

# Effect of data self-collection as an activating teaching method in a statistical software course in medical biometry – a pilot study

## Abstract

**Background:** Biostatistics is an integral part of the studies of human medicine. Students learn the basics of analyzing and interpreting study results. It is important to demonstrate the subject's relevance by means of appropriate measures to maximize learning success. We investigated whether an active involvement of students in the process of data collection may improve test performance and motivation among medical students.

**Methods:** We conducted a pilot study comparing active involvement of students ( $n_1=45$ ) in the process of data collection and standard education ( $n_2=26$ ). All students of this pilot study participated in an observational study assessing their preferences regarding sweets or salty munchies, and students of the experimental group subsequently used this data set during the exercises throughout the semester. Primary and secondary endpoints were examination success and motivation respectively.

**Results:** Superiority of the activating teaching method could not be demonstrated (intervention: 109.0 points (SD 8.8), control: 113.8 points (SD 6.5)). The course ratings were superior in the intervention group (median grade 1 vs. median grade 2 in the control group), although this was not a significant improvement ( $p=0.487$ ).

**Conclusions:** Biostatistics education should incorporate approaches contributing to a better understanding of learning contents. Possible reasons why this pilot study failed to prove superiority of the intervention were a lack of sample size as well as the good grades in the control group. The presented teaching concept has to be evaluated by means of a larger sample enabling more valid conclusions. Furthermore, the considered research question in the experimental group may be changed to a more relevant one for medical practice.

**Keywords:** Activation, biostatistics, software, SPSS

Benjamin Mayer<sup>1</sup>  
Ulrike Braisch<sup>1</sup>  
Marianne Meule<sup>1</sup>  
Andreas Allgoewer<sup>1</sup>  
Silvia Richter<sup>1</sup>  
Rainer Muche<sup>1</sup>

1 Ulm University, Institute for  
Epidemiology and Medical  
Biometry, Ulm, Germany

## Introduction

Biostatistics is an integral part of the studies of human medicine. In Germany, the subject shows up in the curriculum in combination with epidemiology and medical informatics. This so-called cross-sectional subject Q1 is highly important for prospective physicians, since it introduces the basics of planning, conducting, analyzing, interpreting, and reporting studies in medical research. Biostatistics is especially geared to teach the statistical-mathematical principles of standard approaches of data analysis. Teaching epidemiology is primarily focused on essential aspects of observational studies, which are largely applied to investigate the development and dissemination of diseases. Ultimately, the interest of medical informatics is on the appropriate use of information technology methods (e.g. databases or statistical soft-

ware) in order to efficiently manage and analyze medical data sets.

The overriding educational objective of lectures in biostatistics for students of human medicine is to familiarize them with the basic knowledge and terminology of medical statistics. This is essential for an autonomous preparation of scientific work (e.g. dissertation or research article). Moreover, students should be able to understand the content of published articles in order to evaluate their importance appropriately. This is an important skill for the subsequent professional life, no matter in which field of medical research [1]. Of course, physicians who are primarily involved in research projects have to cope with published research articles more often, but also practical physicians should be able to profoundly assess the significance of research findings.

Despite the undeniable importance of biostatistics in practice, it seems that the subject does not enjoy great

popularity among students compared to other, more clinical subjects. The main reasons for that may be a general unpopularity of mathematical disciplines and an insufficient connection to practical relevance. Therefore, it is all the more important to incorporate descriptive examples in the lectures in order to maximize learning success. Ideally, activating teaching methods should be additionally applied. It has been demonstrated that this didactic approach could improve learning success [2], [3]. Also, from the motivational model of Keller and Kopp (*ARCS-attention, relevance, confidence, satisfaction*) it could be derived that successful didactical approaches necessarily have to include specific motivational components [4].

A first step towards more practical relevance has been realized more than ten years ago at our institute. A PC-based seminar in biostatistics has been designed and implemented. Since then a maximum of 76 students per semester (25% of the full cohort) can work on practical exercises in the field of medical statistics [5]. The analyses base on real data from a finalized study in pediatrics. This observational study investigated the prevalence of type 2 diabetes mellitus in obese children and adolescence [6]. By means of this dataset students work on exercises in descriptive statistics, confidence intervals, event time analysis, correlation and regression analysis, and statistical tests. Initially, SAS Analyst was used for the analysis, but in the meantime we switched to the statistical software SPSS. This software is not only used for the exercises, but also for processing the short examinations at the end of each seminar. Students have to generate the results using SPSS and afterwards all solutions are recorded in Microsoft Access input screens. Correction and marking are subsequently done automatically by means of respective SAS programs. The results of all short examinations are cumulated and replace the common exam at the end of the semester.

This initial approach of a PC-based seminar in biostatistics improved the practical relevance [7]. However, due to limited capacity in the PC-labs it is still not possible for the majority of the students to attend the PC-seminar. They have to attend the regular seminar where the exercises and short examinations are handled using pocket calculators. This is of course not appropriate and uncondusive regarding the goal of arousing interest for the subject. The increased practical relevance achieved by implementing the PC-seminar could be further enhanced if the students were actively involved in the process of data collection. This would enable students to subsequently work on exercises which base on their own data. Thus, self-assessment of research data is in the sense of integrating activating teaching methods which are evidently able to increase learning success [2], [8], [9]. Consequently, we developed our idea of the study concept presented in the following.

In the course of diverse university events (e.g. anniversary of Ulm University, Pupils' University, adult training courses) we presented an example study in order to demonstrate the principles of biostatistics to a broad and

non-specialist audience. Interested visitors were able to participate in this observational study which captured primarily the participants' preference for sweets (in German so called "Naschkatzen") or salty munchies ("Nagetiere"). In addition, few basic variables like age, gender, body height and body weights were collected. The analysis of this observational data primarily focused an evaluation of possible differences between the two groups regarding the participants' body mass index (BMI) [10]. The acronym NaNa was used for this example study because of the afore mentioned German labels of both study groups.

Consequently, the NaNa concept provided a promising opportunity regarding the aim of raising practical relevance in teaching biostatistics. On the one hand, students would be involved actively in the data collection process. Thus, basic principles of data collection as well as the associated problems can be discussed while implementing a respective data base. On the other hand, motivation may be increased by the fact that students will work on data which they have collected on their own. This may have positive effects on learning success. Altogether, it seemed possible in this way to emphasize the relevance of biostatistics. To evaluate the effect of this interactive didactic method we conducted a prospective, controlled, two-arm, single center, pilot study with observational data. Specifically, the NaNa-based teaching concept was compared to the current standard of education in the statistical software course in biostatistics. Our research hypothesis was that the NaNa concept is superior with respect to learning success and motivation.

The article is structured as follows: first, the applied didactic approaches (intervention and control) are described along with the collected study variables. Afterwards, the results of both research hypotheses, improvement of learning success and motivation, are presented. The article concludes with a discussion of the takeaways regarding the effect of the didactic intervention as well as the potential of the NaNa concept to be established as a new standard for the statistical software course.

## Material and methods

### Study design, participants and didactic intervention

The study followed a mono-centric, two-arm, controlled design with prospective observational data. There was no formal sample size calculation because of the pilot-character of this study, although the sample size could have calculated based on information about the primary outcome (examination success) in the control group. Instead, all students of human medicine at Ulm University who attended the PC-based seminar on biostatistics during the winter term 2016/17 participated in our study ( $N=71$ ). The students belonged to one of four seminar groups (maximum space for 24 and 14 students, respectively) which were instructed by three lecturers. Nearly all students ( $N_{dataset}=70$ ) attended data collection in the

course of an unscheduled date before the official launch of the seminar. The two larger seminar groups ( $n_1=45$ ) received the NaNa concept as the leading didactic intervention and worked with the self-assessed participants data during the semester. The two smaller groups ( $n_2=26$ ) served as a control intervention and worked with real observational data of a finalized study in pediatrics [6]. Hence, the control group received the standard didactic intervention which has been used for more than 10 years in the respective statistics software course. Students of both interventional groups processed exactly the same exercises during the tutorials and examinations, respectively. Only the variables used in the exercises differed. All students gave written consent to participate in the study. The ethics committee of Ulm University approved the study.

