LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
MEDIZINISCHE FAKULTÄT
INSTITUT FÜR MEDIZINISCHE INFORMATIONSVERARBEITUNG, BIOMETRIE UND EPIDEMIOLOGIE - IBE
MASTERSTUDIENGANG EPIDEMIOLOGIE

Master's Thesis

Implementation and Assessment of a Research Methods and Biostatistics Course for General Practitioners

Prepared by: **Zekeriya Aktürk**

Master's in Epidemiology

Ludwig-Maximilians-University, Institute for Medical Information Processing, Biometry, and Epidemiology

Marchioninistr. 15, 81377 Munich

Director: Prof. Dr. rer. nat. Ulrich Mansmann

Supervisor

Prof. Dr. Marco Roos

Department of General Practice, Augsburg University, Augsburg

Reviewer

Brigitte Strahwald, MSc Mmel

Pettenkofer School of Public Health, Ludwig-Maximilians-University, Munich

Submission: 31.07.2025

Disputation: 16.09.2025

Dedication

To all those who chose the long route of morality over the short route of convenience—

Who are vocal when silence would be safer,

Who remain erect when kneeling would spare them.

To the brave hearts that purchased liberty through exile, persecution, or demise—

To the nameless names lost to the ocean,

To the innocent children killed under debris of war,

To the hands never held and futures never realized.

This work is for you—

Not as comfort,

But as a silent defiance,

A memorial in words,

A promise that your courage will never be forgotten.

May every page be a whisper in the face of oppression,

And every thought be a seed of solidarity.

Acknowledgments

To begin a Master's degree at 55 years old was no decision made lightly, nor an easy feat to achieve. It was a thing done not only with the mind, but with the whole heart, knowing full well that the way would be a burden not only for me, but for the one who shares my life.

To my amazing wife Semra—my soulmate—I am indebted for more than allotted space allows. In choosing this path, I borrowed from our shared time, leaned on your forbearance, and gained courage from your quiet company. I thank you for sitting through study sessions, hearing me practice presentations, and holding ground as I followed the ideals of learning again.

Learning, for me, is simply ambition. It is a religious assignment—a faith issue. Lifelong learning is not a personal slogan but a divine decree written in the rhythm of my life, an assignment I am willing to follow until my dying moment.

This Master's is a hallmark in my second academic track. The first began in 1999, as a young physician starting university life with open eyes and open hands—teaching, research, patient care. I had leadership roles, like the honor of being department head and even dean. But that track was terminated; abruptly and brutally.

In 2016, following the alleged coup in Turkey, I was ousted from my position at one of the country's most renowned health universities. I was among countless victims of a government that sowed fear and shattered lives. I was imprisoned. And yet, the conclusion to the story did not arrive then.

My greatest gratitude is to all who assisted me to flee from that storm—to breathe, to begin anew. I would like to thank most specifically Dr. Britta Zangen, a fierce women's rights activist and compassionate human being from Düsseldorf, who flung open her heart as well as her home. Her letter of invitation was my freedom ticket to Germany during the pandemic.

In my second beginning, Prof. Antonius Schneider opened his arms and invited me into his department. His backing laid the foundation stone of the second half of my life. I am grateful to Prof. Marco Roos also, whose very nature is curiosity and courage. His words;

> *"Sich auf den Weg machen", "Neugierig entdecken", "Manches wagen",*
> *"Aus der Reihe tanzen", and "Mit dem Leben wachsen"*

provided not just inspiration, but rich soil in which to start afresh. He is one of the most visionary academic leaders in general practice.

My heartfelt thanks are also due to my "Kardeş" Dr. Raphael Kunisch, who stood by me during dark times with unwavering dedication. And to Dr. Marc Jamoulle, the visionary Quaternary Prevention innovator and tireless activist who insisted on my freedom when I was gagged in prison—gracias. You maintained the testimony that solidarity doesn't recognize borders.

To my friend, benefactor, and current employer, Dr. med. Sentayehu Assefa—thank you for giving me a space in which to serve once more, to teach, and to grow. In deed and spirit, you continue the legacy of your ancestor, the illustrious King Al-Najashi, who provided refuge to the oppressed. Your clinic at Gersthofen is no longer a place of work—it is an abode of dignity and trust.

I also thank Prof. Ulrich Mansmann and his team, who welcomed me into their department and offered this enriching course. I learned much and would have learned much more if time allowed.

To those who shaped me many years before this tale began—Ayşe Akyüz, Fatma and Yahya Aktürk, Remzi Özkan, Barbara Hartkopf, and Deborah Luxton Ulmer—your encouragement ignited the early coals of my passion and commitment. That fire continues to blaze.

This dissertation is the result of numerous travels—personal, academic, and spiritual. My desire is that it not only enriches the academic life of general practice in Germany, but acts as a compass for others: for the numerous immigrant academics, physicians, and pursuers of truth who were displaced by injustice but will not be defeated by it.

May it be an encouragement on their road, as others have been on mine.

Zekeriya Aktürk

Dinkelscherben, July 2025

Summary

Background

Biostatistics and research methodology are central to evidence-based medical practice but are often under-emphasized in GP training. In order to restore the balance, the "4ALL" course was created—an online, modular, SPSS-linked training course aimed specifically at GPs and medical students with little prior experience of research methods.

Objectives

- To develop and present the 4ALL blended-learning course.
- To evaluate its effectiveness in improving research-related knowledge.
- To assess participant interest, completion rates, and self-reported learning.

Methods

Design: A pre-post pilot test with no control.

Participants: 85 medical doctors and medical students showed interest, 72 enrolled, 22 completed the full course.

Instruments: Pre/post knowledge test, Constructivist Online Learning Environment Survey (COLLES) questionnaire.

Analysis: Descriptive statistics, Wilcoxon tests, polynomial regression, and linear mixed-effects models were performed. A directed acyclic graph (DAG) informed the modelling approach.

Results

Knowledge Gains: Statistically significant increase from pretest (Mean = 60.3 ± 18.7) to posttest (Mean = 79.1 ± 15.8; $p = 0.001$).

COLLES Scores: Overall positive (approx. Mean = 3.0), indicating a positive learning process.

Subgroup Analysis: No significant differences by age, sex, occupation, or experience. Self-rated competence was a predictor of pretest performance.

Regression Results: Polynomial regression revealed a curvilinear trend for posttest score and age. Mixed-effects modeling confirmed statistically significant learning gains and determined interactions with self-rated competence.

Conclusion

The 4ALL course increased the statistical and research awareness of dedicated participants. Although the completion rate was moderate (26%), the positive feedback and strong performance of the graduates underscore the educational value of the course. A DAG-based, parsimonious model supported the application of a linear mixed-effects analysis. Logistical challenges suggest the need for modular reorganization to optimize participant scheduling and improve learning outcomes. The 4ALL course is a valuable educational resource for medical education and general practice in biostatistics and research. It should be further developed and delivered in a modular format. A final test should be administered after three months to evaluate long-term knowledge retention. Given its proven efficacy and acceptability, 4ALL holds promise as a scalable model for research education in primary care.

Strengths and Limitations

Strengths: Pre-post design, mixed-methods analysis, COLLES questionnaire, DAG-informed modeling. Limitations: Small sample size, relatively low completion rate, no control group, limited qualitative feedback.

Keywords: Education, Medical; Biostatistics; General Practice; Physicians, Primary Care; Online Learning; Linear Models; Germany

Content

List of Tables

List of Figures

List of Appendices

Introduction

Biostatistics and research methods are essential in medical education for advancing medical research and high-quality patient care. Medical students and general practitioners (GPs), in particular, require structured training to meet the demands of evidence-based medicine in these areas. Despite the availability of numerous courses in clinical training, there is a persistent gap in methodological education, especially on the practical application of study design, statistical analysis, and interpretation of scientific literature.

A thorough understanding of biostatistical methods is needed for evaluating the quality and accuracy of research findings so that physicians can make informed, evidence-based clinical choices (Schimek, 2025). Moreover, proficiency in research methodology enables medical professionals to design, conduct, and analyze studies independently, hence leading to medical improvements and healthcare quality (Arbeitsgruppe Biostatistik und Medizinische Biometrie, 2025). In addition, biostatistical proficiency is also necessary in the analysis of epidemiologic information, relevant to the identification of health trends and implementation of effective public health and prevention methods (Institut für Epidemiologie, Biostatistik und Prävention, 2025a).

Despite the necessity of such skills, many GPs are unable to receive formal biostatistics and research methodology training. While clinical training is well established, there are often large deficits in statistical and methodological training, limiting physicians' involvement with research and evidence-based practice. Existing programs, such as the biometry compact courses at Universitätsmedizin Frankfurt, aim to bridge this gap through the improvement of research competence in general practice and forging academic communities (Institut für Biostatistik und Mathematische Modellierung, 2025). Similarly, international curricula reflect the same issue; for example, only 31 out of 140 U.S. medical schools offer epidemiology and biostatistics as an independent course (AAMC, 2018). In Germany, the situation is likely comparable, highlighting the need for specialized biostatistics training programs for medical professionals (Arias et al., 2017; Mai et al., 2020). Places such as the University of Zurich have responded by including biostatistics lectures and courses in their medical curriculum (Institut für Epidemiologie, Biostatistik und Prävention, 2025b).

To address this educational gap, the 4ALL course was developed as part of an internship project in MSc Epidemiology to provide a structured, self-paced, and practice-oriented learning experience. Initially, the course was named “AFORMED - Allgemeinärzte in der Forschung: Praxisorientierte Methoden und Datenanalyse”. However, during the course progress and after

receiving feedback, the name was modified to “4ALL - Forschungsmethodik für Allgemeinmedizinerinnen”.

The course includes two fully developed modules, covering research methodology, statistical fundamentals, and SPSS applications. All slides, scripts, and video recordings for SPSS demonstrations have been created during the internship work of the master candidate, and the course was uploaded to the Moodle platform of the Institute of General Practice at the University of Augsburg (https://iam-augsburg.de/).

The course videos are available at the author's YouTube channel (Module 1: https://www.youtube.com/playlist?list=PLa9XLqw8q_5ofjlZtz7qqJkfnV0-hKIkE, Module 2: https://www.youtube.com/playlist?list=PLa9XLqw8q_5p6u5e_g02Al11ryCm0jisd).

This master’s thesis evaluated the effectiveness and applicability of the 4ALL course in a small sample of participants. Using a pretest-posttest design, the study measured knowledge improvement and collected structured participant feedback on course usability. The results will be used to refine and optimize the course for broader implementation. Moreover, the findings will contribute to ongoing efforts to integrate biostatistics training into medical curricula, ensuring that GPs and medical students acquire the methodological competencies needed for independent research and evidence-based clinical decision-making.

Objectives

1. Assess the change in participants’ knowledge levels before and after completing the 4ALL course (primary outcome).
2. Evaluate participant satisfaction and feedback.
3. Identify areas for course improvement based on qualitative and quantitative feedback.

Methodology

Study Design

This study uses a quasi-experimental pretest-posttest design with a single intervention group.

Study Population and Sample Size

Target participants: The course targeted general practitioners, medical students, and early-career researchers interested in research methodology. A course flyer (Appendix 1) was

developed and promoted using different means (e.g., e-mail lists of the institute, social media, learning halls of the Augsburg University Medical Faculty).

Sample size: Approximately 30 participants were aimed as a convenience sample. After promoting the course during April 2025, 85 people showed interest. Of these, 72 registered themselves to the Moodle platform, 65 did some of the course content, such as the pretest, and 22 completed the whole course, including the posttest (Figure 1).

