or is useful to understand) catching-up systems in science. While Canada is classified among "advanced countries" in a ranking of scientific capacity and infrastructure of 76 countries, Poland is classified among "developing countries" (Wagner, 2008, p. 88). Thus, their paper was a test of the accuracy and validity of the first year of publication as a proxy for biological age. From among numerous bibliometric indicators, the year of the first publication was found to be the best single linear estimator of the ages of individual researchers (Nane et al., 2017, p. 726). However, it was found to function particularly well when working with large sets of scholars (from an "averaged" viewpoint). Furthermore, the predictive power at the individual observational case was found to be "relatively limited, especially in some fields" (Nane et al., 2017, p. 726).

The abovementioned paper comes closest to our research in terms of its aims, goals, and approach. However, in our research, we have worked with a much larger data set with

administrative, biographical, and publication and citation data (“Polish Science Observatory,” with Scopus data) in which biological age was provided for over 20,000 scientists. Here, we use STEM and non-STEM fields, 27 individual academic disciplines, unambiguously determined gender, two institutional types, and three major academic positions.

2.4. Methodological Limitations and Practical Implications Thus Far

The use of academic age as a proxy for biological age raises a number of common questions. These pertain to the data set bias (usually Scopus or Web of Science, which come with their own linguistic, geographical, and disciplinary biases) other than those associated with global bibliometric sources that cannot be easily used in global interdisciplinary studies that go beyond single institutions or single countries. Furthermore, gender differences cannot be easily examined without massive gender determination of publication authors. Not only do bibliometric data sets favor certain countries (Anglo-American) over other countries, but it is also that certain techniques (author disambiguation, author gender determination) are much less reliable for certain parts of the world (South East Asia, including China) because of lower probability levels in ascribing individual names to gender and individual author IDs to real scientists. For example, gender determination levels reach high percentages for countries such as Poland and Russia in contrast to China and South Korea (see Gender Probability Scores in Elsevier 2020).

In addition, academic age can be used only for publishing scientists (all non-publishers are excluded from samples) and, in addition, scientists who publish in internationally indexed journals and predominantly in English. Globally, and particularly among countries that are new to the global science system, younger cohorts are generally more present in global data sets than older cohorts, with implications for age structure biases. However, if competing sources of global data are considered (e.g., the Google Scholar database), the bias in favor of relatively young scientists may be even higher (Radicchi & Castellano, 2013).

Furthermore, certain research questions can be examined much more adequately with large data sets so that academic ages are more averaged and much less adequately in individual cases or cases with a small number of observations (Nane et al., 2017, p. 726). The countries that have recently joined global science and the disciplines that continue to prioritize national languages and books and book chapters (rather than journal articles)—such as major segments of the humanities and arts—cannot be reliably examined with the use of academic age as a proxy for biological age; in this case, the age at which a doctoral degree was awarded may be a better methodological choice, data availability permitting (see Guns et al., 2019). However, access to dates of receiving the PhD degree on a massive national scale is technically possible only in several countries—those in which Current Research Information Systems (CRIS) are being implemented, usually in relation to national research assessment exercises (Italy, Norway, and Poland).

2.5. Hypotheses of this Research and the National Context

Our research questions are linked to findings from previous studies that used both biological and academic ages as their variables and are embedded in the Polish national context. However, the research questions are also closely related to the data set at our disposal in which we have reliable data on every individual scientist and their gender, institutional affiliation, date of birth, date of obtaining doctorates, habilitations and professorial titles, as well as full metadata on their lifetime individual publication portfolios from Scopus (using Elsevier's ICSR Lab cloud computing services), including the date of the first publication.

We assume three hypotheses here. First, we assume that the biological age at which the first publication indexed in global databases like Scopus appears for a given scientist differs significantly for men and women: in practice, it may be lower for men (Hypothesis 1). Second, we assume that the biological age at which a researcher publishes for the first time is significantly different in STEMM and non-STEMM disciplines (with a focus on the humanities and social sciences—the HUM and SOC fields): in practice, it may be lower for STEMM disciplines (Hypothesis 2). Third, we assume that the biological age of a researcher at the time of first publication varies by institutional research intensity (particularly between a small group of research-intensive institutions on the one hand and all others on the other): in practice, it may be lower in institutions that are more heavily involved in research (Hypothesis 3). The three research questions, the three hypotheses, and results (confirmation) are presented in Table 1.

Table 1

Research hypotheses and results (summary)

Research Question	Hypothesis	Result
RQ1. What is the relationship between academic age and gender?	Hypothesis 1. We would expect that the age at which the first publication is published differs between man scientists and woman scientists.	Confirmed – lower for man scientists.
RQ2. What is the relationship between academic age and academic discipline?	Hypothesis 2. We would expect that the age at which the first publication is published differs between academic disciplines, especially between STEMM and non-STEMM disciplines.	Confirmed – lower for STEMM disciplines.
RQ3. What is the relationship between academic age and institutional research intensity?	Hypothesis 3. We would anticipate that the age at which the first publication is published differs between institutional types by research intensity, especially between a subset of ten research-intensive universities and all others.	Confirmed – lower for research-intensive universities.

With regard to the national context, the Polish science and higher education systems have been studied intensively in the past few years. For example, Kulczycki et al. (2017)

examined the higher education funding system and Bieliński and Tomczyńska (2018) studied the various manifestations of the ethos of science. Feldy and Kowalczyk (2020) studied how scientists view the system of financing science, and Kulczycki and Korytkowski (2020) examined changing publication patterns. In addition, recent studies focused on higher education reforms (e.g., Shaw, 2019; Antonowicz et al., 2020), international research collaboration (Kwiek, 2020b), and high research productivity (Kwiek, 2018b). Furthermore, gender disparities have been explored through the concept of gender homophily in academic publishing (Kwiek & Roszka, 2021a) and both male scientists and female scientists have been shown to collaborate intensively with men, even in gender-balanced academic disciplines. Further, gender differences in Polish academia have also been studied with reference to major research collaboration types: Kwiek & Roszka (2020) examined international research collaboration by gender and showed that male scientists dominate in this collaboration type, with significant cross-disciplinary differences. Finally, Kwiek & Roszka (2021b) studied what they termed the gender solo research gap, or the differences between males and females, using solo research across their career stages.

3. Data and Methods

3.1. Data Set

We used the integrated “Polish Science Observatory” database (see Kwiek & Roszka 2021a, pp. 4-6 for a more detailed description). Two large databases were merged to produce this database: Database I was an official national administrative and biographical register of all Polish scientists (“Nauka Polska”—“The Polish Science”); Database II was the Scopus publication and citation database. Database I comprised 99,535 scientists employed in the Polish science sector as of November 21, 2017. Only scientists with at least a doctoral degree (70,272) and employed in the higher education sector were selected for further analysis (54,448 scientists). The data used were both demographic and professional, with each scientist identified by a unique ID. Database II, the original Scopus database, included 169,775 names from all 85 universities with internationally visible publications in the decade 2009–2018. Authors in Database II were defined by their institutional affiliations, their Scopus documents, and individual Scopus IDs. The key procedure was to appropriately identify authors with their different individual IDs in the two databases and provide them with a new ID in the integrated “Polish Science Observatory” database. Probabilistic methods of data integration were used (as defined in Enamorado et al., 2019). Database I contained biographical and professional career information on all authors affiliated with the 85 largest Polish universities in the 2009–2018 reference period. Database II contained metadata on 377,886 publications. From among the 377,886 articles in the original Database II, 230,007 were written by the authors included in Database I. Subsequently, only articles written in journals were selected for further analysis, all other publication types being omitted. Therefore, the number of publications in the database was reduced to 158,743 articles and the number of unique authors was reduced to 25,463. Approximately half of the Polish scientists from the higher education sector (53.4%) did not publish a paper indexed in the Scopus database in the reference period—which is in line with previous findings

regarding the distribution of Polish publications (Kwiek 2018a)—with the overwhelming majority of publications belonging to national publication outlets.

3.3. Methods

3.3.1. Determining Gender

There was no need to determine the gender of Polish scientists through, for example, the various probabilistic methods, external data sets with gender-determined observations, or special gender determining software (for example, GenderizeR) (see Wais, 2016), as in our previous research that included the gender of international research collaborators of Polish scientists as one of the variables (Kwiek & Roszka, 2021a). In this case, gender is provided for all scientists with at least a PhD degree from all public science sectors present in our database, including 20,596 scientists in our final sample. It is a binary variable provided by the national registry of scientists. It was not possible to include any other options (based on “other” or “prefer not to disclose” responses) in the registry.

3.3.2. Determining Disciplines

It was necessary to determine the academic disciplines represented by Polish scientists. Every Polish scientist present in our integrated database was ascribed to one of 334 ASJC disciplines at the four-digit level and one of 27 ASJC disciplines at the two-digit level (following Abramo, Aksnes, & D’Angelo (2020), who defined the Web of Science subject category for each Italian and Norwegian professor in their sample). In the ASJC system of disciplines used in Scopus, a journal publication can have one or multiple disciplinary classifications derived from journal disciplinary classifications. The dominant ASJC for each scientist was determined on the basis of all publications (type: article) present in their individual publication portfolios for the period 2009–2018 (the mode being the most frequently occurring value). In the case when there was no single mode, the dominant ASJC was randomly selected. Consequently, we had Polish scientists unambiguously defined by their gender and ASJC discipline, along with the metadata of all their publications indexed in the Scopus database. Three disciplines at the two-digit level were excluded from further analysis as they did not meet an arbitrary minimum threshold of 50 scientists per discipline (GEN, NEURO, and NURS).

3.2.3. Determining Academic Age

Having an individual scientist as the unit of analysis, we calculated the academic age based on the year of the first publication for every scientist in our sample. Our data set provides the date of birth and the dates of receiving all the relevant scientific degrees (doctoral degree for all and habilitation degree and professorship title for a few scientists, depending on the stage of academic career).

The academic age we used ranges from 0 to 40 (as we generally assumed to include scientists in the biological age range of 30–70 years and with doctorates). We obtained the dates of the first publication indexed in the Scopus database for every Polish scientist using

the application programming interface (API) protocol (a set of programming codes that enables data transmission between one software product and another) provided by Scopus. To use API, we applied individual Scopus IDs for every scientist in the sample. Our reference year was 2017 and the difference between 2017 and the year of the first publication was used to indicate the academic age as of 2017.

For example, regardless of the biological age as stated in our data set and derived from a national registry of scientists, if a scientist had their first Scopus-indexed publication in 1997, their academic age in 2017 was 20 years; moreover, if they had their first publication in 2016, their academic age in 2017 was 1 year, thereby implying being at the very beginning of the academic career as defined through publications. It can be roughly assumed that in the natural sciences, the first publication in the database usually occurs during the last years of writing a doctorate—that is, around the age of 30 years in the Polish case (Central Statistical Office GUS, 2020). Thus, in a large simplification, the academic age of 0 would imply about 30 years of biological age, and the academic age of 20 would imply about 50 years of biological age. We also link academic age to the idea of career age as the age spent in science (only twice: in Figure 1 and Table 2). In this sense, in the Polish context, a lifetime academic career, rare as it is, would include 40 years of life: an academic age in the range of 0–40 years (implying a career age of 30–70 years) and a biological age in the range of 30–70 years.

