

Statistical literacy and scientific reasoning & argumentation in physicians

Abstract

Objective: Statistical literacy (SL) of physicians, i.e. the ability to use and interpret statistical numbers in the context of science, is an essential prerequisite for risk estimation and communication. Together with scientific reasoning and argumentation (SRA) skills, SL provides the basis for evidence-based practice. Several studies suggest that in medical students both skills are underdeveloped. The aim of the present study was to investigate these skills in practicing physicians and how these skills were acquired.

Methods: Data collection in $N=71$ physicians was conducted online and as paper pencil. SL was assessed with multiple-choice items. SRA skills evidence evaluation and drawing conclusions were measured with a decision scenario.

Results: Study results indicated that physicians have medium levels of SL ($M=17.58$, $SD=6.92$, max 30 pts.) and SRA (evidence evaluation: $M=7.75$, $SD=1.85$, max 10 pts.; drawing conclusions: $M=37.20$, $SD=5.35$, max 60 pts.). Skills development via autodidactic learning activities ($M=4.78$, $SD=1.13$, range 1-6) was reported significantly more often than development during formal medical education ($M=2.31$, $SD=1.46$), $t(71)=-9.915$, $p<.001$, or in extracurricular activities ($M=3.34$, $SD=1.87$), $t(71)=4.673$, $p<.001$. The active involvement in research seemed decisive: The number of publications and time spent in research significantly correlated with SL, $r(71)=.355$, $p=.002$; respectively $r(71)=.280$, $p=.018$. SRA skills were predicted by the type of MD-thesis, $\beta=-.380$, $p=.016$, and working in research, $\beta=3.355$, $p=.008$.

Conclusion: Active involvement in research activities seems to be a very important factor for the development of both SL and SRA skills. The implementation of systematic fostering of these skills during formal medical education seems warranted.

Keywords: evidence-based practice, scientific reasoning and argumentation, statistical literacy, medical education, postgraduate medical education

1. Background

Following Sackett et al. (1997) regarding the modern definition of evidence-based medicine, scientific reasoning skills are considered essential together with physicians' expertise for best possible decision-making in the best interest of the patient [1], [2], [3]. For the assessment of risks and probabilities and their applicability to specific patients, a basic understanding of statistics and evidence evaluation is necessary. Statistical literacy (SL) is not only the ability to understand statistical information, but also to apply it in decision-making [4]. It comprises the aptitude of critical reflection about statistics as evidence in arguments [5]. Statistical literacy is based on numeracy, the aptitude of mathematical operations [6], and encompasses the ability to use and interpret statistical numbers in the context of science [7], [8], [9] and the ability to explain and critically evaluate them [10],

[11] (working definition for the present study). Furthermore, it is intertwined with scientific reasoning and argumentation skills (SRA), to provide the basis for evidence-based decision-making [12], [13].

Based on the framework by Fischer et al. (2014), SRA can be defined as the competence of comprehending and applying scientific working methods and their results when solving problems [14], [15]. SRA can be described by eight epistemic activities and this study focuses on two of them, evidence evaluation (EE) and drawing conclusions (DC).

Various studies suggest an intertwining of SRA and SL, with Anderson et al. (2013) stating that the latter is needed to evaluate scientific evidence [16] and Franklin et al. (2005) hypothesizing that SL itself encompasses also SRA skills [17].

However, a collective statistical illiteracy has been observed among physicians [4], [18]. Similarly, SRA skills

Felicitas M. Schmidt¹
Jan M. Zottmann¹
Maximilian Sailer²
Martin R. Fischer¹
Markus Berndt^{1,3}

1 University Hospital, LMU Munich, Institute of Medical Education, Munich, Germany

2 University of Passau, Chair of Educational Science, Passau, Germany

3 Walden University, Richard W. Riley College of Education and Leadership, Minneapolis, USA

needed for evidence-based practice [19], such as EE or DC, are underdeveloped [4], [7].

In a study by Anderson et al. (2014), 52% of the participating physicians answered only two (or fewer) of four questions regarding statistical concepts correctly [7]. This is in line with findings by Windish et al. (2007) reporting only 40% of resident physicians demonstrating adequate understanding of biostatistical concepts [20]. Similar results were found by Gigerenzer and Wegwarth (2008), showing that 33% of gynecologists were not aware of the benefits of mammography screening, with 79% being unable to interpret the positive predictive value [21]. Gigerenzer et al. (2008) summarized various studies on the concept of the positive predictive value and its dependence on prevalence, finding that 50% of participants were under the impression that false positive test results in HIV testing do not exist. They also found that only two of 20 urologists have sufficient knowledge about the reliability of a PSA-test. Thus, a lot of measurement tools for the assessment of SL exist [18], [22], [23], which focus commonly on one of the three levels of Watson (1997) or are designed in a certain context [7], [22]. Overall, physicians' SL is not below-average [7]. However, it can be considered comparable to other educated samples [18], [24] and was found superior to that of residents in research training [20] or medical students [25]. The few studies that have assessed medical students' SL are supporting the findings of superiority of physicians [26]. Berndt et al. (2021) compared medical students to those of social sciences and economics and found that medical students in their first years of study scored higher in comparison to social science students and comparable to economics students [27]. This study not only assessed SL, but also the two aforementioned skills EE and DC. Further possible links between SL and SRA have been examined with the Medical Data Interpretation Test [22] where participating physicians scored higher overall than participants with other postgraduate degrees (89 out of 100 score points). Johnson et al. (2014) assessed numeracy of medical students and residents and found students with poor numeracy being more likely to misjudge risks of different treatment alternatives and that the confidence in treatment recommendation increased during medical school [25].

In summary, SL and SRA seem to be underdeveloped in medical students and better developed in physicians. However, the development of SL and SRA may not necessarily happen within formal education and the question remains how, where, and when this development occurs. The present study should contribute to the discussion when and how to best foster SL and SRA skills in lifelong learning of physicians by applying a test instrument comprising various aspects of SL and SRA. It aimed at providing further insights into these skills in physicians and to identify demographic factors and learning opportunities that may be associated with the development of these skills. Our explorative research questions were:

RQ1. Statistical literacy and SRA skills

- To what extent are SL and SRA skills developed in physicians?
- To what extent does SL predict SRA skills of physicians?

RQ2. Education and skills development

- How, where, and when do physicians develop SL and SRA skills?
- Which demographic factors are related to the development of SL and SRA skills?

2. Methods

2.1. Design and sample

Our study followed a quasi-experimental, causal-comparative design with two dependent variables: SL and SRA. We included $N=71$ German-speaking physicians (31 females, 34 males, 6 NA), from different work settings and locations in our study: hospital ($n=44$), outpatient sector ($n=3$), research ($n=8$), study program ($n=2$) and administration ($n=2$), ($n=12$ NA). A MD-thesis, a scientific work as optional part of the medical study program (not equivalent to a PhD thesis), was completed by 58 participants and 9 were currently working on it. Despite the modest sample size, we consider our sample representative with regard to scientific experience, as a MD-thesis is very common in Germany. The mean age of participants was 40 years ($SD=9.59$, $range=26-65$) (see table 1).

2.2. Test instrument

For the assessment of SL and SRA skills, we used an instrument previously developed in the context of a study by Berndt et al. (2021) who initiated the ongoing validation process with 217 economics, social sciences, and medical students from LMU Munich [27]. The test instrument combines multiple choice items to assess SL with a decision scenario [28] to assess the participants' skills in EE and DC. For this study, items on relevant demographic factors were added and piloted with ten medical students from LMU Munich.

2.2.1. Demography

Demographic and biographic parameters of the participants were assessed with a special interest in their working history and environment (hospital, out-patient care, research). Questions were adapted from a study by Epstein et al. [29] and comprised multiple choice items, some with the opportunity to fill in additional free text; five items on the MD-thesis, three items on the professional career, two items on the publication record (type of authorship, number of publications), and three items on the current job description.

Table 1: Description of study group (N=71)

Variable	Options	<i>n</i>	Percentage
Gender	Male	34	47.9%
	Female	31	43.7%
	No answer	6	8.4%
German mother tongue	Yes	63	88.7%
	No	3	4.2%
	NA	5	7.0%
MD in Germany	Yes	68	95.8%
	No	3	4.2%
MD-Thesis	Yes	58	81.7%
	Currently working on	9	12.7%
	No thesis or skipped	4	5.6%
Academic qualification	Habilitation	12	16.9%
	Professorship	7	9.9%
Ever worked as researcher	Yes	36	50.7%
	No	35	49.3%
Working Environment	Hospital	44	62.0%
	Out-patient care	3	4.2%
	Research	8	11.3%
	Study	2	2.8%
	Administration	2	2.8%
	NA	12	16.9%

2.2.2. Statistical literacy

Statistical literacy was measured with multiple choice items based on validated instruments [7], [18], [23], to assess a broad spectrum from basic numeracy to conditional probabilities and statistical concepts. Duplicates and factual knowledge questions were excluded, so that all three levels described by Watson (1997) were covered. Additionally, all items were weighted for difficulty [11]. Internal consistency of the SL test was .82 (Cronbach's α) in our sample with a maximum score of 30 points. All items were framed in a medical context; however, no medical content knowledge was necessary to answer them correctly.

2.2.3. SRA skills

The assessment of SRA skills focused on the two epistemic activities EE and DC with a decision scenario in a medical context (general medicine, out-patient care) and provided two separate overall scores for EE and DC (Cronbach's α for EE items .87; for DC items .74). For EE, participants had to evaluate four pieces of evidence [30], [31], [32], including one authentic pharmaceutical brochure that advertised herbal drugs, in terms of their scientific quality, evidence strength, and relevance for the present situation on a 6-point Likert scale based on the QUESTS criteria [33].

Then, the participants rated the persuasiveness (Likert 1-6) of 20 arguments, which were extracted from the presented evidence and assigned a level of argument strength from 1 (lowest) to 4 (highest). For 13 participants, the evaluation of 1 to 5 arguments out of 20 was missing. In order to avoid dropouts of these cases, the

respective values were imputed from the average of the respective item. The participants' ratings for scientific quality were compared to the ratings of scientific quality by the authors, resulting in a measure of similarity for EE and DC. The range of these measures was from 0-10 (EE score) and 0-60 (DC score) with zero indicating no similarity.