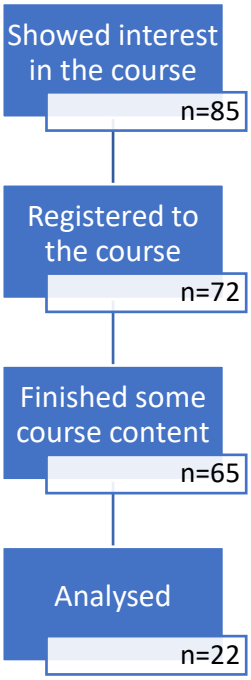


Figure 1: Participant flow diagram

Intervention

The 4ALL Course was delivered via the Moodle platform of the Augsburg University Institute of General Practice at https://iam-augsburg.de/course/view.php?id=17. The course includes two modules covering research methodology, statistical analysis, and SPSS applications. Participants had self-paced access to course materials, quizzes, and practical demonstrations. The course structure with specific objectives of the different modules is given in Appendix 2.

Data Collection

Pretest and Posttest: Designed to measure participants' baseline and post-course knowledge. The test included multiple-choice questions (MCQs) on research methods and biostatistics (Appendix 3).

Participant Feedback Survey: Participant satisfaction with the online learning environment was assessed using the Constructivist Online Learning Environment Survey (COLLES). The COLLES questionnaire (Baker, 2007) assesses the following dimensions using a five-point Likert scale (1-worse, 5-best):

- Course clarity and difficulty level
- Practical applicability
- Preferred learning format (videos, quizzes, case studies)
- Areas for improvement

For the ease of analysis, the different subdomains of the COLLES questionnaire were added together to give a general mean score.

Self-rated competence was queried in a scale from 1 (best) to 5 (worst).

Data Analysis

Quantitative data was analyzed using the R software (R version 2024.12.1). The tidyverse, gt, gtsummary, lme4, broom.mixed, car, and dagitty R packages were used. The full R script is included in Appendix 5.

Descriptive Statistics: Descriptive statistics were calculated for background and demographic variables. For continuous variables (e.g., age, test scores), mean and standard deviation (SD) were reported. Categorical variables were summarized as frequencies and percentages.

Knowledge Improvement: Pretest and posttest scores were compared using the Wilcoxon signed-rank test, due to the small sample size and potential deviations from normality.

Subgroup Analysis: Participants were grouped by key demographic and experience-related variables (e.g., age group, sex, occupation, job experience, self-rated competence, and research experience). Wilcoxon rank-sum tests were used to compare pretest and posttest scores across independent subgroups. Results were tabulated using gt and gtsummary packages in R for elegant formatting.

Polynomial Regression: Second-degree polynomial regression models were constructed separately for pretest and posttest scores to examine curvilinear associations with age. Model fit was evaluated using R^2, adjusted R^2, and p-values.

Linear Mixed-Effects Modeling (LMM): To account for the within-subject design (pretest and posttest measurements per participant), a linear mixed-effects model was applied using the lme4 R-package. The model included Time (Pre vs. Post), Age, Sex, Occupation, Self-Rated Competence, and Research Experience as fixed effects, with interaction terms between Time and each variable. A random intercept for Participant ID accounted for intra-individual variability. The model was informed by a Directed Acyclic Graph (DAG) to ensure parsimony and avoid overadjustment (Figure 4).

Qualitative Feedback Analysis: Manual and thematic analyses were conducted on open-ended feedback on course dropout and completion reasons. Response counts were performed to identify dominant themes, such as time pressures, mismatched expectations, and course difficulty. The qualitative synthesis assumed an inductive approach in accordance with procedures described by Schutt (Schutt, 2018).

Use of AI Assistance: During the development of this thesis, the AI-based tool ChatGPT (OpenAI, 2024) was used to assist with R programming code and English language polishing in parts of the text. All outputs were critically assessed, edited, and confirmed by the author.

Expected Outcomes

The main outcome of the research was the difference in knowledge level, assessed on the basis of pretest and posttest scores. Secondary outcomes were participant satisfaction, assessed through the COLLES questionnaire, and qualitative feedback from open-ended questions. These outcomes were assessed to determine course effectiveness and identify areas for possible improvement.

Ethical Considerations

All respondents gave their informed consent. Confidentiality and anonymity of response were guaranteed in the study. The study protocol was approved by the LMU ethics committee (Appendix 4).

Timeline

The timeline of the study, including the preparatory internship phase is given in Table 1.

Table 1: Study timeline

Phase	Task	Duration
December 2024-March 2025	Internship: Preparation of Module 1 and 2 content of the course	16 weeks
April 2025	Participant recruitment & pretest administration	4 weeks
May – June 2025	Course implementation & posttest administration	8 weeks
July 2025	Data analysis & thesis writing	4 weeks

Results

Of the 62 participants who took the pretest, data for 22 participants were available for complete analysis. The median (min.-max.) age of the analyzed (38.5, 22-61) and not analyzed (35.5, 26-61) participants was not significantly different (Wilcoxon W=394, p=0.238). However, the proportion of females in the not-analyzed group was significantly higher compared to the analyzed participants (Table 2).

Table 2: Comparison of sex between analyzed and not-analyzed participants

	Analyzed	Not-analyzed	Chi-square, p
Female	12 (54.5%)	20 (66.7%)	4.071, p=0.044
Male	10 (45.5%)	10 (33.3%)	

Several participants who registered for the course but did not complete it cited time constraints as the primary reason for their withdrawal. Statements such as “Keine Zeit,” “zeitlich nicht

möglich," "zu beschäftigt," and "im Urlaub" highlight how scheduling conflicts and general busyness were the most common barriers to participation. In a few cases, participants mentioned that the course did not meet their expectations or that they had misunderstood the course description. One participant expressed a lack of confidence in their ability to understand the course content, while another simply stated a lack of interest in participating. Notably, at least one individual indicated continued interest and asked to be informed about future course offerings.

Descriptive Findings

The participants were mostly young adults with an age span from 22 to 61 years (Table 3).

Table 3: Demographic characteristics of the participants

Characteristic	**N = 22**[1]
Age	40.1 ± 10.7
Sex	
Female	10 (45%)
Male	12 (55%)
Occupation	
Medical Doctor	14 (64%)
Resident Medical Doctor	5 (23%)
Medical student	3 (14%)
Job experience (years)	11.9 ± 10.9
Experience in research	
Yes	12 (60%)
No	10 (40%)
[1] Mean ± SD; n (%)	

The most common sources to hear about the course were the Moodle platform and friends (Table 4).

Table 4: Sources of information about the course.

Characteristic	**N = 22**[1]
Heard via Moodle platform	22 (100%)
Heard from friends	15 (68%)
Heard via IAM email/newsletter	5 (23%)
Heard via social media	2 (9.1%)
Heard via other sources	3 (14%)
[1] n (%)	

In the 22 analyzed participants, the greatest motivation for participating was to improve their statistical expertise and research skills (85%), followed by a desire for research (75%) and to gain a deeper understanding of clinical studies and guidelines (55%) (Table 5).

Table 5: Motivations to join the course

Characteristic	**N = 22**[1]
Improve knowledge of statistics and research	17 (85%)
General interest in research	15 (75%)
Better interpretation of studies/guidelines	11 (55%)
Prepare for scientific activity	7 (35%)
Support for projects or thesis	6 (30%)
[1] n (%)	

Self-rated competence in statistics and research was most commonly rated as 3 out of 5, reported by 40% of participants. Overall, 80% rated themselves between 2 and 4, indicating a moderate level of confidence (Table 6).

Table 6: Self-rated competence in statistics and research

Characteristic	**N = 22**[1]
Self-rated competence*	
2	4 (20%)
3	8 (40%)
4	6 (30%)
5	2 (10%)
[1] **n (%)**	

*1=best, 5=worst

The number of publications varied widely, reflecting a heterogeneous group regarding research experience. While 20% had no publications, one reported 85 publications (Table 7).

Table 7: Distribution of the total number of publications

Characteristic	N = 22[1]
Number of publications	
0	2 (20%)
1	2 (20%)
2	1 (10%)
4	2 (20%)
10	1 (10%)
20	1 (10%)
85	1 (10%)
[1] **n (%)**	

The mean scores for the seven COLLES sub-modules ranged from 2.87 to 3.21, indicating a generally moderate to moderately positive perception of the learning environment.

The COLLES scores were fairly consistent across sub-modules, with only minor variations. The highest average was observed in COLLES Module 2.3 (Advanced Statistical Principles) (3.21 ± 0.86), while the lowest was in Module 2.1 (Hypothesis Testing; Categorical Outcome) (2.87 ± 0.73). Overall, the similarity of means and standard deviations suggests a uniform participant experience across the different aspects of the course modules (Table 8).

Table 8: Distribution of the COLLES scores in the sub-modules

Characteristic	N = 22[1]
COLLES Module 1.1	3.12 ± 0.84
COLLES Module 1.2	3.03 ± 0.94
COLLES Module 1.3	2.94 ± 1.01
COLLES Module 1.4	2.92 ± 0.81
COLLES Module 2.1	2.87 ± 0.73
COLLES Module 2.2	2.96 ± 0.84
COLLES Module 2.3	3.21 ± 0.86
[1] **Mean ± SD**	

Course Expectations

The majority of participants expected to improve their statistical competencies, especially in biostatistics, followed by a better understanding of research methods and scientific literature interpretation. Many had expressed concrete goals such as conducting their own analyses or writing academic papers. There was also a strong interest in applying evidence-based practices in general practice and gaining confidence in evaluating clinical guidelines and studies (Table 9).

Table 9: Tallying of the expectations from the course

Category	N	Example Statements
Improving or refreshing knowledge in biostatistics	22	"Improve my knowledge of biostatistics," "Better understand statistics"
Understanding or applying research methodology	19	"Deepen my knowledge of research methods," "Conduct my own study"
Improving the interpretation and evaluation of studies	16	"Better interpret studies," "Critically evaluate results"
Conducting or understanding scientific research independently	12	"Perform my own analyses," "Work scientifically," "Prepare a doctoral thesis"
Learning to use SPSS or other software	6	"Use SPSS effectively," "Work with statistical software"
General interest / knowledge gain / refresher	6	"Increase knowledge," "Update knowledge," "Refresh statistics skills"
Making evidence-based decisions in daily clinical practice	5	"More confidence in evaluating guidelines," "Assess strength of evidence"
Participating in research projects or networks	2	"Join a research network," "Collaborate on projects"
Better understanding of pharmaceutical studies/information	1	"Better assess pharmaceutical representations"

Univariate Analyses

Comparing the Pretest score among the different subgroups of participants, no statistically significant difference was found. Overall, the levels of knowledge at the baseline (as indicated by pretest scores) were consistent across all subgroups being studied, showing a relatively homogeneous starting point for the course participants, regardless of age, gender, professional status, or work experience. Only, the pretest scores were significantly higher in the better self-rated competence group, compared to the worse self-rated group (Table 10).

Table 10: Univariate comparison of the pretest scores

Characteristic	Subgroup	Pretest Score	p-value
	Group	Mean ± SD	
Age (years)	≤40	61.1 ± 17.9	0.764
	>40	58.8 ± 21.2	
Sex	Female	59.1 ± 17.9	0.872
	Male	61.2 ± 20.2	
Occupation	Doctor	61.4 ± 19.2	0.570
	Other	58.4 ± 19.0	
Self-rated competence	<4**	66.2 ± 12.1	0.025
	≥4	49.3 ± 23.3	
Job experience	≤10	59.2 ± 18.2	0.381
	>10	61.5 ± 20.3	
Research experience	No	58.6 ± 19.5	0.910
	Yes	59.9 ± 19.4	

*Mann-Whitney U Test **1=excellent, 5=bad

Although none of the posttest differences were statistically significant, certain trends suggest higher posttest performance among male participants and doctors. These findings may warrant further exploration with a larger sample (Table 11).