3.3. Sample

We began with a set of 25,463 scientists (14,886 male scientists and 10,577 female scientists—58.5% and 41.5%, respectively). This set included all scientists who were full-time employed in the higher education sector and who had at least a single article indexed in the Scopus database in the period 2009–2018 and who had at least a doctoral degree. Generally, 95%–97% of scientists in higher education have been employed full-time in the previous decade (GUS, 2020). The sample includes all internationally visible (through publication type: article and in an international database type: Scopus) Polish academic scientists.

We defined the sample for this research in the following steps: we began with all scientists from all public science sectors present in our “Observatory” data set—99,935; then, we moved on to scientists with doctorates, which limited the number of our observations to 70,272; then, we moved on to scientists with doctorates and employed in higher education—54,448; among these, there were 32,937 scientists with publications of any type in Scopus in the period 2009–2018. Finally, there were 25,463 scientists with journal articles only.

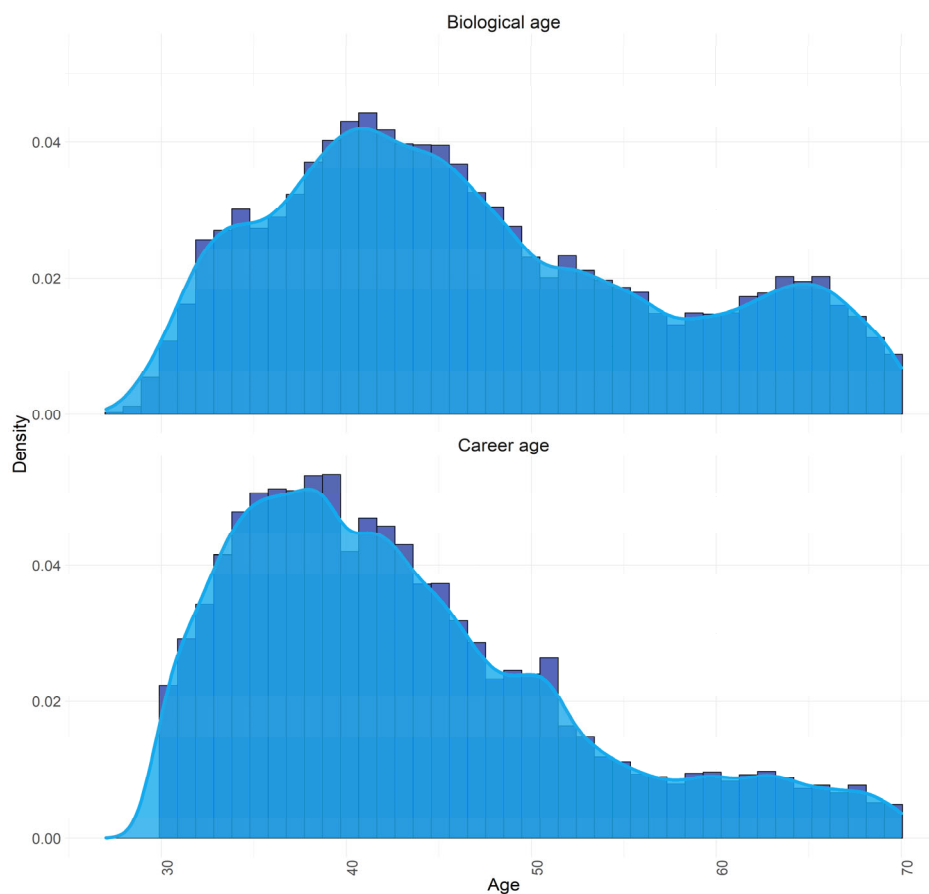
We decided to include information on the entire productivity of individual authors in the database thus created (the original database contained bibliometric information only for the decade 2009–2018, and the reference period for the demographic and professional data of the authors was November 2017); in particular, we decided to include the year of the first publication. Such information was obtained from the Scopus API database for 21,285 individuals. Moreover, excluding individuals older than 70 years of age from the database and those for whom an incorrect date of first publication was found led to a reduction of the effective sample size to 20,596 authors ($N = 20,596$).

Figure 1 depicts the distribution of the sample by biological age (upper panel) and career age (bottom panel) for illustrative purposes: the Polish academic profession is relatively young, with the majority of scientists in the 30–45 age group and a substantial proportion of scientists who are older than 50 years of age (and older than 60 years of age). The distribution of the sample in terms of career age (derived from academic age to make the comparison clearer) looks similar: the majority of scientists have been publishing for no more than 15 years. The long tail on the right shows that although the proportion of older scientists is substantial (by biological age), they have been publishing internationally for a much shorter duration than their biological age indicates. Only a small minority of scientists (all academic disciplines combined) have been publishing for 25, 30, or 40 years. The generational difference between younger and older scientists and their publishing is clearly shown by the displacement of the sample to the left along the x-axis.

The descriptive statistics of the biological and career ages (used here for the sake of clarity and calculated as the academic age plus 30) of the sample are provided in Table 1, with major parameters such as gender, academic position (assistant, associate, and full professors), institutional type (research-intensive IDUB type and the rest), and ASJC academic disciplines. Career age here and below is used for the purpose of clarity only.

Figure 1

Distribution of scientists' biological age (top) and career age (bottom) and the Kernel density plot for all academic fields combined ($N = 20,596$).



Several aspects are required to be emphasized here: the biological median age of 45 years is higher than the median career age of 41 years, and both are higher for males than for females. The difference between the two ages increases with subsequent academic positions: while for assistant professors it is small, for full professors it is substantial (two and seven years, respectively). This implies that the cohort of junior scientists began publishing internationally earlier than the cohort of senior scientists and particularly earlier than current full professors. In addition, the cross-disciplinary differentiation is clear: for STEMM disciplines, the discrepancies between median biological age and median career age are much lower than that for non-STEMM disciplines. Scientists in STEMM disciplines begin publishing internationally much earlier than scientists in non-STEMM disciplines. In major STEMM disciplines such as BIO (biochemistry, genetics, and molecular biology) there is no discrepancy at all and in CHEM, COMP, MATH, PHYS, and MED, the discrepancy is in the range of merely 1–2 years. In contrast, in such major non-STEM disciplines as BUS (business, management, and accounting), ECON (economics, econometrics, and finance), HUM (arts and humanities), and SOC (social sciences), the discrepancy reaches 8–10 years.

Clearly, for major social science and humanities fields, the discrepancy is fundamentally higher than for major natural science fields, which clearly indicates that for the currently employed academic workforce, biological age can be inferred for STEMM disciplines with marginal error; however, for non-STEMM disciplines, the error is substantial and allocating scientists to major career stages—beginning, early, middle, and late—can be misleading.

Table 2

Descriptive statistics of biological and career age by gender, academic position, institutional type, and academic discipline (N = 20,596).

		Biological age			Career age		
		Mean	SD	Median	Mean	SD	Median
Gender	Total	47.0	10.6	45.0	43.2	9.3	41.0
	Female scient.	45.1	9.7	43.0	41.2	8.2	39.0
	Male scientists	48.4	11.0	46.0	44.6	9.9	42.0
Academic position	Assistant Prof.	41.3	7.8	40.0	39.0	6.2	38.0
	Associate Prof.	50.8	8.4	49.0	45.6	8.4	45.0
	Full Professor	61.3	6.8	63.0	54.4	10.3	56.0
IDUB	IDUB	46.6	10.8	44.0	43.9	9.7	42.0
	Rest	47.2	10.5	45.0	42.9	9.2	41.0
ASJC discipline	AGRI	47.6	10.4	46.0	42.0	7.9	40.0
	BIO	45.3	10.3	43.0	44.6	9.3	43.0
	BUS	46.3	9.7	44.0	36.3	6.6	34.0
	CHEM	45.8	11.1	44.0	46.1	9.8	45.0
	CHEMENG	47.3	11.7	45.0	43.7	10.4	41.0
	COMP	46.5	10.5	44.0	44.1	9.0	42.0
	DEC	48.1	11.2	43.5	41.6	11.8	37.5
	DENT	45.0	9.8	45.0	39.9	8.8	39.0

EARTH	48.2	11.2	46.0	43.8	10.1	42.0
ECON	44.3	9.2	42.0	36.0	5.7	35.0
ENER	46.9	12.0	44.0	39.8	9.0	37.0
ENG	47.7	11.3	45.0	43.4	9.4	41.0
ENVIR	47.1	10.2	45.0	42.1	8.1	40.0
HEALTH	48.8	10.4	49.0	38.1	4.8	37.0
HUM	46.7	9.6	45.0	36.2	6.3	34.0
IMMU	45.7	9.3	45.0	45.7	8.1	45.0
MATER	46.6	10.9	44.0	45.4	9.6	43.0
MATH	47.1	11.3	45.0	45.5	9.8	43.0
MED	47.9	9.8	47.0	45.1	8.9	44.0
PHARM	44.5	10.3	42.0	43.9	9.1	42.0
PHYS	48.2	11.5	46.0	48.6	10.4	46.0
PSYCH	44.6	10.5	41.5	37.0	6.2	36.0
SOC	45.4	9.6	44.0	36.7	6.5	35.0
VET	47.3	10.4	46.0	44.5	7.8	43.5

The gender distribution of our sample of Scopus publishers is approximately 40/60 (42.5% of female scientists and 57.5% of male scientists) and is similar to the current distribution of higher education personnel (46.1% and 53.9% in 2019) (GUS, 2020). Approximately half the scientists in our sample are middle-aged (in the 40–54 age bracket; 47.3%) and over half of them are assistant professors (55.7.0%; see the structure of the sample in Data Appendix, Table 5). We employed a common international terminology of academic careers with three types of professors employed in the higher education system rather than a complicated system of Polish academic degrees, titles, and positions. Consequently, all assistant professors in our sample hold doctoral degrees, all associate professors hold habilitations (postdoctoral degrees necessary to begin independent academic careers), and all full professors hold professorship titles.

3.4. Variables

We use numerical and categorical variables which are of a biological, demographic, administrative, and institutional nature. The list of variables, their brief descriptions, and the data sources (Observatory, Scopus or Ministry) are presented in Table 3.

