In search for an open-science: Traumatic Stress Researcher’s data-sharing behaviour in relation to Biological Age, Professional Age and Career Stage

Bachelor Thesis

By:

Hakan Culha (s2300559)

1st Supervisor: L.I.M. Lenferink PhD (Lonneke)

2nd Supervisor: Mark Tempelman

Behavioural, Management and Social Sciences (BMS) Department

Positive Clinical Psychology and Technology

University of Twente

2022

APA-Version: 7

Abstract

Background. Sharing one's precious data is a practice that is both still uncommon and yet extremely valuable to scientific progress. Many researchers agree on the importance of data sharing, yet across the disciplines there seems to be a disagreement on the actual sharing practice. Fostering the understanding of why that is, this study is exploring age, as a biological and professional dimension, and career stage as predictors of data sharing behaviour.

Methods. The sample consist of 193 participants who were recruited by a means of a purposive sampling method. This study utilizes a binary logistic regression analysis that compares a univariate to a multivariate model. The independent variables in the equation are biological age (BA), as a categorical variable that accounts for the time a person has been alive, professional age (PA), as a continuous variable that accounts for time spent in a respective field and career stage (CS), as a categorical variable, that accounts for the level of expertise. Data sharing behaviour (DSB) is treated as a binary dependent variable reflecting if a researcher had shared data at least once at any point in time and serves as the outcome variable. Lastly, PA*CS were tested for an interaction effect.

Results. PA turned out to be significantly positively associated with having shared data in both the univariate and multivariate model. Different to BA and CS, who were both significantly associated with DSB in the univariate model, were insignificant and weaker associated in the multivariate model. Moreover, there was no significant moderation effect found for PA*CS on DSB.

Discussion. Overall, PA seems to be a good fit for a model that tries to explore the individual predictors of DSB. Interestingly enough, the relationship between BA and DSB might need further research in order to include possible confounding variables. CS on the other hand might be a moderating variable that influences the strength of the association between PA and DSB, although in this study results were non-significant, therefore further research is needed to solidify either claim. Lastly it is suggested to expand the exploration of correlates of DSB to personality traits, sense of community theory and other factors that might accompany BA, PA, and CS.

Keywords. Age, professional age, years of research, experience, career stage, data sharing behaviour, traumatic stress field,

1. Introduction

It is commonly believed that elder people are much wiser as younger folks, as they have simply lived for longer. Unsurprisingly then it is common for them to teach younger people the 'ropes', meaning sharing what they know. In science, this is not as straightforward. Sharing one's knowledge can go further than just teaching people how to do a specific task. In the scientific field, sharing knowledge may also be seen as sharing one's precious data, which is then to be understood as *data sharing*. Data sharing can be roughly defined as providing the raw data through uploading it into a data repository or sharing it via personal means upon request (Global Collaboration on Traumatic Stress, n.d.). Nowadays it is easier than ever to share data, as with the implementation of the world wide web, the world became more connected. Researchers from across the globe can team up and work together on projects or stay in touch with the newest discoveries as long as they have access to the internet. In fact, nowadays it has even become increasingly more accessible to read various articles online (Lawrence, 2001, p. 521), which enables that the collection and sharing of data can have a faster and greater impact on the scientific community (Ferguson et al., 2014).

Following this line of thought, more researchers agree that inaccessibility to data, meaning to not share data, is one major roadblock on their way to answer some more complex scientific questions (Tenopir et al., 2011; Tenopir et al., 2020). Empirical sciences all the more have always relied on available data for progress, which is why a scarcity of it may lead to major setbacks (Houtkoop et al., 2018). Not only is it slowing down the progress, but also lowering the general transparency and validity of results. The unavailability of data denies others the possibility to re-evaluate results through re-analysing and re-interpreting the data, which in turn could have the potential to increase the reliability of the past work on which newer discoveries can rely on as a fundament (Tedersoo et al., 2021). This creates an interesting conundrum nonetheless, as the general acceptance of sharing data is rising, yet this rise is not reflected directly in the actual data sharing practices of researchers (Tenopir et al., 2020). This is especially relevant for the field of psychology, as in this field data sharing compared to other fields is still uncommon (Houtkoop et al., 2018). One of the fields in psychology where data sharing can be beneficial is that of the traumatic stress field. Early traumatic experiences can develop into more complex disorders, for some it may even turn out to be a lifelong problem with traumatic stress being associated to lasting changes in the brain (Bremner, 2006). It is estimated that most individuals (70%) on a global scale will encounter a traumatic event at least once in their lifetime (Olff et al., 2019).

A deeper understanding of all the intricates of the complexity that traumatic stress events can have, is set out to be greatly beneficial in the long run. By combing multiple small data sets and analysing it through various angles, a better conceptualisation of follow up disorders such as PTSD for example, can be constructed, expanding its understanding beyond just fear to, for example, symptoms of emotional dysregulations or disassociations (Olff et al., 2019). This idea is not foreign in the traumatic stress field as in an article by Nancy Kassam-Adams and Miranda Olff (2020), they found that by merging data, intervention effectiveness and trauma-related cognitive models can be improved. Another innovative idea is to use machine learning approaches (ML) for stress pathology diagnostic and predictive models that utilises great amounts of data from various sources (Schultebraucks & Galatzer-Levy, 2019). Overall, it also honours the impact of each individual data point collected as it maximises the effect it can have (Brakewood & Poldrack, 2013).