Table 11: Univariate comparison of the posttest scores

Characteristic	Subgroup	Posttest Score	p-value
	Group	Mean ± SD	
Age (years)	≤40	81.2 ± 15.7	0.365
	>40	75.5 ± 16.3	
Sex	Female	75.8 ± 16.9	0.254
	Male	81.8 ± 15.0	
Occupation	Doctor	80.6 ± 13.07	0.664
	Other	76.5 ± 19.6	
Self-rated competence	<4	83.3 ± 7.3	0.238
	≥4	70.0 ± 22.5	
Job experience	≤10	77.6 ± 15.2	0.539
	>10	80.3 ± 16.8	
Research experience	No	75.8 ± 18.4	0.521
	Yes	79.4 ± 15.2	

*Mann-Whitney U Test, **1=excellent, 5=bad

A scatter graph was drawn to analyze the relationship between test scores and age. Both the posttest scores and pretest scores appear to follow a non-linear trend in relation to age. A rise in both scores from younger participants up to around 35–40 years was observed, while there was

a gradual decline in both scores beyond that age. Across nearly all ages, posttest scores are consistently higher than pretest scores, which confirms the effectiveness of the course. The gap between pre- and posttest appears wider among participants aged 25–35, indicating greater learning gains in this group. The highest scores for both tests are observed in participants around their late 30s to early 40s, potentially reflecting an optimal balance of academic background and professional maturity. Beyond age 45, scores decline (Figure 2).

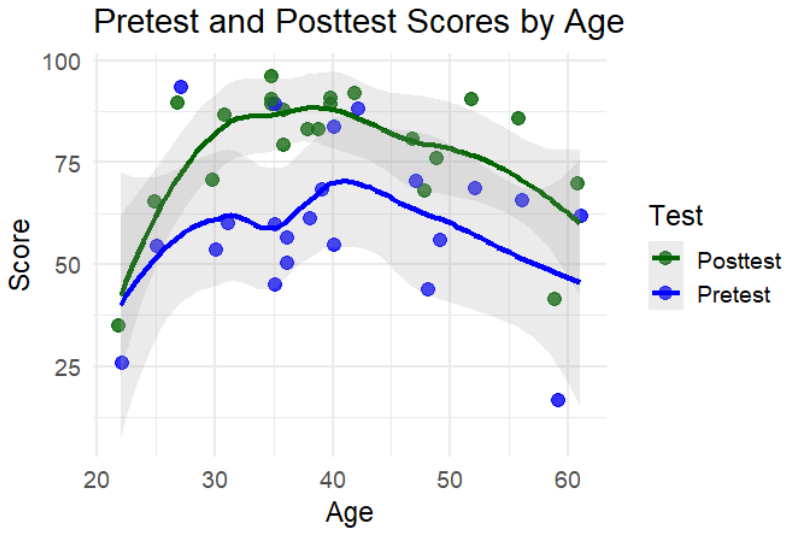


Figure 2: Scatter graph showing pretest and posttest scores by age

A polynomial regression was fit to further investigate the relationship between age and the test scores. Both predictors (age-linear and age-quadratic) are marginally significant, suggesting a non-linear relationship between age and pretest scores. While this curvilinear pattern is visually supported and backed by the quadratic term in the regression, the results are not statistically significant, likely due to the small sample size (Table 12).

Table 12: Age-pretest polynomial regression

Characteristic	Beta	95% CI	p-value
(Intercept)	-33	-147, 80	0.5
Age	4.9	-0.69, 11	0.082
Age²	-0.06	-0.13, 0.01	0.072
R^2	0.166		
Adjusted R^2	0.078		
Sigma	18.0		
Statistic	1.88		
p-value	0.2		
Df	2		
Log-likelihood	-93.2		
AIC	194		
BIC	199		
Deviance	6,154		
Residual df	19		
No. Obs.	22		
CI = Confidence Interval			

A significant quadratic relationship was found between age and posttest scores ($p < 0.001$). Scores increased with age up to a certain point and then declined, as indicated by the positive linear term ($\beta = 7.6$) and the negative quadratic term ($\beta = –0.09$). This model showed a strong fit ($R^2 = 0.515$), suggesting that age meaningfully shaped post-course test performance (Table 13).

Table 13: Age-posttest polynomial regression

Characteristic	Beta	95% CI	p-value
(Intercept)	-69	-142, 3.9	0.062
Age	7.6	4.0, 11	<0.001
Age²	-0.09	-0.13, -0.05	<0.001
R^2	0.515		
Adjusted R^2	0.464		
Sigma	11.6		
Statistic	10.1		
p-value	<0.001		
df	2		
Log-likelihood	-83.4		
AIC	175		
BIC	179		
Deviance	2,535		
Residual df	19		
No. Obs.	22		
CI = Confidence Interval			

Main Outcome Analysis

Participants showed a significant improvement in their test scores following the course. The mean posttest score was 79.1 ± 15.8, compared to a mean pretest score of 60.3 ± 18.7. This difference was statistically significant (p = 0.001, Wilcoxon rank sum test) (Table 14, Figure 3).

Table 14: Pretest vs. posttest comparison

Characteristic	Posttest N = 22[1]	Pretest N = 22[1]	p-value[2]
Test Score	79.1 ± 15.8	60.3 ± 18.7	0.001

[1] Mean ± SD

[2] Wilcoxon rank sum test

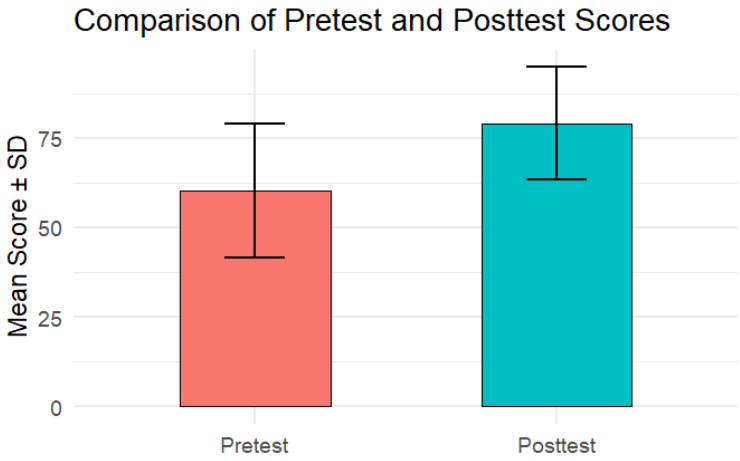


Figure 3: Comparison of the pretest and posttest scores

Model Selection Guided by a Directed Acyclic Graph (DAG)

To identify an appropriate and statistically stable model for analyzing changes in knowledge scores, we first constructed a Directed Acyclic Graph (DAG) to visualize the conceptual relationships among key variables (Figure 4). The DAG was developed to avoid overadjustment and clarify the minimal sufficient adjustment set, considering the relatively small sample size

(N = 22 participants, each with pre- and posttest scores). The model code for the DAG is given in Appendix 6.

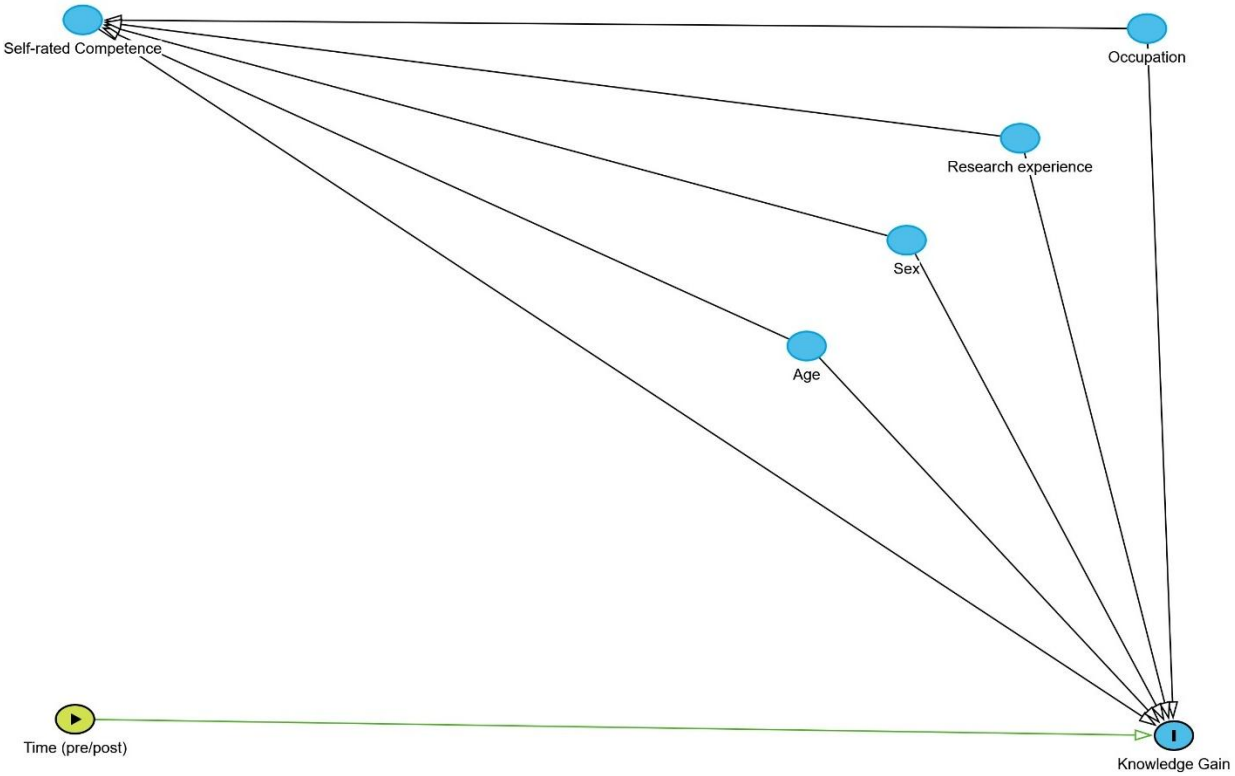


Figure 4: Directed acyclic graph (DAG)

In this theoretical model, Age, Sex, and Occupation were predicted to influence Self-Rated Competence and Knowledge Gain directly or indirectly. Self-Rated Competence was considered a substitute for prior perceived ability and perhaps a mediator of demographic effects. Time (Pre/Post) was the principal intervention variable and a direct cause of knowledge gain. From this DAG, we found that Job Experience could be omitted from the model due to likely collinearity with Age. We also considered prior research experience as an alternative background variable. While not a confounder of the Time–Knowledge Gain, it may influence self-reported skill and learning attainment, and was therefore investigated as a potential predictor or effect modifier within the model. We included age, sex, prior research experience, and self-perceived ability in the model to modify effects and improve model fit, not as confounders (since the DAG indicated no adjustment to estimate the time effect was necessary) but to explore effect modification and improve model fit. These covariates were selected on theoretical grounds and their potential to explain heterogeneity in knowledge acquisition.

We therefore fit the following parsimonious linear mixed-effects model using the lme4 package in R:

*model_lmm <- lmer(Score ~ Time * (Age + Sex + Occupation + SelfRatedGroup + ResearchExperience) + (1 | ParticipantID), data = data_long)*

Mixed Effects Model Analysis

A linear mixed-effect model was fitted to assess the effect of time (pre/post), age, sex, profession, self-rated ability, and prior research experience on participants' knowledge scores. The model included a random intercept for participant ID to adjust for repeated measures.

The overall time effect did not achieve statistical significance (Estimate = 8.24, p = 0.49), nor were there significant interaction effects. Surprisingly, there was a significant difference in the case of self-rated competence: self-raters who rated themselves less competent before taking the course (self-rating ≥ 4) received significantly lower actual test scores (Estimate = –20.16, p = 0.03) compared to those who rated themselves more competent. No important effects were found for age, gender, job, or prior research experience (Table 15). While the unadjusted pretest-posttest score comparison was statistically significantly better, the time effect in the adjusted linear mixed-effects model was not statistically significant—most likely due to model complexity (overfitting) and sample size.