Table 3

The variables used in the analysis

No.	Variable	Description	Source
1.	Biological age	Numerical variable. Biological age as provided by the national registry of scientists (N = 99,935). Age in full years as of 2017 is used.	Observatory
2.	Age group	Categorical variable. Three major age groups are used: young (39 and younger; N = 5,708), middle-aged (40–54; N = 9,746), and older (55 and older; N = 5,142) scientists.	Observatory
3.	Gender	Categorical variable, as provided by the national registry of scientists “The Polish Science” (<i>Nauka Polska</i> , N = 99,935). No other options are possible in the registry.	Observatory

No.	Variable	Description	Source
4.	Academic position	Categorical variable. Three Polish degrees used as proxies of academic positions: doctoral degree only (assistant professor; N = 11,470); habilitation degree (associate professor; N = 6,242); and professorship title (full professor; N = 2,884). All scientists without doctoral degrees and from outside of the higher education sector were excluded from the analysis. Four career stages are used: beginning (<5 years of academic experience-since the first publication in Scopus), early career (5–14 years), middle career (15–29 years), and late career (30 years and more).	Observatory
5.	Discipline	Categorical variable. All scientists ascribed to one of 27 Scopus ASJC disciplines. Dominant disciplines were used (N = 20,596).	Scopus
6.	STEMM disciplines	Categorical variable. STEM disciplines: AGRI (agricultural and biological sciences); BIO (biochemistry, genetics, and molecular biology); CHEMENG (chemical engineering); CHEM (chemistry); COMP (computer science); DEC (decision science); EARTH (Earth and planetary sciences); ENER (energy); ENG (engineering); ENVIR (environmental science); GEN (biochemistry, genetics, and molecular biology); IMMU (immunology and microbiology); MATER (materials science); MATH (mathematics); NEURO (neuroscience); NURS (nursing); PHARM (pharmacology, toxicology, and pharmaceuticals); PHYS (physics and astronomy); and MED (medicine).	Scopus
7.	Non-STEMM disciplines	Categorical variable. Non-STEM disciplines: BUS (business, management, and accounting); DENT (dentistry); ECON (economics, econometrics, and finance); HEALTH (health professions); HUM (arts and humanities); PSYCH (psychology); SOC (social sciences); and VET (veterinary).	Scopus
8.	Research-intensive institution	Categorical variable. The 10 institutions (from among 85 examined) are the IDUB (or “Excellence Initiative–Research University”) institutions selected in 2019.	Ministry

4. Results

4.1. Example

Before going further, we provide an example: the first publication of researcher A (a male in the ECON discipline who obtained his PhD in 1995, his habilitation in 2005, and his professorship in 2012) was, according to the Scopus database, published in 2015. The following assumption (based on selected global papers discussed in section 2.2) is made in this regard: 2015 is the beginning of an academic career for this scientist in the sense of beginning to publish, his academic age is 0. Therefore, in 2017, he must have a biological age of 32. However, our “Observatory” database using his date of birth indicates 52 years as his biological age in 2017.

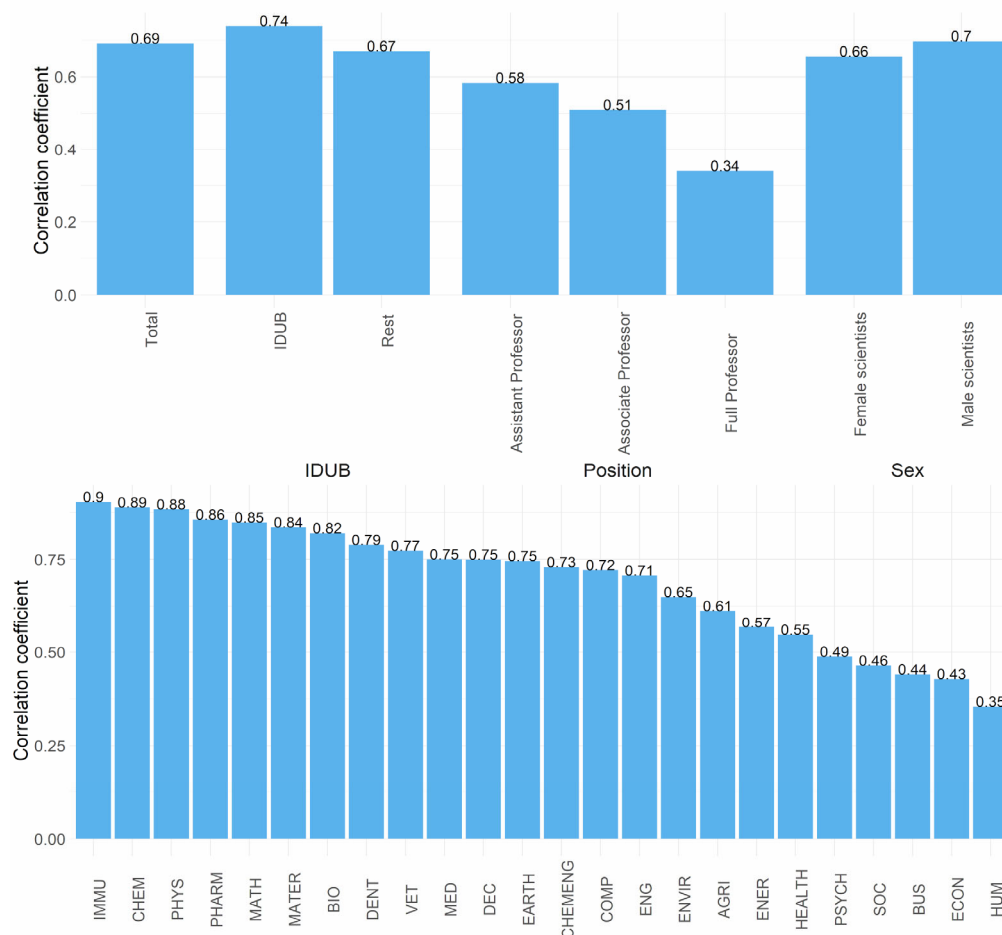
Thus, an inconsistency arises: using academic age as a proxy for biological age, scientist A is young and just beginning his scientific career (academic age = 2, career stage: beginning); whereas using biological age, we conclusively find that scientist A is older (52 years old, career stage: middle). Our task below is to estimate this mismatch between academic age and biological age on a large scale of the entire national science system, depending on selected parameters.

4.2. Correlation of Biological and Academic Ages

In order to analyze the influence of biological age (in the range of 30–70 years) on academic age (in the range of 0–40 years), a linear correlation analysis between these variables was performed (Figure 2). Regardless of the approach of the correlations presented, a positive relationship was observed in each case. All Pearson's linear correlation coefficients are significantly different from zero (at a significance level of $\alpha = 0.05$). Another clearly visible pattern is the usually strong or very strong correlation observed for disciplines belonging to the STEMM cluster. For most STEMM disciplines, a stronger correlation was observed than for all disciplines together (i.e., Total). The scatter plots (Figure 3) clearly indicate that the vast majority of individuals publishing in STEMM disciplines publish their first article at a relatively young age (the points shift to the right on the X axis). Moreover, the highest correlations were observed for STEMM disciplines such as chemistry ($r = 0.89$), physics and astronomy ($r = 0.88$), mathematics ($r = 0.85$) or medicine ($r = 0.75$).

Figure 2.

Correlation coefficient for Polish scientists between academic age and biological age, by institutional type (research-intensive IDUB institutions and the rest), academic position, gender, and ASJC discipline (N = 20,596).



In contrast, disciplines such as business (BUS), management and accounting, arts and humanities (HUM), and social science (SOC) (representing the non-STEMM field)—

relatively abundantly represented by Polish scientists—are characterized by a relatively low strength of the relationship between biological and academic ages. The value of the correlation coefficient does not exceed 0.5 in any case and remains in the range from 0.36 for humanities to 0.49 for social sciences; the scatter analysis (Figure 3) indicates a clear shift of points to the left on the X axis.

Thus, the results reveal that most representatives of non-STEMM disciplines, despite their relatively advanced biological age, are relatively young in terms of their academic age. Their first publication in the Scopus database—that is, their publication debut in the global scientific arena—clearly begins later than in the case of STEMM scientists.

This observation is consistent with common intuitions and previous surveys (Kwiek, 2015a, 2020b), according to which it takes a shorter amount of time for representatives of the sciences in Poland to participate in the global circulation of science than for representatives of the humanities, social sciences, or economics (among other reasons due to the expansion of the private sector between 1990 and 2006 as well as on which much of the energy of non-STEMM scientists had been focused, Kwiek & Szadkowski 2019).

Smaller differences in correlations were observed for the other independent variables: university type in terms of research intensity (research-intensive IDUB institutions vs. the rest), academic position, and gender. For faculty at research-intensive universities (IDUB), the correlation is higher than for the rest ($r = 0.74$ vs. $r = 0.67$), but for both, the correlation coefficient is close to the correlation for all observations ($r = 0.69$). Moreover, the correlation is higher for men than that for women ($r = 0.70$ vs. $r = 0.66$).

Furthermore, the differences in academic positions reveal (Figure 4) that although the correlation coefficients are significantly lower, the differences in the slope of the regression curve clearly show that scientists in lower positions are generally academically younger than their colleagues in higher positions (shift of points to the left on the X-axis), which is probably due to the correlation of position with biological age. Simultaneously, as the academic seniority increases, the strength of the correlation clearly decreases (from $r = 0.58$ for assistant professors to only $r = 0.34$ for full professors).

These results clearly show the diversity of scientists in terms of when they begin to publish: current full professors, the oldest and highest in the academic hierarchy, are the latest to begin; and the youngest and at the beginning of their academic career, scientists only with a doctorate (assistant professors), find international publishing more natural. These findings confirm the results of earlier surveys and interviews that reveal radical generational differences in Polish science (Kwiek, 2015b) and survey results linking “internationalists” and “locals” in research to age and academic generations (Kwiek, 2020b). Remarkably, an earlier survey research (conducted on 4,000 returned questionnaires) is strongly corroborated by the detailed large-sample research presented here.

Intergenerational differences in international publishing are due to distinctly different starting conditions for different groups of researchers: different opportunities for international collaboration and different institutional requirements at successive stages of academic career development, particularly growing after 2010, when two series of structural higher education reforms (2010–2012 and 2016–2018) were initiated. Faculty at the level of assistant professors and, thus, also predominantly young faculty, have been embarking on their academic careers for a decade now, under radically better financial conditions and

changed political and social realities than their colleagues who are currently associate or full professors. The inter-cohort differences clearly reveal the evolution of the Polish system of science.

Figure 3

Scatter analysis for Polish scientists between academic age (range 0–40 years) and biological age (range 30–70 years) by ASJC discipline (N = 20,596).