Imminently, data sharing is nonetheless not a main goal in and of itself. In science, data sharing is only as useful as its secondary function, which is to foster further discoveries and innovations through reuse (Wilkinson et al., 2016); a lack thereof can be seen as a huge obstacle in this pursuit (Tenopir et al., 2015). Thus, the apparent lack of sharing data in the field of traumatic stress is to be taken more seriously, as reuse of data depends on available data first, which if not available defeats the whole point.

Factors Influencing Data Sharing

This gives rise to various papers on several factors that influence data sharing behaviour, one most prevalent in psychology is the fear that discoveries might get invalidated by alternative analyses (Houtkoop et al., 2018). As established before, this fear runs counter to the idea of empirical sciences yet remains understandable in the greater sense that scientific recognition of the researcher and their career is important in this community to strive. Additionally, individual-level predictors such as perceived career benefits, perceived effort, scholarly altruism as well as institutional-level predictors such as regulative pressure by journals or normative pressure are all found to be significantly related to data sharing behaviour across all disciplines (Kim & Stanton, 2015). Other papers also address generational and demographical factors such as age and career stage, wherein career stage is found to be significantly affecting data sharing behaviour, with later career stages sharing less data than earlier career stages (Campbell et al., 2019), and age is also found to be significantly affecting the researchers perception of sharing data, with younger people being more in favour of making their data available, postulating a position that sharing and reusing data is incredibly important for progress and their ability to answer scientific questions in comparison to older people (Campbell et al., 2019; Tenopir et al., 2011).

Current Study

Puzzling nonetheless is that although younger people tend to be more in favour of sharing data, in reality older people are more active when it comes to actual data sharing behaviour (Campbell et al., 2019; Tenopir et al., 2015). This creates a problem in research and is to be further investigated, illustrating a gap in knowledge. Additionally, one paper investigating this phenomenon found out that, as we age people seem to be less worried about recognition in the workplace and getting promotional opportunities, but instead become more concerned with their own work and focus on autonomy (Tenopir et al., 2015).

Combining this with a study about generational knowledge sharing, it is suggested that age needs to be further differentiated (Widén et al., 2020). Age is to be seen as in (1) biological age, which represents the time a person has been alive measured in years and (2) professional age, which represents the time spent in a respective work field. This is in line with the findings that experience in a work field is positively associated with data sharing behaviour, with researchers that used secondary data for their own research (Zhu, 2019).

The following hypotheses were formulated to better understand the research question "To what extend is biological age, professional age and career stage related to data sharing behaviour in the traumatic stress field?". First, whether professional age is positively associated with data sharing behaviour (H1). Second, if is negatively associated with data sharing behaviour (H2), as sharing is often in association with career opportunities and recognition in the workplace. Third, an interesting question is also how professional age ties in with career stage, as later career stages are associated with lower data sharing behaviours yet experience in a work field is associated with higher data sharing behaviour (Campbell et al., 2019; Zhu 2019) thus to test whether higher or lower professional ages interact with higher or lower career stages, an interaction term will be added.

2. Methods

Participants and Procedure

The study population consisted of researchers that worked within the traumatic stress field at some point in their career, including people in training. Participation included completing a survey online. The online survey has been translated into the following 7 languages: English, Japanese, Spanish, French, Korean, Brazilian Portuguese and Arabic.

The majority of the sample was recruited in collaboration with Lonneke Lenferink, Nancy Kassam-Adams, and the Global Collaboration on Traumatic Stress. All participants gave informed consent before each test and the study was approved by the ethics committee of IRB of the Children's Hospital of Philadelphia. The study is carried out in multiple languages with each survey having their own link. The Global Collaboration on Traumatic Stress and their scientific societies helped to spread the survey on a global scale, via online announcements, social media, word of mouth and personal communications. The survey has been uploaded into a web-based online survey tool, REDCap.com (Research Electronic Data Capture). During the collection of participants, a "recruitment letter" has been used by the 4 other researchers, which can be found in the Appendix (A). The rest of the sample were participants collected by four other researchers that are writing their papers alongside this one. These participants were contacted via a recruitment letter sent by email. Each researcher was asked to successfully email at least 35 participants. A document to report the names of each participant that has been asked to participate was created in order to avoid potential overlap between the researchers. Some emails were not sent as the email address was no longer in use, changed or due to other circumstances such as an automatic response that explained that the participant may not be available at the time.

The participants were able to click on the link provided in the recruitment letter to get redirected to the main page and from there on the participants could select one of the available languages that the survey has been translated to. Doing so will redirect the participant to the respectively translated versions of the survey. On each section, instructions were given on how to fill in the survey. After the demographics part, a definition of data sharing, data re-use and metadata was provided to further guide the participants and avoid ambiguity with what it means to share or re-use data. At the end, space was provided to share any additional comments regarding their views on data sharing or re-use.