## Data assessment

Data assessment was pseudonymized. Available study participant ( $N=70$ ) gave information about demographic and health-related variables. Furthermore, variables describing the consumption behavior of the offered snacks were assessed, in particular type of snack (sweets or salty munchies), frequency of snack consumption and main reason for the consumption. There was a wide choice of both offered snack groups in order to cover all preferences.

Relevant demographic variables included age, gender, body height, body weight, and origin. Health-related variables involved physical activity, blood pressure, smoking status, chronic diseases, and allergies. The collected data were used to create exercises for the tutorials and examinations, respectively. All exercises were directly guided by those of the standard didactic intervention.

## Outcomes

The primary objective for an assessment of the didactic intervention's effect was the cumulated sum of points from all single examinations during the semester and the accompanied grades, respectively. The results of the lecture's evaluation by the students was a secondary objective of the trial.

## Statistical analysis

First, a comparison of both collectives regarding demographic variables was conducted descriptively. Continuous variables were described using mean, standard deviation (SD), median, and quartiles (where appropriate). Frequencies were calculated for categorical variables and the chi-square test or Fisher's exact test were used, respectively, to compare them subsequently. The unpaired t-test was used to analyze the primary endpoint. From earlier semesters it could be expected that the cumulative sum of points will be normally distributed. The Mann-Whitney-U test was used for the evaluation of the secondary endpoint (acceptance) due to its ordinal scale level. Cohen's

effect size ( $d$ ) has been additionally calculated in order to express the strength of the effect sizes [11]. A  $p \leq 0.05$  was considered significant, whereas all results were interpreted in an explorative manner. Only single missing values occurred in the final data set, whereas no missing values were present in variables concerning both the primary and secondary outcome. The statistical software R (version 3.2.1, <http://www.r-project.org>) was used for the analyses.

## Results

Study participants showed no difference compared to those who did not attend the PC-based seminar in biostatistics during the winter term 2016/17 with respect to the distribution of age and gender. On average, they were 23.9 years old (SD 2.9) and 43% were male (see Table 1). The majority attended the 6<sup>th</sup> or 7<sup>th</sup> semester (77%) and decided to take sweets (76%) in the course of data collection. Independent of their choice for either snack groups 80% stated "enjoyment" as the main reason for the consumption. There was no difference between the intervention and control groups regarding age and gender, but with respect to the distribution of semesters ( $p=0.0003$ ).

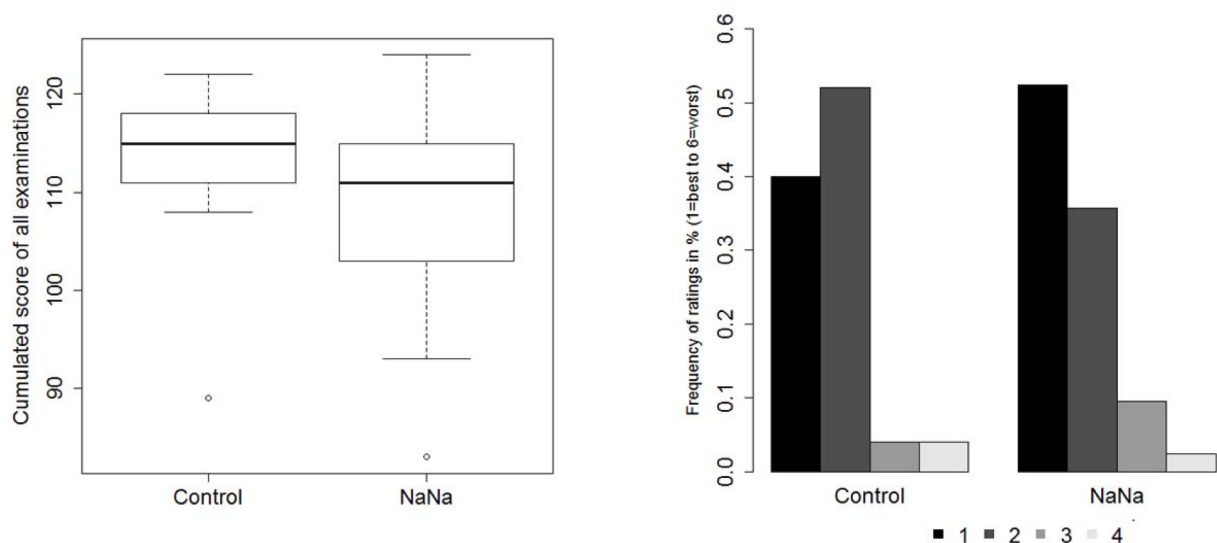
The primary research hypothesis of a superiority of the NaNa concept with respect to improved learning success and motivation could not be demonstrated by means of the collected study data. On the contrary, students of the control group were even slightly superior with a mean of 113.8 points (SD 6.5) from all examinations compared to 109.0 points (SD 8.8) on average for students of the NaNa group ( $p=0.012$ , see Figure 1). This corresponds to an effect size of  $d=0.62$  indicating a moderate effect. Of course, this was also confirmed by an evaluation of the resulting grades (1.6 (SD 0.6) in the control group vs. 2.1 (SD 0.7) in the NaNa group, i.e.  $d=0.77$ ,  $p=0.001$ ). Nevertheless, the interventional group scored the seminar slightly better (median score 1) than the control group (median score 2), whereas the scoring based on regular school grades from 1=best to 6=worst. This difference was not statistically significant ( $d=0.16$ ,  $p=0.487$ ). The same tendency was also observed when looking at more detailed questions referring specific aspects of the lecture evaluation. The median score for "I would attend the lecture again." was 6 ("complete agreement") in the NaNa group and 5 ("agreement") in the control group. The question "I have learned a lot" was scored equally in both groups with a median of 5 ("agreement").

## Discussion

All subjects in the curriculum of the studies of human medicine which are not primarily focused on clinics, as e.g. biostatistics, have to deal with the same problems. Often missing practical relevance is mentioned. Since many of these subjects are naturally focused on funda-

**Table 1: Sample characteristics**

Variable	Full cohort (N=70)	Intervention group (n <sub>1</sub> =45)	Control group (n <sub>2</sub> =25)
Age (years) [mean (SD)]	23.9 (2.9)	23.7 (2.7)	24.2 (3.1)
Gender (male) [n (%)]	30 (43%)	12 (48%)	18 (40%)
Semester [n (%)]			
- 6th	18 (26%)	18 (40%)	0 (0%)
- 7th	36 (51%)	19 (42%)	17 (68%)
- 8th	6 (9%)	2 (4%)	4 (16%)
- 9th	10 (14%)	6 (13%)	4 (16%)
BMI (kg/m <sup>2</sup> ) [mean (SD)]	22.5 (2.8)	22.4 (2.6)	22.8 (3.3)
Waist to height ratio (cm/m) [mean (SD)]	0.45 (0.05)	0.46 (0.04)	0.44 (0.06)
Snack [n (%)]			
- Sweets	53 (76%)	35 (78%)	18 (72%)
- Salty munchies	17 (24%)	10 (22%)	7 (28%)
Frequency snack consumption [n (%)]			
- never	18 (26%)	12 (27%)	6 (24%)
- 1-3x per week	29 (41%)	20 (44%)	9 (36%)
- more often than 3x per week	23 (33%)	13 (29%)	10 (40%)
Physical activity (hours per week) [n (%)]			
- 0-2	12 (17%)	5 (11%)	7 (28%)
- 2-4	35 (50%)	25 (56%)	10 (40%)
- 4-6	11 (16%)	9 (20%)	2 (8%)
- more than 6	12 (17%)	6 (13%)	6 (24%)
Blood pressure (mmHg) [mean (SD)]			
- diastolic	81.9 (9.1)	83.1 (10.0)	79.8 (6.5)
- systolic	128.9 (12.7)	130.9 (14.6)	125.4 (7.4)
Chronic diseases (yes) [n (%)]	5 (7%)	3 (7%)	2 (8%)
Allergy (yes) [n (%)]	17 (24%)	13 (29%)	4 (16%)

**Figure 1: Comparison of study groups with respect to attained sum of points (left) and evaluation (grades 1-6) of the seminar (right)**

mentals (e.g. biochemistry, physics) or methodological aspects, education in these fields should comprise practical elements. These may include innovative technical devices, e.g. virtual reality glasses in cardiology, patient dummies with integrated measurement sensors in emergency medicine, or the application of didactic apps which support the students [12], [13]. Alternatively, activating teaching methods empirically lead to a better understanding of the teaching contents [2], [3]. Our NaNa learning approach explicitly addressed the attention and satisfaction aspects of Keller and Kopp's ARCS model [4]. Also the confidence aspect was considered by means of the repeated examinations (self-control) during the semester. However, the relevance aspect may not be perfectly incorporated in our study. Students in the standard educational group worked on a data example (prevalence of type 2 diabetes mellitus) of higher medical relevance compared to students of the experimental group (differences between students preferring sweets instead of salty munchies).