Table 15: Linear Mixed Effects Model – Fixed Effects Summary

Predictor	Estimate	SE	t value	95% CI Lower	95% CI Upper
(Intercept)	86.77	23.30	3.72	47.07	126.48
Time: Posttest	8.24	17.00	0.48	-21.03	37.50
Age	-0.35	0.51	-0.67	-1.22	0.53
Sex: Male	-4.48	8.58	-0.52	-19.10	10.14
Occupation: Other	-3.42	11.22	-0.30	-22.54	15.69
Self-rated competence ≥4	-20.16	9.15	-2.20	-35.74	-4.57
Research experience: Yes	-2.66	9.76	-0.27	-19.29	13.98
Time × Age	0.04	0.37	0.11	-0.60	0.69
Time × Sex: Male	8.64	6.26	1.38	-2.14	19.42
Time × Occupation: Other	-0.95	8.18	-0.12	-15.04	13.14
Time × Self-rated competence ≥4	7.40	6.67	1.11	-4.08	18.89
Time × Research experience: Yes	2.33	7.12	0.33	-9.93	14.59

SE: Standard error

The random effects component of the model indicated substantial variability between participants. The estimated variance of the random intercept (ParticipantID) was 234.6,

corresponding to a standard deviation of 15.3, suggesting moderate between-subject differences in baseline knowledge levels. The residual variance (within-subject or unexplained variance) was estimated at 84.9 (SD = 9.2) (Table 16). These results justify the inclusion of a random intercept to account for repeated measures and individual heterogeneity in test performance.

Table 16: Linear Mixed Effects Model – Random Effects Summary

Random Effect	Group	Variance	Std. Dev.
Intercept	ParticipantID	234.6	15.3
Residual (error)	—	84.9	9.2

The intra-class correlation (ICC) was calculated as;

$$\frac{Variance\ random\ intercept}{Variance\ random\ intercept + Residual\ variance}$$

= 234.5/(234.6+84.9)

= 0.7339 (73.4%)

This means, roughly 73% of the total variance in test scores can be attributed to differences between participants.

Discussion

Key Findings

This pilot evaluation of the 4ALL course demonstrated that participants who completed the program showed a significant improvement in their knowledge scores, with mean posttest results substantially higher than pretest scores. While baseline knowledge levels were relatively homogeneous across age, sex, professional role, and experience groups, self-perceived competence was positively related to actual performance. The existence of a significant overall learning gain was confirmed by the pre–post comparison analysis.

While the linear mixed-effects model did not find a statistically significant impact of time, it allowed the examination of participant characteristics and their potential influence on learning outcomes, revealing a significant positive correlation between self-perceived competence and performance. However, the course completion rate was modest, suggesting challenges in sustaining long-term participation. Despite this, participant satisfaction (as reflected in

COLLES scores) was generally positive, pointing to the course's value for learning and potential for future refinement and scaling.

General Discussion of the Findings

Biostatistics and research methodology are fundamental to medical professionals since they not only provide a basis for critically reading scientific publications but also for carrying out independent studies. This is particularly so for general practitioners (GPs), who operate at the interface of clinical practice and research. An in-depth understanding of biostatistical techniques makes the physician competent to assess the quality and validity of study results, which is essential for making evidence-based patient treatment decisions (Schimek, 2025).

Despite the growing importance of evidence-based medicine, there is still a knowledge gap in biostatistics and research training, particularly among GPs and medical students. While clinical training opportunities are abundant, structured courses focusing on methodological competencies are scarce. This gap can hinder the ability of GPs to engage in research and apply statistical reasoning in daily practice. Universities and institutions such as the Institute for Biostatistics and Mathematical Modeling at UCT Frankfurt and the University of Zurich have recognized this need and have introduced specialized courses in biostatistics and epidemiology to support medical students and professionals in developing essential methodological skills.

The 4ALL course was designed to address a recognized training gap by offering a structured, practical, and accessible learning device for GPs and medical students. It blended theoretical foundations with practical SPSS use to improve participants' ability to analyze epidemiological data, critically evaluate scientific literature, and conduct independent research. This master's thesis rigorously tested the course's usability and effectiveness in a pretest–posttest design and facilitated participant feedback.

The results of this evaluation contribute to a broader understanding of how biostatistics training and research methodology can be successfully integrated into medical education. The findings inform future course enhancements and highlight best practices for teaching methodological competencies to medical professionals. Given the growing importance of research literacy for GPs, particularly with regard to evidence-based medicine and public health, the 4ALL course can potentially become an effective teaching method. The project can therefore serve as a model for creating future educational courses for general practice and other disciplines.

The analysis of subgroup differences revealed that baseline knowledge levels (scores on the pretest) tended to be similar for participants regardless of age, sex, job title, job experience level, or research experience. This homogeneity suggests a relatively even starting point for

learners. The only statistically significant difference at pretest was found in relation to self-rated competence: participants who rated themselves more competent scored significantly higher. This accords with earlier findings that self-assessment, while imperfect, can moderately reflect actual performance in adult and professional learning settings (Athanasou, 2007).

Self-rated competence was significantly associated with pretest scores, suggesting that participants had a reasonably accurate sense of their baseline knowledge. However, this association disappeared in the posttest, indicating that the course helped participants with lower self-rated competence to catch up with their peers. This pattern suggests that the course may have successfully equalized knowledge across different self-perception levels, which is a desirable educational outcome.

On the other hand, the posttest scores—albeit not significant—provided some intriguing trends. The male participants and medical doctors tended to achieve higher scores, as did those with prior research experience or more job experience. Importantly, though, the nature of the self-rated competence effect was maintained: participants who rated themselves more competent (lower scores) scored higher on the posttest as well. This alignment between self-perception and performance suggests metacognitive accuracy, rather than overconfidence. These findings are supported by available research that has shown self-assessment to be a valid measure of actual competence, especially by motivated students and professionals currently active in seeking to improve their skills (Eva et al., 2004).

As a whole, these subgroup trends (though mostly non-significant) are documented to point toward possible differences in how students benefit from the course. With a larger sample size, these patterns may be clearer and should be explored further to inform individualized learning strategies and targeted support in future implementations.

The non-linear pattern between age and test scores suggests that age should not be treated as a purely linear predictor of performance. Educational interventions might benefit from age-sensitive adaptations, particularly to better engage older participants.

An important insight from our analysis is the emergent curvilinear effect of age on posttest performance, in contrast to only a marginal trend in the pretest. The quadratic regression model for posttest scores revealed a significant inverted-U relationship, with participants in their late 30s to early 40s showing the highest gains—reflected in a strong model ($R^2 = 0.52$, $p < 0.001$). By contrast, the pretest model was only borderline significant ($R^2 = 0.17$, $p = 0.072$), indicating a weaker association between age and baseline knowledge.

This pattern aligns well with cognitive aging literature emphasizing that many cognitive abilities follow an inverted-U trajectory, peaking in mid-adulthood before gradually declining. Studies have reported that executive function, working memory, and probabilistic learning all rise through adolescence into adulthood and then diminish in older age (Berthelot et al., 2019). For example, implicit sequence learning remains stable through early adulthood and declines notably after 45–60 years (Nemeth et al., 2013).

Age not being a strong predictor at baseline but becoming a significant factor post-intervention suggests that age may influence the capacity to integrate and apply learning rather than merely reflect pre-existing knowledge. Within adult education disciplines, this supports the idea that learners benefit optimally with interventions when they are at their best cognitive window, which is often midlife (Nemeth et al., 2013). Interestingly, adults in the late 30s to early 40s tend to possess a balance of cognitive mastery, experience (cognitive reserve), and flexibility, making this age most conducive to learning complex skills.

The significant increase in test scores from pretest to posttest suggests that the 4ALL course had a positive impact on participants' knowledge acquisition. The improvement from a mean score of 60.3 ± 18.7 to 79.1 ± 15.8 ($p = 0.001$, Wilcoxon test) reflects a meaningful learning gain during the intervention period.

This result aligns with previous studies demonstrating the effectiveness of structured, interactive, and context-sensitive educational programs in improving statistical and research-related competencies among healthcare professionals (Norman & Schmidt, 1992). Specifically, it is adult learners who are best supported when learning is problem-centered, directly applied to their job roles, and actively reinforced, all of which tenets were incorporated into the design of the 4ALL course (Taylor & Hamdy, 2013).

Furthermore, the use of a non-parametric test (Wilcoxon rank-sum) adds strength to the finding, given that it adjusts for the small sample size and potential non-normal distribution of test scores. That the course yielded such a measurable effect within a short timeframe also supports its potential as a scalable and effective training intervention for early-career health professionals.

The linear mixed-effects model was chosen to appropriately account for the repeated measures design—each participant contributed both pre- and posttest scores. Such models are well-established in educational and psychological research for handling within-subject correlations and accommodating missing data or unbalanced designs (Gordon, 2019). The inclusion of a random intercept for ParticipantID effectively captured individual differences in baseline

knowledge, with between-subject variance ($\sigma = 234.6$) explaining approximately 73% of total variance, justifying its use.

Our parsimonious model, guided by a DAG approach and limited by sample size, included Time, Age, Sex, Occupation, Self-Rated Performance, and Research Experience, with interaction terms for differential effects over time. This strategy follows best practices for mixed models, which recommend including time-by-variable interactions to maintain power and validity (Schuler, 2022).

The fixed effects revealed a statistically significant negative association between self-rated competence and test performance: participants who rated themselves worse (i.e., higher scores on the 1–5 scale) performed worse on the test (Estimate = –20.16, p = 0.03). This positive alignment between self-assessment and actual performance suggests a relatively good level of metacognitive accuracy among participants — those who felt less confident indeed performed less well. This contrasts with common findings of overconfidence (as seen in the Dunning–Kruger effect (Dunning, 2011)), and instead supports studies reporting accurate self-assessment in motivated or academically engaged learners (Eva & Regehr, 2005).

Applicability of the Course

The overall interest in the 4ALL course was promising: following the promotional phase in April 2025, a total of 85 individuals expressed interest, 72 registered on the Moodle platform, and 65 engaged with the course to some extent (e.g., completed the pretest). However, only 22 participants (26%) completed the entire course, including the posttest assessment. This small rate of completion indicates the key problem with the course as it stands: despite clear demand, sustained participation and active engagement were difficult to achieve.

Several non-completers of the course provided feedback that pressures of time were the most common barrier. Reasons such as “keine Zeit,” “im Urlaub,” and “zu beschäftigt” were frequently mentioned, which means that the timing and organization of the course likely do not fit into the professional and private schedule of the target learners. Other participants also spoke about unfilled expectations, content misunderstandings, or self-efficacy deficits to master the material.

These findings suggest the need for structural modification to improve course usability and applicability. An applied answer would be to deliver the course in smaller, bite-sized portions—perhaps dividing it into weekend modules, aligning with the existing seven sub-modules. This format could make flexible participation achievable, reduce perceptions of time commitments, and enable participants to sense incremental progress.

In spite of the modest completion rate, the learning outcomes among completers were decidedly favorable. Posttest scores were significantly higher than pretest scores, and the COLLES ratings revealed overall favorable learning experience in all submodules, with average values grouping around 3.0 on a 5-point scale. These findings support the pedagogical validity and utility of the course content and format for motivated learners.

Lastly, the 4ALL course has much potential for effect among active clinicians. However, applicability and sustainability can be enhanced through modular reorganization, greater clarity of expectations, and focused follow-up with interested but inactive registrants. These changes will not only boost course completion but also enable more participants to be able to avail themselves of the desired learning outcomes.