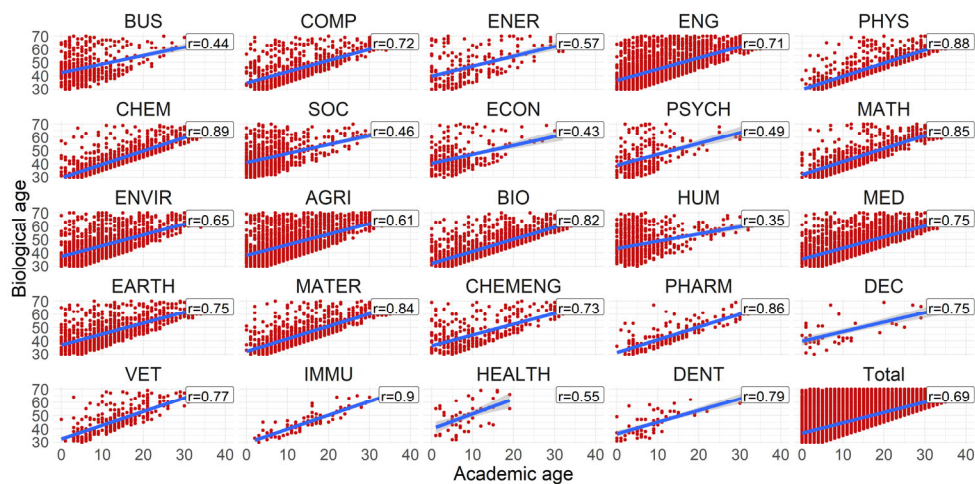
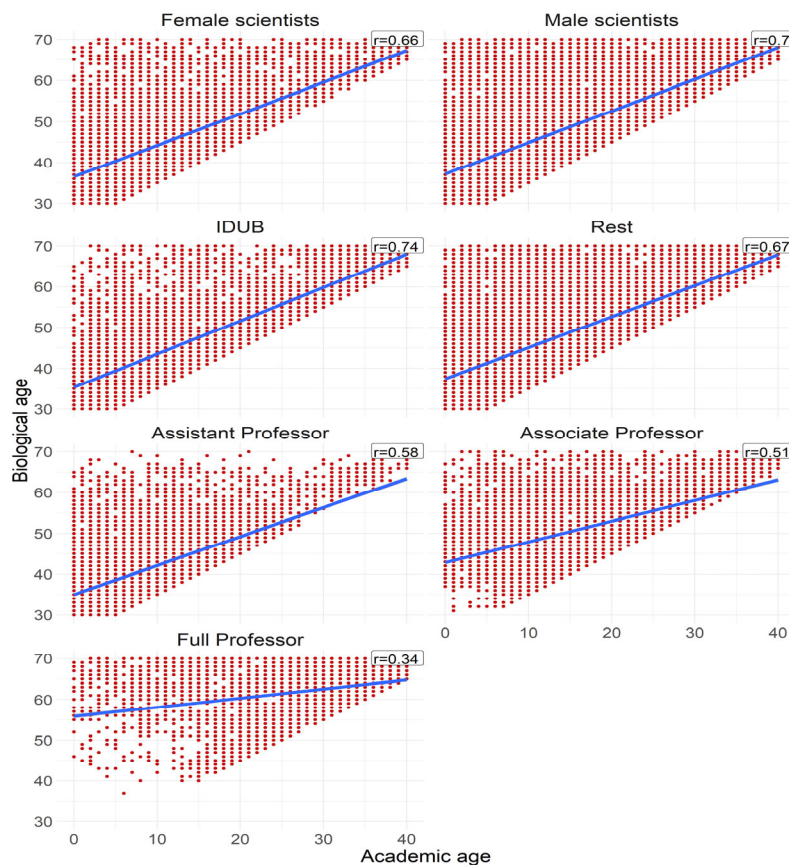


Figure 4.

Scatter analysis for Polish scientists between academic age (range 0–40 years) and biological age (range 30–70 years) by gender (top panel) institutional type (research-intensive IDUB institutions vs. the rest, middle panel), and academic position (bottom panels) (N = 20,596).



Analyzing the distribution of biological age by individual academic ages for all scientists—regardless of discipline, institutional type, academic position, and gender—there is an increasing but fading trend of the medians of biological age and a decreasing variation of biological age with increasing academic age (Figure 5). For gender and institutional type (IDUB vs. rest), no significant differences in this process are noted, while disciplines (Figure 6) and academic position (Figure 7) have a strong influence on the shape of the biological age distribution.

For the STEMM disciplines, the trend of medians in many cases (e.g., BIO, CHEM, COMP, EARTH, ENG, ENVIR, MATER, MATH, MED, or PHYS) appears not to be fading but largely linear, once again highlighting the strong association between academic age and biological age in the STEMM field. Although academic careers evolve with the age of the scientist, for non-STEMM fields, it appears that the trend does not occur at all in certain cases (particularly in large disciplines such as ECON, HUM, or PSYCH) or is barely noticeable (BUS, SOC). Similarly, the variability in the distribution of biological age, which decreases with academic age for STEMM disciplines, appears to be unchanged or changing in a disordered manner. In terms of academic positions, the distribution of biological age for individual years of academic age appears to be similar to the general pattern (i.e., somewhat resembling a logistic curve—fading growth) for assistant professors and associate professors, but behaves rather differently for full professors. Among full professors, there is a clear dominance of scientists who are older in biological age for almost every year of academic age.

Figure 5.

Distribution of biological age for Polish scientists for individual academic years of age ($N = 20,596$).

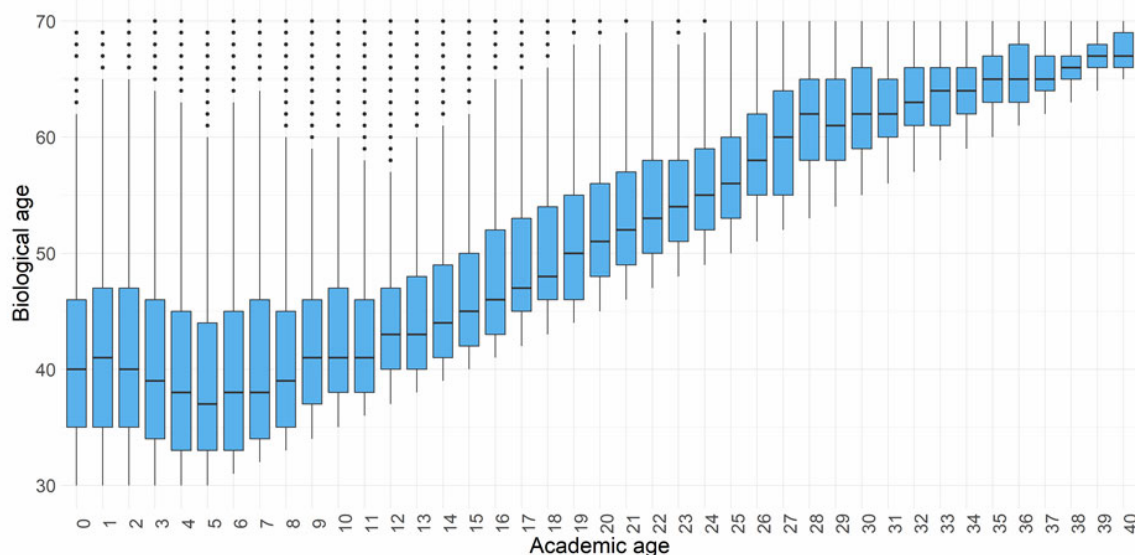
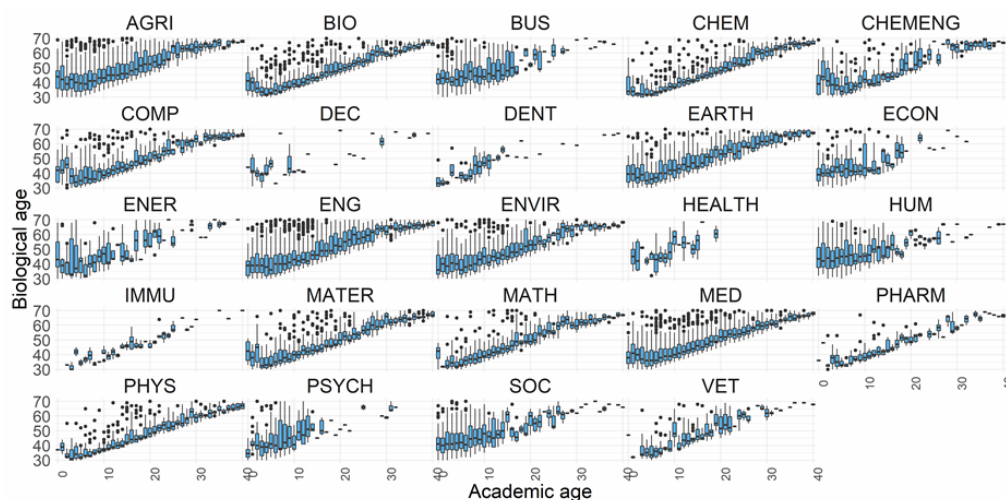
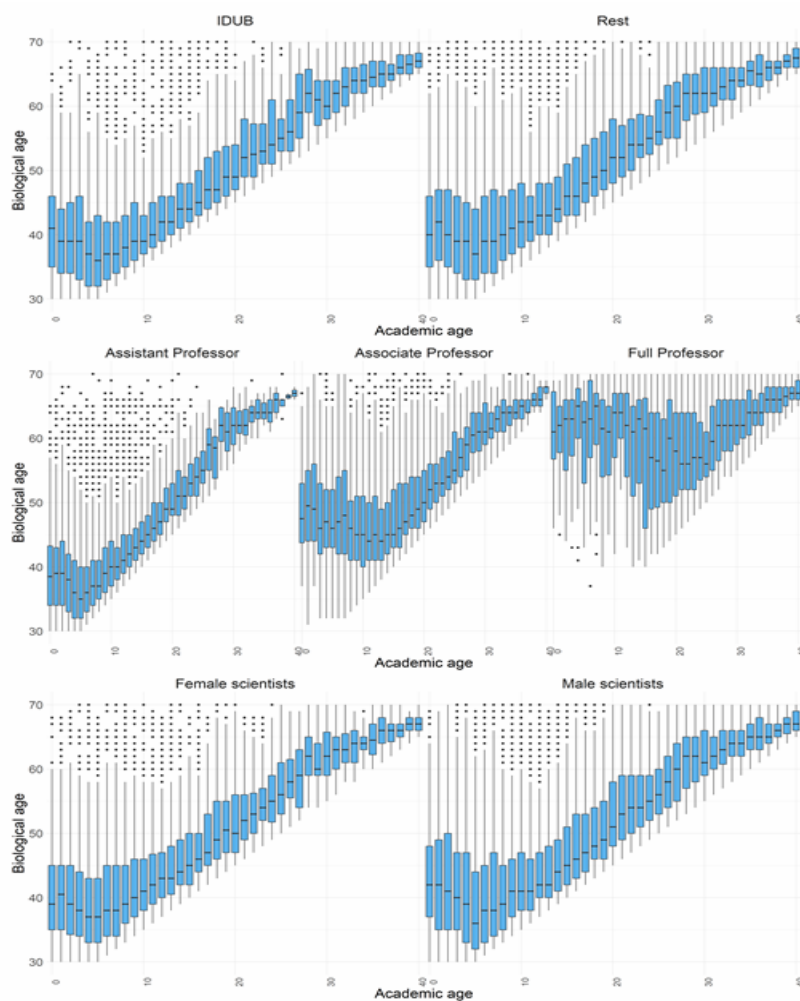


Figure 6.

Distribution of biological age for Polish scientists for individual academic years of age by academic disciplines (N = 20,596).