Materials/Measure

The materials used consist of a survey, see appendix (B), that used adapted portion from other research (Kim & Stanton, 2013; Kim & Stanton, 2015; Kim & Yoon, 2017) and newly developed items. The survey begins with a small insert in which the reasons and the relevant stakeholders are explained for this research. There is a disclaimer that the results will be shared on the Global Collaboration website alongside with the final data set which will be made available upon request. The survey is used for multiple studies, with each having their own unique subsets of focal points, henceforth, not all items will be needed for each individual analysis. The important items and the ones that will ultimately be used for this paper are the ones that assess the researchers' data sharing behaviour (DSB), their biological age (BA), their professional age (PA) and their career stage (CS).

Data sharing behaviour (DSB) has been assessed using a total of 6 items. The items measure the frequency of the researcher's data sharing behaviour with the options of having (1) "Never" shared data, having shared data (2) "1 or 2 times" or having shared (3) "More than 2 times". A frequency distribution for each variable has been created to get an overview of the overall variance of the items. In order to avoid computational issues due to small group sizes, the answer options were collapsed into a dichotomous variable with the options of either having (0) “Never” shared or having shared (1) “at least once”. All responses were later combined to get a final impression on whether the researcher shared at least once, no matter the means of sharing. Meaning that if a person responded with at least a “having shared data 1 or 2 times” on any item they would get a 1. If they responded with “never” having shared data on all items, they would be assigned a 0. It has been decided to so as the means of sharing (webspace, university repository, personal email etc.) is of no importance to this study, as the goal is not to investigate the amount of change per unit in the different methods of sharing, but simply the distinction of having shared data or not in relation to the different independent variables. A full list of the items can be found in the appendix (B), with questions starting with “How often have you…” followed by, “Deposited/Uploaded your data…?”, “Been personally asked to share data for an article you published?” and lastly “Provided data (in response to a request) via personal communication methods? (e.g., email or file share)?” respectively.

The remaining three variables, BA, PA, and CS were collected in the demographics section at the start of the survey. The variable *biological age* has been assessed by asking “ has been categorized by the data collectors and it was not possible to revert it into a numerical variable. The categories are as follows: a value of 1 corresponds to 20-29, a value of 2: 30-39, 3: 40-49, 4: 50-59, 5: 60-69, 6: 70 and above, making a total of 6 different age groups to be further analysed. The last two categories were combined into 5: 60+, as 70 and above consisted of only 6 participants.

Professional age was collected on a numerical scale in a continuous fashion by asking “How many years have you been conducting research in this discipline? (including research conducted during your training, e.g., masters, doctoral, or any post- graduate/professional research)?”. The variable did not need any modifications as it can be used the way it was collected and will be used together with biological age and career stage to further differentiate the age dimension in predicting data sharing behaviour.

Career Stage has been categorized by the data collectors into three segments: Trainee, Junior and Senior. These three categories are a collection of the 8 different selection options and the option ‘other’ on the online survey. Trainee consists of “Master students”, “Doctoral/PhD students” and “post-Doctoral degree trainee”. Junior consists of “Research Scientists”, “Instructor” and “Assistant Professor/Lecturer”. Senior consists of “Assistant Professor” and “Full Professor”. The option “Other” was qualitatively assessed and the responses were all found to be related to Juniors in this sample and also put into the same category. The categories were coded in reversed order with the value of 1 corresponding to seniors and in order to align it with the other variables it has been recoded. The value of 1 now corresponds to trainee, 2 to junior and lastly senior is assigned the value of 3.

Data Analysis

To analyse the data set, the statistical program SPSS has been used. A descriptive frequency table of the relevant items revealed the number of missing responses per item, with the missing responses appearing to be missing completely at random (MCAR), as the online survey was not forcing participants to answer every question before submitting. Therefore, it was opted for a complete case analysis, meaning to delete participants list wise, excluding a total of 25 participants from the analysis. Outliers were checked for, and no extreme values were deleted as they were all true extremes, with the values correspond with what you would expect. An example would be that a person in the age group 5, meaning being 60+ years old also conducted research for around 40 years. No measures to account for collinearity were implemented, as the variables that are concerned are dummy coded with three or more categories (Johnston et al., 2017).

With the dependent variable (DV) being dichotomous and the independent variables (IVs) being both continuous and categorical, a logistic regression has been chosen. No dummy coding was needed, as SPSS allows you to specify categorical variables in a logistic regression and creates them automatically. At first a univariate logistic regression for each IV was used to investigate the independent relationship and to later compare the results with previous research and with the final model. In the final model, the three IVs were entered into the model simultaneously in predicting the likelihood of having shared data to see if the individual relationships change. Additionally, an interaction term (PA*CS) is added to see if there is a moderation effect between CS and PA.

3. Results

Characteristics of the sample

A sample of originally 218 participants were gathered, after deleting participants listwise, the sample size changed to 193. Gender wise most participants were female (59.6%) with the average age ranging from 30-39 (34.7%). Most of the participants 76 (39.4%) were from Europe and North America 52 (26.9%). Additionally, the highest level of education completed has been gathered and categorized, showing that the majority of the participants, 79 (40.9 %) are juniors. The majority of the participants were primarily focused on research in the Traumatic Stress Field (81.9%), filled in the survey in English (68.9%) and had an average publication count of 16 papers (*SD*=22.5) in the last 5 years. On average the participants in this sample conducted around 14 (*SD=11.2*) years of research in the traumatic stress field and most reported to have shared data at least once (70%). A full list of the descriptives can be found in Table 1.