Strengths and Limitations

This initial evaluation of the 4ALL course has several important strengths. First, the study employed a pre–post design with repeated testing for knowledge, allowing us to measure learning gains objectively. Use of standardized educational measures, such as the COLLES questionnaire, contributed depth to the evaluation by assessing participants' perception of the online learning environment along dimensions of relevance, reflection, and interactivity. Additionally, the analysis employed both descriptive and inferential statistics, including Wilcoxon tests and a linear mixed-effects model, to account for within-subject variation and potential interactions over time. The use of a Directed Acyclic Graph (DAG) to guide variable selection for the regression model further reflects a methodologically robust and transparent approach to controlling for potential confounding.

Another strength is heterogeneity in the participant group, including medical doctors, resident doctors, and medical students with different levels of prior research experience. The high posttest scores among completers, together with moderate-to-high COLLES ratings, suggest both course and participants were effective and enjoyed by those completing it in full.

Yet some limitations should be noted. The complete pre–post sample size was small, limiting generalizability of findings and power of subgroup analyses and regression models. Completion rate was also low, with only 26% of those registering initially completing the course. Although reasons for non-completion were gathered, these were mainly self-reported and open-text, without standardized measurement of barriers. This risks introducing bias and reducing interpretability.

While the causal model (DAG) supported the inclusion of multiple predictors, the limited sample size in this pilot study constrains the statistical power of the linear mixed-effects model.

Consequently, the estimates from the model should be interpreted as exploratory rather than confirmatory. Future studies with larger cohorts are needed to validate these effects with higher precision.

Moreover, there was no control group, which prevents attribution of knowledge gains solely to the intervention. While the pre–post design offers some insight into learning outcomes, future implementations would benefit from a waitlist control design. Lastly, although the COLLES tool provided valuable feedback, we did not collect qualitative feedback in-depth (e.g., interviews or open-ended reflections) that might have provided a richer understanding of participant experiences.

Despite these limitations, the 4ALL pilot project represents an important step toward structured, accessible, and evidence-based research education for medical professionals. The findings will inform future iterations of the course, especially in terms of delivery format and participant engagement.

Significance and Future Implications

This study provides valuable insights into the effectiveness of a structured research methodology course for general practitioners and medical students. Findings have the potential to inform refinements and contribute to scaling 4ALL for broader implementation in medical education.

This research aligns with higher education's commitment to advancing evidence-based medical education (Maggio et al., 2013) and could serve as a pilot for future studies exploring e-learning interventions in medical training.

Conclusion

The 4ALL course successfully addressed a critical gap in research education for medical professionals by offering an accessible, practice-oriented, and evidence-based training program. Despite the modest course completion rate, the significant improvement in posttest scores among completers and the overall positive feedback on the learning environment demonstrate the course’s pedagogical effectiveness and practical relevance.

The findings suggest that the 4ALL course is particularly well-suited to support general practitioners and medical students who require statistical and methodological competencies but may lack access to structured training opportunities. Given its modular design, digital format, and integration of applied SPSS instruction, the course represents a valuable educational resource that aligns well with the demands of primary care, research, and teaching. As such, the

course holds strong potential for broader implementation within the Institute of General Practice, not only for training early-career researchers but also for professional development activities among practicing physicians.

To ensure the long-term impact of the course, a follow-up assessment should be done 3-6 months after course completion, using a repeat posttest. This will help evaluate the retention of knowledge over time and provide further insights into the sustainability of the learning gains. Measuring long-term retention is essential for determining whether such educational interventions truly support enduring competency and integration of knowledge into professional practice.

Nonetheless, the evaluation also points to areas for refinement. Structural modifications—such as delivering the course in shorter, modular units and improving communication around course expectations—will be essential to enhance participant engagement and retention. By addressing these practical challenges and incorporating participant feedback, the course can evolve into a sustainable and widely applicable training tool.

In conclusion, the 4ALL course is a promising educational initiative that can contribute meaningfully to strengthening research capacity in general medicine. With targeted improvements and a commitment to evaluating both short- and long-term outcomes, it may serve as a model for similar programs in other medical faculties and institutions focused on building methodological literacy among healthcare professionals.

Data Availability

The dataset used in this study is freely available to anyone upon request from the author.

References

AAMC. (2018). *Number of medical schools including topic as an independent course or part of an integrated course: Biostatistics*. Curriculum Reports. https://www.aamc.org/data-reports/curriculum-reports/data/content-documentation-independent-course-or-part-integrated-course

Arbeitsgruppe Biostatistik und Medizinische Biometrie. (2025). *Biostatistik und Medizinische Biometrie*. Universität Bielefeld. https://www.uni-bielefeld.de/fakultaeten/medizin/fakultaet/arbeitsgruppen/biostatistik/

Arias, A., Peters, O. A., & Broyles, I. L. (2017). New curricular design in biostatistics to prepare residents for an evidence-based practice and lifelong learning education: A pilot approach. *International Endodontic Journal*, *50*(10), 999–1010. https://doi.org/10.1111/iej.12714

Athanasou, J. (2007). Self-Assessments of Skills in an Adult Educational and Vocational Guidance. *The Educational and Developmental Psychologist*, *24*(1), 36–44. https://doi.org/10.1017/S0816512200029084

Baker, J. (2007). Constructivist online learning environment survey. In *Handbook of Research on Electronic Surveys and Measurements* (S. 299–301). IGI Global Scientific Publishing.

Berthelot, G., Johnson, S., Noirez, P., Antero, J., Marck, A., Desgorces, F.-D., Pifferi, F., Carter, P. A., Spedding, M., Manoux, A. S., & Toussaint, J.-F. (2019). The age-performance relationship in the general population and strategies to delay age related decline in performance. *Archives of Public Health*, *77*(1), 51. https://doi.org/10.1186/s13690-019-0375-8

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology* (Bd. 44, S. 247–296). Elsevier. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Eva, K. W., Cunnington, J. P. W., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in*

Health Sciences Education : Theory and Practice, *9*(3), 211–224. https://doi.org/10.1023/B:AHSE.0000038209.65714.d4

Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine :, Journal of the Association of American Medical Colleges..*, *80*(10 Suppl), S46-54. https://doi.org/10.1097/00001888-200510001-00015

Gordon, K. R. (2019). How Mixed-Effects Modeling Can Advance Our Understanding of Learning and Memory and Improve Clinical and Educational Practice. *Journal of Speech, Language, and Hearing Research : JSLHR*, *62*(3), 507–524. https://doi.org/10.1044/2018_JSLHR-L-ASTM-18-0240

Institut für Biostatistik und Mathematische Modellierung. (2025). *Kompaktkurse zur Biometrie*. Universitätsmedizin Frankfurt. https://www.unimedizin-ffm.de/statistikkurse

Institut für Epidemiologie, Biostatistik und Prävention. (2025a). *Forschungsmethoden*. Universität Zürich. https://ebpi.uzh.ch/de/research_methods_DE.html

Institut für Epidemiologie, Biostatistik und Prävention. (2025b). *Vorlesungen und Kurse in Biostatistik*. Universität Zürich. https://www.ebpi.uzh.ch/de/teaching/uzhprograms/faculties/teachingbiostats.html

Maggio, L. A., Tannery, N. H., Chen, H. C., ten Cate, O., & O'Brien, B. (2013). Evidence-based medicine training in undergraduate medical education: A review and critique of the literature published 2006–2011. *Academic Medicine*, *88*(7), 1022–1028.

Mai, D. H., Taylor-Fishwick, J. S., Sherred-Smith, W., Pang, A., Yaworsky, J., Whitty, S., Lafever, A., Mcilvain, C., Schmitt, M., Rogers-Johnson, M., Pace, A., & Dobrian, A. D. (2020). Peer-Developed Modules on Basic Biostatistics and Evidence-Based Medicine Principles for Undergraduate Medical Education. *MedEdPORTAL : The Journal of*

Teaching and Learning Resources, *16*, 11026. https://doi.org/10.15766/mep_2374-8265.11026

Nemeth, D., Janacsek, K., & Fiser, J. (2013). Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Frontiers in computational neuroscience*, *7*, 147.

Norman, G. R., & Schmidt, H. G. (1992). The psychological basis of problem-based learning: A review of the evidence. *Academic medicine*, *67*(9), 557–565.

Schimek, M. G. (2025). *Evidenzbasierte Medizin und Biostatistik*. Institut für Med. Informatik, Statistik und Dokumentation. https://user.medunigraz.at/michael.schimek/Wiss_GL_2013_2014/EBM_Einfuehrung_MGS.pdf

Schuler, A. (2022). Mixed models for repeated measures should include time-by-covariate interactions to assure power gains and robustness against dropout bias relative to complete-case ANCOVA. *Therapeutic Innovation & Regulatory Science*, *56*(1), 145–154.

Schutt, R. K. (2018). *Investigating the social world: The process and practice of research*. Sage publications. https://us.sagepub.com/sites/default/files/upm-binaries/43454_10.pdf

Taylor, D. C., & Hamdy, H. (2013). Adult learning theories: Implications for learning and teaching in medical education: AMEE Guide No. 83. *Medical teacher*, *35*(11), e1561–e1572.

Appendices

Course Flyer

Appendix 1: Course flyer

Einladung zur Teilnahme am Online-Kurs

„Bessere Entscheidungen im Praxisalltag durch evidenzbasierte Medizin:
Forschung und Biostatistik für die Hausarztpraxis“

Liebe Kolleginnen, Kollegen und Studierende,

wie oft stehen wir im Praxisalltag vor medizinischen Leitlinien oder aktuellen Studien und fragen uns: „Wie verlässlich sind diese Ergebnisse?“ oder „Kann ich diese Empfehlung auf meine Patienten anwenden?“

Um fundierte und sichere Entscheidungen treffen zu können, sind Kenntnisse in Forschungsdesign und statistischer Analyse unerlässlich. Doch während es viele Fortbildungen zu klinischen Themen gibt, fehlen praxisnahe Angebote zu Biostatistik und Forschungsmethoden – dabei sind sie die Grundlage der evidenzbasierten Medizin.

Daher laden wir euch herzlich zu einem neuen **Online-Kurs** ein.

Warum dieser Kurs?

✓ **Praktische Anwendung für den Praxisalltag und eigene Forschung:** Ihr lernt nicht nur, wie man Studien kritisch bewertet und wissenschaftliche Erkenntnisse in die Patientenversorgung integriert, sondern auch, wie ihr eigene klinische Fragestellungen in wissenschaftliche Studien umsetzt – von der Planung einer Studie über die Datenanalyse bis hin zur Interpretation der Ergebnisse.

✓ **Fallbeispiele & aktuelle Studien:** Wir besprechen jede Woche Online die Anwendung der Methoden im Alltag.

✓ **Effizient & flexibel:** Mit kurzen Videolektionen, interaktiven Übungen und wöchentlichen Besprechungen lässt sich der Kurs ideal in den Alltag integrieren.

Kursdetails:

Zeitraum: 1. Mai – 30. Juni 2025
Plattform: Moodle, https://iam-augsburg.de/
Kosten: Keine Kursgebühr
Anmeldung: zekeriya.aktuerk@med.uni-augsburg.de
Zielgruppe: Studierende, Ärztinnen und Ärzte in Weiterbildung, niedergelassene Kolleginnen und Kollegen

Kursstruktur:

70 Kurzvideos (Gesamtlaufzeit < 6 Stunden)
Quizfragen & Selbstübungen zur Vertiefung
Wöchentliche Online-Sitzung (Fr 16:00 – 17:00 Uhr) mit Diskussion einer aktuellen Studie oder eines Praxisbeispiels
Kursbewertungsformulare zur Weiterentwicklung des Programms
Teilnahmezertifikat des Instituts für Allgemeinmedizin

Diese Fortbildung ist eine Gelegenheit, eure Kompetenzen in Forschungsmethodik und Biostatistik zu erweitern – praxisnah, kompakt und völlig flexibel.