Lectures in the field of medical statistics should interconnect the theoretical aspects of study planning, data analysis and interpretation to practice-oriented research questions, data and instruments, respectively, to a special degree. Of course, a major goal is enabling students to critically assess the validity of research articles themselves, but also providing experience with available statistical software should be striven for. The already developed PC-seminar in biostatistics, serving as a framework for the current study, addresses these requirements. Also the medical licensure act 2004 for German physicians defined exactly those demands [5], [14]. The presented NaNa concept extended the well-established PC-seminar and included a self-collection of empirical study data with subsequent analysis in order to increase motivation among the participating students. By incorporating the students already in the data collection process it was expected that this also would have a positive effect on learning success. However, the results showed that there was no significant improvement with respect to both outcomes of the study. In fact, the control group even achieved higher cumulative scores with 113.8 points against 109.0 points in the NaNa group, which corresponds to a moderate effect size. Consequently, students of the control group on average got better grades (1.6 in the control group vs. 2.1 in the NaNa group). In light of the number of participants who have been investigated in the course of this study, it is not possible at that point to conclude that this is a systematic effect. A more comprehensive evaluation of the differences found in our pilot study would require to have larger studies providing more statistical evidence. However, based on mean cumulative score of 110 points (SD 8.9) during the past years (corresponds to an overall grade 2), a total sample size of N=160 students would be required to demonstrate a significant effect of the NaNa intervention assuming the students to improve their results to 114 points (overall grade 1.5) assuming a power of 80% and a two-sided type 1 error of 5%. This could be realized by either extend-

ing the observation on multiple semesters or by a multicentric approach. Such a study, however, has to be scheduled as a cluster-randomized study in order to prevent possible differences of the involved universities [15]. Regarding the effect of the NaNa concept on student's acceptance there was a slight tendency of the interventional group being superior. The overall rating of the seminar based on regular school grades revealed a higher median score in the NaNa group compared to the control group receiving the standard didactic approach. Likewise, students of the NaNa group answered more convincing to the question whether they would attend the seminar again.

## Limitations

It was not possible to implement our study, as initially planned, as a cluster-randomized study allocating the four seminar groups to either interventions. This was primarily due to organizational reasons (class schedule), but especially because of the limited capacity of the available PC-labs. Both seminar groups taking place on Tuesday provide 24 working places, whereas the seminar groups on Thursday and Friday have only 14 working places available. Randomization would have had the potential drawback that accidentally both smaller seminar groups were chosen to apply the NaNa concept. Thus, the theoretically available outcome variables for the interventional group would have been decreased from potentially 48 to 28 values. In light of the fact that the cumulative sum of points following the standard educational concept was constant over the last years, we thought it was rather acceptable to have only 28 values for the control group. Moreover, the comparative analyses summarized in Table 1 showed no difference between both collectives, except for the distribution of semesters. This difference is not surprising, since students of the 6<sup>th</sup> semester were not able to attend the PC-seminar on Thursday and Friday (both control groups) because of their specific class schedule.

Overall, the sample size in this pilot study is certainly falling short of deriving valid statements from the conducted analyses. Especially the results of the control group referring to the acceptance of the course are not representative relying on only 26 students. However, the control groups' results regarding learning success are indeed valid since they perfectly match with the cumulative sum of points from earlier semesters.

With respect to the small effects found in this study it is important to note that the grades in biostatistics have been on a high level in the past years. Therefore, it was clear at the outset that distinct improvements in learning success can hardly be achieved. As noted above, a distinctively larger sample would be required in order to demonstrate an improvement in learning success when applying the NaNa concept. Moreover, all study participants attended the data collection in order to get a preferable large data set. As a result of this, however,

also students of the control group have experienced activating elements.

## Conclusion

The present intervention study was not able to confirm the predefined research hypotheses assuming the activating didactic concept to be superior to the standard approach. As expected, it was only found that motivation and qualitative evaluation of the seminar showed a slight tendency of being increased in the NaNa group. However, more extensive data are required to validly assess the effect of self-assessment as an activating teaching method on learning success and acceptance of the subject biostatistics. Future studies may include additional outcome measures, e.g. motivation score tools [16] enabling a more comprehensive evaluation of the teaching concept. Taking into account all the arguments raised before and in spite of the unverifiable impact of the applied intervention on learning success, education in biostatistics should perpetuate implementing practical examples in order to give students an understanding of the subject's relevance.

## Competing interests

This study was funded by the AG Lehrforschung of the Medical Faculty, Ulm University. There are no conflicts of interest for any of the authors.

## References

- Kumar D, Angurana R. Impact of epidemiological and statistical skills training in undergraduates. *Med Educ.* 2011;45(11):1138. DOI: 10.1111/j.1365-2923.2011.04129.x
- Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R. A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Med Educ.* 2000;34(2):120-125. DOI: 10.1046/j.1365-2923.2000.00574.x
- Krummenauer F, Weiler U. Gambling kick or content motivation - what is really initialized by the introduction of software into medical biometry lessons? *GMS Z Med Ausbild.* 2005;22(3):Doc56. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000056.shtml>
- Keller JM, Kopp TW. An application of ARCS model of motivational design. In: Reigeluth CM. (Hrsg), *Instructional theories in action. Lessons illustrating selected theories and models.* Hillsdale, NJ: Erlbaum, 1987. S.289-320.
- Muche R, Babik T. Auswahl und Einbindung einer Statistiksoftware im "Lehrprojekt Biometrie" an der Universität Ulm. *GMS Med Inform Biom Epidemiol.* 2008;4(1):Doc2. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/mibe/2008-4/mibe000061.shtml>
- Wabitsch M, Hauner H, Hertrampf M, Muche R, Hay B, Mayer H, Kratzer W, Debatin KM, Heinze E. Type II diabetes mellitus and impaired glucose regulation in Caucasian children and adolescents with obesity living in Germany. *Intern J Obes Relat Metab Disord.* 2004;28(2):307-313. DOI: 10.1038/sj.ijo.0802555
- Muche R, Janz B, Einsiedler B. Ein (halb-)automatisiertes Prüfungstool für semesterbegleitende Prüfungen in einem Statistiksoftwarekurs im Medizinstudium. In: Rendtel U, Schirnbacher P, Kao O, Lesener WF, Minkenberg R (Hrsg). *Proceedings der 14. Konferenz der SAS®-Anwender in Forschung und Entwicklung (KSFE).* Aachen: Shaker Verlag; 2010. S.175-187.
- Schukajlow S, Leiss D, Pekrun R, Blum W, Mueller M, Messner R. Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educ Stud Mathemat.* 2011;79(2):215. DOI: 10.1007/s10649-011-9341-2
- Mayer B, Ring C, Muche R, Rothenbacher D, Schmidt-Straßburger U. A blended learning module in an online master study programme in oncology. *Educ Health.* 2015;28(1):1-5. DOI: 10.4103/1357-6283.161951
- Mayer B, Danner B. Von Naschkatzen und Nagetieren – Eine interaktive Einführung in die Medizinische Biometrie mit der NANA-Studie. In: Rauch G, Muche R, Vonthein R (Hrsg). *Zeig mir Biostatistik! Ideen und Material für einen guten Biometrie-Unterricht.* Berlin: Springer Verlag, 2014. S.3-14. DOI: 10.1007/978-3-642-54336-4\_1
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2. Edition. Hillsdale: Lawrence Erlbaum Associates; 1988.
- Snashall E, Hindocha S. The Use of Smartphone Applications in Medical Education. *Open Med J.* 2016;3:322-327. DOI: 10.2174/1874220301603010322
- Ventola C. Mobile Devices and Apps for Health Care Professionals: Uses and Benefits. *P T.* 2014;39(5):356-364.
- German Federal Ministry for Health. *Medical Licensure Act. Federal Law Gazette.* Berlin: German Federal Ministry for Health; 2002.
- Dreyhaupt J, Mayer B, Keis O, Oechsner W, Muche R. Cluster-randomized studies in educational research: principles and methodological aspects. *GMS J Med Educ.* 2017;34(2):Doc26. DOI: 10.3205/zma001103
- Cook DA, Castillo RM, Gas B, Artino AR Jr. Measuring achievement goal motivation, mindsets and cognitive load: validation of three instruments' scores. *Med Educ.* 2017;51(10): 1061-1074. DOI: 10.1111/medu.13405