Demographic Characteristics Table
Table 1

Demographic Characteristics of Participants

Characteristic	Sample		
	N	*%*	*M (SD)*
Gender			
Male	74	38.3	
Female	115	59.6	
Non-binary	1	0.5	
Prefer not to say	2	1.0	
Missing	1	0.5	
Age			
20-29	25	13.4	
30-39	63	34.7	
40-49	49	26.2	
50-59	30	16	
60-69	20	10.7	
Region			
Africa	4	2.1	
Asia	23	11.9	
Australia	12	6.2	
Europe	76	39.4	
Middle East	4	2.1	
North America	52	26.9	
South America	19	9.8	
Missing	3	1.6	
Career-Stage			
Trainee	56	29.9	
Junior	79	40.9	
Senior	58	30.1	
Traumatic Stress field as primary research focus?			
Yes	158	81.9	
No	34	17.6	
Missing	1	0.5	
Survey Language			
English	133	68.9	
Japanese	22	6.2	
Brazilian Portuguese	18	9.3	
French	12	6.2	
Spanish	8	4.1	
Publication Count in the last 5 years			16.8 (22.5)
Professional Age			14.6 (11.2)
Data Sharing Behaviour (0-1)			0.7 (0.5)

Note. N = 193.

The univariate association between DSB and BA

A univariate logistic regression was performed to ascertain the effects of BA, PA and CS on the likelihood that participants have shared data at least once respectively (see Table 2).

The logistic regression model for BA on DSB is found to be statistically significant, $\chi2(4)$ = 17.22 ($p < .01$). The model explained 12.1% (Nagelkerke R^2) of the variance in DSB. An increase in biological age was associated with a higher likelihood to have shared data at least once, with the odds of people around 40-49 being 6.67 (*$p<.001$)* times more likely and people around 50-59 being 7.5 ($p<.01$) times more likely than people around 20-29 years to have shared data at least once.

The univariate association between DSB and PA

The logistic regression model for PA on DSB is found to be statistically significant, $\chi 2(1)$ = 17.22 ($p<.001$). The model explained 16.1% (Nagelkerke R^2) of the variance in DSB. An increase in professional age was associated with higher odds to have shared data at least once. An increase in years of research is found to be associated with an 9% increase in the odds to have shared data ($p<.001$)*.* Therefore, accepting hypothesis 1.

The univariate association between DSB and CS

The logistic regression model for CS on DSB is found to be statistically significant, $\chi 2(2)$ = 27.38 ($p<.001$). The model explained 18.8% (Nagelkerke R^2) of the variance in DSB. An increase in career stage was associated with a higher likelihood to have shared data at least once, with people in the junior stage being 3.92 ($p<.001$) times more likely and people in the senior stage being 9.03 ($p<.001$) times more likely to have shared data at least once compared to trainees. Therefore, rejecting Hypothesis 2.

The multivariate association between DSB and BA,PA and CS

A multivariate logistic regression was performed to ascertain the effects of BA, PA and CS on the likelihood that participants have shared data at least once when entered simultaneously in a model.

The multivariate logistic regression model is found to be statistically significant $\chi 2(7)$ = 37.42 ($p<.001$). The model explained 25.1% (Nagelkerke R^2) of the variance in DSB. The results show differences between the strength and significance of the contribution of several independent variables when results of the multivariate analyses are compared with results of the univariate analyses (see Table 2).

With all variables in the equation, BA and CS positive association is weaker and no longer statistically significant. Professional Age was found to be statistically significant (p<.05) with a 95% CI [1.05, 1.17], showing that the odds ratio to have shared data is 9% higher with an increase in professional age. People around 40-49 show an 36% ($p<.645$) increase in the odds to have shared data compared to people around 50-59 ($B=1.13$, $p<.877$), both in comparison to reference group. A change of direction, although not significant ($p<.100$), has been found between people aged 60+ and 20-29 years old, with the 60+ having reduced odds by 83% ($B=0.17$, $SE = 1.08$) to have shared data at least once.

Concludingly, hypothesis 1 can be accepted in a multivariate model. Given the change in significance for career stage, hypothesis 2 failed to be rejected in the multivariate model.

Interaction Effect of PA on CS

With an interaction term added to the equation, the odds ratio for career stage decreases with higher ranks and the odds ratio for professional age to have shared data increased per additional year. The interaction showed a negative association between PA*CS, although non-significant (*p<0.157*). Therefore, it was opted to not implement it into the final model.

Table 2

Univariate and Multivariate Logistic regression analysis on the likelihood of having shared data at least once with BA,PA and CS as predictors

Data Sharing Behaviour				
	Univariate		Multivariate	
Variables	*Exp(B)*	*SE*	*Exp(B)*	*SE*
Age (20-29 as reference)				
30-39	2.80*	.49	1.27	.54
40-49	6.67***	.55	1.36	.66
50-59	7.50**	.64	1.13	.80
60-69	5.00**	.62	.17	1.08
Career Stage (Trainee as reference)				
Junior	3.92***	.38	2.14	.45
Senior	9.03***	.48	3.38	.70
Professional Age	1.09***	.02	1.09*	.04

*Note. N = 193. * p<.050 ** p<.010 ***p<.001.*

4. Discussion

Tying it all back to the start, the question whether or not data sharing can be seen as "passing on the torch" by elders and whether that analogy can safely be translated towards data sharing behaviour can be better understood. Previous research illustrates a gap between the general perception of data sharing and actual data sharing behaviours. Therefore, the objective of this study was to examine the relationship between biological age, professional age, and career stage in regard to data sharing behaviour. Novel about this study is the implementation of professional age alongside biological age and career stage, as a study about knowledge sharing and a study about experience in the work field both emphasised the importance of it over simply relying on biological age for inferences (Widén et al., 2020; Zhu 2019).