Interessiert? Dann meldet euch unter zekeriya.aktuerk@med.uni-augsburg.de an oder schreibt bei Fragen!

Wir freuen uns auf eure Teilnahme und spannende Diskussionen!

Mit besten Grüßen,
Kursleitung – Institut für Allgemeinmedizin, Universität Augsburg

***Der Kurs ist von der Bayerischen Landesärztekammer (BLÄK) mit 134 CME-Punkten anerkannt und zertifiziert.**

Appendix 2: Course structure and objectives

Course Structure and Objectives

The course consists of 70 short video lessons (total duration: ~6 hours). The course is divided into the following parts:

- Module 1.1 - Research Methods
- Module 1.2 - Basic Statistical Theory
- Module 1.3 - Descriptive Statistics
- Module 1.4 - Data Entry and Analysis Preparation
- Module 2.1 - Hypothesis Testing; Categorical Outcome
- Module 2.2 - Hypothesis Testing; Numerical Outcome
- Module 2.3 - Advanced Statistical Principles

Quiz questions to review knowledge follow after each unit. Self-study exercises are prepared for practical application. Additionally, weekly live sessions with discussions on current studies or practical examples are offered.

Aims and Objectives of the Different Modules

Module 1.1 - Research Methods

After completing Module 1.1, participants will have a basic understanding of research methods and study designs. They will be able to formulate clinical questions, research scientific literature, select suitable study designs, and evaluate their level of evidence. This knowledge forms the basis for a critical evaluation of research findings and their own research projects.

Learning Objectives

After completing this module, participants will be able to:

- Assess their current prior knowledge of research methods.
- Identify areas in which they need to deepen their knowledge.
- Apply the PICO method to formulate clear, research-relevant questions.
- Structure clinical questions to facilitate literature research.
- Understand the difference between null and alternative hypotheses.
- Formulate scientific hypotheses precisely.
- Distinguish between testable and untestable hypotheses.
- Effectively search for scientific articles in PubMed.

- Use relevant keywords and search strategies.
- Evaluate the quality and relevance of retrieved publications.
- Use the features of Clinical Queries to specifically search for clinical studies.
- Use MeSH terms for a more precise and systematic literature review.
- Distinguish between qualitative, quantitative, and mixed-methods approaches.
- Evaluate the advantages and disadvantages of different research approaches.
- Recognize different classifications and terms for research types.
- Correctly interpret terms from the research literature.
- Identify different study designs based on real-world research questions.
- Assign case studies to different research methods.
- Understand the levels of evidence in scientific studies.
- Evaluate the significance of different study designs using the hierarchy of evidence.
- Select the appropriate study design for a given research question.
- Understand the differences between experimental and observational studies.
- Understand the design and methodology of case-control studies.
- Apply the calculation and interpretation of odds ratios (ORs).
- Evaluate the strengths and weaknesses of case-control studies.
- Understand the design and applications of cohort studies.
- Calculate and interpret the relative risk (RR) for evaluating epidemiological associations.
- Analyze the advantages and disadvantages of prospective and retrospective cohort studies.
- Explain the different phases of clinical trials (phases 1–4).
- Understand the importance of randomized controlled trials (RCTs) for evidence-based medicine.
- Evaluate methods for minimizing bias in clinical trials.

List of video recordings:

1. Probability rules
2. Normal distribution
3. Other distributions
4. Sampling
5. Sampling examples
6. Sample variation

Module 1.2 - Basic Statistical Theory

After completing Module 1.2, participants will have a solid understanding of basic statistical concepts and their application in medical research. They will be able to calculate probabilities, interpret distributions, choose appropriate sample selection procedures, and consider sample variation.

Learning Objectives

After completing this module, participants will be able to:

- Understand and apply the basic rules of probability (addition and multiplication).
- Apply probability concepts to medical questions.
- Distinguish between conditional probabilities and independent events.
- Explain the properties of the normal distribution and understand their significance in medical statistics.
- Interpret standard normal distributions and calculate Z-scores.
- Recognize the practical relevance of the normal distribution for medical data.
- Identify different theoretical distributions (chi-square, t-, F-, binomial, and Poisson distributions).
- Understand the application of these distributions in various medical research scenarios.
- Explain the differences between symmetric and asymmetric distributions.
- Define different sampling methods (e.g., random, systematic, stratified sampling).
- Understand the importance of sample size and its influence on statistical results.
- Recognize bias caused by improper sampling.
- Analyze real-world applications of different sampling methods.
- Evaluate the advantages and disadvantages of probabilistic and non-probabilistic samples.
- Explain the importance of a representative sample for the generalizability of study results.
- Understand the causes and effects of sample variability.
- Interpret concepts such as standard deviation, standard error, and confidence intervals.
- Recognize the influence of sample size and variability on statistical results.

List of video recordings:

1. Probability rules
2. Normal distribution

3. Other distributions
4. Sampling
5. Sampling examples
6. Sample variation

Module 1.3 - Descriptive Statistics

After completing Module 1.3, participants will have a solid understanding of descriptive statistics and data visualization. They will be able to describe data in a structured manner, calculate statistical parameters, create tables and graphs, and interpret them correctly. This forms the basis for further statistical analyses and evidence-based medical decisions.

Learning Objectives

After completing this module, participants will be able to:

- Understand the differences between nominal, ordinal, and metric variables.
- Recognize which statistical methods are appropriate for different data types.
- Select the appropriate variable scale in SPSS for data entry and analysis.
- Calculate and interpret the mean, median, and mode.
- Understand when each location parameter is most appropriate.
- Evaluate the influence of outliers on location parameters.
- Calculate measures of variability: range, interquartile range, variance, and standard deviation.
- Understand the importance of coefficients of variation for comparing different data sets.
- Apply measures of dispersion in the interpretation of medical data.
- Calculate confidence intervals and understand their meaning.
- Explain the dependence of confidence intervals on sample size and variability.
- Understand the difference between point estimates and interval estimates.
- Recognize how confidence intervals are used to evaluate study results.
- Understand the relationship between confidence intervals and statistical significance.
- Avoid critical errors in the interpretation of confidence intervals.
- Describe and interpret frequency distributions.
- Understand the concepts of unimodality, multimodality, skewness, and kurtosis.
- Recognize the differences between symmetric, right-skewed, and left-skewed distributions.

- Apply various transformations to fit data to a normal distribution.
- Recognize the need for data transformations for specific statistical tests.
- Avoid common errors in data fitting.
- Perform transformations in SPSS to normalize and standardize data.
- Evaluate the effects of transformations on statistical tests.
- Correctly apply transformations in medical research.
- Understand the criteria for a good scientific table.
- Correctly create tables for different types of statistical data.
- Avoid errors in table formatting and interpretation.
- Create and correctly format tables with SPSS.
- Prepare SPSS results for scientific publications.
- Design tables to provide statistically accurate and easily understandable information.
- Identify and correctly use histograms, box plots, bar charts, and scatter plots.
- Interpret graphs to provide a clear, understandable presentation of data.
- Avoid common errors in graphically presenting data.
- Create graphs in SPSS and adapt them to the needs of scientific work.
- Use different graph types correctly and explain their meaning.
- Design data visualizations to clearly convey key statistical insights.

List of video recordings:

1. Data types
2. Measures of central tendency
3. Parameters of spread
4. Confidence intervals
5. Interpretation of confidence intervals
6. Frequency distribution
7. Data transformation
8. Data transformation with SPSS
9. Scientific tables
10. Scientific tables with SPSS
11. Graphs
12. Graphs with SPSS

Module 1.4 - Data Entry and Analysis Preparation

After completing Module 1.4, participants will have a solid understanding of data preparation and entry in SPSS. They will be able to enter data accurately, correctly define variable types, clean up erroneous data, identify outliers, and perform basic descriptive analyses. This is an essential foundation for further statistical analyses.

Learning Objectives

After completing this module, participants will be able to:

- Enter data correctly into SPSS and select the correct format for variables.
- Understand the difference between data views and variable views in SPSS.
- Organize datasets efficiently and avoid sources of data entry errors.
- Identify and correctly use the different variable types in SPSS (nominal, ordinal, metric).
- Understand how variable types influence the selection of statistical tests.
- Create coding strategies for categorical data (e.g., dummy variables).
- Recognize common data errors such as missing values, typos, or incorrect coding.
- Apply methods for identifying and correcting erroneous data.
- Understand basic principles of data cleaning for reliable statistical analyses.
- Apply SPSS functions for error control and data cleaning.
- Use syntax-based methods for data cleaning in SPSS.
- Review data quality and ensure that incorrect input does not distort analysis results.
- Understand the difference between outliers and extreme values.
- Apply methods for detecting outliers (e.g., boxplots, z-scores, IQR method).
- Evaluate the impact of outliers on statistical analyses.
- Identify outliers in SPSS and interpret their significance for the analysis.
- Apply methods for handling outliers (transformation, Winsorization, exclusion).
- Evaluate the influence of extreme values on measures of location and dispersion.
- Conduct descriptive analyses in SPSS (e.g., means, median, standard deviation).
- Correctly interpret and graphically present the results of SPSS analyses.
- Correctly prepare tables and charts for scientific reports.

List of video recordings:

1. Data entry in SPSS

2. Variable types in SPSS
3. Error control and data cleaning
4. Error control and data cleaning in SPSS
5. Outliers and extreme values
6. Outliers and extreme values in SPSS
7. Descriptive statistics with SPSS

Module 2.1 - Hypothesis Testing; Categorical Outcome

This module forms the foundation for advanced statistical testing and is essential for evidence-based medical practice. After completing Module 2.1, participants will have a solid understanding of hypothesis testing with categorical data. They will be able to correctly apply and interpret the chi-square test, the Fisher exact test, the McNemar test, and the binomial test. They will be able to use SPSS to analyze categorical data and understand test results, as well as recognize when to use each test and identify potential biases in the analysis.

Learning Objectives

After completing this module, participants will be able to:

- Understand the basic principles of hypothesis testing.
- Clearly define the difference between the null hypothesis (H_0) and the alternative hypothesis (H_1).
- Explain the meaning of p-values, significance levels (α), and statistical power.
- Understand the types of errors in hypothesis testing (Type I and Type II errors).
- Apply the chi-square test for independence.
- Understand the prerequisites for applying the chi-square test.
- Calculate expected values and the chi-square statistic.
- Interpret test results and identify statistically significant relationships.
- Perform chi-square tests in SPSS.
- Correctly interpret SPSS output (e.g., contingency tables, significance values, expected frequencies).
- Identify sources of error and potential biases in the application of the test.
- Understand when to use the Fisher exact test instead of the chi-square test.
- Interpret the calculation of the exact p-value.
- Understand the advantages and limitations of the Fisher Exact Test.

- Perform the Fisher Exact Test correctly in SPSS.
- Interpret SPSS output and recognize when it is necessary to use it.
- Formulate results for reporting in scientific papers.
- Understand the application of the McNemar test for dependent categorical variables.
- Explain the difference between independent and dependent samples in categorical data.
- Explain the prerequisites and interpretations of the McNemar test.
- Perform the McNemar test correctly in SPSS.
- Interpret SPSS output and make hypotheses decisions.
- Evaluate when the McNemar test is preferable to another method.
- Apply the binomial test to analyze a single categorical variable.
- Explain the meaning of the null distribution and the interpretation of the test.
- Conduct binomial tests and chi-square goodness-of-fit tests in SPSS.
- Interpret results and document them correctly for scientific reports.
- Avoid common errors in application and make the correct test choice.