2.3. Procedure

The study was completed by the participants either online with LamaPoll [<https://www.lamapoll.de/>], a survey tool optimized for mobile applications, or as paper pencil (return rate 16.5% online and 66.7% paper pencil). Average duration was approximately 45 minutes. Participants were invited via mailing lists and personal contacts.

2.4. Statistical analyses

Statistical analyses were performed with IBM SPSS 25. Descriptive and frequency data were computed for primary analysis and Cronbach's alpha for internal consistency. Extensive outlier analyses were conducted and all required prerequisites for statistical analyses, such as normal distribution and homoscedasticity, were tested. *T*-tests, one-factorial ANOVAs, and linear regression models were calculated to assess differences and the association of demographic factors with SL and SRA. Probability values less than .05 were considered significant. Data in natural verbal language (free text in demography section) underwent independent thematic analysis by two authors to extract common themes.

3. Results

We included 71 completed questionnaires (see table 1). The entire data set was checked for univariate outliers. Skewness and kurtosis for all variables was within the ± 2 range [34]. The prerequisites for *t*-tests and ANOVA were fulfilled, unless indicated otherwise below.

3.1. Statistical literacy and SRA skills

The 71 physicians' average score in SL was $M=17.58$, $SD=6.92$, with a range of 5 to 30 out of 30 attainable points (59%). On average, physicians evaluated the evidences concordantly with the authors' evaluation, EE score: $M=7.75$, $SD=1.85$ (77%). The ratings for argument quality were in accordance with the authors' rating, DC score: $M=37.20$, $SD=5.35$ (62%). Statistical literacy and DC were significantly inversely correlated, $r(71)=-.272$, $p=.022$. However, no correlation was found between SL and EE, $r(71)=.198$, $p=.098$, nor EE and DC, $r(71)=.138$, $p=.256$.

3.2. Education and skills development

We explored how, where, and when physicians developed scientific skills (see figure 1). Significantly more participants indicated to have acquired scientific skills in an autodidactic manner ($M=4.78$, $SD=1.13$, Likert 1-6 scale) rather than during their study program ($M=2.31$, $SD=1.46$, Likert 1-6 scale), $t(71)=-9.915$, $p<.001$, or in extracurricular activities ($M=3.34$, $SD=1.87$, Likert 1-6 scale), $t(71)=4.673$, $p<.001$. In a free-text box, participants added various other learning opportunities, such as massive open online courses, higher education, workshops, and learning through peer reviews and feedback (see figure 2).

Having completed or working on the MD-thesis showed no effects on SL, EE, or DC. However, these results have to be treated carefully, as the prerequisites for ANOVA were not fulfilled in our sample with only four participants without a MD-thesis. The fostering of critical scrutiny of study results presented by other researchers during the preparation of the MD-thesis was positively correlated with SL, $r(71)=.271$, $p=.033$.

Regarding the postgraduate phase, a one-factorial ANOVA showed a significant main effect of having worked in research on SL, $F(1,70)=12.737$, $p=.001$, partial $\eta^2=.156$ and the type of authorship in publications, $F(5,71)=3.886$, $p=.004$, partial $\eta^2=.230$.

Time spent in research was significantly associated with better SL, $r(71)=.28$, $p=.018$, as was the number of publications, $r(71)=.36$, $p=.002$.

Regarding SRA, linear regression models revealed that the corresponding score in EE increased by $\beta=.314\pm.150$, $p=.041$, when the Likert value of the MD-thesis supervisor's content-related support was increased by one point. Additionally, the form of the MD-thesis (experimental, clinical, empirical, statistical, or literature review) was associated with EE, with experimental and clinical design

being positively related to EE skills, $\beta=-.380\pm.154$, $p=.016$, $R^2=.187$, $F(1,59)=4.353$, $p=.041$. DC was higher when participants indicated to have already worked in research, $\beta=3.355\pm 1.229$, $p=.008$, $R^2=.314$, $F(1,68)=7.448$, $p=.008$.

4. Discussion

4.1. Research question 1: Statistical literacy and SRA skills

We found average statistical literacy of physicians (59%), a rather high-level EE score (77%) and a medium-level DC score (62%). SL did not predict the SRA skills of physicians.

Due to the focus on SL rather than the combination of basic numeracy and SL [27], our test instrument discriminated well, and we did not find any ceiling effects as observed in other educated samples [7], [35]. A comparison to other studies assessing SL of physicians is not easily done as every test covers a different range of SL. Schmidt et al. (2017) focused on knowledge of 18 different statistical tests among pathologists and observed a rather low level of SL [36]. Anderson et al. (2014) did not create an overall score but distinguished between fact, concept, and relation questions and found altering levels of SL [7]. A study with Greek residents also concentrated on knowledge questions and reported a rather low SL [26]. The EE and DC scores of German medical students we had previously examined with a similar instrument [27] were almost on the same level as the physicians' scores in the present study. Riegelman and Hoveland (2012) found that residents struggled when critical reflection upon research was required [37], whereas the physicians in our study showed medium to high levels of SRA skills.

EE and SL scores were not correlated. DC and SL were inversely correlated. In contextual frameworks, SL has been regarded as a prerequisite for SRA [38] and in a Dutch community-based study, more numerate participants showed enhanced performance in SRA due to increased evaluation of pros and cons in decision-making and evaluation of judgments [39]. As evidence has not been predominantly presented in numerical or statistical terms, the missing link of EE and SL was expected, but the antithetical relationship of DC and SL was not. Future research could incorporate statistical information in decision scenarios in order to further analyze this connection in practicing physicians.

4.2. Research question 2: Education and skills development

We explored how, where, and when physicians developed SL and SRA skills. They indicated to have acquired scientific skills mostly in an autodidactic manner, in higher education outside of their medical study program, or in extracurricular activities.

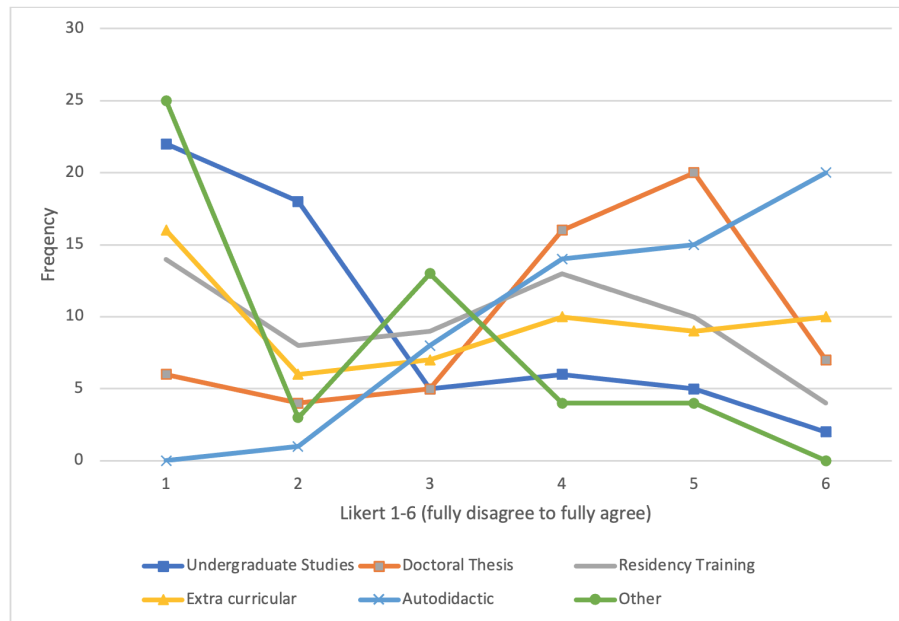


Figure 1: In which context have you acquired scientific skills? (N=71 physicians)

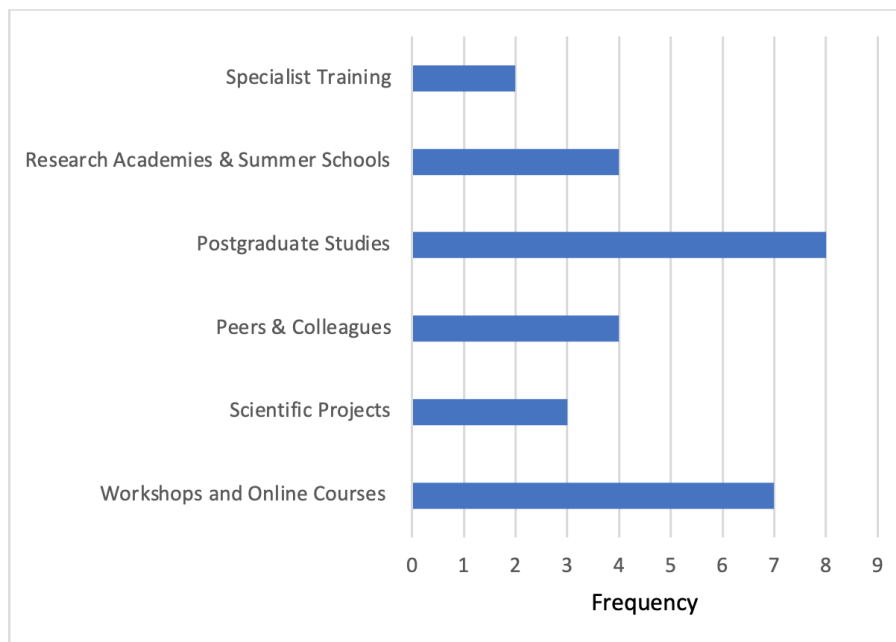


Figure 2: In which context have you acquired scientific skills? Free text analysis category Other (n=28 physicians)

Better SL was associated with the fostering of critical scrutiny of study results during the time spent working on the MD-thesis, in research or having worked in research, the number of publications, and the type of authorship. Our findings are in line with Schmidt et al. (2017), who found that having an advanced degree other than MD or statistic courses were positively associated with SL. A study with physicians, residents, and final year medical students in Thailand showed – not surprisingly – that having recently completed a statistical workshop led to higher SL scores [40]. However, additional courses are often hard to integrate in medical training. A study showed that 37% of American Ob-Gyn residents do not receive formal training [16], while another study with

neurology residents observed a lack of acceptance for interventions on SL [41].