Hypothesis testing, BA,PA and CS on DBS

Looking back at the first hypothesis (H1) it can be said that as professional age increases, the likelihood of a person to have shared data at least once increases as well. The direction and strength of the relationship is fairly constant even when biological age and career stage are held constant in the multivariate model. This positive association is in line with Zhu's (2019) paper that examined how people with more experience in a work field are also more likely to engage in data sharing behaviour.

Regarding the second hypothesis (H2), in a univariate model, where career stage does not have to "compete" for variance in the sample with other factors it seems to have a positive association yet in the context of the multivariate model it failed to be rejected. Nonetheless, in this sample with an increase in career stage, the odds ratio to have shared at least once increases. Not much can be said about the relationship of data sharing behaviour and individual level predictors as this study is fairly limited by the way DSB was conceptualised, with no temporal measurements. This means that the researcher could have shared the data at any point in their career. If a study were to measure DSB in terms of frequency then care would be needed in regard to some of the temporal stages in a career, as you cannot stay a master student indefinitely, much in contrast to a senior, working as a full professor. Furthermore, interesting for future research would be the relationship between perceived career benefits as a significant predictor (Kim & Stanton, 2015; Kim, 2017) moderated by career stage. Another great point for exploration would be to relate biological age and possible biological factors (Widén et al., 2020) that accompany old age with later career stages (Tenopir et al., 2015), as older people supposedly share more data (Campbell et al., 2019; Tenopir et al., 2015) yet later career stages are associated with sharing less (Campbell et al., 2019).

Moving on to the interaction term it can be said to be non-significant, hence no moderation effect of the two variables could be determined. Further research regarding the differences of sharing behaviours among the groups needs to be done. As the concept DSB in this study simple measured if a researcher shared data at least once at any point in their career and not the frequency of shared data with regard to their respective career stage, these results should be taken with caution.

Biological age had the most surprising results. Although not significant, the results suggest that the relationship between BA and having shared data is not linear. People ranged 40-49 had the highest odd ratio of having shared data when compared to people aged 20-29. Interestingly enough the odds ratio seems to decrease after this age with people aged 60+ having the lowest likelihood of having shared data at least once. This is in contrast with the literature stating that although younger people tend to be more in favour of sharing data (Campbell et al., 2019; Tenopir et al., 2011), older people are actually more active when it comes to sharing data (Tenopir et al., 2015), yet agrees with

Tenopir et al. (2015) that the age groups between 40-49 would benefit from a closer examination in a continuous manner to better understand how it differs to the older groups.

Limitations and Future Research

An advantage from using the multivariate regression model is that the individual relations could be tested when "competing" for one another and studying the data at different levels. This revealed that the direction of the relationship may change when controlled for another variable, emphasizing the importance of using multivariable models with variables that may be of a confounding nature (Wijngaards-de Meij et al., 2005). In this case biological age's contribution towards having shared data in the traumatic stress field may be superseded by professional age. Another explanation for the non-significant relationship may be due to co-dependency of the variables. Professional age can only increase in parallel to biological age, meaning that further research regarding the collinearity of the two variables is needed to further differentiate the statistical power of one variable over another. Additionally, it might give another explanation as to why biological age is no longer statistically significant, as collinearity can make some variables insignificant when they should in fact be significant (Daoud, 2017).

This implies for future research that, in order to better understand generational differences of older generation vs younger generation, age alone may not seem to suffice to account for difference in data sharing behaviour. Most research mentioned in this paper made associations about age groups and data sharing by utilising only biological age as a decisive factor. Conversely, "age [should] be treated in a much more nuanced way" (Widén et al., 2020), including experiences collected in a work field and level of expertise. Generation wise it could be argued that the younger people see data sharing as merely a means to an end, given that younger people are more likely to play with the idea of putting restriction on their data, with restrictions here, mostly meaning possibility for collaborations, credibility, and increased reputation as well as co-authorship (Tenopir et al., 2015). Another way to put this is that young people may try to become "rich" in citation, as they are nowadays seen as "currency" in the academic field (Masic, 2014) and a great way to build a career. Although these findings were related to generational differences in knowledge sharing, it may be suggested to do the same for data sharing behaviour. However, the intricate mechanism of the relationship between years of experience and level of expertise is not yet clear, further research needs to examine in what way these two interact. To illustrate a point, a person can have 15 years of experience in playing the piano, yet these years may be spent in pressing down a single note and not being able to learn a single piece, while another person could have only 3 years of experience and may be able to play various complex pieces.

There might be other potential variables that might change the relationships we see here, as it has been suggested that age is linked to neuroplasticity and personality changes, with years of research being linked to sharing behaviour, but might also be linked to connections that are built over the years and with it, a sense of community (Widén et al., 2020). Career stage might also be linked to a sense of personal identity (Paloniemi 2006; Stewart et al. 2017 as cited in Widén et al., 2020), meaning that core beliefs that make up how you perceive yourself might be another factor that could influence data sharing behaviour.