List of video recordings:

1. General approach to hypothesis testing
2. Categorical outcome - independent measurements (chi-square test)
3. Chi-square analysis with SPSS
4. Categorical outcome - independent measurements (Fisher's exact test)
5. Fisher's exact test with SPSS
6. Categorical outcome - two dependent measurements (McNemar test)
7. McNemar test with SPSS
8. Categorical outcome - one measurement (binomial test and chi-square analysis)
9. Binomial test and chi-square analysis with SPSS

Module 2.2 - Hypothesis Testing; Numerical Outcome

This module lays the foundation for evidence-based analysis of numerical data in medical research. After completing Module 2.2, participants will have a solid understanding of hypothesis testing for numerical data. They will be able to calculate and interpret correlations, analyze differences between two or more groups using t-tests, Mann-Whitney U tests, ANOVA, and Kruskal-Wallis tests, evaluate time-dependent measurements using repeated-measures

ANOVA and the Friedman test, and perform the given tests in SPSS and correctly interpret the results.

Learning Objectives

After completing this module, participants will be able to:

- Understand the differences between Pearson and Spearman correlations.
- Know the prerequisites for applying correlation analyses.
- Correctly interpret the results of a correlation analysis.
- Correctly calculate correlations in SPSS.
- Interpret SPSS output and choose correctly between Pearson and Spearman.
- Avoid common errors in correlation analysis.
- Apply the one-sample t-test.
- Understand the test's assumptions (e.g., normal distribution of the data).
- Correctly interpret test results (t-value, degrees of freedom, p-value).
- Correctly perform one-sample t-tests in SPSS.
- Prepare test results for scientific reports.
- Correctly apply the t-test for independent samples (Student's t-test).
- Verify the assumptions for the t-test (normal distribution, homogeneity of variance).
- Use the Mann-Whitney U test as a nonparametric alternative to the t-test.
- Understand the differences between parametric and nonparametric tests.
- Conduct independent t-tests and Mann-Whitney U tests in SPSS.
- Correctly interpret SPSS output (Levene's test, p-value, effect size).
- Apply the dependent samples t-test to analyze repeated measures.
- Review the prerequisites for the paired t-test.
- Use the Wilcoxon test as a nonparametric alternative to the t-test.
- Understand the differences between the two tests and justify their use.
- Correctly conduct dependent samples tests in SPSS.
- Correctly interpret results and identify differences between groups.
- Apply one-way ANOVA.
- Check the assumptions of ANOVA (normal distribution, homogeneity of variance).
- Use the Kruskal-Wallis test as a nonparametric alternative to ANOVA.
- Perform the tests correctly in SPSS and interpret the results.
- Apply post-hoc tests to identify significant group differences.

- Apply the repeated measures ANOVA.
- Understand the assumptions of ANOVA for repeated measurements (e.g., sphericity).
- Use the Friedman test as a nonparametric alternative to repeated measures ANOVA.
- Perform the repeated measures ANOVA and Friedman test correctly in SPSS.
- Correctly interpret multivariate and within-subject effects.
- Prepare the results for scientific reports.

List of video recordings:

1. Two metric or ordinal variables (correlation analysis)
2. Pearson and Spearman correlation analysis with SPSS
3. Metric outcome - 1 sample (One-sample t-test)
4. Single-sample t-test - SPSS
5. Metric outcome - 2 independent measurements (Independent samples t-test (Student's t-test))
6. Student's t-test - SPSS
7. Metric outcome - 2 independent measurements (Mann-Whitney U-test)
8. Mann-Whitney U-test with SPSS
9. Metric outcome - 2 dependent measurements (Dependent samples t-test)
10. Dependent groups t-test with SPSS
11. Metric outcome - 2 dependent measurements (Wilcoxon test)
12. Wilcoxon test with SPSS
13. Metric outcome - More than two groups (One-way ANOVA - Kruskal-Wallis test)
14. One-way ANOVA with SPSS
15. Kruskal-Wallis Test with SPSS
16. Metric Outcome - More than 2 Dependent Measures (General Linear Model, Repeated Measures
17. Repeated Measures ANOVA and Friedman Analysis)
18. Repeated Measures ANOVA with SPSS
19. Two-Way Repeated Measures Analysis of Variance - SPSS
20. Friedman Analysis with SPSS

Module 2.3 - Advanced Statistical Principles

This module lays the foundation for precise and methodologically sound statistical analysis in medical research. After completing Module 2.3, participants will have in-depth knowledge of statistical error sources, sample size, and correction methods for multiple testing. They will be able to identify and minimize types of error in hypothesis testing, calculate and interpret sample sizes for statistical tests, conduct practical sample size analyses using Russ Lenth's Power App, and apply appropriate correction methods for multiple comparisons.

Learning Objectives

After completing this module, participants will be able to:

- Understand the differences between Type I errors (α errors) and Type II errors (β errors).
- Explain the impact of the significance level (α) on the risk of error.
- Analyze the influence of sample size and statistical power on hypothesis testing.
- Identify strategies for reducing error in statistical analyses.
- Explain the importance of sample size for statistical tests.
- Understand the impact of small and large samples on statistical power.
- Analyze the dependence of effect size, variance, and significance level on sample size.
- Calculate the optimal sample size for various tests using Russ Lenth's Power App.
- Interpret the parameters of a sample calculation (effect size, power, significance level).
- Understand practical application scenarios for sample size calculations.
- Understand the problems of multiple hypothesis testing (increasing the α error).
- Recognize different methods for correcting for multiple comparisons (Bonferroni, Tukey, Tamhane)
- Select the appropriate correction method for a given research question.

List of video recordings:

1. Hypothesis testing errors
2. Sample size
3. Sample size calculation with RussLenthsApp
4. Correction for multiple hypothesis tests

Appendix 3: Pretest and Posttest

Pretest and Posttest

1.	Forschungsmethoden	Bitte wählen Sie das Outcome von der folgenden Fragestellung: „Gibt es einen signifikanten Unterschied zwischen einer kalorienbasierten Diät und einer auf dem glykämischen Index basierenden Diät bei der Verringerung des Body-Mass-Index adipöser Männer?“ kalorienbasierter Diät glykämischer Index basierenden Diät Body-Mass-Index adipöse Männer	Body-Mass-Index
2.	Fragestellung (PICO)	Welche der folgenden Komponenten gehört <u>nicht</u> zu einer gut formulierten PICO-Frage? a) Population b) Intervention c) Confounder d) Outcome	Confounder
3.	Hypothesen formulieren	*Richtig oder falsch:* Eine Nullhypothese (H_0) besagt, dass ein statistisch signifikanter Effekt vorliegt.	F

4.	Literaturrecherche (PubMed)	Welche Methode ist am effektivsten, um gezielt nach randomisierten kontrollierten Studien (RCTs) in PubMed zu suchen? a) MeSH-Begriffe verwenden b) Den Filter „Randomized Controlled Trial“ aktivieren c) Nur im Titel suchen d) Die Suchbegriffe mit "NOT" kombinieren	Den Filter „Randomized Controlled Trial“ aktivieren
5.	Arten der Forschung	Welche Forschungsmethode wird verwendet, um Ursache-Wirkungs-Zusammenhänge am besten zu untersuchen? a) Fall-Kontroll-Studie b) Kohortenstudie c) Experimentelle Studie d) Querschnittsstudie	Experimentelle Studie
6.	Arten der Forschung	*Richtig oder falsch:* Eine Querschnittsstudie ist immer retrospektiv.	F
7.		Bitte ordnen Sie die folgenden Stichwörter als „qualitativ“, „quantitativ“ oder „beides“ Interventionsstudien - Quantitativ Interview - Qualitativ Focus Gruppen - Qualitativ Kohortenstudien - Quantitativ	

		Querschnittsstudie - Beides Fall-Kontroll-Studie - Quantitativ	
8.	Benennungen bezüglich Arten der Forschung	Welche Aussage trifft auf eine Längsschnittstudie zu? a) Sie erfasst Daten zu einem einzelnen Zeitpunkt b) Sie beobachtet dieselben Personen über einen längeren Zeitraum c) Sie ist immer retrospektiv d) Sie kann keine Kausalzusammenhänge untersuchen	B
9.		Welche Studienarten können retrospektiv sein? Fall-Kontroll-Studien Kohortenstudien Querschnittsstudien Systematische Reviews und Meta-Analysen Randomisierte Kontrollierte Studien	 A B C D
10.		Welcher Studiendesign wird unten beschrieben? Erwachsene Patienten, die zu einer Allgemeinmedizinpraxis kommen, werden zufällig in zwei Gruppen eingeteilt.	

		Gruppe 1: Zusätzlich zur routinemäßigen Behandlung führt der Arzt ein Gespräch über die Wichtigkeit von Bewegung und die Teilnehmer erhalten ein Informationsblatt über die Bedeutung von Bewegung. Gruppe 2: Zusätzlich zur routinemäßigen Behandlung erhalten die Patienten ein Informationsblatt über die Umwelt. Am Studienanfang und bei der nächsten Visite werden alle Teilnehmer über ihre täglichen Schritte befragt. Querschnittstudie Randomisiert Doppelblind Einzelblind Prospektiv Fall-Kontroll-Studie	
11.	Evidenzpyramide	Welche Studie hat die höchste Evidenzstufe? a) Fall-Kontroll-Studie b) Kohortenstudie c) Meta-Analyse d) Expertenmeinung	Meta-Analyse

12.	Fall-Kontroll-Studien	Warum sind Fall-Kontroll-Studien besonders anfällig für Recall Bias?	
13.		„Welche Unterschiede im Risiko gibt es bezüglich Rauchens zwischen Patienten mit Lungenkrebs und gesunden Kontrollpersonen?“ Welche der folgenden Primärstudientypen ist für die obige Fragestellung am besten geeignet? Fall-Kontroll-Studie Kohortenstudie Randomisierte kontrollierte Studie (RCTs) Meta-Analyse	Fall-Kontroll-Studie
14.		Welche der Aussagen bezüglich des Zitats aus der Brewster- Studie 2019 (Creatine Kinase and Blood Pressure: A Systematic Review) (https://pubmed.ncbi.nlm.nih.gov/30970679/) ist/sind korrekt? “…CK was a strong predictor of treatment failure in the general population, with an adjusted odds ratio of 3.7 [1.2 to 10.9]”. Eine Einheit CK erhöht das Therapieversagen 3,7 Fach. Der Zusammenhang von CK und Therapieversagen ist signifikant. Das Konfidenzintervall des Odds Ratios ist breit. Die Daten kommen von einer Fallkontrollstudie.	