In the present study, better EE was associated with having been responsible for a research project (e.g. the MD-thesis) with experimental or clinical design and having content-related support by the supervisor. These findings are in line with the subjective impression of German medical graduates with a MD-thesis who rated their scientific skills higher compared to those working on it [29]. However, the participants in the study by Epstein et al. (2018) did not feel confident enough to conduct research on their own. This is particularly important because having already worked in research was associated with a higher SL and DC score in the present study and in Schmidt et al. (2017). Moreover, Epstein et al. (2018) found that

medical graduates self-estimate their scientific skills after medical school as rather low. In the United States, only 68.1% of medical students in their final year participated in research during medical school and only 42% had (co-)authored a paper submitted for publication. It seems important that medical students become involved in research projects and the subsequent publication of findings during the completion of their MD-thesis, as this might enhance their SL and SRA skills in the long run.

4.3. Strengths and limitations

This study built upon an innovative approach by the authors to assess SL and SRA skills in university students [27]. The inclusive approach of SL assessment allowed a better description of the actual skills. However, it comes with the disadvantage of limited comparability with prior research. As the participant group of practicing physicians is not easily recruited, we considered the sample size of $N=71$ to be satisfactory. While generalizability is potentially limited, our sample seems representative for the German-speaking medical education system which produces large numbers of medical doctoral degrees. The addition of numerous demographic variables yielded insights on how, where, and when scientific skills were acquired and helped to identify potential associated factors. Due to the broad age range in our sample, participants may have been exposed to different learning experiences in formal medical training and, depending on the place of study, may also have studied in reformed curricula. This could potentially have influenced their skills development and lead to further individual differences. In our study, we did not gather data on specific study programs, courses, and their descriptions, in which physicians might have acquired their skills. Already, the test instrument used in this study may be considered extensive and time consuming for physicians, as was indicated by 11 participants in their feedback.

5. Conclusion

We assessed SL and SRA skills in German-speaking physicians together with a thorough analysis of demographic variables. The active involvement in research apparently plays an important role in the development of these skills and might in consequence enhance evidence-based practice. As most participants indicated to have acquired these skills post-graduate and in an autodidactic manner, we argue to formalize and intensify the acquisition of these skills in medical study programs. Medical education curricula should include more statistical training and aim to get students involved in research more often, e.g. by offering inquiry-based learning [42] where students conduct research projects independently and are fully responsible for all phases of the research process.

Funding

This work was supported by the German Federal Ministry of Education and Research (BMBF, grant no. 01PB14004C) and an intramural grant of the Förderverein WiFoMed of the Medical Faculty of LMU Munich.

Acknowledgements

The authors would also like to thank Wolfgang Gaissmaier for his valuable comments and suggestions to the present study.

Competing interests

The authors declare that they have no competing interests.

References

1. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull.* 2009;135(6):943-973. DOI: 10.1037/a0017327
2. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. *Ann Behav Med.* 2008;35(3):261-274. DOI: 10.1007/s12160-008-9037-8
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Was ist Evidenz-basierte Medizin und was nicht? *Munch Med Wochenschr.* 1997;139(44):28-29. DOI: 10.1055/b-0036-140841
4. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest.* 2008;8(2):53-96. DOI: 10.1111/j.1539-6053.2008.00033.x
5. Shield M. Statistical literacy: thinking critically about statistics. *Significance.* 1999;1(1):15-20.
6. Peters E. Beyond Comprehension: The Role of Numeracy in Judgments and Decisions. *Curr Dir Psychol Sci.* 2012;21(1):31-35. DOI: 10.1177/0963721411429960
7. Anderson BL, Gigerenzer G, Parker S, Schulkin J. Statistical literacy in obstetricians and gynecologists. *J Healthc Qual.* 2014;36(1):5-17. DOI: 10.1111/j.1945-1474.2011.00194.x
8. Walker HM. Statistical literacy on the social science. *Am Stat.* 1951;5(1):6-12. DOI: 10.1080/00031305.1951.10481912
9. Wallman KK. Enhancing statistical literacy: enriching our society. *J Am Statl Ass.* 1993;88(421):1-8. DOI: 10.1080/01621459.1993.10594283
10. Ben-Zvi D, Garfield J. Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In: Ben-Zvi D, Garfield J, editors. *The challenge of developing statistical literacy, reasoning and thinking.* Dordrecht, The Netherlands: Kluwer; 2004. p.3-25. DOI: 10.1007/1-4020-2278-6
11. Watson JM. Assessing Statistical Thinking Using the Media. In: Gal I, Garfield JB, editors. *The Assessment Challenge in Statistics Education.* Amsterdam: IOS Press and The International Statistical Institute; 1997. p.107-121.
12. Sedlmeier P, Gigerenzer G. Teaching Bayesian Reasoning in Less Than Two Hours. *J Exp Psychol Gen.* 2001;130(3):380-400. DOI: 10.1037/0096-3445.130.3.380

13. Fischer F, Kollar I, Ufer S, Sodian B, Hussmann H, Pekrun R, Neuhaus B, Dorner B, Pankofer S, Fischer M, Strijbos JW, Heene M, Eberle J. Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Front Learn Res.* 2014;2(3):28-45. DOI: 10.14786/flr.v2i2.96
14. Hetmanek A, Engelmann K, Opitz A, Fischer F. Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. In: Fischer F, Chinn CA, Engelmann K, Osborne J, editors. *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge.* New York: Routledge; 2018. p.203-226.
15. Rudolph JL, Horibe S. What do we mean by science education for civic engagement? *J Res Sci Teach.* 2016;53(6):805-820. DOI: 10.1002/tea.21303
16. Anderson BL, Williams S, Schulkin J. Statistical literacy of obstetrics-gynecology residents. *J Grad Med Educ.* 2013;5(2):272-275. DOI: 10.4300/JGME-D-12-00161.1
17. Franklin C, Kader G, Mewborn D, Moreno J, Peck, R, Perry M, Schaeffler R. Guidelines for Assessment and Instruction in Statistics Education (GAISE) report: A Pre-K-12 curriculum framework. Alexandria: VA: American Statistical Association; 2005.
18. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001;21(1):37-44. DOI: 10.1177/0272989X0102100105
19. Monahan J. *Statistical literacy: A prerequisite for evidence-based medicine.* Los Angeles, CA: SAGE Publications Sage; 2007.
20. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA.* 2007;298(9):1010-1022. DOI: 10.1001/jama.298.9.1010
21. Gigerenzer G, Wegwarth O. Risikoabschätzung in der Medizin am Beispiel der Krebsfrüherkennung. *Z Evid Fortbild Qual Gesundheitswes.* 2008;102(9):513-519. DOI: 10.1016/j.zefq.2008.09.008
22. Schwartz LM, Woloshin S, Welch HG. Can patients interpret health information? An assessment of the medical data interpretation test. *Med Decis Making.* 2005;25(3):290-300. DOI: 10.1177/0272989X05276860
23. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment Dec Making.* 2012;7(1):25-47. DOI: 10.1037/t45862-000
24. Okamoto M, Kyutoku Y, Sawada M, Clowney L, Watanabe E, Dan I, Kawamoto K. Health numeracy in Japan: measures of basic numeracy account for framing bias in a highly numerate population. *BMC Med Inform Decis Mak.* 2012;12:104. DOI: 10.1186/1472-6947-12-104
25. Johnson TV, Abbasi A, Schoenberg ED, Kellum R, Speake LD, Spiker C, Foust A, Kreps A, Ritenour CW, Brawley OW, Master VA. Numeracy among trainees: are we preparing physicians for evidence-based medicine? *J Surg Educ.* 2014;71(2):211-215. DOI: 10.1016/j.jsurg.2013.07.013
26. Msaouel P, Kappos T, Tasoulis A, Apostolopoulos AP, Lekkas I, Tripodaki ES, Keramaris NC. Assessment of cognitive biases and biostatistics knowledge of medical residents: a multicenter, cross-sectional questionnaire study. *Med Educ Online.* 2014;19:23646. DOI: 10.3402/meo.v19.23646
27. Berndt M, Schmidt FM, Sailer M, Fischer F, Fischer MR, Zottmann JM. Investigating Statistical Literacy and Scientific Reasoning & Argumentation in Medical-, Social Sciences-, and Economics Students. *Lern Individual Diff.* 2021;86:101963. DOI: 10.1016/j.lindif.2020.101963
28. Trempler K, Hetmanek A, Wecker C, Kieseewetter J, Wermelt M, Fischer F, Fischer M, Gräsel C. Nutzung von Evidenz im Bildungsbereich - Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und der Bewertung von Studien. *Z Pädagogik.* 2015;61:144-166.
29. Epstein N, Huber J, Gartmeier M, Berberat P, Reimer M, Fischer M. Investigating on the acquisition of scientific competencies during medical studies and the medical doctoral thesis. *GMS J Med Educ.* 2018;35(2):Doc20. DOI: 10.3205/zma001167
30. Simon D. Chronische Herzinsuffizienz. *Apothek Umschau.* 2016. Zugänglich unter/available from: <http://www.apotheken.de/gesundheits-heute-news/article/chronische-herzinsuffizienz/>
31. Neue Empfehlungen zum Einsatz von Weißdornpräparaten bei Herzinsuffizienz. *Ärztblatt.de.* 2017. Zugänglich unter/available from: <https://www.aerzteblatt.de/fachgebiete/kardiologie/news?nid=72259>
32. Meissner T. Weißdorn: Effekt auf das Endothel im Fokus. *Ärztezeitung.* 2017.
33. Harden M, Grant J, Buckley G, Hart R. BEME guide no. 1: best evidence medical education. *Med Teach.* 1999;21(6):553-526. DOI: 10.1080/01421599978960
34. Tabachnick BG, Fidell LS. *Using multivariate statistics.* 4th ed. Needham Heights, MA: Allyn & Bacon; 2001.
35. Hanoch Y, Miron-Shatz T, Cole H, Himmelstein M, Federman AD. Choice, numeracy and physician-in-training performance: The case of Medicare part D. *Health Psychol.* 2010;29:454-459. DOI: 10.1037/a0019881
36. Schmidt RL, Chute DJ, Colbert-Getz JM, Firpo-Betancourt A, James DS, Karp JK, Miller DC, Milner DA Jr, Smock KJ, Sutton AT, Walker BS, White KL, Wilson AR, Wojcik EM, Yared MA, Factor RE. Statistical Literacy Among Academic Pathologists: A Survey Study to Gauge Knowledge of Frequently Used Statistical Tests Among Trainees and Faculty. *Arch Pathol Lab Med.* 2017;141(2):279-287. DOI: 10.5858/arpa.2016-0200-OA
37. Riegelman RK, Hovland K. Scientific Thinking and Integrative Reasoning Skills (STIRS): Essential outcomes for medical education and for liberal education. *Peer Rev.* 2012;14(4):10.
38. Watson JM, Callingham R. Statistical literacy: A complex hierarchical construct. *Stat Educ Res J.* 2003;2(2):3-46.
39. Ghazal S, Cokely ET, Garcia-Retamero R. Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment Dec Making.* 2014;9(1):15-34.
40. Laopaiboon M, Lumbiganon P, Walter SD. Doctors' statistical literacy: a survey at Srinagarind Hospital, Khon Kaen University. *J Med Ass Thailand.* 1997;80(2):130-137.
41. Leira E, Granner M, Torner J, Callison R, Adams H. Education research: the challenge of incorporating formal research methodology training in a neurology residency. *Neurology.* 2008;70(20):e79-e84. DOI: 10.1212/01.wnl.0000312281.64033.36
42. Mieg H, editor. *Inquiry-Based Learning - Undergraduate Research.* Cham: Springer Nature; 2019. DOI: 10.1007/978-3-030-14223-0