### Corresponding author:

PD Dr. Benjamin Mayer

Ulm University, Institute for Epidemiology and Medical Biometry, Schwabstr. 13, D-89075 Ulm, Germany, Phone: +49 (0)731/502-6896, Fax: +49 (0)731/502-26902 [benjamin.mayer@uni-ulm.de](mailto:benjamin.mayer@uni-ulm.de)

### Please cite as

Mayer B, Braisch U, Meule M, Allgoewer A, Richter S, Muche R. *Effect of data self-collection as an activating teaching method in a statistical software course in medical biometry – a pilot study.* *GMS J Med Educ.* 2018;35(1):Doc9. DOI: 10.3205/zma001156, URN: <urn:nbn:de:0183-zma0011565>

### This article is freely available from

<http://www.egms.de/en/journals/zma/2018-35/zma001156.shtml>

Received: 2017-07-31

Revised: 2017-10-26

Accepted: 2017-12-04

Published: 2018-02-15

**Copyright**

©2018 Mayer 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/>.

# Effekt einer Selbsterhebung von Daten als aktivierende Lehrmethode in einem Statistiksoftwarekurs in medizinischer Biometrie – eine Pilotstudie

## Zusammenfassung

**Hintergrund:** Die Medizinische Biometrie ist ein zentraler Bestandteil des Studiums der Humanmedizin. Die Studierenden lernen die Grundlagen der Analyse und Interpretation von Studienergebnissen. Die Relevanz des Fachgebietes sollte durch die Anwendung geeigneter Lehrmaßnahmen veranschaulicht werden, um den Lernerfolg zu maximieren. Wir untersuchten, ob eine aktive Beteiligung der Studierenden im Rahmen des Datenerhebungsprozesses den Erfolg im Testat und die Motivation bei Studierenden der Humanmedizin erhöhen kann.

**Methoden:** Wir führten eine Pilotstudie durch, bei der eine aktive Beteiligung von Studierenden im Datenerhebungsprozess ( $n_1=45$ ) verglichen wurde mit einem Standardansatz der Lehre ( $n_2=26$ ). Alle Studierenden dieser Pilotstudie nahmen Teil an einer Beobachtungsstudie, in deren Rahmen die Vorliebe für Süßigkeiten oder salzige Knabberereien erhoben wurde. Die Studierenden der Interventionsgruppe arbeiteten anschließend während der Seminarübungen des gesamten Semesters mit diesem Datensatz. Die primären und sekundären Endpunkte der Studie waren der Erfolg im Testat sowie die Motivation.

**Ergebnisse:** Eine Überlegenheit der aktivierenden Lehrmethode konnte nicht gezeigt werden (Interventionsgruppe: 109.0 Punkte (SD 8.8), Kontrollgruppe: 113.8 Punkte (SD 6.5)). Die Bewertungen des Seminars waren in der Interventionsgruppe leicht überlegen (mediane Note 1 vs. mediane Note 2 in der Kontrollgruppe), jedoch stellte dies keine signifikante Verbesserung dar ( $p=0.487$ ).

**Schlussfolgerungen:** Die Lehre im Fach Medizinische Biometrie sollte didaktische Ansätze integrieren, die zu einem besseren Verständnis der Lehrinhalte beitragen können. Mögliche Gründe, weshalb eine Überlegenheit des betrachteten Lehrkonzeptes hier nicht gezeigt werden konnte, könnten eine zu geringe Fallzahl sowie die guten Bewertungen der Kontrollgruppe gewesen sein. Das vorgestellte Lehrkonzept sollte auf der Basis einer größeren Stichprobe evaluiert werden, um zu validen Schlussfolgerungen zu gelangen. Zudem sollte die zugrunde gelegte Forschungsfrage in der Interventionsgruppe geändert werden in eine für die medizinische Praxis relevantere.

**Schlüsselwörter:** Aktivierung, Medizinische Biometrie, Software, SPSS

## Einleitung

Die Medizinische Biometrie ist ein zentraler Bestandteil des Studiums der Humanmedizin. In Deutschland ist das Fach in Kombination mit den Fächern Epidemiologie und Medizinische Informatik im so genannten Querschnittsfach Q1 im Curriculum implementiert. Das Querschnittsfach hat insofern eine hohe Relevanz für zukünftige Mediziner, da es die Grundlagen der Studienplanung, -durchführung, -auswertung und -interpretation in der medizinischen Forschung einführt. Die Medizinische

Biometrie zielt dabei insbesondere auf die statistisch-methodischen Grundlagen ab, welche den Standardauswertungsmethoden zugrunde liegen. Die Lehre im Fach Epidemiologie hat ihren Fokus primär auf die zentralen Aspekte von Beobachtungsstudien ausgerichtet, in deren Rahmen überwiegend die Entwicklung und Verbreitung von Krankheiten untersucht werden. Das Interesse der Medizinischen Informatik gilt schließlich dem geeigneten Einsatz informationstechnologischer Methoden (z.B. Datenbanken und statistische Software) im Rahmen einer effizienten Verarbeitung und Auswertung medizinischer Daten.

Benjamin Mayer<sup>1</sup>  
Ulrike Braisch<sup>1</sup>  
Marianne Meule<sup>1</sup>  
Andreas Allgoewer<sup>1</sup>  
Silvia Richter<sup>1</sup>  
Rainer Muche<sup>1</sup>

<sup>1</sup> Universität Ulm, Institut für  
Epidemiologie und  
Medizinische Biometrie, Ulm,  
Deutschland



Das übergeordnete Lehrziel der Vorlesung Medizinische Biometrie für Studierende der Humanmedizin ist sie mit den Grundlagen und der Terminologie des Faches vertraut zu machen. Dies ist eine Grundvoraussetzung für die eigenständige Erstellung wissenschaftlicher Arbeiten (z.B. Dissertation oder Fachartikel). Zudem sollten die Studierenden dazu befähigt werden den Inhalt publizierter Fachartikel verstehen und deren inhaltliche Relevanz bewerten zu können. Dies stellt eine wichtige Fertigkeit ihres späteren Berufslebens dar, unabhängig davon in welchem Bereich der medizinischen Forschung sie arbeiten werden [1]. Zwar werden in der Forschung aktive Mediziner häufiger damit konfrontiert sich mit Fachartikeln auseinanderzusetzen, jedoch sollten auch praktisch tätige Mediziner in der Lage sein die Bedeutung neuer Forschungsergebnisse valide einzuschätzen.

Trotz der unbestreitbaren Bedeutung der Medizinischen Biometrie in der wissenschaftlichen Praxis, scheint es als würde dem Fach keine große Sympathie entgegengebracht werden im Vergleich zu anderen, eher klinisch orientierten Fächern. Hauptsächlich kann dies vielleicht auf eine grundsätzliche Unbeliebtheit mathematischer Fächer und einer nicht ausreichend erkennbaren praktischen Relevanz zurückgeführt werden. Deshalb ist es umso wichtiger anschauliche Beispiele in die Lehrveranstaltungen zu integrieren, um den Lernerfolg zu maximieren. Idealerweise sollten zusätzlich aktivierende Lehrmethoden eingesetzt werden, um diesen Prozess zu unterstützen. Es konnte in zahlreichen Studien belegt werden, dass dieser didaktische Ansatz den Lernerfolg verbessern kann [2], [3]. Ebenso kann aus dem Motivationsmodell nach Keller und Kopp (ARCS) abgeleitet werden, dass erfolgreiche didaktische Ansätze notwendigerweise spezifische motivierende Komponenten beinhalten müssen [4].