**Figure 7.**

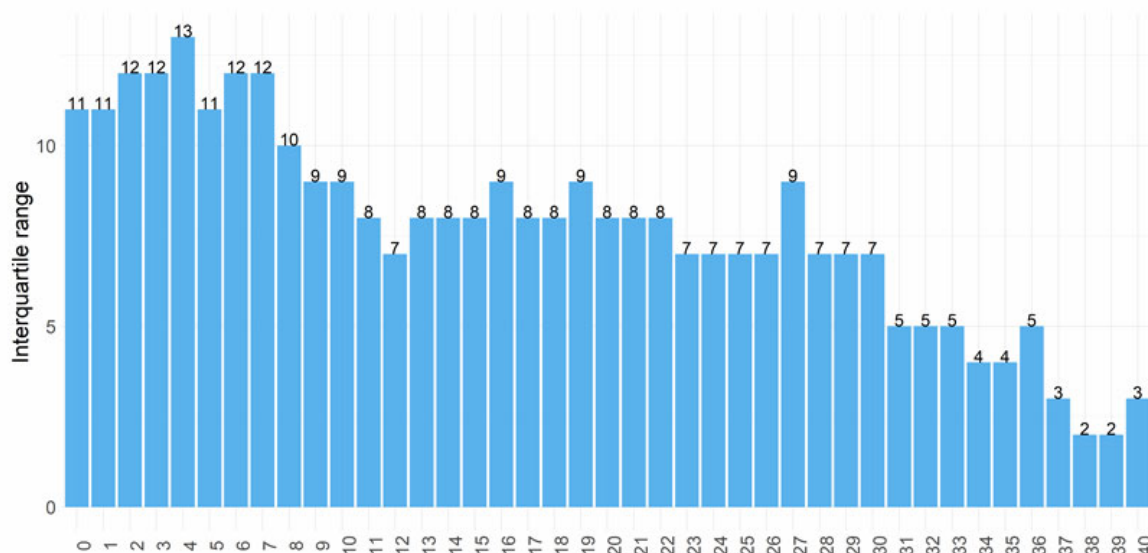
Distribution of biological age for Polish scientists for individual academic years of age by institutional research intensity (IDUB), academic position, and gender (N = 20,596).



The interquartile ranges (Figure 8) decrease as academic age increases, and they vary between 11 years and 13 years for the low academic ages to 2–3 years for the oldest ages. In other words, variability decreases as academic age increases. While for the younger cohorts the deviation from the median age for the middle 50% of academics is approximately ± 6 years, it is only ± 2 years for the older cohorts.

Figure 8.

Biological age interquartile range for Polish scientists for individual academic years of age (range: 0–40 years) ($N = 20,596$).



4.3. Contingency Analysis

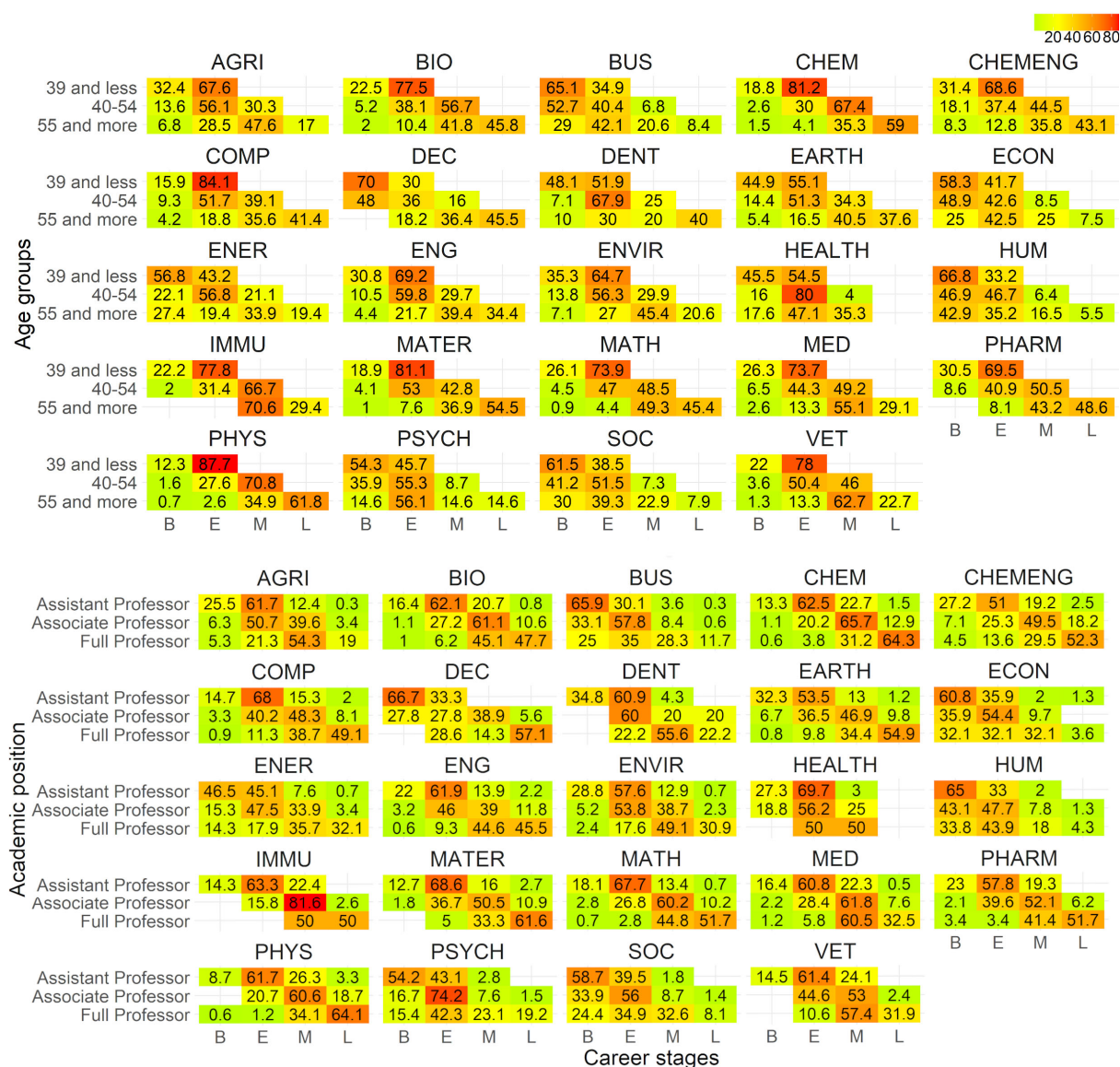
The conclusions obtained from the correlation analysis (section 4.2) are confirmed by the contingency analysis of academic career stages (four stages) with academic position (three positions) and age groups (four age groups). We allocated all scientists to four career stages or years of academic publishing (academic age brackets): beginning (B, less than five years of academic experience since the first publication in Scopus), early career (E, 5–14 years), middle career (M, 15–29 years), and late career (L, 30 and over). In very approximate terms, if academic age is 0 years (beginning of publication career in the sense of first publication = 30 years), then those in the beginning stage are aged 30–34 years, those in the early career stage are aged 35–44 years, those in the middle career stage are aged 45–59 years, and those in the late career stage are aged 60 years and over. However, in all analyses, we use a well-defined and strictly determined academic age range of 0–40 years for each researcher.

An example of the differences between STEMM and non-STEMM disciplines is very well illustrated in the comparison between CHEM (chemistry) and HUM (arts and humanities) disciplines, both relatively populous (the upper panel in Figure 9 presents age groups and the lower panel presents academic positions). In chemistry, the vast majority of assistant professors (62.5%) are in the early stage of their careers. On the other hand, a majority of associate professors (65.7%) are in the middle stage, while a majority of full professors (64.3%) are in the late stage. It is evident that the correlation of successive promotions and stages of scientific career with the advancement of own scientific work is

understood as a bracket of academic age. Chemistry is an excellent example of a discipline in which academic positions correspond with the publication trajectories of scientists: full professors mostly publish much longer than other categories of scientists, as is expected. The contingency analyses by academic position (bottom panel) and by age group (top panel) are similar in this case, with the highest percentage of young scientists under the age of 40 (81.2%) being in the early career stage; this is consistent with the idea that younger generations entering the academic profession have been well internationalized in research.

Figure 9

Contingency analysis of Polish scientists between biological age and academic age groups by academic position (upper panel) and ASJC discipline (lower panel). Four career stages are analyzed: B—beginning, E—early career, M—middle career, and L—late career. The contingency level increases as the colors change from green to red (N = 20,596).



In contrast, this division is not clear-cut in the humanities. We can observe a shift in the academic career stage (defined by the date of first publication) toward a younger age (Figure 9, upper panel) and earlier academic career stage (Figure 8, lower panel). Thus, among assistant professors at HUM, two-thirds (65.0%) are at the initial (beginning) stage of

their career rather than the early stage, as was expected; but most importantly, the subsequent positions do not indicate a clear distinction among career stages—that is, one cannot identify a stage clearly dominated by associate professors or full professors. Scientists in these positions are mainly in the beginning and early stages of their careers, rather than in middle and late career stages, as suggested in the model (confirmed for CHEM above).

This implies that in the non-STEMM disciplines such as HUM arts and humanities (as in BUS business, management, and accounting; ECON economics, econometrics, and finance; PSYCH psychology; and SOC social sciences), Polish scientists began publishing their articles in the international circulation radically later than those in chemistry, which is compared here, but also in disciplines such as BIO and PHYS. It also implies that in the humanities, it is mainly the youngest scientists and those at the earliest stage of their scientific careers who publish internationally.

Analyzing the contingency between academic position and academic career stages (Figure 9, lower panel), the evident pattern is that scientists in almost all non-STEMM disciplines are delayed by at least one academic career stage compared to STEMM disciplines. Moreover, in non-STEMM disciplines, associate professors and full professors do not dominate at all in the late stages of their careers, which implies that they publish internationally for shorter periods of time than one might assume.

An analogous pattern is observed for the contingency of age groups and academic careers (Figure 9, upper panel): the STEMM disciplines clearly show advancement to later academic career groups with increasing age group, while for the non-STEMM disciplines this relationship is not obvious. We observe a significant shift over time for non-STEMM disciplines—the first articles of non-STEMM scientists are published late, often only in the phase of working as a full professor. This has one implication: the academic age in the case of non-STEMM disciplines does not keep up with the biological age, and full professors are often at the same stage of academic age as assistant professors. Thus, the inference of biological age of scientists in STEMM disciplines proves to be an adequate approximation of reality, while the same inference in non-STEMM disciplines proves to lead to erroneous conclusions and distorted results.

4.4. Modeling Approach: Linear Regression Model

The correlation analysis of academic age and biological age with particular independent variables in the disjointed (two-dimensional) approach presented above leads to the observation of interesting patterns, but only the combined (multivariate) impact of all of the variables on biological age enables the provision of a full picture of the studied phenomenon.

In order to conduct multivariate analysis, a linear multivariate regression model was created, where the dependent variable was biological age and the independent variables were (1) academic age, (2) gender, (3) institutional type, (4) academic discipline, and (5) academic position. The resulting model explained 61.5% of the variation in biological age and the standard error was 6.59 years (i.e., when determining an age from the estimated model, we are off the mark by an average of 6.59 years; the relative error was 14%). All interpretations

are subject to the *ceteris paribus* assumption and the significance level used was $\alpha=0.05$ (Table 4).