Corresponding author:

Dr. phil. Markus Berndt
 University Hospital, LMU Munich, Institute of Medical Education, Pettenkoferstr. 8a, D-80336 Munich, Germany
 Markus.Berndt@med.uni-muenchen.de

Please cite as

Schmidt FM, Zottmann JM, Sailer M, Fischer MR, Berndt M. Statistical literacy and scientific reasoning & argumentation in physicians. *GMS J Med Educ.* 2021;38(4):Doc77.
DOI: 10.3205/zma001473, URN: urn:nbn:de:0183-zma0014734

This article is freely available from

<https://www.egms.de/en/journals/zma/2021-38/zma001473.shtml>

Received: 2020-04-27

Revised: 2020-12-05

Accepted: 2021-01-25

Published: 2021-04-15

Copyright

©2021 Schmidt et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Statistikverständnis und wissenschaftliches Denken & Argumentieren bei praktizierenden Ärzt*innen

Zusammenfassung

Ziel: Statistikverständnis (SV) von ÄrztInnen, d. h. die Fähigkeit statistische Zahlen im wissenschaftlichen Kontext zu verwenden und zu interpretieren, ist eine wichtige Voraussetzung für Risikoabschätzung und -kommunikation. Statistikverständnis bildet in Verbindung mit der Fähigkeit zum wissenschaftlichen Denken und Argumentieren (WDA) die Grundlage für eine evidenzbasierte medizinische Praxis. Mehrere Studien legen nahe, dass beide Fähigkeiten bei Medizinstudierenden nicht ausreichend entwickelt sind. Ziel der vorliegenden Studie war die Untersuchung dieser Fähigkeiten bei praktizierenden ÄrztInnen und wie diese Fähigkeiten erworben wurden.

Methoden: Die Daten von $N=71$ ÄrztInnen wurden online und in Papierform erhoben. Statistikverständnis wurde anhand von Multiple-Choice-Fragen gemessen, die WDA-Aktivitäten Evidenzen bewerten und Schlussfolgerungen ziehen anhand eines Entscheidungsszenarios.

Ergebnisse: Die Studienergebnisse zeigten bei ÄrztInnen ein mittleres Niveau im SV ($M=17.58$, $SD=6.92$, max. 30 Pkt.) und WDA (Evidenzen bewerten: $M=7.75$, $SD=1.85$, max. 10 Pkt.; Schlussfolgerungen ziehen: $M=37.20$, $SD=5.35$, max. 60 Pkt.). Die Entwicklung entsprechender Fähigkeiten durch autodidaktische Lernaktivitäten ($M=4.78$, $SD=1.13$, Spannweite 1–6) wurde signifikant häufiger angegeben als während der formalen medizinischen Ausbildung ($M=2.31$, $SD=1.46$), $t(71)=-9.915$, $p<.001$ oder durch außercurriculäre Aktivitäten ($M=3.34$, $SD=1.87$), $t(71)=4.673$, $p<.001$. Die aktive Beteiligung an Forschung schien eine entscheidende Rolle zu spielen: Die Anzahl der Veröffentlichungen und die Zeit der Tätigkeit in der Forschung korrelierten signifikant mit dem SV, $r(71)=.355$, $p=.002$; respektive $r(71)=.280$, $p=.018$. WDA-Fähigkeiten wurden vorhergesagt durch die Art der medizinischen Doktorarbeit, $\beta=-.380$, $p=.016$, sowie durch die Tätigkeit in der Forschung, $\beta=3.355$, $p=.008$.

Schlussfolgerung: Das aktive Mitwirken an Forschungsaktivitäten scheint ein sehr wichtiger Faktor für die Entwicklung sowohl des SV als auch des WDA zu sein. Dies spricht für die Implementierung einer systematischen Förderung dieser Fähigkeiten in die formale medizinische Ausbildung.

Schlüsselwörter: evidenzbasierte Praxis, wissenschaftliches Denken und Argumentieren, Statistikverständnis, medizinische Ausbildung, postgraduale medizinische Weiterbildung

1. Wissenschaftlicher Hintergrund

Wissenschaftliche Argumentationsfähigkeiten gelten in Anlehnung an die moderne Definition evidenzbasierter Medizin nach Sackett et al. (1997) zusammen mit ärztlichem Fachwissen als essenziell für die bestmögliche Entscheidungsfindung im Interesse von PatientInnen [1], [2], [3]. Für die Beurteilung von Risiken und Wahrscheinlichkeiten und deren Anwendung für PatientInnen ist ein grundlegendes Verständnis von Statistik und die Fähigkeit

Evidenzen zu bewerten erforderlich. Bei Statistikverständnis (SV, engl. statistical literacy) handelt es sich nicht nur um die Fähigkeit statistische Informationen zu verstehen, sondern diese auch in die Entscheidungsfindung einfließen zu lassen [4]. Hierzu gehört, Statistiken als Evidenzgrundlage der wissenschaftlichen Argumentation kritisch reflektieren zu können [5]. Statistikverständnis basiert auf einem entsprechenden Zahlenverständnis und dem Beherrschen mathematischer Operationen [6] und umfasst die Fähigkeit, statistische Zahlen im wissenschaftlichen Kontext zu verwenden und zu interpretieren [7], [8], [9], sowie die Fähigkeit, statistische Zahlen zu erklären

Felicitas M. Schmidt¹
Jan M. Zottmann¹
Maximilian Sailer²
Martin R. Fischer¹
Markus Berndt^{1,3}

1 Klinikum der Universität München, LMU München, Institut für Didaktik und Ausbildungsforschung in der Medizin, München, Deutschland

2 Universität Passau, Lehrstuhl für Erziehungswissenschaft, Passau, Deutschland

3 Walden University, Richard W. Riley College of Education and Leadership, Minneapolis, USA

und kritisch zu bewerten [10], [11] (Arbeitsdefinition für die vorliegende Studie). Zudem ist SV eng mit Fähigkeiten zum wissenschaftlichen Denken und Argumentieren (WDA, engl. scientific reasoning and argumentation) verbunden und schafft so die Grundlage für eine evidenzbasierte Entscheidungsfindung [12], [13].

Ausgehend vom Rahmenkonzept von Fischer et al. (2014) kann WDA als die Kompetenz definiert werden, wissenschaftliche Arbeitsmethoden und die damit verbundenen Ergebnisse zu verstehen und bei der Problemlösung anzuwenden [14], [15]. Wissenschaftliches Denken und Argumentieren kann anhand von acht epistemischen Aktivitäten beschrieben werden. Der Schwerpunkt dieser Studie liegt auf den beiden Aktivitäten Evidenzen bewerten (EB, engl. evidence evaluation) und Schlussfolgerungen ziehen (SZ, engl. drawing conclusions).

Verschiedene Studien legen nahe, dass WDA und SV eng miteinander verknüpft sind, wobei SV laut Anderson et al. (2013) zum Bewerten wissenschaftlicher Evidenz erforderlich ist [16]. Gemäß Franklin et al. (2005) umfasst SV selbst auch WDA-Fähigkeiten [17].

Jedoch wurde bei ÄrztInnen ein weit verbreitetes Defizit hinsichtlich statistischer Kenntnisse beobachtet [4], [18]. In ähnlicher Weise sind WDA-Fähigkeiten wie EB und SZ, die für eine evidenzbasierte Praxis [19] erforderlich sind [4], [7], unterentwickelt.