Ein erster Schritt in Richtung mehr Praxisnähe wurde an unserem Institut bereits vor mehr als zehn Jahren gemacht. Damals wurde ein PC-basiertes Seminar in Medizinischer Biometrie entwickelt und in die reguläre Lehre implementiert. Seither können maximal 76 Studierende pro Semester (25% der Gesamtkohorte) Übungsaufgaben in einem praxisorientierten Setting der Medizinischen Biometrie bearbeiten [5]. Die Aufgaben und Analysen basieren auf einem realen Datensatz einer bereits abgeschlossenen Studie aus dem Bereich Pädiatrie. Im Rahmen der betreffenden Beobachtungsstudie wurde die Prävalenz von Typ 2 Diabetes mellitus in adipösen Kindern und Jugendlichen untersucht [6]. Unter Verwendung dieser Daten bearbeiten die Studierenden verschiedene Übungsaufgaben in den Bereichen deskriptive Statistik, Konfidenzintervalle, Ereigniszeitanalyse, Korrelation und Regression, sowie statistische Tests. Ursprünglich wurde die Software SAS Analyst für die Auswertungen verwendet, mittlerweile wird jedoch mit der Statistiksoftware SPSS gearbeitet. Die Software wird dabei nicht nur für die Bearbeitung der Übungsaufgaben eingesetzt, sondern auch für die Kurztestate am Ende eines jeden Seminars. Die Studierenden erzeugen ihre Analyseergebnisse mit SPSS und übertagen diese dann in speziell dafür entwickelte

Microsoft Access Eingabemasken. Die Korrektur und Benotung der Kurztestate erfolgt anschließend dann automatisiert mit Hilfe entsprechender SAS-Programme. Die Ergebnisse aller Kurztestate werden kumuliert und ersetzen am Semesterende die gewöhnliche Examensnote.

Dieser initiale Ansatz eines PC-basierten Seminars in Medizinischer Biometrie konnte die praktische Relevanz verbessern [7]. Allerdings ist es für die Mehrzahl der Studierenden aus Kapazitätsgründen in den PC-Pools nicht möglich das PC-Seminar zu besuchen. Damit besuchen 75% der Gesamtkohorte der Studierenden das reguläre Seminar, in dessen Rahmen die Übungen und Testate mit Hilfe von kleineren Rechenaufgaben unter Verwendung eines Taschenrechners durchgeführt werden. Dies ist natürlich nicht angemessen und förderlich für das Ziel einer Erhöhung des Interesses für das Fachgebiet. Die durch die Implementierung des PC-Seminars erreichte Steigerung der praktischen Relevanz könnte unter Umständen weiter gesteigert werden, wenn die Studierenden bereits in den Prozess der Datenerhebung involviert wären. Dies würde es ihnen ermöglichen im Anschluss an die Datengenerierung mit den eigenen Daten zu arbeiten, was insgesamt damit dem Konzept der aktivierenden Lehrmethoden entspricht. Diese sind nachweislich in der Lage den Lernerfolg zu erhöhen [2], [8], [9]. Demzufolge wurde von uns das hier vorgestellte Studienkonzept entwickelt.

Im Rahmen verschiedener universitärer Veranstaltungen (z.B. Jubiläum der Universität, „Schüleruniversitäten“, Erwachsenen-/Seniorenweiterbildung) nutzten wir bereits eine Beispielstudie, um die Grundlagen der Medizinstatistik einem breiten, nicht-spezialisierten Publikum zu demonstrieren. Interessierten Teilnehmern wurde die Teilnahme an einer kleinen Beobachtungsstudie angeboten, deren Ziel darin bestand die Präferenz der Teilnehmer hinsichtlich Süßigkeiten („Naschkatzen“) oder salzigen Knabberereien („Nagetiere“) zu erfassen. Zusätzlich wurden wenige andere Basisvariablen wie Alter, Geschlecht, sowie Körpergröße und Körpergewicht erhoben. Die Auswertung der Beispielstudie zielte hauptsächlich auf den möglichen Unterschied im Body Mass Index (BMI) zwischen den beiden Gruppen ab [10]. Das im Folgenden verwendete Studienakronym NaNa resultierte dabei aus den zuvor genannten Gruppenbeschreibungen von Naschkatzen und Nagetieren.

Demzufolge stellte das NaNa Studienkonzept eine vielversprechende Möglichkeit dar die praktische Relevanz des Seminars in Medizinischer Biometrie zu verbessern. Einerseits konnten damit die Studierenden in den Prozess der Datenerhebung involviert werden, so dass die grundlegenden Aspekte dieses Prozesses mit den damit assoziierten Problem diskutiert werden konnten. Zum anderen erhofften wir uns eine Motivationssteigerung auf Seiten der Studierenden aufgrund der Tatsache, dass sie im Rahmen des Seminars mit den eigenen Daten arbeiten können. Dies könnte auch einen positiven Effekt auf den Lernerfolg haben. Um den Effekt dieser aktivierenden Lehrmethode evaluieren zu können, führten wir

eine Pilotstudie auf der Basis von Beobachtungsdaten durch. Im Speziellen wurde das NaNa-Lehrkonzept verglichen mit der bis dahin als Standard implementierten Lehrmethode im Softwarekurs Medizinische Biometrie. Unserer Forschungshypothese entsprechend sollte das NaNa-Konzept überlegen sein in Bezug auf Lernerfolg und Motivation.

Der Artikel ist wie folgt strukturiert: Zunächst werden die eingesetzten didaktischen Ansätze (Intervention und Kontrolle) vorgestellt und die erhobenen Studiendaten beschrieben. Anschließend erfolgt eine Evaluierung der beiden Forschungshypothesen, d.h. Verbesserung des Lernerfolgs und Erhöhung der Motivation. Der Artikel schließt mit einer kritischen Diskussion der Kernaussagen der Analysen im Hinblick auf den Effekt der didaktischen Intervention und des Potentials des NaNa-Konzeptes als Standardlehrmethode im betreffenden Statistiksoftwarekurs eingesetzt zu werden.

## Material und Methoden

### Studiendesign, Studienteilnehmer und didaktische Intervention

Die Studie folgte einem monozentrischen, zweiarmigen, kontrollierten Design mit prospektiven Beobachtungsdaten. Aufgrund des Pilotcharakters der Studie erfolgte in der Planungsphase keine formale Abschätzung der Stichprobengröße, obwohl eine Berechnung auf Basis entsprechender Vordaten zum Lernerfolg (Testatergebnisse) in der Kontrollgruppe hätte durchgeführt werden können. Stattdessen wurden alle Studierenden der Humanmedizin der Universität Ulm, die am PC-Seminar der Medizinischen Biometrie im Wintersemester 2016/17 teilgenommen haben, in die Studie eingeschlossen ( $N=71$ ). Die Studierenden besuchten eine von vier möglichen Seminargruppen (maximal 24 bzw. 14 Teilnehmer pro Kurs), welche von 3 verschiedenen Dozenten unterrichtet wurden. Fast alle Studierenden ( $N_{dataset}=70$ ) nahmen an der Datenerhebung teil, welche im Rahmen eines außerplanmäßigen Termins vor offiziellem Start des Seminars stattfand. Die zwei größeren Seminargruppen ( $n_1=45$ ) wurden mit dem interventionellen NaNa-Lehrkonzept unterrichtet und arbeiteten demnach während des Kurses mit den eigens erhobenen Daten aller Studienteilnehmer. Die zwei kleineren Gruppen ( $n_2=26$ ) fungierten als Kontrollgruppe und wurden mit dem etablierten Standardlehrkonzept unterrichtet, so dass deren Seminar- und Testaufgaben auf realen Beobachtungsdaten einer pädiatrischen Studie basierten [6]. Die Studierenden beider Vergleichsgruppen bearbeiteten exakt dieselben Aufgaben während der Übungen und Testate. Einzige Ausnahme waren die Variablen, welche für die Berechnungen verwendet wurden. Alle Studierenden gaben ihr schriftliches Einverständnis zur Teilnahme an der Studie. Zudem lag ein positives Votum der Ethikkommission Ulm vor.