According to the model, if academic age increases by one year, biological age increases by 0.6 years on average. This is also the variable with the strongest influence on the dependent variable (the standardized coefficient is 0.523, which is the largest among all independent variables). The characteristic with the second strongest effect is academic position. The position of full professor has a positive impact compared to the position of assistant professor (reference category; the increase in age is on average 11 years greater); simultaneously, the position of associate professor is also characterized by a high impact (increase on average 5.4 years). This relationship applies only to the studied population of scientists publishing in the Scopus database, and not all Polish scientists. Moreover, the predictor associated with working at universities other than the 10 research universities (reference category) positively influences biological age (on average by 1.3 years). Importantly, gender does not significantly affect the prediction of biological age.

A few rather interesting findings in the model come from the analysis of the effect of individual academic disciplines on predicted biological age. Scientists publishing in BUS, DEC, and ENER have similar biological age as those publishing in HUM (which is a reference category). Only assignment to the HEALTH discipline has a significantly positive effect on biological age (by two years on average). Publishing in the vast majority of disciplines has a negative effect on biological age compared to HUM (except for the small HEALTH discipline, the small DEC discipline, and the large BUS and ENER disciplines). Assignment to the PHYS, CHEM, IMMU, BIO, PHARM, MATH, MATER, COMP, and MED disciplines has the strongest negative effect on biological age (4–6 years on average). These disciplines belong to the traditional STEMM field, which clearly indicates an earlier start to the academic career (as measured by academic age). The seven other STEMM disciplines (i.e., AGRI, ENVIR, CHEMENG, DENT, ENG, EARTH) also have a negative impact compared to HUM by an average of 2–2.5 years. In contrast, the non-STEMM domains (particularly the largest, such as ECON, PSYCH, SOC) also show a negative impact, but clearly smaller than the others (1–1.5 years). This analysis demonstrates the overwhelming supremacy of STEMM disciplines in internationally visible scholarly production.

The analysis of standardized coefficients reveals that the most important predictor of biological age is academic age, the value for which was as high as 0.523 and it was much higher than the second most influential factor—position (associate professor had a value of 0.234, while full professor 0.357). Therefore, these two characteristics are definitely the strongest determinants of biological age. Among the disciplines, CHEM, BIO, MED, PHYS, MATER, and MATH (0.13, 0.13, 0.12, 0.11, 0.11, 0.09 respectively) were characterized by relatively high, although much lower than previously mentioned, values of standardized coefficients, which indicates a strong influence of STEMM disciplines on age. The remaining disciplines were characterized by the value of the standardized coefficient significantly below 0.1. Moreover, IDUB did not turn out to be a strong predictor, with a value of only 0.06. The phenomenon of collinearity (significant correlation of the vector of independent variables) did not occur in the model—the values of VIF coefficients in almost every case were lower than 2 (with 4 as a value allowing to state the occurrence of significant interdependence).

Table 4

Linear model coefficients (dependent variable: biological age; reference categories: gender—female, institutional type: research-intensive IDUB, ASJC discipline—HUM, and academic position—assistant professor) (N = 20,596).

R² = 0.615, SE = 6.59	Estimate	Standardized	SE	t-value	Pr(> t)	VIF
(Intercept)	38.269	-	0.250	152.998	<0.001	-
Academic age	0.595	0.523	0.007	90.076	<0.001	1.867
Gender: Male scientists	0.094	0.004	0.101	0.934	0.350	1.179
Institutional research intensity: Rest	1.253	0.055	0.103	12.206	<0.001	1.089
ASJC discipline: AGRI	-2.460	-0.072	0.278	-8.846	<0.001	1.585
ASJC discipline: BIO	-5.352	-0.129	0.299	-17.876	<0.001	1.226
ASJC discipline: BUS	0.290	0.004	0.371	0.780	0.436	1.512
ASJC discipline: CHEM	-5.983	-0.131	0.312	-19.182	<0.001	1.159
ASJC discipline: CHEMENG	-2.709	-0.034	0.416	-6.517	0.000	1.383
ASJC discipline: COMP	-3.768	-0.070	0.336	-11.231	<0.001	1.018
ASJC discipline: DEC	-1.910	-0.008	1.001	-1.909	0.056	1.027
ASJC discipline: DENT	-2.616	-0.014	0.853	-3.068	0.002	1.417
ASJC discipline: EARTH	-2.118	-0.042	0.322	-6.576	0.000	1.123
ASJC discipline: ECON	-1.724	-0.019	0.457	-3.775	0.000	1.102
ASJC discipline: ENER	-0.818	-0.008	0.495	-1.653	0.098	2.086
ASJC discipline: ENG	-2.162	-0.069	0.275	-7.875	0.000	1.521
ASJC discipline: ENVIR	-2.515	-0.060	0.297	-8.463	<0.001	1.022
ASJC discipline: HEALTH	2.039	0.010	0.937	2.177	0.029	1.367
ASJC discipline: IMMUN	-5.787	-0.037	0.720	-8.040	0.000	1.044
ASJC discipline: MATER	-4.705	-0.105	0.309	-15.243	<0.001	1.535
ASJC discipline: MATH	-4.912	-0.090	0.336	-14.606	<0.001	1.356
ASJC discipline: MED.	-3.698	-0.122	0.273	-13.568	<0.001	2.052
ASJC discipline: PHARM	-5.534	-0.053	0.514	-10.765	<0.001	1.094
ASJC discipline: PHYS	-5.663	-0.105	0.338	-16.744	<0.001	1.441
ASJC discipline: PSYCH	-1.533	-0.015	0.490	-3.130	0.002	1.098
ASJC discipline: SOC	-0.878	-0.015	0.336	-2.614	0.009	1.332
ASJC discipline: VET	-4.115	-0.046	0.456	-9.031	<0.001	1.119
Academic position: associate professor	5.407	0.234	0.114	47.567	<0.001	1.279
Academic position: full professor	10.923	0.357	0.171	63.708	<0.001	1.764

In addition, three models were estimated for the age of obtaining PhD, habilitation, and full professorship using the same independent variables. However, these models had a negligible goodness of fit to the empirical data (R^2 ranged from 0.005 for the model for the age of PhD to 0.09 for the age of professorship). In fact, the only variable that showed a significant association with each model was academic age, but the magnitude of the coefficient for this variable did not exceed 0.07, thereby implying that an increase in academic age by one year only marginally affects the increase in age at degree completion. In addition, other characteristics such as gender and field significantly affected the age of degree attainment only for the models for habilitation and professorship. These models are not analyzed in this paper.

5. Discussion and Conclusions

Our detailed examination of the entire population of Polish academic scientists visible in the last decade in global science and holding at least a PhD ($N = 20,569$) clearly indicates that using academic age as a proxy for biological age in academic career studies works well for STEMM disciplines.

However, for non-STEMM disciplines (particularly for HUM, SOC, ECON, or BUS), this usage performs dramatically worse. This negative conclusion is particularly important for systems that are only recently (in the last one or two decades) more widely visible in global academic journals—that is, for countries classified as “developing” (which, apart from Poland, also include Portugal, Slovakia, Bulgaria, Romania, Iran, and Turkey, among others) and as “lagging” (such as Nigeria or Indonesia) (Wagner, 2008, p. 88).

Thus, the differences in the usefulness of the usage of academic age in examining academic careers at the micro-level of the individual scientist apply to varying extents to different clusters of countries: on the one hand, there are countries that form the core of global science and on the other hand there are countries that form its peripheries. The idea of core and peripheries (Olechnicka et al., 2018) dates back to the work of Immanuel Wallerstein (1976), but it is increasingly criticized today on the basis of global publication patterns. For example, Marginson (2021) indicates the growing role of newcomers in the global circulation of knowledge, which changes the traditional core-periphery picture and challenges the traditional narratives about where science is produced.

For countries from the scientifically “advanced” cluster, constituting the core of the global circulation of knowledge from the outset, the academic age used as a proxy for biological age works very well for all disciplines (as in Quebec, Canada) (Nane et al., 2017; Costas et al., 2015). In the case of the second cluster of scientifically “developing” countries, represented here by Poland, the proxy works well only for scientists in STEMM disciplines. For STEMM scientists, it works in the 85%–90% range—that is, STEMM scientists tend to operate according to globally valid patterns in which the relationship between academic age and biological age is highly predictable. Thus, for example, for scientists working in chemistry, the correlation between academic age and biological age is very high (with correlation coefficient $r = 0.89$), which is similar to physics and astronomy ($r = 0.88$) and mathematics ($r = 0.85$). The correlation is radically lower for non-STEMM scientists for whom the proxy works only in the 35%–50% range: it is the lowest for HUM ($r = 0.35$), followed by BUS ($r = 0.42$), ECON ($r = 0.43$), and SOC ($r = 0.49$).

The Polish case confirms the existence of significant differences in global publication patterns in the last three decades between scientifically “developing” countries and “advanced” countries (Wagner, 2008, p. 88). Poland is a good example of a country that is beyond the traditional academic centers of science, where the number of publications in global academic journals has been systematically increasing. While Western countries have been functioning in common English-language global science networks for several decades, and with radically increasing intensity since 1989, Poland and other scientifically developing countries have been participating in these networks on a larger scale only for the last 10–15 years. Geopolitical differences have resulted in the post-communist countries in the European Union (EU), including Poland (and other poorer countries of the world), becoming massively visible in Scopus or Web of Science journals relatively recently. This is clearly

evidenced by the publication data. In contrast, Western countries have been operating on a large scale (and through a high percentage of their scientists) in the indexed journal system since the very outset. These geopolitical differences have practical implications for the use of academic age in micro-level studies on scientists.

Our research confirms that the globalization of science is associated with significant cross-disciplinary differentiation between STEMM and non-STEMM disciplines. The micro-level data that we analyzed at the individual level suggest a delayed participation of social scientists and humanists in global science networks as opposed to the continuing presence of natural scientists in them. Thus, while science is globalizing fast, it is globalizing unequally: faster in STEMM fields than in non-STEMM fields, with practical implications for predicting biological age from academic age in academic career studies.

Our study supports the idea that it is worthwhile to collect complete data at the individual level: for detailed analyses of academic career in its various dimensions (research productivity, international research collaboration, international mobility), it is useful to have birth dates of all researchers from all science sectors, as in the Polish POL-on system currently under expansion. In such systems, national analyses do not require proxies for biological age—that is, there is no need to use academic age. It is useful to establish and develop full administrative, biographical, and publication databases in the form of Current Research Information Systems (CRIS are institutional and national systems with high standards of coverage, quality, and standardization, usually containing information regarding researchers and research groups, their projects, funding, outputs and outcomes) (see Sivertsen, 2019, p. 667).

Our research also suggests that in scientifically developing countries, academic age as a proxy for biological age must be used more cautiously than in advanced countries and ideally must only be used for STEMM disciplines. The inconsistencies between the two ages in non-STEMM disciplines are greater there and the correlations are radically lower, thereby making the proxy an inadequate analytical tool.