Additionally, as this paper only investigated data sharing behaviour and not reuse it must be mentioned that some researchers might differ in their sharing and reusing quantity, with some identifying more as a “reuser” and some identifying more as a “data sharer” (Curty et al., 2017). Sharing data has shown its use, with the ability to combine smaller data sets into larger ones, and thus rendering the limitations of a smaller sample size, allowing for more complex analysis and promises for greater discoveries (Ferguson et al., 2014).

Concluding, as we move forwards, we need to be aware of the factors that are guiding our judgment of how we as a scientific community should act. The implementation of the internet opened up many possibilities, we can now connect with researchers worldwide by the click of a button, share, discuss and work together on ideas and methods from the comforts of our own homes. At the heart of any science is data, without data there is no science, yet collecting data is a timely act, and time is not a resource we can get back. It is up to us how we manage this situation, whether we think sharing data might run the risk of ruining our chances for success or whether contributing to possible ground-breaking discoveries is all the success that is needed.

References:

Brakewood, B., & Poldrack, R. A. (2013). The ethics of Secondary Data Analysis: Considering the application of Belmont principles to the sharing of Neuroimaging Data. *NeuroImage*, *82*, 671–676. https://doi.org/10.1016/j.neuroimage.2013.02.040

Bremner, J. D. (2006). Traumatic stress: Effects on the brain. *Dialogues in Clinical Neuroscience*, *8*(4), 445–461. https://doi.org/10.31887/dcns.2006.8.4/jbremner

Campbell, H. A., Micheli-Campbell, M. A., & Udyawer, V. (2019). Early career researchers embrace data sharing. *Trends in Ecology & Evolution*, *34*(2), 95–98. https://doi.org/10.1016/j.tree.2018.11.010

Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. *PLOS ONE*, *12*(12). https://doi.org/10.1371/journal.pone.0189288

Daoud, J. I. (2017). Multicollinearity and regression analysis. *Journal of Physics: Conference Series*, *949*, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., & Martone, M. E. (2014). Big data from small data: Data-sharing in the 'long tail' of neuroscience. *Nature Neuroscience*, *17*(11), 1442–1447. https://doi.org/10.1038/nn.3838

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data Sharing in Psychology: A Survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10.1177/2515245917751886

Johnston, R., Jones, K., & Manley, D. (2017). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, *52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Kassam-Adams, N., & Olff, M. (2020). Embracing data preservation, sharing, and re-use in Traumatic Stress Research. *European Journal of Psychotraumatology*, *11*(1), 1739885. https://doi.org/10.1080/20008198.2020.1739885

Kim, Y. (2017). Fostering scientists' data sharing behaviors via data repositories, journal supplements, and Personal Communication Methods. *Information Processing & Management*, *53*(4), 871–885. https://doi.org/10.1016/j.ipm.2017.03.003

Kim, Y., & Stanton, J. M. (2015). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *67*(4), 776–799. https://doi.org/10.1002/asi.23424

Kim, Y., & Yoon, A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *68*(12), 2709–2719. https://doi.org/10.1002/asi.23892

Lawrence, S. (2001). Online or invisible. *Nature*, *411*(6837), 521. http://www.m-hikari.com/online.pdf

Masic, I. (2014). Plagiarism in scientific research and publications and how to prevent it. *Materia Socio Medica*, *26*(2), 141. https://doi.org/10.5455/msm.2014.26.141-146

Olff, M., Amstadter, A., Armour, C., Birkeland, M. S., Bui, E., Cloitre, M., Ehlers, A., Ford, J. D., Greene, T., Hansen, M., Lanius, R., Roberts, N., Rosner, R., & Thoresen, S. (2019). A decennial review of psychotraumatology: What did we learn and where are we going? *European Journal of Psychotraumatology*, *10*(1), 1672948. https://doi.org/10.1080/20008198.2019.1672948

Paloniemi, S. (2006). Experience, competence and Workplace Learning. *Journal of Workplace Learning*, *18*(7/8), 439–450. https://doi.org/10.1108/13665620610693006

Schultebraucks, K., & Galatzer-Levy, I. R. (2019). Machine learning for prediction of posttraumatic stress and resilience following trauma: An overview of basic concepts and recent advances. *Journal of Traumatic Stress*, *32*(2), 215–225. https://doi.org/10.1002/jts.22384

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, *8*(1). https://doi.org/10.1038/s41597-021-00981-0

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0134826

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data Sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, *15*(3). https://doi.org/10.1371/journal.pone.0229003

The Global Collaboration on Traumatic Stress. (n.d.). *International Survey on Data Sharing and Re-use in Traumatic Stress Research*. Retrieved September 27 , 2022, from https://redcap.chop.edu/surveys/?s=HAFYJPRNF9

Widén, G., Ahmad, F., Sivén, T., & Ivantsova, E. (2020). Understanding generational differences in knowledge sharing. In A. Garcia-Perez, & L. Simkin (Eds.), *Proceedings of the 21st European Conference on Knowledge Management: A Virtual Conference hosted by Coventry University, UK* (pp. 841-849). Academic conferences international. https://urn.fi/URN:NBN:fi-fe202201148122