### Datenerhebung

Die Datenerhebung erfolgte pseudonymisiert. Die an der Datenerhebung anwesenden Studienteilnehmer ( $N=70$ ) machten Angaben zu demographischen und gesundheitsbezogenen Variablen. Zudem wurde das Konsumverhalten der angebotenen Snacks erhoben, insbesondere die Art des Snacks (süß oder salzig), die Konsumhäufigkeit und der Hauptgrund für den Konsum. In beiden Snackgruppen gab es ein breites Angebot, um möglichst alle Präferenzen abzudecken.

Die demographischen Variablen umfassten Alter, Geschlecht, Körpergröße, Körpergewicht und Herkunftsland. Die gesundheitsbezogenen Merkmale waren körperliche Aktivität, Blutdruck, Raucherstatus, chronische Erkrankungen und Allergien. Die erhobenen Merkmale wurden genutzt um die Übungs- und Testaufgaben entsprechend zu konzipieren. Alle Aufgaben orientierten sich dabei strikt an denen der Kontrollgruppe.

### Endpunkte

Die primäre Zielgröße der Studie zur Erfassung des Effektes der didaktischen Intervention war die kumulierte Punktzahl aus allen sechs Testaten während des Semesters bzw. der damit einhergehenden Note. Die Ergebnisse der Evaluierung des Seminars durch die Studierenden war die sekundäre Zielgröße unserer Studie.

### Statistische Analyse

Ein Vergleich der beiden Studienkollektive in Bezug auf die demographischen Variablen erfolgte zunächst deskriptiv. Für kontinuierliche Merkmale wurden Mittelwert, Standardabweichung (SD), Median und Quartile berechnet. Für kategoriale Merkmale wurden Häufigkeiten berechnet, die anschließend mit Hilfe des Chi-Quadrat-Tests bzw. des exakten Test nach Fisher weitergehend verglichen wurden. Ein ungepaarter t-Test wurde für die Analyse des primären Endpunktes verwendet. Aus früheren Semestern konnte erwartet werden, dass die kumulierten Punkte einer Normalverteilung folgen. Der Mann-Whitney-U-Test wurde eingesetzt zum Vergleich der Kollektive im Hinblick auf die ordinalskalierte, sekundäre Zielgröße (Akzeptanz bzw. Motivation). Zusätzlich wurde jeweils Cohen's  $d$  berechnet, um die Stärke des Effektes auszudrücken [11]. Ein  $p \leq 0.05$  wurde als statistisch signifikant betrachtet, wobei alle Ergebnisse in explorativem Sinne interpretiert werden. Nur vereinzelt zeigten sich fehlende Werte in den Daten, insbesondere für die primären und sekundären Endpunkte war die Datenerhebung vollständig. Die Statistiksoftware R (Version 3.2.1, <http://www.r-project.org>) wurde für die Analysen verwendet.

### Ergebnisse

Die Studienteilnehmer zeigten keine Unterschiede hinsichtlich der Alters- und Geschlechterverteilung im Ver-

Tabelle 1: Kollektivbeschreibung

Merkmal	Gesamt (N=70)	NaNa-Gruppe (n <sub>1</sub> =45)	Kontrollgruppe (n <sub>2</sub> =25)
Alter (Jahre) [MW (SD)]	23.9 (2.9)	23.7 (2.7)	24.2 (3.1)
Geschlecht (männlich) [n (%)]	30 (43%)	12 (48%)	18 (40%)
Semester [n (%)]			
- 6.	18 (26%)	18 (40%)	0 (0%)
- 7.	36 (51%)	19 (42%)	17 (68%)
- 8.	6 (9%)	2 (4%)	4 (16%)
- 9.	10 (14%)	6 (13%)	4 (16%)
BMI (kg/m <sup>2</sup> ) [MW (SD)]	22.5 (2.8)	22.4 (2.6)	22.8 (3.3)
Waist to height ratio (cm/m) [MW (SD)]	0.45 (0.05)	0.46 (0.04)	0.44 (0.06)
Snack [n (%)]			
- Süßigkeiten	53 (76%)	35 (78%)	18 (72%)
- Salzige Knabbereien	17 (24%)	10 (22%)	7 (28%)
Häufigkeit Snackkonsum [n (%)]			
- Nie	18 (26%)	12 (27%)	6 (24%)
- 1-3x pro Woche	29 (41%)	20 (44%)	9 (36%)
- Mehr als 3x pro Woche	23 (33%)	13 (29%)	10 (40%)
Körperliche Aktivität (h pro Woche) [n (%)]			
- 0-2	12 (17%)	5 (11%)	7 (28%)
- 2-4	35 (50%)	25 (56%)	10 (40%)
- 4-6	11 (16%)	9 (20%)	2 (8%)
- Mehr als 6	12 (17%)	6 (13%)	6 (24%)
Blutdruck (mmHg) [MW (SD)]			
- Diastolisch	81.9 (9.1)	83.1 (10.0)	79.8 (6.5)
- Systolisch	128.9 (12.7)	130.9 (14.6)	125.4 (7.4)
Chronische Erkrankungen (ja) [n (%)]	5 (7%)	3 (7%)	2 (8%)
Allergien (ja) [n (%)]	17 (24%)	13 (29%)	4 (16%)

MW=Mittelwert; SD=Standardabweichung; BMI=Body Mass Index; h=Stunden

gleich zu jenen Studierenden, die während des Wintersemesters 2016/17 nicht am PC-Seminar teilgenommen haben. Sie waren durchschnittlich 23.9 Jahre alt (SD 2.9) und 43% waren männlich (siehe Tabelle 1). Die Mehrzahl besuchte das 6. oder 7. Fachsemester (77%) und entschied sich für einen süßen Snack (76%) während der Datenerhebung. Unabhängig vom gewählten Snack gaben 80% der Studienteilnehmer „Genuss“ als Hauptgrund des Konsums an. Hinsichtlich Alter und Geschlecht konnten keine Unterschiede der beiden Vergleichskollektive nachgewiesen werden, jedoch im Hinblick auf die Verteilung der Fachsemester ( $p=0.0003$ ).

Die primären Forschungshypothesen einer Überlegenheit des NaNa Konzeptes in Bezug auf eine Erhöhung des Lernerfolgs und der Motivation konnte auf Basis der erhobenen Daten nicht gezeigt werden. Vielmehr zeigten die Studierenden der Kontrollgruppe sogar etwas bessere Ergebnisse bezüglich Lernerfolg mit durchschnittlich

113.8 Punkten (SD 6.5) über alle Testate hinweg im Vergleich zu 109.0 Punkten (SD 8.8) für Studierende der NaNa-Gruppe ( $p=0.012$ , siehe Abbildung 1). Dieses Resultat zeigt sich in einer Effektgröße von  $d=0.62$ , welche einen moderaten Effekt darstellt. Als eine logische Konsequenz waren damit auch die Noten der Kontrollgruppe im Mittel besser (1.6 (SD 0.6)) als in der NaNa-Gruppe (2.1 (SD 0.7)), d.h.  $d=0.77$  und  $p=0.001$ . Jedoch bewerteten die Studierenden der Interventionsgruppe den Kurs etwas besser (Median 1) als die Studierenden der Kontrollgruppe (Median 2), wobei die Bewertung sich an den regulären Schulnoten 1=beste Bewertung bis 6=schlechteste Bewertung orientierte. Dieser Unterschied war jedoch nicht signifikant ( $p=0.487$ ). Dieselbe Tendenz konnte beobachtet werden bei spezifischeren Fragen der Kursbewertung. Der mediane Score für die Frage „Ich würde die Lehrveranstaltung wieder besuchen“ lag bei 6 („vollständige Zustimmung“) in der NaNa-Gruppe und bei