Furthermore, the Polish case refers to several parallel processes that are evident in global science: (1) presence on the peripheries of global science for geopolitical reasons; (2) low but improving individual research productivity; (3) delayed entry into global science and globally indexed academic journals; and (4) permanent underfunding of the academic research system (in the first two decades following 1989). These processes are characteristic of much of the world and affect, to varying degrees, both post-communist countries in Europe and Central Asia, as well as Latin American and African countries.

Academic institutions across the world have complete data on the biological age of their academic staff. In contrast, the availability of such detailed data at the macro level of countries remains limited, thereby impeding the possibility of rigorous analysis of the academic profession in demographic terms. Our study can be replicated for individual institutions, provinces (or other administrative districts), and smaller subpopulations of researchers (such as beneficiaries of grant programs). It can be replicated wherever reliable data on the biological age of researchers and reliable historical publication data are in place; thus, the results for other institutional, geographic, and national contexts can be compared with ours, thereby leading to a more comprehensive picture. Global comparative studies of

academic careers require reliable demographic data on a national scale, the production of which this study strongly encourages.

Acknowledgments

We gratefully acknowledge the support of the Ministry of Education and Science through its Dialogue grant 0022/DLG/2019/10 (RESEARCH UNIVERSITIES). We also gratefully acknowledge the assistance of the International Center for the Studies of Research (ICSR) Lab, a cloud-based service provided by Elsevier for research purposes. We are particularly grateful to Kristy James, Data Scientist in the Lab, for her continuous support. We are also grateful to Łukasz Szymula, an excellent doctoral student in the Center for Public Policy Studies.

References

- Abramo, G., Aksnes, D. W., & D'Angelo, C. A. (2020). Comparison of research productivity of Italian and Norwegian professors and universities. *Journal of Informetrics*, 14(2), 101023.
- Abramo, G., C. A. D'Angelo, Murgia, G. (2016). The combined effect of age and seniority on research performance of full professors. *Science and Public Policy*, 43(3), 301–319.
- Abramo, G., D'Angelo, C.A., Solazzi, M. (2011). The relationship between scientists' research performance and the degree of internationalization of their research. *Scientometrics*, 86, 629–643.
- Aksnes, D. W., Rørstad, K., Piro, F. N., & Sivertsen, G. (2011) Are female researchers less cited? A large scale study of Norwegian researchers. *Journal of the American Society for Information Science and Technology*, 62(4), 628–636.
- Antonowicz, D., Kulczycki, E., & Budzanowska, A. (2020). Breaking the deadlock of mistrust? A participative model of the structural reforms in higher education in Poland. *Higher Education Quarterly*, On-line first February 14, 2020, <https://doi.org/10.1111/hequ.12254>.
- Aref S., Zagheni E., West J. (2019) The demography of the peripatetic researcher: evidence on highly mobile scholars from the web of science. In I. Weber et al. (eds) *Social Informatics. SocInfo 2019. Lecture Notes in Computer Science* (vol 11864). Springer.
- Badar, K., M. Hite, J., & F. Badir, Y. (2014). The moderating roles of academic age and institutional sector on the relationship between co-authorship network centrality and academic research performance. *Aslib Journal of Information Management*, 66(1), 38–53. doi:10.1108/ajim-05-2013-0040
- Barnard, H., Cowan, R., Fernandez de Arroyabe Arranz, M., & Muller, M. (2015). The role of global connectedness in the development of indigenous science in middle income countries. In Daniele Archibugi & Andrea Filippetti (Eds.), *The Handbook of Global Science, Technology, and Innovation* (pp. 386–410). Wiley-Blackwell.
- Bieliński, J., & Tomczyńska, A. (2018). The ethos of science in contemporary Poland. *Minerva*, 57(2), 151–173.
- Bonaccorsi, A., Daraio, C. (2003). Age effects in scientific productivity. The case of the Italian National Research Council (CNR). *Scientometrics*, 58(1), 49–90.
- Chan, H.F., Torgler, B. (2020). Gender differences in performance of top cited scientists by field and country. *Scientometrics*, 125, 2421–2447.
- Cole, S. (1979). Age and scientific performance. *American Journal of Sociology*, 84(4), 958–977.
- Coomes, O.T., Moore, T., Paterson, J., Breau, S., Ross, N.A. & Roulet, N. (2013). Academic performance indicators for departments of geography in the United States and Canada. *The Professional Geographer*, 65(3), 433–450.
- Costas, R., Bordons, M. (2005). Bibliometric indicators at the micro-level: Some results in the area of natural resources at the Spanish CSIC. *Research Evaluation*, 14(2), 110–120
- Costas, R., Nane, GF., & Larivière, V. (2015). Is the year of first publication a good proxy of scholars academic age? In A.A. Salah, Y. Tonta, A.A. Akdag Salah (Eds.), *Proceedings of the 15th international conference on scientometrics and informetrics* (pp. 988–998). Istanbul: Bogaziçi University Printhouse.

- Costas, R., van Leeuwen, T.N, Bordons, M. (2010). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: the effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology*, 61(8), 1564–1581
- Costas, R., van Leeuwen, T.N. & Bordons, M. (2010). Self-citations at the meso and individual levels: Effects of different calculation methods. *Scientometrics*, 82, 517–537.
- Costas, R, Bordons, M. (2007). A classificatory scheme for the analysis of bibliometric profiles at the micro level. *Proceedings of ISSI 2007: 11th international conference of the ISSI, Vols I and II*, p.226–230.
- Edwards, D., & Smith, T. F. (2009). Supply issues for science academics in Australia: now and in the future. *Higher Education*, 60(1), 19–32. doi:10.1007/s10734-009-9283-0
- Elsevier, (2020). *The researcher journey through a gender lens*. Amsterdam: Elsevier.
- Enamorado, T., Fifield, B., Imai, K., (2019). Using a probabilistic model to assist merging of large-scale administrative records, *American Political Science Review*, 113(2), 353–371.
- Enamorado, T., Fifield, B., Imai, K., 2019, Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review*, 113(2), 353–371.
- Feldy, M., & Kowalczyk, B. (2020). The ethos of science and the perception of the Polish system of financing science. *European Review*, 28(4), 599–616.
- Fellegi I. P., Sunter A. B., 1969, A theory for record linkage. *Journal of the American Statistical Association* 64, 1183–1210.
- Gingras Y., Larivière V., Macaluso B., Robitaille J.P. (2008). The effects of aging on researchers' publication and citation patterns. *PLoS One*. 3(12), e4048. doi: 10.1371/journal.pone.0004048. Epub 2008 Dec 29. PMID: 19112502; PMCID: PMC2603321.
- Gu, X., & Blackmore, K. (2017). Quantitative study on Australian academic science. *Scientometrics*, 113(2), 1009–1035. doi:10.1007/s11192-017-2499-3
- Gui, Q., Liu, C., & Du, D. (2019). Globalization of science and international scientific collaboration: A network perspective. *Geoforum*, 105, 1–12. doi:10.1016/j.geoforum.2019.06.017
- Gui, Q., Liu, C., & Du, D. (2019). Globalization of science and international scientific collaboration: A network perspective. *Geoforum*, 105, 1–12. doi:10.1016/j.geoforum.2019.06.017
- Guns, R., Eykens, J., and Engels, T. C. E. (2019). To what extent do successive cohorts adopt different publication patterns? Peer review, language use, and publication types in the social sciences and humanities. *Frontiers in Research Metrics and Analytics* 3. doi: 10.3389/frma.2018.0003
- GUS (2020). *Higher Education Institutions and Their Finances in 2019*. Warsaw: GUS (Central Statistical Office).
- Harron K., Dibben C., Boyd J., Hjern A., Azimaee M., Barreto M. L., Goldstein H., (2017). Challenges in administrative data linkage for research. *Big Data & Society*, July–December 2017, 1–12.
- Herzog T. N., Scheuren F. J., Winkler W. E., (2007). *Data Quality and Record Linkage Techniques*, Springer.
- Kulczycki, E., & Korytkowski, P. (2020). Researchers publishing monographs are more productive and more local-oriented. *Scientometrics*, 125, 1371–1387.
- Kulczycki, E., Korzeń, M., & Korytkowski, P. (2017). Toward an excellence-based research funding system: Evidence from Poland. *Journal of Informetrics*, 11(1), 282–298.
- Kwiek, M. (2015a). The internationalization of research in Europe. A quantitative study of 11 national systems from a micro-level perspective. *Journal of Studies in International Education*, 19(2), 341–359.
- Kwiek, M. (2015b). Academic generations and academic work: Patterns of attitudes, behaviors and research productivity of Polish academics after 1989. *Studies in Higher Education*, 40(8), 1354–1376.
- Kwiek, M. (2016). The European research elite: A cross-national study of highly productive academics across 11 European systems. *Higher Education*, 71(3), 379–397.
- Kwiek, M. (2018a). High research productivity in vertically undifferentiated higher education systems: Who are the top performers? *Scientometrics*, 115(1), 415–462.
- Kwiek, M. (2018b). International research collaboration and international research orientation: Comparative findings about european academics. *Journal of Studies in International Education*, 22(1), 1–25.
- Kwiek, M. (2019). *Changing European academics. A comparative study of social stratification, work patterns and research productivity*. London and New York: Routledge.
- Kwiek, M. (2020a). What large-scale publication and citation data tell us about international research collaboration in Europe: Changing national patterns in global contexts. *Studies in Higher Education*, 45, online first April 10, 2020, 1–21.
- Kwiek, M. (2020b). Internationalists and locals: International research collaboration in a resource-poor system.