In einer Studie von Anderson et al. (2014) beantworteten 52% der teilnehmenden ÄrztInnen nur zwei (oder weniger) von vier Fragen zu statistischen Konzepten richtig [7]. Dies steht im Einklang mit den Ergebnissen von Windish et al. (2007), laut denen nur 40% der AssistenzärztInnen biostatistische Konzepte angemessen verstehen [20]. Zu ähnlichen Ergebnissen kamen Gigerenzer und Wegwarth (2008), denen zufolge 33% der GynäkologInnen den Nutzen des Mammografie-Screenings nicht kannten und 79% den positiven prädiktiven Wert nicht interpretieren konnten [21]. Gigerenzer et al. (2008) fassten verschiedene Studien zum Konzept des positiven prädiktiven Werts und seiner Abhängigkeit von der Prävalenz zusammen und fanden heraus, dass 50% der Teilnehmenden nicht an die Möglichkeit falsch positiver HIV-Testergebnisse glauben. Sie fanden außerdem heraus, dass nur zwei von 20 UrologInnen ausreichende Kenntnisse über die Zuverlässigkeit eines PSA-Tests haben. In der Folge existiert eine Vielzahl von Messinstrumenten zur Erhebung des SV [18], [22], [23], die sich in der Regel auf eine der drei Ebenen von Watson (1997) konzentrieren oder nur für einen bestimmten Kontext konzipiert sind [7], [22]. Insgesamt ist das SV von ÄrztInnen nicht unterdurchschnittlich ausgeprägt [7]. Es kann als vergleichbar mit anderen akademischen Stichproben [18], [24] angesehen werden und war dem SV von AssistenzärztInnen in der Forschungsausbildung [20] oder von Medizinstudierenden [25] nachweislich überlegen. Die wenigen Studien, in denen das SV von Medizinstudierenden untersucht wurde, stützen das Untersuchungsergebnis der diesbezüglichen Überlegenheit von ÄrztInnen [26]. Berndt et al. (2021) verglichen Medizinstudierende mit Studierenden der So-

zial- und Wirtschaftswissenschaften und fanden heraus, dass Medizinstudierende in ihren ersten Studienjahren im Vergleich zu Studierenden der Sozialwissenschaften besser und auf vergleichbarem Niveau wie Studierende der Wirtschaftswissenschaften abschnitten [27]. In dieser Studie wurde nicht nur SV erhoben, sondern auch die beiden oben genannten WDA-Aktivitäten EB und SZ. Weitere mögliche Zusammenhänge zwischen SV und WDA wurden mit dem Medical Data Interpretation Test [22] untersucht, bei dem teilnehmende ÄrztInnen insgesamt besser als Teilnehmende mit anderen postgradualen Abschlüssen abschnitten (89 von 100 Punkten). Johnson et al. (2014) untersuchten das Zahlenverständnis von Medizinstudierenden und AssistenzärztInnen und stellten fest, dass Studierende mit schlechtem Zahlenverständnis die Risiken verschiedener Behandlungsalternativen häufiger falsch einschätzen, wobei das Vertrauen in die eigene Behandlungsempfehlung während des Medizinstudiums zunahm [25].

Zusammenfassend lässt sich sagen, dass SV und WDA bei Medizinstudierenden unterentwickelt, bei ÄrztInnen jedoch besser entwickelt zu sein scheinen. Die Entwicklung von SV und WDA muss aber nicht unbedingt innerhalb der formalen Ausbildung erfolgen, und es stellt sich die Frage, wie, wo, und wann diese Entwicklung stattfindet. Die vorliegende Studie soll einen Beitrag zur Diskussion leisten, wann und wie SV und WDA im lebenslangen Lernen von ÄrztInnen am besten gefördert werden können. Hierzu wurde ein Testinstrument eingesetzt, das verschiedene SV- und WDA-Aspekte umfasst. Ziel der Studie war es, weitere Erkenntnisse über diese Fähigkeiten bei ÄrztInnen zu gewinnen und demografische Faktoren und Lerngelegenheiten zu identifizieren, die möglicherweise mit der Entwicklung dieser Fähigkeiten in Zusammenhang stehen. Unsere explorativen Forschungsfragen lauteten:

FF1. Statistikverständnis und WDA

- In welchem Umfang sind SV und WDA bei ÄrztInnen entwickelt?
- In welchem Ausmaß sagt der SV-Score die WDA-Fähigkeiten von ÄrztInnen voraus?

FF2. Entwicklung von SV und WDA

- Wie, wo, und wann entwickeln ÄrztInnen SV und WDA?
- Welche demografischen Faktoren stehen mit der Entwicklung von SV und WDA in Zusammenhang?

2. Methoden

2.1. Design und Stichprobe

Unsere Studie folgte einem quasi-experimentellen, kausal-vergleichenden Design mit zwei abhängigen Variablen: SV und WDA. Wir schlossen deutschsprachige ÄrztInnen ($N=71$; 31 Frauen, 34 Männer, 6 k.A.) aus verschiedenen Arbeitsumfeldern und Orten in unsere Studie ein: Klinik ($n=44$), ambulante Versorgung ($n=3$), Forschung ($n=8$),

Tabelle 1: Beschreibung der Untersuchungsgruppe (N=71)

Variable	Antwortmöglichkeiten	n	Prozentzahlen
Geschlecht	Männlich	34	47.9%
	Weiblich	31	43.7%
	Keine Angabe	6	8.4%
Deutsch als Muttersprache	Ja	63	88.7%
	Nein	3	4.2%
	Keine Angabe	5	7.0%
Medizinstudium in Deutschland	Ja	68	95.8%
	Nein	3	4.2%
Medizinische Doktorarbeit	Ja	58	81.7%
	Aktuell laufend	9	12.7%
	Keine oder abgebrochen	4	5.6%
Akademische Qualifikation	Habilitation	12	16.9%
	Professur	7	9.9%
Als WissenschaftlerIn gearbeitet	Ja	36	50.7%
	Nein	35	49.3%
Arbeitsumfeld	Klinik	44	62.0%
	Ambulante Versorgung	3	4.2%
	Forschung	8	11.3%
	Studium	2	2.8%
	Verwaltung	2	2.8%
	Keine Angabe	12	16.9%

Studienprogramm ($n=2$) und Verwaltung ($n=2$), ($n=12$ k.A.). Eine medizinische Dissertation, eine wissenschaftliche Arbeit als optionaler Teil des Medizinstudiums (nicht äquivalent mit einer PhD-Arbeit), wurde von 58 Teilnehmenden abgeschlossen und von 9 als aktuell laufend angegeben. Trotz unserer relativ kleinen Stichprobengröße sehen wir diese hinsichtlich der wissenschaftlichen Erfahrung für repräsentativ, da medizinische Dissertationen in Deutschland sehr verbreitet sind. Das Durchschnittsalter der Teilnehmenden betrug 40 Jahre ($SD=9.59$, Spannweite=26–65) (siehe Tabelle 1).

2.2. Testinstrument

Für die Messung von SV und WDA haben wir ein Instrument verwendet, das zuvor im Rahmen einer Studie von Berndt et al. (2021) entwickelt wurde [27]. Das Testinstrument kombiniert Multiple-Choice-Fragen zur Erhebung des SV mit einem Entscheidungsszenario [28] zur Erhebung der WDA-Aktivitäten EB und SZ der Teilnehmenden. Für die vorliegende Studie wurden Items zu relevanten demografischen Faktoren hinzugefügt und mit zehn Medizinstudierenden der LMU München pilotiert.

2.2.1. Demografie

Demografische und biografische Parameter der Teilnehmenden wurden mit besonderem Interesse an ihrem beruflichen Werdegang und ihrem Arbeitsumfeld (Klinik, ambulante Versorgung, Forschung) erfasst. Die aus einer Studie von Epstein et al. [29] adaptierten Items waren Multiple-Choice-Fragen. Teilweise bestand hierbei die Möglichkeit, zusätzlichen Freitext einzugeben. Fünf Multiple-Choice-Fragen bezogen sich auf die medizinische

Dissertation, drei Fragen auf den beruflichen Werdegang, zwei Fragen auf die bisherigen Veröffentlichungen (Art der Autorenschaft, Anzahl der Veröffentlichungen), und drei Fragen auf das aktuelle Stellenprofil.

2.2.2. Statistikverständnis

Das SV wurde mit Multiple-Choice-Items gemessen, die auf validierten Instrumenten [7], [18], [23] basierten. Auf diese Weise konnte ein breites Spektrum ausgewertet werden, von grundlegenden Zahlenverständnis bis hin zum Rechnen mit bedingten Wahrscheinlichkeiten und dem Umgang mit statistischen Konzepten. Doppelte Items und reine Faktenwissensfragen wurden herausgenommen, sodass letztendlich alle drei von Watson (1997) beschriebenen Ebenen abgedeckt wurden. Zusätzlich wurden alle Items nach Schwierigkeit gewichtet [11]. Die interne Konsistenz des SV-Tests betrug in unserer Stichprobe bei einer maximalen Punktzahl von 30 Punkten .82 (Cronbachs α). Alle Items waren in einem medizinischen Kontext formuliert, wobei zur richtigen Beantwortung kein medizinisches Fachwissen erforderlich war.

2.2.3. Wissenschaftliches Denken und Argumentieren

Die Erhebung der WDA-Fähigkeiten konzentrierte sich auf die beiden epistemischen Aktivitäten EB und SZ in einem Entscheidungsszenario im medizinischen Kontext (Allgemeinmedizin, ambulante Versorgung). Hierbei gab es zwei separate Scores für EB und SZ (Cronbachs α betrug für EB-Items .87 und für SZ-Items .74). Für den EB-Score mussten die Teilnehmenden vier kurze Artikel (Evidenzen) [30], [31], [32] auf einer sechs Punkte umfassenden Likert-Skala gemäß den QUEST-Kriterien [33]

hinsichtlich ihrer wissenschaftlichen Qualität, Evidenzstärke, und Relevanz für die vorliegende Situation bewerten, darunter eine authentische pharmazeutische Broschüre, in der für ein pflanzliches Arzneimittel geworben wird.

Anschließend bewerteten die Teilnehmenden die Überzeugungskraft (Likert 1-6) von 20 Argumenten, die im Vorfeld von den Autoren aus den präsentierten Evidenzen extrahiert wurden, und vergaben eine Einschätzung der Argumentstärke von 1 (sehr schwach) bis 4 (sehr stark). Bei 13 Teilnehmenden fehlten bis zu fünf Bewertungen der insgesamt 20 Argumente. Um diese Datensätze weiterhin verwenden zu können, wurden die Werte aus dem Durchschnitt der jeweiligen Bewertungen durch Imputation ergänzt. Aus den durch die Teilnehmenden vorgenommenen Bewertungen der wissenschaftlichen Qualität wurde mit einer unabhängigen Bewertung der wissenschaftlichen Qualität seitens der Autoren ein Übereinstimmungsmaß für EB und SZ errechnet. Dieses bewegte sich für EB zwischen 0–10 (EB-Score) und für SZ zwischen 0–60 (SZ-Score), wobei Null auf keine Übereinstimmung hinwies.