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The Fair Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18

Wijngaards-de Meij, L., Stroebe, M., Schut, H., Stroebe, W., van den Bout, J., van der Heijden, P., & Dijkstra, I. (2005). Couples at risk following the death of their child: Predictors of grief versus depression. *Journal of Consulting and Clinical Psychology*, *73*(4), 617–623. https://doi.org/10.1037/0022-006x.73.4.617

Zhu, Y. (2019). Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, *46*(1), 41–52. https://doi.org/10.1177/0165551518823174

Appendices

Appendix A

Recruitment message

Dear [Name],

My name is … and I am a psychology student at the University of Twente and currently working on my Bachelor thesis. In cooperation with Lonneke Lenferink, Nancy Kassam-Adams, and the Global Collaboration on Traumatic Stress, we are conducting an international survey to better understand traumatic stress researchers' opinions and experiences regarding data sharing and data re-use. Therefore, we are recruiting traumatic stress researchers at any career stage (including trainees) to share opinions and experiences by participating in the following survey. The survey will take approximately 10 min to complete.

The results of this global survey will be shared on the Global Collaboration website (https://www.global-psychotrauma.net/) and in scientific publications and it will help us to create tools and resources for traumatic stress researchers. The final dataset from this survey will be available upon request for use by other researchers.

Participation is voluntary and there are no known risks or personal benefits to you from participating in this study.

As the survey is available in multiple languages (English, Japanese, Spanish, French, Portuguese, Korean, and Arabic), we would kindly ask to participate if you are proficient in one of the available languages.

If you have questions about the survey, the study, or the study dataset, please contact the study team at childtraumadata@chop.edu.

Follow this link to the survey:

https://www.global-psychotrauma.net/data-sharing

Thank you for your participation.

Regards,

Hakan Culha

University of Twente, NL

Appendix B

International Survey

International Survey on Data Sharing and Re-use in Traumatic Stress Research

The Global Collaboration on Traumatic Stress, a coalition of 11 scientific societies in the field of traumatic stress, is conducting a survey to better understand traumatic stress researchers' opinions and experiences regarding data sharing and data re-use. Results of this global survey will be shared on the Global Collaboration website (https://www.global-psychotrauma.net/), and will help us create tools and resources for traumatic stress researchers. The final dataset from this survey will be available upon request for use by other researchers.

If you are a traumatic stress researcher at any career stage (including trainees) we invite you to share your opinions and experiences by participating in this survey. The survey is anonymous, and your participation is voluntary. There are no known risks or personal benefits to you from participating in thisstudy.

If you have questions about the survey, the study, or the study dataset, please contact the study team at childtraumadata@chop.edu.

By continuing to the survey, you are consenting to participate in this study.

THANK YOU for your participation.

==

Part 1 - So that we can describe the respondents to this survey, please tell us a bit about yourself.

1.What is your academic / research discipline? CHECK ALL THAT APPLY

- Psychology
- Psychiatry
- Medicine – other than psychiatry - specify:

- Nursing
- Social Work
- Public Health
- Education
- Other – Specify: ________________________________

2. How many years have you been conducting research in this discipline? (include research conducted during your training, e.g., masters, doctoral, or any post-graduate/professional research)

3. What is your current job title / academic rank / trainee status? If multiple apply, select highest rank.
   - Full Professor
   - Associate Professor
   - Assistant Professor / Lecturer
   - Instructor
   - Research scientist
   - Post-doctoral trainee
   - Doctoral/PhD student
   - Masters student
   - Other – Specify: ________________________________

4. In the last 5 years, how many publications involving research data have you published (including those as first author or co-author)?

5. How many of these publications involved analyses of research data collected by others outside you / your research team / your co-authors?

6.Is trauma / traumatic stress your primary research focus? Yes / No

SKIP PATTERN – If no ITEM 6 then go to ITEM 7

If yes – go to ITEM 8

7.What is your primary area of research?

8.What types of trauma have been included in your research? CHECK ALL THAT APPLY

- Acute/Single trauma
- Child Abuse/Maltreatment
- Chronic/Repeated Trauma
- Community Violence
- Death/Bereavement
- Disaster
- Intimate Partner Violence
- Medical Trauma
- Racism / Historical Trauma
- Rape/Sexual Assault
- Refugee/Displacement Experiences
- Secondary / Vicarious Traumatization in Professionals / Helpers
- Terrorism
- Torture
- War / Post-Conflict Settings – Civilians
- War – Military/Peacekeepers/Veterans
- Other(s) – Specify: ________________________

9.What populations have been included? CHECK ALL THAT APPLY

- Adults
- Adolescents
- Children

10.What types of data have you collected? CHECK ALL THAT APPLY

- Data from surveys / questionnaires
- Data from standard interviews

- Qualitative data
- Intensive longitudinal (EMA / ESM) data
- Experimental task performance data
- Genetic data
- Biological / physiological data (other than genetic)
- Data retrieved from health / medical records
- Data from other non-research records or sources (administrative data, online / social media data)
- Other – Specify: ______________________________

11. What is your age in years?

12. How do you identify your gender?
    - Male
    - Female
    - Non-binary
    - Other
    - Prefer not to say

13. Do you consider yourself to be of an ethnic / cultural background that is under-represented amongst researchers in the discipline / research community in which you work?”