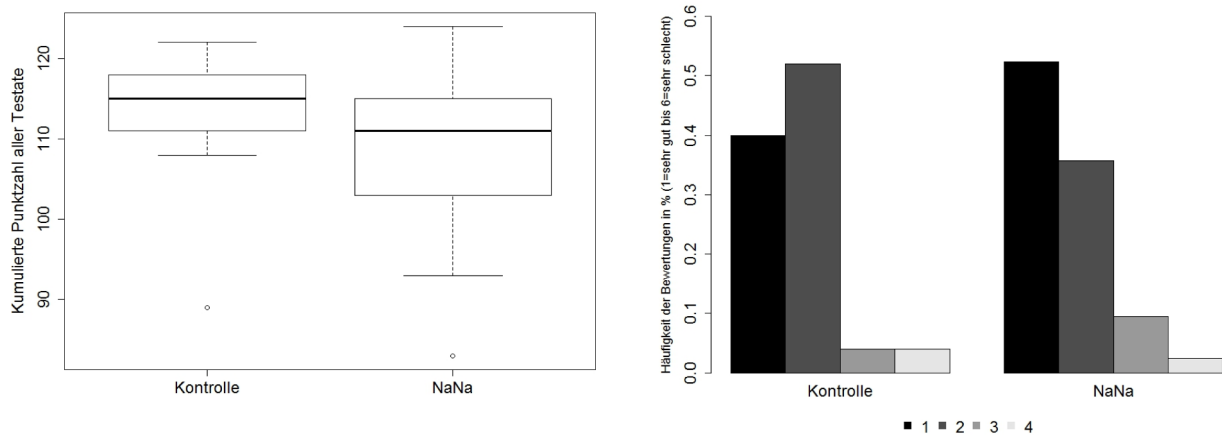


Abbildung 1: Vergleich der Studiengruppen hinsichtlich der erreichten Gesamtpunktzahl in den Testaten (links) und der Evaluation des Seminars (Noten 1-6, rechts)

5 („Zustimmung“) in der Kontrollgruppe. Die Frage „Ich habe viel gelernt“ wurde in beiden Gruppen im Median mit 5 („Zustimmung“) bewertet.

## Diskussion

Nahezu alle Fächer des Curriculums im Studiengang Humanmedizin, die wie die Medizinische Biometrie nicht primär einen klinischen Fokus haben, sehen sich mehr oder weniger mit denselben Problemen konfrontiert. Oftmals wird fehlende Praxisnähe als Grund genannt. Da viele dieser Fächer natürlicherweise auf die Vermittlung von Grundlagen (z.B. Biochemie, Physik) oder methodische Aspekte ausgerichtet sind, sollte die Lehre in diesen Fächern praxisnahe Elemente beinhalten. Dies können beispielsweise technische Geräte zur Wissensvermittlung wie Virtual Reality Brillen in der Kardiologie, Patientendummies mit integrierten Messsensoren in der Notfallmedizin oder entsprechende Lern-Apps sein, welche die Studierenden unterstützen [12], [13]. Zudem ist der positive Effekt aktivierender Lehrmethoden auf das Verständnis der Lehrinhalte nachweislich belegt [2], [3]. Unser NaNa Lehrkonzept adressiert explizit die *attention* und *satisfaction* Aspekte von Keller und Kopp's ARCS-Modell [4]. Außerdem wurde auch der *confidence* Aspekt berücksichtigt durch die wiederholten Prüfungen (Selbstkontrolle) während des Semesters. Allerdings wurde der relevanz Aspekt in unserer Studie nicht optimal umgesetzt. Die Studierenden der Kontrollgruppe arbeiteten mit realen Daten (Prävalenz von Typ 2 Diabetes mellitus) und für die Praxis relevanteren Fragestellungen im Vergleich zu den Studierenden der Interventionsgruppe (Unterschiede zwischen Studierenden die Süßigkeiten präferieren anstelle salziger Knabbereien).

Lehrveranstaltungen in Medizinischer Biometrie sollten die theoretischen Aspekte von Studienplanung, Datenanalyse und Interpretation in Bezug setzen zu praxisorientierten Fragestellungen, Daten und Methoden. Natürlich ist dabei ein wichtiger Anspruch die Studierenden dazu zu befähigen selbstständig die Validität von Forschungsergebnissen und -artikeln bewerten zu können. Vor allem

in der heutigen Zeit sollte den Studierenden jedoch auch ermöglicht werden erste Erfahrungen mit Statistiksoftware zu sammeln, da dies für die Praxis eine große Relevanz hat. Das bereits entwickelte PC-Seminar in Medizinischer Biometrie, das die Rahmenbedingung für die vorgestellte Lehrstudie festlegte, adressiert diese Anforderungen. Auch im Rahmen der Änderung der Approbationsordnung für Ärzte im Jahre 2004 wurden diese Anforderungen explizit gestellt [5], [14]. Das vorgestellte NaNa-Lehrkonzept erweitert das etablierte PC-Seminar und schließt eine Selbsterhebung empirischer Daten inklusive deren anschließender Auswertung mit ein, um die Motivation der Studierenden möglichst zu steigern. Durch die Involvierung der Studierenden bereits im Prozess der Datenerhebung wurde erhofft, dass dies einen positiven Effekt auf den Lernerfolg hat. Allerdings zeigten die Ergebnisse der Studie, dass keine signifikante Verbesserung hinsichtlich beider Zielgrößen durch die Lehrintervention erzielt werden konnte. Tatsächlich erreichten die Studierenden der Kontrollgruppe sogar eine höhere kumulierte Punktzahl von durchschnittlich 113.8 Punkten gegenüber 109.0 Punkten in der NaNa-Gruppe, was einen moderaten Effekt darstellte. Dem entsprechend erzielten die Studierenden der Kontrollgruppe im Mittel auch bessere Noten (1.6 in der Kontrollgruppe gegenüber 2.1 in der NaNa-Gruppe). Im Hinblick auf die Anzahl der in die Studie eingeschlossenen Teilnehmer ist es an dieser Stelle jedoch nicht möglich von einem systematischen Effekt zu sprechen. Eine umfangreichere Evaluation der in unserer Pilotstudie gefundenen Resultate wäre notwendig auf der Basis größerer Studienkollektive, die mehr statistische Evidenz mit sich bringen. Allerdings wäre unter Betrachtung der mittleren kumulierten Punktzahl von 110 Punkten (SD 8.9) in den letzten zehn Jahren (entspricht einem Notendurchschnitt von 2.0) eine Gesamtfallzahl von N=160 Studierenden notwendig, um eine signifikante Verbesserung der NaNa-Gruppe auf durchschnittlich 114 Punkte (Notendurchschnitt 1.5) zu erreichen (unter der Annahme eines zweiseitigen Typ 1 Fehlers von 5% und einer Power von 80%). Dies könnte durch eine Ausweitung der Studiendauer über mehrere Semester hinweg erreicht werden, oder aber über einen multizentrischen Ansatz, der jedoch

als Cluster-randomisierte Studie durchgeführt werden müsste, um für mögliche Differenzen zwischen den teilnehmenden Universitäten korrigieren zu können [15]. In Bezug auf den Effekt des NaNa-Konzeptes auf die Akzeptanz der Studierenden gegenüber unserem Fach konnte die leichte Tendenz einer Überlegenheit der Interventionsgruppe festgestellt werden, welche jedoch statistisch nicht signifikant war. Die Gesamtbewertung des Seminars auf Basis von Schulnoten führte zu einem höheren Median in der NaNa-Gruppe im Vergleich zur Kontrollgruppe. Ebenso antworteten die Studierenden der NaNa-Gruppe überzeugender auf Fragen, ob sie das Seminar noch einmal besuchen würden.

## Limitationen

Entgegen unserer ursprünglichen Intention war es nicht möglich Cluster-randomisierte Studie zu implementieren, in deren Rahmen die vier Seminargruppen zufällig auf die zwei Lehrkonzepte randomisiert worden wären. Die ist unter anderem auf organisatorische Gründe zurückzuführen (Stundenplan der betreffenden Fachsemester), vor allem jedoch auch aufgrund der limitierten Kapazität in den zur Verfügung stehenden PC-Pools. Beide Seminare am Dienstag bieten Platz für maximal 24 Studierende, während für die Seminare am Donnerstag und Freitag nur 14 Arbeitsplätze verfügbar sind. Eine Clusterrandomisierung hätte den potentiellen Nachteil gehabt, dass per Zufall die beiden kleineren Seminargruppen mit dem NaNa-Konzept unterrichtet worden wären. Damit hätten sich die theoretisch möglichen 48 Beobachtungen für die Zielgrößen in der Interventionsgruppe auf nur 28 Beobachtungen reduziert. Aufgrund der Tatsache, dass die kumulierte Punktzahl unter dem Standardlehransatz in den letzten Jahren konstant war, hielten wir es für angemessener die kleinere Fallzahl in der Kontrollgruppe zu haben. Zudem zeigten die Analysen in Tabelle 1 keinen Unterschied zwischen beiden Vergleichsgruppen mit Ausnahme der Fachsemesterverteilung. Dieser Unterschied wiederum ist nicht überraschend, da die Studierenden des 6. Semesters aufgrund ihres Stundenplans zur Dienstags (Interventionsgruppe) besuchen können. Insgesamt ist die Fallzahl dieser Pilotstudie sicherlich zu niedrig, um valide Aussagen aus den Analysen abzuleiten. Insbesondere die Ergebnisse der Kontrollgruppe in Bezug auf die Akzeptanz des Seminars sind mit nur 26 Studierenden nicht repräsentativ. Allerdings bestätigten die Ergebnisse der Kontrollgruppe im Hinblick auf den Lernerfolg die Beobachtungen der letzten Jahre.