2.3. Ablauf

Die Studie wurde von den Teilnehmenden entweder online mit LamaPoll [<https://www.lamapoll.de/>], einem für Mobilgeräte optimierten Umfrage-Tools, oder auf Papierfragebögen ausgefüllt (Rücklaufquote online 16.5% und Papierfragebögen 66.7%). Die durchschnittliche Dauer betrug etwa 45 Minuten. Die Einladung der Teilnehmenden erfolgte über Mailinglisten und persönliche Kontakte.

2.4. Statistische Analysen

Statistische Analysen erfolgten mit IBM SPSS 25. Für die Primäranalyse wurden deskriptive und Häufigkeitsdaten berechnet und für die interne Konsistenz Cronbachs Alpha. Es erfolgten umfangreiche Ausreißeranalysen. Weiterhin wurden alle für die statistischen Analysen erforderlichen Voraussetzungen wie Normalverteilung und Homoskedastizität geprüft. T-Tests, einfaktorische Varianzanalysen, und lineare Regressionsmodelle wurden verwendet, um Unterschiede zu berechnen und den Zusammenhang zwischen demografischen Faktoren und SV und WDA zu beurteilen. Wahrscheinlichkeitswerte kleiner als .05 wurden als signifikant angesehen. In verbaler Sprache vorliegende Daten (Freitext im Abschnitt zur Demografie) wurden von zwei der Autoren einer separaten thematischen Analyse unterzogen, um häufig vorkommende Themen zu extrahieren.

3. Ergebnisse

Wir schlossen 71 ausgefüllte Fragebögen ein (siehe Tabelle 1). Der gesamte Datensatz wurde auf univariate Ausreißer geprüft. Schiefe und Wölbung lagen bei allen Variablen innerhalb der ± 2 Spannweite [34]. Sofern fol-

gend nicht anders angegeben, waren die Voraussetzungen für t-Tests und Varianzanalysen erfüllt.

3.1. Statistikverständnis und WDA

Der durchschnittliche SV-Score der 71 teilnehmenden ÄrztInnen lag bei $M=17.58$, $SD=6.92$ (59%), mit einer Spannweite von 5 bis 30 bei 30 erreichbaren Punkten. Die ÄrztInnen bewerteten die vier Evidenzen im Durchschnitt übereinstimmend mit der Bewertung der Autoren, EB-Score: $M=7.75$, $SD=1.85$ (77%). Die Bewertungen für die Qualität der Argumente entsprachen der jeweiligen Bewertung der Autoren, SZ-Score: $M=37.20$, $SD=5.35$ (62%). Zwischen SV und SZ bestand eine signifikante, inverse Korrelation, $r(71)=-.272$, $p=.022$. Eine Korrelation zwischen SV und EB, $r(71)=.198$, $p=.098$ und zwischen EE und DC, $r(71)=.138$, $p=.256$ wurde jedoch nicht gefunden.

3.2. Entwicklung von Fähigkeiten

Wir untersuchten, wie, wo, und wann ÄrztInnen wissenschaftliche Fähigkeiten entwickelten (siehe Abbildung 1). Signifikant mehr Teilnehmende gaben an, wissenschaftliche Fähigkeiten autodidaktisch ($M=4.78$, $SD=1.13$, Likert-Skala von 1 bis 6) versus während des Studiums ($M=2.31$, $SD=1.46$, Likert-Skala von 1 bis 6), $t(71)=-9.915$, $p<.001$ oder bei außercurriculären Aktivitäten ($M=3.34$, $SD=1.87$, Likert-Skala von 1 bis 6), $t(71)=4.673$, $p<.001$ erworben zu haben. In einem Freitextfeld ergänzten die Teilnehmenden verschiedene weitere Lerngelegenheiten, z. B. Massive Open Online Courses, Hochschulkurse, Seminare, und Lernen durch Peer-Reviews und Peer-Feedback (siehe Abbildung 2).

Eine abgeschlossene oder in Arbeit befindliche medizinische Doktorarbeit zeigte keine Auswirkungen auf SV, EB, oder SZ. Diese Ergebnisse müssen jedoch mit Vorsicht interpretiert werden, da in unserer Stichprobe, bei der nur vier Teilnehmende keine medizinische Dissertation vorzuweisen hatten, die Voraussetzungen für eine Varianzanalyse nicht erfüllt waren. Die Förderung des kritischen Hinterfragens von Studienergebnissen anderer Forscher bei der Vorbereitung der eigenen medizinischen Doktorarbeit zeigte eine positive Korrelation mit dem SV, $r(71)=.271$, $p=.033$.

Bezüglich der postgradualen Phase zeigte eine einfaktorische Varianzanalyse einen Haupteffekt der Forschungserfahrung, $F(1,70)=12.737$, $p=.001$, partielles $\eta^2=.156$ und einen Haupteffekt der Art der Autorenschaft in Veröffentlichungen, $F(5,71)=3.886$, $p=.004$, partielles $\eta^2=.230$. Die Zeit der Tätigkeit in der Forschung korrelierte signifikant mit besserem SV, $r(71)=.28$, $p=.018$. Dasselbe galt für die Anzahl der Veröffentlichungen, $r(71)=.36$, $p=.002$. Hinsichtlich des WDA ergaben lineare Regressionsmodelle, dass der entsprechende EB-Score um $\beta=.314\pm.150$, $p=.041$ anstieg, wenn der Likert-Wert der inhaltlichen Unterstützung durch den Doktorvater bzw. die Doktorin um einen Punkt anstieg. Zusätzlich korrelierte die Art der medizinischen Dissertation (experimentell, klinisch,

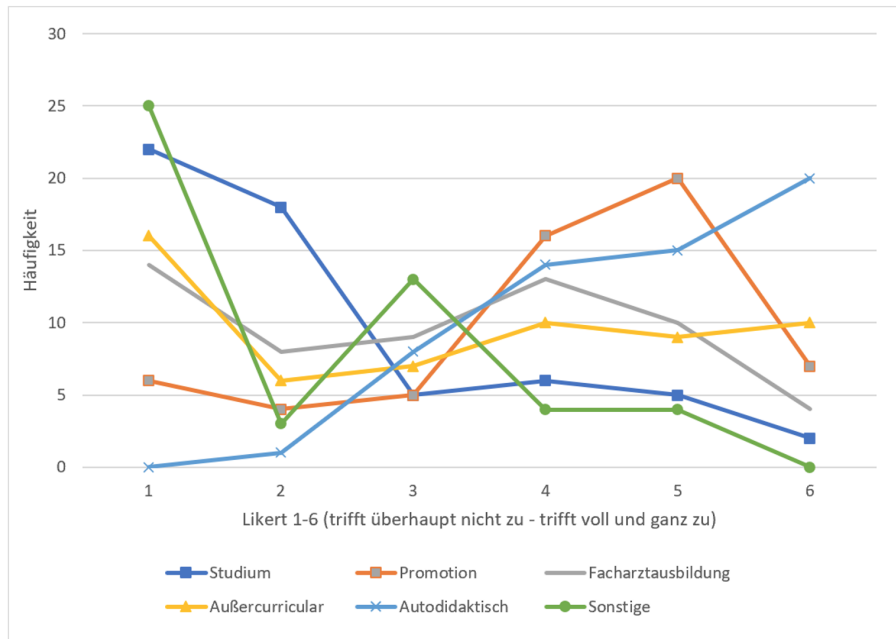


Abbildung 1: Wo haben Sie Ihrer Meinung nach Wissenschaftskompetenz erworben? (N=71 ÄrztInnen)

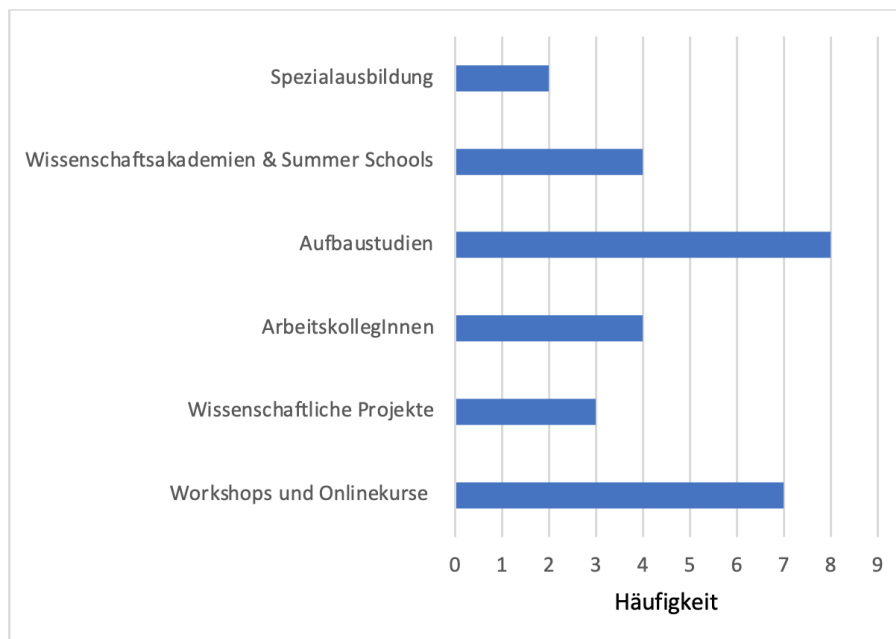


Abbildung 2: Wo haben Sie Ihrer Meinung nach Wissenschaftskompetenz erworben? Freitextanalyse der Kategorie Sonstige (n=28 ÄrztInnen)

empirisch, statistisch, oder Literaturstudie) mit dem EB-Score, wobei ein experimentelles und klinisches Design positiv mit dem EB-Score in Zusammenhang stand, $\beta = -.380 \pm .154$, $p = .016$, $R^2 = .187$, $F(1,59) = 4.353$, $p = .041$. Der SZ-Score war höher, wenn die Teilnehmenden angaben, bereits in der Forschung gearbeitet zu haben, $\beta = 3.355 \pm 1.229$, $p = .008$, $R^2 = .314$, $F(1,68) = 7.448$, $p = .008$.