Yes / No / Prefer not to say

14. In what country do you live and work? [DROP DOWN LIST – SEE LIST AT END OF THIS DOC]

Part 2 - Please indicate to what extent you agree with the following statements, thinking about the institutions and research communities that you are part of.

IN MY RESEARCH COMMUNITY ....

RESPONSES FOR THIS SECTION: 1, Strongly Disagree | 2, Moderately Disagree | 3, Slightly Disagree | 4, Neutral | 5, Slightly Agree | 6, Moderately Agree | 7, Strongly Agree | -99, Don't Know

15. It is expected that researchers would share data.

16. Researchers share data even if not required by policies.

17. Many researchers are currently participating in data sharing.

18. Public funding agencies require researchers to share data.

19. Journals require researchers to share data.

20. Researchers can easily access metadata about existing data sources.

21. Researchers have the tools they need to share appropriate metadata along with their data.

22. Data repositories are available for researchers to deposit / share their data.

23. Researchers can easily access data repositories to request / acquire data for re-use.

24. It is difficult to publish work that is based in data re-use, i.e. new analyses of data collected by others.

25. Re-using data for new / secondary analyses has led to advances in the field.

Part 3 - Thinking about YOUR OWN VIEWS AND EXPERIENCES, please indicate the extent to which you agree with the following statements.

RESPONSES FOR THIS SECTION: 1, Strongly Disagree | 2, Moderately Disagree | 3, Slightly Disagree | 4, Neutral | 5, Slightly Agree | 6, Moderately Agree | 7, Strongly Agree

26. I am willing to help other researchers within my institution / research community by sharing data.

27. I am willing to help other researchers outside my institution / research community by sharing data.

28. I can earn academic 'credit' such as more citations by sharing data.

29. Data sharing would be helpful in my academic career.

30. Sharing data is an ethical obligation as a researcher.

31. Sharing data honors the contributions of research participants.

32. Sharing data has a high risk of violating the rights of research participants.

33. There is a high probability of losing publication opportunities if I share data.

34. Data sharing may cause my research ideas to be stolen by other researchers.

35. My shared data may be misused or misinterpreted by other researchers.

36. I believe that the overall riskiness of sharing data is high.

37. Sharing data involves too much time for me (e.g. to organize / annotate).

38. I would find data sharing difficult to do.

39. I have adequate time and funding for any effort that may be required in sharing my data.

40. I include statements about data sharing in my participant consent forms.

41. My institution's ethics committee / IRB makes it hard for me to share research data gathered in IRB approved studies.

42. When I begin a project, I organize the data to enable later data re-use and sharing.

43. I feel prepared (via training or experience) to manage my data in a way that facilitates re-use and sharing.

44. I know how to de-identify / anonymize my data so that it can be shared.

45. I know how to clearly document how my raw data was processed / cleaned for analysis.

46. Re-using other researchers' data can improve the quality of my overall program of research.

47. Re-using other researchers' data reduces the time/cost/effort I spend on my research.

48. If I re-use other researchers' data, I worry that I might misinterpret the data.

49. If I re-use other researchers' data, I worry that I might not be able to publish with that data.

50. Re-using other researchers' data requires too much time and effort to locate data sets.

51. Re-using other researchers' data requires too much time and effort to access (or get permission to use) data sets.

52. Re-using other researchers' data requires too much time and effort to process data sets for a new study.

Part 4 - How often have you...

RESPONSES FOR THIS SECTION: Never | 1 or 2 times | More than 2 times

53. Deposited your data, RELATED TO AN ARTICLE YOU PUBLISHED, into an institutional repository (i.e. repository maintained by a journal, university, funder, national data archive, etc)?

54. Uploaded your data, RELATED TO AN ARTICLE YOU PUBLISHED, into a "public" Web space (e.g. PsyArxiv, MedArxiv, OSF)?

55. Deposited your data / dataset, NOT IN CONNECTION TO A SPECIFIC PUBLICATION, into an institutional repository?

56. Uploaded your data / dataset, NOT IN CONNECTION TO A SPECIFIC PUBLICATION, into a “public” Web space?

57. Been personally asked to share data for an article you published?

58. Provided data (in response to a request) via personal communication methods? (e.g., email or fileshare)?

59. Downloaded or requested data from a repository for your own analyses / research?

60. Directly requested data from another researcher / research team for use in your own work?

61. Collaborated with other researchers to combine (your & their) data for new analyses / new work?

62. Published results of work that included use of others' data?

Part 5 – Any additional comments?

62. Please share any additional comments about your views or experiences regarding datasharing or data re-use: OPEN TEXT FIELD

Portions of this survey were adapted from the following studies:

Kim, Y. (2013). *Institutional and Individual Influences on Scientists' Data Sharing Behaviors* (Doctoral Dissertation). surface.syr.edu/it_etd/85/.

Kim, Y., & Stanton, J. M. (2016). Institutional and Individual Factors Affecting Scientists'

Data-Sharing Behaviors: A Multilevel Analysis. *Journal of the Association for Information Science and Technology*, *67*(4), 776–799. https://doi.org/10.1002/asi.23424

Kim, Y., & Yoon, A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *68*(12), 2709–2719. https://doi.org/10.1002/asi.23892

END OF SURVEY