Es ist bezüglich der in dieser Studie gefundenen, kleinen Effekte wichtig zu erwähnen, dass die Noten im Fach Medizinische Biometrie in den letzten Jahren per se auf einem hohen Level waren. Demzufolge war es von Beginn an klar, dass bedeutsame Verbesserungen des Lernerfolgs kaum erreicht werden können. Wie bereits erwähnt wäre eine sehr viel größere Fallzahl notwendig gewesen, um eine Verbesserung des Lernerfolgs durch das NaNa-Konzept erreichen zu können. Letztlich nahmen auch alle Studierenden an der Datenerhebung teil, um einen

möglichst umfangreichen Datensatz für die NaNa-Gruppe erzeugen zu können. In Folge dessen machten aber auch alle Studierenden die Erfahrung mit aktivierenden Elementen in der Lehre.

## Schlussfolgerung

Die vorgestellte Studie war nicht in der Lage die vorab definierten Forschungshypothesen einer Überlegenheit von Lehrmethoden mit aktivierenden Elementen gegenüber dem Standardlehriansatz zu bestätigen. Erwartungsgemäß konnte nur eine leichte Tendenz in Bezug auf eine verbesserte Motivation der Studierenden in der NaNa-Gruppe festgestellt werden, nicht jedoch im Hinblick auf eine Verbesserung des Lern- und Prüfungserfolgs. Allerdings sind umfangreichere Daten notwendig, um valide Aussagen über die Effektivität der Datenselbsterhebung als aktivierendes Lehrelement in Medizinischer Biometrie treffen zu können. Zukünftige Studien sollten zusätzlich weitere Zielgrößen mit einschließen, wie zum Beispiel andere Scores zur Erfassung der Motivation [16], um eine noch umfassendere Evaluierung des Lehrkonzeptes vornehmen zu können. Werden all die zuvor genannten Argumente zusammengenommen und in Anbetracht des nicht eindeutig bewertbaren Effektes der untersuchten Lehrintervention auf den Lernerfolg, sollte die Lehre in Medizinischer Biometrie weiterhin versuchen eine möglichst hohe Praxisrelevanz durch die Einbindung von anschaulichen Beispielen zu erzeugen, um den Studierenden die Relevanz des Fachgebiets zu vermitteln.

## Interessenkonflikt

Diese Studie wurde finanziell unterstützt durch die AG Lehrforschung der Medizinischen Fakultät der Universität Ulm. Alle Autoren des Artikels erklären, dass keine Interessenkonflikte bestehen.

## Literatur

1. Kumar D, Angurana R. Impact of epidemiological and statistical skills training in undergraduates. *Med Educ.* 2011;45(11):1138. DOI: 10.1111/j.1365-2923.2011.04129.x
2. Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R. A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Med Educ.* 2000;34(2):120-125. DOI: 10.1046/j.1365-2923.2000.00574.x
3. Krummenauer F, Weiler U. Gambling kick or content motivation - what is really initialized by the introduction of software into medical biometry lessons? *GMS Z Med Ausbild.* 2005;22(3):Doc56. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000056.shtml>
4. Keller JM, Kopp TW. An application of ARCS model of motivational design. In: Reigeluth CM. (Hrsg), *Instructional theories in action. Lessons illustrating selected theories and models.* Hillsdale, NJ: Erlbaum, 1987. S.289-320.

5. Muche R, Babik T. Auswahl und Einbindung einer Statistiksoftware im "Lehrprojekt Biometrie" an der Universität Ulm. *GMS Med Inform Biom Epidemiol*. 2008;4(1):Doc2. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/mibe/2008-4/mibe000061.shtml>
6. Wabitsch M, Hauner H, Hertrampf M, Muche R, Hay B, Mayer H, Kratzer W, Debatin KM, Heinze E. Type II diabetes mellitus and impaired glucose regulation in Caucasian children and adolescents with obesity living in Germany. *Intern J Obes Relat Metab Disord*. 2004;28(2):307-313. DOI: 10.1038/sj.ijo.0802555
7. Muche R, Janz B, Einsiedler B. Ein (halb-)automatisiertes Prüfungstool für semesterbegleitende Prüfungen in einem Statistiksoftwarekurs im Medizinstudium. In: Rendtel U, Schirmbacher P, Kao O, Lesener WF, Minkenberg R (Hrsg). *Proceedings der 14. Konferenz der SAS®-Anwender in Forschung und Entwicklung (KSFE)*. Aachen: Shaker Verlag; 2010. S.175-187.
8. Schukajlow S, Leiss D, Pekrun R, Blum W, Mueller M, Messner R. Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educ Stud Mathemat*. 2011;79(2):215. DOI: 10.1007/s10649-011-9341-2
9. Mayer B, Ring C, Muche R, Rothenbacher D, Schmidt-Straßburger U. A blended learning module in an online master study programme in oncology. *Educ Health*. 2015;28(1):1-5. DOI: 10.4103/1357-6283.161951
10. Mayer B, Danner B. Von Naschkatzen und Nagetieren – Eine interaktive Einführung in die Medizinische Biometrie mit der NANA-Studie. In: Rauch G, Muche R, Vonthein R (Hrsg). *Zeig mir Biostatistik! Ideen und Material für einen guten Biometrie-Unterricht*. Berlin: Springer Verlag, 2014. S.3-14. DOI: 10.1007/978-3-642-54336-4\_1
11. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2. Edition. Hillsdale: Lawrence Erlbaum Associates; 1988.
12. Snashall E, Hindocha S. The Use of Smartphone Applications in Medical Education. *Open Med J*. 2016;3:322-327. DOI: 10.2174/1874220301603010322
13. Ventola C. Mobile Devices and Apps for Health Care Professionals: Uses and Benefits. *P T*. 2014;39(5):356-364.
14. German Federal Ministry for Health. *Medical Licensure Act*. Federal Law Gazette. Berlin: German Federal Ministry for Health; 2002.
15. Dreyhaupt J, Mayer B, Keis O, Oechsner W, Muche R. Cluster-randomized studies in educational research: principles and methodological aspects. *GMS J Med Educ*. 2017;34(2):Doc26. DOI: 10.3205/zma001103
16. Cook DA, Castillo RM, Gas B, Artino AR Jr. Measuring achievement goal motivation, mindsets and cognitive load: validation of three instruments' scores. *Med Educ*. 2017;51(10): 1061-1074. DOI: 10.1111/medu.13405

**Korrespondenzadresse:**

PD Dr. Benjamin Mayer  
 Universität Ulm, Institut für Epidemiologie und  
 Medizinische Biometrie, Schwabstr. 13, 89075 Ulm,  
 Deutschland, Tel.: +49 (0)731/502-6896, Fax: +49  
 (0)731/502-26902  
[benjamin.mayer@uni-ulm.de](mailto:benjamin.mayer@uni-ulm.de)

**Bitte zitieren als**

*Mayer B, Braisch U, Meule M, Allgoewer A, Richter S, Muche R. Effect of data self-collection as an activating teaching method in a statistical software course in medical biometry – a pilot study. GMS J Med Educ. 2018;35(1):Doc9. DOI: 10.3205/zma001156, URN: urn:nbn:de:0183-zma0011565*

**Artikel online frei zugänglich unter**

<http://www.egms.de/en/journals/zma/2018-35/zma001156.shtml>

**Eingereicht:** 31.07.2017

**Überarbeitet:** 26.10.2017

**Angenommen:** 04.12.2017

**Veröffentlicht:** 15.02.2018

**Copyright**

©2018 Mayer 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/>.