4. Diskussion

4.1. Forschungsfrage 1: Statistikverständnis und WDA

Wir fanden bei ÄrztInnen einen durchschnittlichen SV-Score (59%), einen relativ hohen EB-Score (77%) und einen mittleren SZ-Score (62%). Der SV-Score gab keinen Aufschluss über die WDA-Fähigkeiten von ÄrztInnen. Durch den Fokus auf SV und nicht auf die Kombination von grundlegenden Zahlenverständnis und SV [27] differenzierte unser Testinstrument gut und es konnten keine

Deckeneffekte, wie sie in anderen akademischen Stichproben beobachtet wurden [7], [35], nachgewiesen werden. Ein Vergleich mit anderen Studien, die das SV von ÄrztInnen beurteilen, ist nicht einfach, da jeder Test einen unterschiedlichen SV-Bereich abdeckt. Schmidt et al. (2017) untersuchten bei PathologInnen das Wissen zu 18 verschiedenen statistischen Tests und fanden ein relativ niedriges SV [36]. Anderson et al. (2014) erstellten keinen Gesamtscore, sondern unterschieden zwischen Fakten-, Konzept- und Relationsfragen und ermittelten unterschiedliche Niveaus von SV [7]. Eine Studie mit AssistenzärztInnen in Griechenland konzentrierte sich ebenfalls auf Wissensfragen und berichtete ein relativ niedriges SV-Niveau [26]. Die EB- und SZ-Scores der deutschen Medizinstudierenden, die wir zuvor mit einem ähnlichen Instrument untersucht hatten [27], lagen fast auf dem gleichen Niveau wie die Scores der ÄrztInnen in der vorliegenden Studie. Riegelman und Hoveland (2012) fanden heraus, dass AssistenzärztInnen Probleme hatten, wenn kritische Reflexion zu Forschungsinhalten erforderlich war [37], während die ÄrztInnen in unserer Studie mittlere bis hohe WDA-Fähigkeiten zeigten. Zwischen EB- und SV-Scores bestand keine Korrelation. Zwischen SZ und SV fanden wir eine inverse Korrelation. In kontextuellen Rahmenmodellen wurde SV als Voraussetzung für WDA angesehen [38]. In einer niederländischen, auf die Allgemeinbevölkerung bezogenen Studie, zeigten die Teilnehmenden mit höherem Zahlenverständnis besseres WDA, das sich aus einer stärkeren Abwägung von Vor- und Nachteilen bei der Entscheidungsfindung und genaueren Bewertung der eigenen Einschätzungen ergab [39]. Da die Informationen der vier Evidenzen in unserer Studie nicht vorwiegend numerisch oder statistisch dargestellt wurden, war von der fehlenden Verknüpfung zwischen EB und SV auszugehen, jedoch nicht von der gegensätzlichen Beziehung von SZ und SV. Künftige Forschung könnte statistische Informationen in Entscheidungsszenarien inkludieren, um diesen Zusammenhang bei praktizierenden ÄrztInnen genauer zu analysieren.

4.2. Forschungsfrage 2: Entwicklung von SV und WDA

Wir untersuchten, wie, wo, und wann ÄrztInnen SV und WDA-Fähigkeiten entwickelten. Sie gaben an, wissenschaftliche Fähigkeiten meist autodidaktisch, in Universitätsseminaren außerhalb des Medizinstudiums, oder in außercurriculären Aktivitäten erworben zu haben. Besseres SV korrelierte mit der Förderung der kritischen Reflexion von Studienergebnissen während der medizinischen Promotion, Arbeit in der Forschung oder mit früherer Forschungserfahrung, sowie mit der Anzahl von Veröffentlichungen und der Art der Autorenschaft. Unsere Ergebnisse stehen im Einklang mit den Ergebnissen von Schmidt et al. (2017), denen zufolge andere Weiterqualifikationen oder absolvierte Statistikurse über das Medizinstudium hinaus, positiv mit dem SV korrelierten. Eine Studie mit ÄrztInnen, AssistenzärztInnen und Medizinstu-

dierenden im Praktischen Jahr in Thailand zeigte wenig überraschend, dass ein kürzlich absolviertes Statistikseminar zu höheren SV-Scores führte [40]. Zusatzkurse lassen sich allerdings oft nur schwer in die medizinische Ausbildung integrieren. Eine Studie zeigte, dass 37% der amerikanischen AssistenzärztInnen im Bereich Gynäkologie und Geburtshilfe keine formale Statistikausbildung erhalten [16], während eine andere Studie mit AssistenzärztInnen im Bereich Neurologie eine mangelnde Akzeptanz hinsichtlich statistikbezogener Fortbildungsmaßnahmen beobachtete [41].

In der vorliegenden Studie fand sich ein Zusammenhang zwischen besserem EB-Score und entsprechender Verantwortungserfahrung für ein Forschungsprojekt (z.B. die medizinische Doktorarbeit) mit experimentellem oder klinischem Design sowie wenn eine gute inhaltliche Unterstützung durch den Doktorvater bzw. die Doktormutter bestanden hatte. Diese Erkenntnis deckt sich mit dem subjektiven Eindruck deutscher promovierter Medizinabsolventen, die ihre wissenschaftlichen Fähigkeiten höher einschätzen als noch an der Promotion arbeitende AbsolventInnen [29]. Die Teilnehmenden der Studie von Epstein et al. (2018) fühlten sich jedoch für eigene Forschungstätigkeiten nicht sicher genug. Dies erscheint besonders wichtig, da vorhandene Forschungserfahrung in der vorliegenden Studie und bei Schmidt et al. (2017) mit einem höheren SV und SZ-Score korrelierte. Darüber hinaus fanden Epstein et al. (2018) heraus, dass Medizinabsolventen ihre wissenschaftlichen Fähigkeiten nach dem Medizinstudium selbst als relativ gering einstufen. In den USA waren nur 68,1% der Medizinstudierenden in ihrem letzten Jahr während des Medizinstudiums an Forschungsprojekten beteiligt, und nur 42% waren Autoren bzw. Koautoren einer zur Veröffentlichung eingereichten Arbeit. Es erscheint wichtig, dass Medizinstudierende während der Promotion in weitere Forschungsprojekte und die anschließende Veröffentlichung der Ergebnisse eingebunden werden, da dies langfristig ihr SV und WDA verbessern könnte.

4.3. Stärken und Grenzen

Diese Studie baute auf einem innovativen Ansatz der Autoren zur Messung von SV und WDA bei Universitätsstudierenden [27] auf. Der integrative Ansatz der Messung des SV erlaubte eine bessere Beschreibung der realen Fähigkeiten. Allerdings besteht hier der Nachteil einer eingeschränkten Vergleichbarkeit mit früheren Forschungsarbeiten. Da die Teilnehmendengruppe der praktizierenden ÄrztInnen nicht leicht zu rekrutieren war, hielten wir die Stichprobengröße von $N=71$ für zufriedenstellend. Obwohl die Generalisierbarkeit potenziell begrenzt ist, ist unsere Stichprobe repräsentativ für den deutschsprachigen Raum, in dessen medizinischen Ausbildungssystem eine große Anzahl medizinischer Dokortitel erworben wird. Die zusätzliche Berücksichtigung zahlreicher demografischer Variablen lieferte Aufschluss darüber, wie, wo, und wann wissenschaftliche Fähigkeiten

erworben wurden und ermöglichte die Identifikation potenzieller einflussnehmender Faktoren.

Aufgrund des breiten Altersspektrums unserer Stichprobe ist davon auszugehen, dass die Teilnehmenden unterschiedliche Lernerfahrungen in der formalen medizinischen Ausbildung gemacht haben und je nach Studienort auch in reformierten Medizinstudiengängen studiert haben. Dies könnte sich auf die Entwicklung der untersuchten Fähigkeiten ausgewirkt haben und weitere individuelle Unterschiede begünstigen. In unserer Studie wurden keine Daten über spezifische Studienprogramme, Seminare und die zugehörigen Beschreibungen erhoben, in denen die ÄrztInnen möglicherweise ihre Fähigkeiten erworben haben. Dies ist dem Umstand geschuldet, dass das in dieser Studie verwendete Testmaterial im Feedback von 11 teilnehmenden ÄrztInnen bereits als sehr umfangreich und zeitaufwendig bewertet wurde.

5. Schlussfolgerung

Wir untersuchten SV und WDA bei deutschsprachigen ÄrztInnen und nahmen in diesem Zusammenhang eine eingehende Analyse demografischer Variablen vor. Die aktive Beteiligung an Forschung spielt offenbar eine wichtige Rolle bei der Entwicklung dieser Fähigkeiten und könnte als Konsequenz eine Verbesserung der evidenzbasierten Praxis nach sich ziehen. Da die meisten Teilnehmenden angaben diese Fähigkeiten postgradual und autodidaktisch erworben zu haben, plädieren wir dafür, den Erwerb dieser Fähigkeiten im Rahmen des Medizinstudiums zu formalisieren und zu intensivieren. Die Lehrpläne der medizinischen Ausbildung sollten mehr statistisches Training beinhalten und darauf abzielen, Studierende häufiger in aktive Forschung einzubeziehen, z. B. durch das Angebot von forschendem Lernen [42], bei dem die Studierenden eigenständig Forschungsprojekte durchführen und für alle Phasen des Forschungsprozesses in vollem Umfang verantwortlich sind.

Förderung

Diese Arbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF, Förderkennzeichen O1PB14004C) und durch ein internes Förderprogramm des Fördervereins WiFoMed der Medizinischen Fakultät der LMU München unterstützt.

Danksagung

Die Autor*innen bedanken sich außerdem bei Wolfgang Gaissmaier für seine wertvollen Anmerkungen und Anregungen zur vorliegenden Studie.