- Scientometrics*, 124, 57–105.
- Kwiek, M., & Szadkowski, K. (2019) Higher Education Systems and Institutions: Poland. In P.N. Texteira & J.C. Shin (Eds), *International Encyclopedia of Higher Education Systems (1–20)*. Springer.
- Kwiek, M., & Roszka, W. (2020). Gender disparities in international research collaboration: A large-scale bibliometric study of 25,000 university professors. *Journal of Economic Surveys*, online first November 18, 2020. <https://doi.org/10.1111/joes.12395>.
- Kwiek, M., Roszka, W. (2021a). Gender-based homophily in research: A large-scale study of man-woman collaboration. *Journal of Informetrics*, 15(3), article 101171. 1–38.
- Kwiek, M., Roszka, W. (2021b). Are Female Scientists Less Inclined to Publish Alone? The Gender Solo Research Gap. Under reviews, available from ArXiv: <https://arxiv.org/abs/2101.03220#>.
- Kyvik, S. (1990). Age and scientific productivity. Differences between fields of learning. *Higher Education*, 19, 37–55.
- Kyvik, S., & Aksnes, D. W. (2015). Explaining the increase in publication productivity among academic staff: A generational perspective. *Studies in Higher Education*, 40, 1438–1453.
- Kyvik, S., & Olsen, T. B. (2008). Does the aging of tenured academic staff affect the research performance of universities? *Scientometrics*, 76(3), 439–455.
- Larivière, V., Vignola-Gagné, E., Villeneuve, C. et al. (2011). Sex differences in research funding, productivity and impact: an analysis of Québec university professors. *Scientometrics*, 87, 483–498.
- Lehman, H. C. (1953). *Age and achievement*. Princeton: Princeton University Press.
- Levin, S., Stephan, P.E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, March 1991, 114–132.
- Liao C.H. (2017). Reopening the black box of career age and research performance. In J. Zhou, G. Salvendy, (Eds), *Human Aspects of IT for the Aged Population. Applications, Services and Contexts. ITAP 2017. Lecture Notes in Computer Science* (vol 10298). Springer. https://doi.org/10.1007/978-3-319-58536-9_41
- Marginson, S. (2021) What drives global science? The four competing narratives, *Studies in Higher Education*, doi: [10.1080/03075079.2021.1942822](https://doi.org/10.1080/03075079.2021.1942822)
- Marginson, S. 2018. *The new geo-politics of higher education*. Oxford: CGHE Working Paper no. 34.
- Marginson, S. and Xu, X. 2021. *Moving beyond centre-periphery science: towards an ecology of knowledge*. CGHE Working Paper no. 63, April 2021.
- Milojević S. (2012). How are academic age, productivity, and collaboration related to citing behavior of researchers? *PLoS One*, 7(11), e49176. doi: 10.1371/journal.pone.0049176. Epub 2012 Nov 7. PMID: 23145111; PMCID: PMC3492318.
- OECD (2021). Gross domestic spending on R&D (indicator). Accessed October 12, 2021. <https://doi.org/10.1787/d8b068b4-en>
- Olechnicka, A., Ploszaj, A., & Celinska-Janowicz, D. (2019). *The geography of scientific collaboration*. London and New York: Routledge.
- Pelz, D.C., Andrews, F.W. (1976). *Scientists in organizations*. New York: Wiley.
- Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2014). Within- and between-department variability in individual productivity: the case of economics. *Scientometrics*, 102(2), 1497–1520.
- Petersen, A.M. (2015). On the impact of super ties in scientific careers. *Proceedings of the National Academy of Sciences Aug 2015*, 112 (34), E4671-E4680. doi: 10.1073/pnas.1501444112
- Radicchi, F., Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, 97, 627–637.
- Robinson-Garcia, N., Costas, R., Sugimoto, C.R., Larivière, V., Nane, G.F. (2020). Task specialization across research careers. *eLife*, 9, e60586 doi: [10.7554/eLife.60586](https://doi.org/10.7554/eLife.60586)
- Rørstad, K., & Aksnes, D.W. (2015). Publication rate expressed by age, gender and academic position—A large-scale analysis of Norwegian academic staff. *Journal of Informetrics*, 9, 317–333.
- Sayers A., Ben-Shlomo Y., Blom A. W., Steele F. (2016). Probabilistic record linkage. *International Journal of Epidemiology*, 45(3), 954–964.
- Shaw, M. A. (2019). Strategic instrument or social institution: Rationalized myths of the university in stakeholder perceptions of higher education reform in Poland. *International Journal of Educational Development*, 69, 9–21.
- Simoes, N., Crespo, N. (2020). A flexible approach for measuring author-level publishing performance. *Scientometrics*, 122, 331–355.

- Sivertsen G. (2019). Developing current research information systems (CRIS) as data sources for studies of research. In W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (Eds), *Springer Handbook of Science and Technology Indicators*. Springer.
- Stephan, P. (2012). *How economics shapes science*. Cambridge, MA: Harvard University Press.
- Stephan, P. E., Levin, S. G. (1992). *Striking the mother lode in science: the importance of age, place, and time*. New York: Oxford University Press.
- Stern N. (1978). Age and achievement in mathematics: A case-study in the sociology of science. *Social Studies of Science*, 8(1), 127-140.
- Sugimoto, C. R., Sugimoto, T. J., Tsou, A., Milojevic', S., & Larivie're, V. (2016). Age stratification and cohort effects in scholarly communication: A study of social sciences. *Scientometrics*, 109(2), 997–1016.
- Sugimoto, C. R., Sugimoto, T. J., Tsou, A., Milojevic', S., & Larivie're, V. (2016). Age stratification and cohort effects in scholarly communication: A study of social sciences. *Scientometrics*, 109(2), 997–1016. doi:10.1007/s11192-016-2087-y.
- Van den Besselaar, P., & Sandström, U. (2016) Gender differences in research performance and its impact on careers: A longitudinal case study. *Scientometrics*, 106(1), 143–162.
- Wagner, C. S. (2008). *The new invisible college. Science for development*. Brookings Institution Press.
- Wais, K. (2016). Gender Prediction Methods Based on First Names with genderizeR. *The R Journal*, 8(1), 17–37
- Wang, D., & Barabási, A. (2021). *The science of science*. Cambridge: Cambridge University Press. doi:10.1017/9781108610834
- Wang, W., Yu, S., Bekele, T. M., Kong, X., & Xia, F. (2017). Scientific collaboration patterns vary with scholars' academic ages. *Scientometrics*, 112(1), 329–343.
- Wallerstein, I. (1976). Semi-peripheral countries and the contemporary world crisis. *Theory and Society*, 3, 461–83.
- Wildgaard, L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. *Scientometrics*, 104, 873–906.
- Zhang, L., & Glänzel, W. (2012). Where demographics meets scientometrics: Towards a dynamic career analysis. *Scientometrics*, 91(2), 617–630. doi:10.1007/s11192-011-0590-8
- Zuckerman, H., Merton, R.K. (1973). Age, aging, and age structure in science. In M.W. Riley, M. Johnson, A. Foner (Eds), *Aging and Society, Vol. 3. A Sociology of Age Stratification*. Russell Sage Foundation.

Data Appendix

Table 5

Structure of the sample, all Polish internationally visible university professors, by gender, age group, academic position, and discipline (Young scientists indicate those aged 39 and younger, Middle-aged indicates those aged 40–54, and Older indicates those aged 55 and over).

		Female			Male			Total		
		N	% col	% row	N	% col	% row	N	% col	% row
Age group	Young	2828	32.3	49.5	2880	24.3	50.5	5708	27.7	100.0
	Middle-aged	4368	49.9	44.8	5378	45.4	55.2	9746	47.3	100.0
	Older	1564	17.9	30.4	3578	30.2	69.6	5142	25.0	100.0
	Total	8760	100.0	42.5	11836	100.0	57.5	20596	100.0	100.0
Academic position	Assistant Prof.	5557	63.4	48.4	5913	50.0	51.6	11470	55.7	100.0
	Associate Prof.	2443	27.9	39.1	3799	32.1	60.9	6242	30.3	100.0
	Full Professor	760	8.7	26.4	2124	17.9	73.6	2884	14.0	100.0
	Total	8760	100.0	42.5	11836	100.0	57.5	20596	100.0	100.0
Discipline (ASJC) – STEM	AGRI	1215	13.9	54.0	1037	8.8	46.0	2252	10.9	100.0
	BIO	900	10.3	61.7	558	4.7	38.3	1458	7.1	100.0
	CHEM	612	7.0	52.1	562	4.7	47.9	1174	5.7	100.0
	CHEMENG	155	1.8	40.6	227	1.9	59.4	382	1.9	100.0
	COMP	141	1.6	17.1	684	5.8	82.9	825	4.0	100.0
	DEC	20	0.2	43.5	26	0.2	56.5	46	0.2	100.0
	EARTH	333	3.8	34.9	622	5.3	65.1	955	4.6	100.0

	ENER	60	0.7	26.0	171	1.4	74.0	231	1.1	100.0
	ENG	412	4.7	15.0	2336	19.7	85.0	2748	13.3	100.0
	ENVIR	735	8.4	51.9	680	5.7	48.1	1415	6.9	100.0
	IMMU	72	0.8	75.8	23	0.2	24.2	95	0.5	100.0
	MATER	417	4.8	34.2	801	6.8	65.8	1218	5.9	100.0
	MATH	213	2.4	26.4	595	5.0	73.6	808	3.9	100.0
	MED	1620	18.5	55.2	1314	11.1	44.8	2934	14.2	100.0
	PHARM	148	1.7	69.8	64	0.5	30.2	212	1.0	100.0
	PHYS	147	1.7	17.6	690	5.8	82.4	837	4.1	100.0
	BUS	280	3.2	53.0	248	2.1	47.0	528	2.6	100.0
	DENT	48	0.5	73.8	17	0.1	26.2	65	0.3	100.0
	ECON	143	1.6	50.4	141	1.2	49.6	284	1.4	100.0
	HEALTH	20	0.2	37.7	33	0.3	62.3	53	0.3	100.0
	HUM	405	4.6	51.0	389	3.3	49.0	794	3.9	100.0
	PSYCH	148	1.7	62.7	88	0.7	37.3	236	1.1	100.0
	SOC	381	4.3	50.8	369	3.1	49.2	750	3.6	100.0
	VET	135	1.5	45.6	161	1.4	54.4	296	1.4	100.0
	Total	8760	100.0	42.5	11836	100.0	57.5	20596	100.0	100.0

Bio:

Marek Kwiek is Professor and Chairholder, UNESCO Chair in Institutional Research and Higher Education Policy, and Director of the Institute for Advanced Studies in Social Sciences and Humanities (IAS), University of Poznan, Poland. His research area is quantitative studies of science and higher education research and policy. Professor Kwiek has published approximately 220 papers and book chapters and 4 sole-authored monographs and he publishes extensively in major academic journals. His most recent papers are in “*Higher Education*” (2021), “*Journal of Informetrics*” (2021), “*Studies in Higher Education*” (2021), “*Scientometrics*” (2020), and “*Journal of Economic Surveys*” (2020). His recent monograph, *Changing European Academics: A Comparative Study of Social Stratification, Work Patterns, and Research Productivity* was published by Routledge (2019). He has been a Principal Investigator or Country Team Leader in 25 international higher education research projects. He has served as an international higher education policy expert for the European Commission, USAID, OECD, World Bank, UNESCO, Council of Europe, and the European Parliament. Moreover, he spent three years at North American universities, including the University of Virginia, UC Berkeley, National Endowment for Democracy in Washington, DC, and McGill University. He is also an editorial board member for *Higher Education Quarterly*, *European Educational Research Journal*, *British Educational Research Journal*, and *European Journal of Higher Education* as well as for a Springer book series, *Higher Education Dynamics*. In 2018, he was elected an ordinary member of the European Academy of Sciences and Arts (EASA) in Salzburg, Austria; moreover, in 2021, he was elected an ordinary member of the *Academia Europaea* in London, the European Academy of Sciences. Contact: kwiekm@amu.edu.pl. www.cpp.amu.edu.pl; www.ias.amu.edu.pl

Dr. Wojciech Roszka is an assistant professor in the Department of Statistics, the Institute of Informatics and Quantitative Economics, Poznan University of Economics and Business. His area of research interest and expertise includes integration of data from different sources with specific emphasis on probabilistic record linkage, statistical matching, data fusion, and microsimulation modelling. He uses these methods primarily in the analysis of the higher education market, the labor market, and demographics. He is also involved in projects related to marketing research, time series forecasting and methods of multivariate statistics. He is a member of the Polish Statistical Society and works in the Center for Public Policy Studies of the University of Poznan.