Interessenkonflikt

Die Autor*innen erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Literatur

1. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull.* 2009;135(6):943-973. DOI: 10.1037/a0017327
2. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. *Ann Behav Med.* 2008;35(3):261-274. DOI: 10.1007/s12160-008-9037-8
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Was ist Evidenz-basierte Medizin und was nicht? *Munch Med Wochenschr.* 1997;139(44):28-29. DOI: 10.1055/b-0036-140841
4. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest.* 2008;8(2):53-96. DOI: 10.1111/j.1539-6053.2008.00033.x
5. Shield M. Statistical literacy: thinking critically about statistics. *Significance.* 1999;1(1):15-20.
6. Peters E. Beyond Comprehension: The Role of Numeracy in Judgments and Decisions. *Curr Dir Psychol Sci.* 2012;21(1):31-35. DOI: 10.1177/0963721411429960
7. Anderson BL, Gigerenzer G, Parker S, Schulkin J. Statistical literacy in obstetricians and gynecologists. *J Healthc Qual.* 2014;36(1):5-17. DOI: 10.1111/j.1945-1474.2011.00194.x
8. Walker HM. Statistical literacy on the social science. *Am Stat.* 1951;5(1):6-12. DOI: 10.1080/00031305.1951.10481912
9. Wallman KK. Enhancing statistical literacy: enriching our society. *J Am Statl Ass.* 1993;88(421):1-8. DOI: 10.1080/01621459.1993.10594283
10. Ben-Zvi D, Garfield J. Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In: Ben-Zvi D, Garfield J, editors. *The challenge of developing statistical literacy, reasoning and thinking.* Dordrecht, The Netherlands: Kluwer; 2004. p.3-25. DOI: 10.1007/1-4020-2278-6
11. Watson JM. Assessing Statistical Thinking Using the Media. In: Gal I, Garfield JB, editors. *The Assessment Challenge in Statistics Education.* Amsterdam: IOS Press and The International Statistical Institute; 1997. p.107-121.
12. Sedlmeier P, Gigerenzer G. Teaching Bayesian Reasoning in Less Than Two Hours. *J Exp Psychol Gen.* 2001;130(3):380-400. DOI: 10.1037/0096-3445.130.3.380
13. Fischer F, Kollar I, Ufer S, Sodian B, Hussmann H, Pekrun R, Neuhaus B, Dorner B, Pankofer S, Fischer M, Strijbos JW, Heene M, Eberle J. Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Front Learn Res.* 2014;2(3):28-45. DOI: 10.14786/flr.v2i2.96
14. Hetmanek A, Engelmann K, Opitz A, Fischer F. Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. In: Fischer F, Chinn CA, Engelmann K, Osborne J, editors. *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge.* New York: Routledge; 2018. p.203-226.
15. Rudolph JL, Horibe S. What do we mean by science education for civic engagement? *J Res Sci Teach.* 2016;53(6):805-820. DOI: 10.1002/tea.21303

16. Anderson BL, Williams S, Schulkin J. Statistical literacy of obstetrics-gynecology residents. *J Grad Med Educ.* 2013;5(2):272-275. DOI: 10.4300/JGME-D-12-00161.1
17. Franklin C, Kader G, Mewborn D, Moreno J, Peck, R, Perry M, Schaeffler R. Guidelines for Assessment and Instruction in Statistics Education (GAISE) report: A Pre-K-12 curriculum framework. Alexandria: VA: American Statistical Association; 2005.
18. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001;21(1):37-44. DOI: 10.1177/0272989X0102100105
19. Monahan J. Statistical literacy: A prerequisite for evidence-based medicine. Los Angeles, CA: SAGE Publications Sage; 2007.
20. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA.* 2007;298(9):1010-1022. DOI: 10.1001/jama.298.9.1010
21. Gigerenzer G, Wegwarth O. Risikoabschätzung in der Medizin am Beispiel der Krebsfrüherkennung. *Z Evid Fortbild Qual Gesundheitswes.* 2008;102(9):513-519. DOI: 10.1016/j.zefq.2008.09.008
22. Schwartz LM, Woloshin S, Welch HG. Can patients interpret health information? An assessment of the medical data interpretation test. *Med Decis Making.* 2005;25(3):290-300. DOI: 10.1177/0272989X05276860
23. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment Dec Making.* 2012;7(1):25-47. DOI: 10.1037/t45862-000
24. Okamoto M, Kyutoku Y, Sawada M, Clowney L, Watanabe E, Dan I, Kawamoto K. Health numeracy in Japan: measures of basic numeracy account for framing bias in a highly numerate population. *BMC Med Inform Decis Mak.* 2012;12:104. DOI: 10.1186/1472-6947-12-104
25. Johnson TV, Abbasi A, Schoenberg ED, Kellum R, Speake LD, Spiker C, Foust A, Kreps A, Ritenour CW, Brawley OW, Master VA. Numeracy among trainees: are we preparing physicians for evidence-based medicine? *J Surg Educ.* 2014;71(2):211-215. DOI: 10.1016/j.jsurg.2013.07.013
26. Msaouel P, Kappos T, Tasoulis A, Apostolopoulos AP, Lekkas I, Tripodaki ES, Keramaris NC. Assessment of cognitive biases and biostatistics knowledge of medical residents: a multicenter, cross-sectional questionnaire study. *Med Educ Online.* 2014;19:23646. DOI: 10.3402/meo.v19.23646
27. Berndt M, Schmidt FM, Sailer M, Fischer F, Fischer MR, Zottmann JM. Investigating Statistical Literacy and Scientific Reasoning & Argumentation in Medical-, Social Sciences-, and Economics Students. *Lern Individual Diff.* 2021;86:101963. DOI: 10.1016/j.lindif.2020.101963
28. Trempler K, Hetmanek A, Wecker C, Kiesewetter J, Wermelt M, Fischer F, Fischer M, Gräsel C. Nutzung von Evidenz im Bildungsbereich - Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und der Bewertung von Studien. *Z Pädagogik.* 2015;61:144-166.
29. Epstein N, Huber J, Gartmeier M, Berberat P, Reimer M, Fischer M. Investigating on the acquisition of scientific competencies during medical studies and the medical doctoral thesis. *GMS J Med Educ.* 2018;35(2):Doc20. DOI: 10.3205/zma001167
30. Simon D. Chronische Herzinsuffizienz. *Apothek Umschau.* 2016. Zugänglich unter/available from: <http://www.apotheken.de/gesundheit-heute-news/article/chronische-herzinsuffizienz/>
31. Neue Empfehlungen zum Einsatz von Weißdornpräparaten bei Herzinsuffizienz. *Ärztblatt.de.* 2017. Zugänglich unter/available from: <https://www.aerzteblatt.de/fachgebiete/kardiologie/news?nid=72259>
32. Meissner T. Weißdorn: Effekt auf das Endothel im Fokus. *Ärztezeitung.* 2017.
33. Harden M, Grant J, Buckley G, Hart R. BEME guide no. 1: best evidence medical education. *Med Teach.* 1999;21(6):553-526. DOI: 10.1080/01421599978960
34. Tabachnick BG, Fidell LS. Using multivariate statistics. 4th ed. Needham Heights, MA: Allyn & Bacon; 2001.
35. Hanoch Y, Miron-Shatz T, Cole H, Himmelstein M, Federman AD. Choice, numeracy and physician-in-training performance: The case of Medicare part D. *Health Psychol.* 2010;29:454-459. DOI: 10.1037/a0019881
36. Schmidt RL, Chute DJ, Colbert-Getz JM, Firpo-Betancourt A, James DS, Karp JK, Miller DC, Milner DA Jr, Smock KJ, Sutton AT, Walker BS, White KL, Wilson AR, Wojcik EM, Yared MA, Factor RE. Statistical Literacy Among Academic Pathologists: A Survey Study to Gauge Knowledge of Frequently Used Statistical Tests Among Trainees and Faculty. *Arch Pathol Lab Med.* 2017;141(2):279-287. DOI: 10.5858/arpa.2016-0200-0A
37. Riegelman RK, Hovland K. Scientific Thinking and Integrative Reasoning Skills (STIRS): Essential outcomes for medical education and for liberal education. *Peer Rev.* 2012;14(4):10.
38. Watson JM, Callingham R. Statistical literacy: A complex hierarchical construct. *Stat Educ Res J.* 2003;2(2):3-46.
39. Ghazal S, Cokely ET, Garcia-Retamero R. Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment Dec Making.* 2014;9(1):15-34.
40. Laopaiboon M, Lumbiganon P, Walter SD. Doctors' statistical literacy: a survey at Srinagarind Hospital, Khon Kaen University. *J Med Ass Thailand.* 1997;80(2):130-137.
41. Leira E, Granner M, Torner J, Callison R, Adams H. Education research: the challenge of incorporating formal research methodology training in a neurology residency. *Neurology.* 2008;70(20):e79-e84. DOI: 10.1212/01.wnl.0000312281.64033.36
42. Mieg H, editor. *Inquiry-Based Learning - Undergraduate Research.* Cham: Springer Nature; 2019. DOI: 10.1007/978-3-030-14223-0

Korrespondenzadresse:

Dr. phil. Markus Berndt
 Klinikum der Universität München, LMU München, Institut für Didaktik und Ausbildungsforschung in der Medizin, Pettenkoflerstr. 8a, 80336 München, Deutschland
 Markus.Berndt@med.uni-muenchen.de

Bitte zitieren als

Schmidt FM, Zottmann JM, Sailer M, Fischer MR, Berndt M. Statistical literacy and scientific reasoning & argumentation in physicians. *GMS J Med Educ.* 2021;38(4):Doc77.
 DOI: 10.3205/zma001473, URN: urn:nbn:de:0183-zma0014734

Artikel online frei zugänglich unter

<https://www.egms.de/en/journals/zma/2021-38/zma001473.shtml>

Eingereicht: 27.04.2020

Überarbeitet: 05.12.2020

Angenommen: 25.01.2021

Veröffentlicht: 15.04.2021

Copyright

©2021 Schmidt et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.