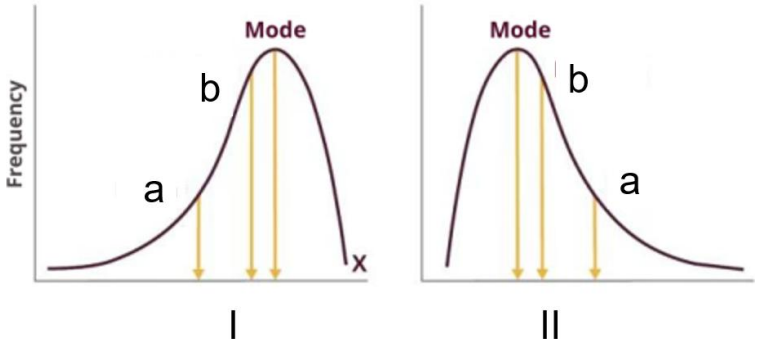
15.	Kohortenstudien	Welche Aussage zur Kohortenstudie trifft zu? Sie beginnt mit einer Exposition und verfolgt das Outcome Sie vergleicht rückblickend Fälle und Kontrollen Sie ist immer experimentell Sie liefert keine Inzidenzraten	
16.	Klinische Studien	Welche klinische Studienphase dient dazu, Langzeiteffekte und Sicherheit zu überwachen? a) Phase I b) Phase II c) Phase III d) Phase IV	
17.		Ein Substanz erhält die Zulassung zur Verwendung als Arzneimittel nach Abschluss der Phase-……..-Studien im Rahmen klinischer Studien. 0 I II III IV	

18.		In einer Kohortenstudie wurde der Einfluss von Medikamentenadhärenz (Risikofaktor) auf Apoplex (Outcome) untersucht. Die Analyse ergab ein Relatives Risiko (RR) von 2,9 (95% Konfidenzintervall: 1,3-6,6). Welche der folgenden Aussagen sind korrekt? Der Einfluss von Medikamentenadhärenz auf Apoplex ist signifikant. Das Konfidenzintervall ist relativ breit. Menschen mit Medikamentenadhärenz haben im Vergleich zu Menschen ohne Adhärenz ein 2,9-fach geringeres Apoplex-Risiko. Das 99% Konfidenzintervall für das RR ist weiter im Vergleich zu 95% Konfidenzintervall.	
19.	Wahrscheinlichkeitsregeln	Wenn die Wahrscheinlichkeit für Erkrankung A 0,2 beträgt und für Erkrankung B 0,3, und beide unabhängig sind, wie groß ist die Wahrscheinlichkeit, beide Erkrankungen zusammen zu haben? a) 0,06 b) 0,25 c) 0,5 d) 0,1	0,06
20.	Normalverteilung	Welche zwei Parameter bestimmen die Form einer Normalverteilung?	
21.		*Richtig oder falsch:* Eine Normalverteilung hat immer eine Schiefe (Skewness) von 1.	F

22.	Andere Verteilungen	Welche Verteilung wird typischerweise für diskrete Daten verwendet? a) Normalverteilung b) Binomialverteilung c) Exponentialverteilung d) t-Verteilung	Binomial
23.		Ordnen Sie jeder der folgenden Verteilungen die passenden Parameter zu: Parameter: A) Freiheitsgrade von Stichprobengrößen B) Freiheitsgrade von Kontingenztabellen C) Freiheitsgrade von Zähler und Nenner D) Erwartungswert λ (Erwartungswert = Varianz) E) n (Anzahl der Versuche) und π (Wahrscheinlichkeit eines Erfolgs pro Versuch) F) Mittelwert μ Verteilungen: t-Verteilung Chi-Quadrat-Verteilung F-Verteilung	

		Binomialverteilung Poissonverteilung	
24.		Stellen Sie sich vor, Sie führen eine Studie an einem großen Krankenhaus durch, um den durchschnittlichen Blutdruck der Patienten zu ermitteln. Sie müssen eine Stichprobe auswählen. Welche der folgenden Stichprobenmethoden wäre in dieser Situation am besten geeignet? A) Sie nehmen alle Patienten, die gerade im Wartezimmer sitzen. B) Sie teilen die Patienten in Gruppen auf, basierend auf Alter und ziehen dann aus jeder Altersgruppe eine Zufallsstichprobe. C) Sie wählen zufällig drei Krankenhausabteilungen aus und ziehen alle Patienten aus diesen Abteilungen für die Studie heran. D) Sie setzen Quoten für die Altersgruppen fest und rekrutieren Patienten so lange, bis die Quoten für jede Altersgruppe erfüllt sind.	B
25.	Sampling (Stichprobenziehung)	Welche Methode ist eine Zufallsstichprobe? a) Gelegenheitsstichprobe b) Schneeballstichprobe c) Geschichtete Zufallsstichprobe d) Quotenstichprobe	

26.	Variation des Samples	Welche zwei Faktoren beeinflussen die Variation einer Stichprobe am meisten?	M,SD
27.	Datentypen	*Richtig oder falsch:* Ordinale Daten haben eine sinnvolle Reihenfolge, aber unbekannte Abstände zwischen den Kategorien.	R
28.	Maße der zentralen Tendenz	Welches Maß ist am besten für eine rechtsschiefe Verteilung geeignet? a) Mittelwert b) Median c) Modus d) Standardabweichung	Median
29.		Welches Skalenniveau liegt für die folgenden Variablen vor? Geschmacksrichtung von Speiseeis Nominalskala Abstand zwischen zwei Gebäuden in cm Verhältnisskale (Ratioskala, metrisch) Schulnoten auf einer Skala von 1 bis 6 Ordinalskala (geordnete Kategorien) gemessener Intelligenzquotient Intervallskala (metrisch)	
30.	Häufigkeitsverteilung	Welche Grafik eignet sich am besten zur Darstellung einer normalverteilten Variable? a) Balkendiagramm b) Histogramm	B

		c) Boxplot d) Kreisdiagramm	
31.		Welche Aussagen treffen auf die Berechnung des Medians (Md) zu? Der Median kann bei Daten berechnet werden, die nominalskaliert vorliegen. **Die Berechnung des Medians unterscheidet sich je nachdem, ob eine gerade oder ungerade Anzahl an Datenpunkten vorliegt.** **Der Median ist ein Maß zentraler Tendenz.** Der Median ist ein Ausreißer-sensitives Maß.	
32.		Welche der Aussage/n über die folgenden Grafiken ist/sind korrekt?  **II ist rechts verzerrt** a steht für Median	

		b steht für Mittelwert bei II ist Median>Mittelwert **Bei II ist Skewness > 0**	
33.		Welche Datentransformation wird in PubMed-publizierten Artikeln am häufigsten verwendet? A) Logarithmische Transformation B) Quadratische Transformation C) Quadratwurzel Transformation D) Reziproke Transformation	A
34.		Welche der folgenden Variablen sind numerisch (stetig/skaliert)? Alter (in Jahren) Einkommen (in €) Bildungsniveau (keine/Mittelschule/Studium) Hämoglobin-Wert im Blut (mg/dl)	ABD
35.		Welche der folgenden Maßnahmen gehören zur Datenbereinigung? **Identifikation von Ausreißern** **Ersetzen fehlender Werte durch den Mittelwert**	

		Erstellung eines neuen Modells zur Datenanalyse **Entfernen doppelter Einträge**	
36.	Chi-Quadrat-Test	Welche Voraussetzung muss für ein Chi-Quadrat-Test erfüllt sein? a) Alle erwarteten Zellhäufigkeiten müssen >5 sein b) Die Stichprobe muss normalverteilt sein c) Die Variablen müssen kontinuierlich sein d) Die Gruppen müssen gleich groß sein	A
37.		Welcher der folgenden Tests ist am besten geeignet, um den Zusammenhang zwischen zwei kategorischen unabhängigen Variablen zu analysieren? A) t-Test für unabhängige Stichproben B) Chi-Quadrat-Test C) Pearson-Korrelation D) Regressionsanalyse	B
38.		Welche der folgenden Szenarien deuten auf abhängige Messungen? Untersuchung des Zusammenhangs zwischen Geschlecht und Raucherstatus: Geschlecht (m/w/d), Raucherstatus (Raucher/Nichtraucher) Vergleich von Behandlungsgruppen hinsichtlich Heilung: Behandlungsart (Medikament A/Medikament B), Heilung (Ja/Nein)	

		Vorher-Nachher-Vergleich des Raucherstatus: Raucherstatus vor der Therapie (Raucher/Nichtraucher), Raucherstatus nach der Therapie (Raucher/Nichtraucher) **Untersuchung von Präferenzen vor und nach einer Informationskampagne: Präferenz vor der Kampagne (Produkt A/Produkt B), Präferenz nach der Kampagne (Produkt A/Produkt B)**	
39.		Welcher statistische Test ist für die folgenden Szenarien geeignet? Ordnen Sie die Szenarien den Tests zu: Szenarien: 1. Wir haben eine kategoriale Variable mit zwei Kategorien (z. B. „Sollten wir eine Pause machen? Ja/Nein“) und möchten testen, ob sich die Kategorien um mehr als 50 % unterscheiden. 2. Wir haben eine kategoriale Variable mit zwei Kategorien (z. B. Raucherstatus Ja/Nein) und möchten prüfen, ob das Ergebnis von einem bekannten gesellschaftlichen Durchschnitt (z. B. 25%) abweicht. 3. Wir haben eine kategoriale Variable mit mehr als zwei Kategorien (z. B. „Welchen Dozent bevorzugen Sie? Kunisch/Trompke/Zeiser“) und möchten testen, ob die Kategorien gleich verteilt sind. 4. Tests: 5. A) Chi-Quadrat-Anpassungstest (Goodness-of-Fit-Test) 6. B) Binomial-Test 7. C) Einfache Signifikanzprüfung (One-Proportion-Test)	
40.	Korrelationen	*Richtig oder falsch:* Die Pearson-Korrelation misst sowohl die Stärke als auch die Richtung eines linearen Zusammenhangs.	R

41.		Welche der folgenden Aussagen beschreibt das Hauptziel der ANOVA korrekt? Die ANOVA prüft, ob die Mittelwerte von mehr als zwei Gruppen gleich sind. Die ANOVA prüft, ob die Varianzen innerhalb der Gruppen signifikant unterschiedlich sind. Die ANOVA prüft, ob die Mittelwerte aller Beobachtungen unabhängig von der Gruppenzugehörigkeit gleich sind. Die ANOVA prüft, ob mindestens eine Gruppe signifikant unterschiedliche Ergebnisse im Vergleich zu einer Referenzgruppe zeigt.	
42.	Student’s t-Test	Der gepaarte t-Test wird verwendet, wenn: a) Zwei Gruppen unabhängig sind b) Messungen von denselben Personen vor und nach einer Intervention stammen c) Die Stichproben unterschiedlich groß sind d) Die Varianzen heterogen sind	B
43.	Wilcoxon-Test	Welcher Test ist die nicht-parametrische Alternative zum gepaarten t-Test? a) Chi-Quadrat-Test b) Wilcoxon-Test c) McNemar-Test d) Kruskal-Wallis-Test	B

44.	ANOVA	Welcher Test ist eine Alternative zur ANOVA, wenn die Normalitätsannahme verletzt ist? a) Wilcoxon-Test b) Kruskal-Wallis-Test c) McNemar-Test d) Chi-Quadrat-Test	B
45.	Repeated Measures ANOVA	Welche zusätzliche Annahme muss bei einer ANOVA mit Messwiederholungen überprüft werden? a) Sphärizität b) Homoskedastizität c) Multikollinearität d) Autokorrelation	A
46.	SPSS – Dateneingabe & Variablentypen	*Richtig oder Falsch: In SPSS beziehen sich „metrische“ Variablen auf kategorische Daten.*	F
47.	Multiple Testkorrekturen	Welche der folgenden Methoden wird verwendet, um Typ-I-Fehler in multiplen Vergleichen zu reduzieren? a) Fisher’s Exact Test b) Bonferroni-Korrektur c) Pearson-Korrelation d) Kruskal-Wallis-Test	B

48.	Fehler beim Hypothesentest	*Richtig oder falsch:* Der Typ-I-Fehler (α-Fehler) tritt auf, wenn die Nullhypothese zu Unrecht abgelehnt wird.	R
49.		Welche grafische Darstellung eignet sich am besten zur Darstellung kategorialer Variablen? a) Kreisdiagramm b) Streudiagramm c) Balkendiagramm	A
50.		Warum sind p-Wert-Korrekturen bei multiplen Hypothesentests notwendig? Um die Wahrscheinlichkeit von Typ-I-Fehlern bei einzelnen Tests zu reduzieren. **Um die Wahrscheinlichkeit von Typ-I-Fehlern über alle Tests hinweg zu kontrollieren.** Um die Wahrscheinlichkeit von Typ-II-Fehlern zu erhöhen. Um sicherzustellen, dass die Ergebnisse klinisch relevant sind. Um die Ergebnisse zwischen verschiedenen Studien vergleichbar zu machen.	

Appendix 4: Ethical approval

Ethical Approval

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

ETHIKKOMMISSION BEI DER LMU MÜNCHEN

Ethikkommission · Pettenkoferstr. 8 · 80336 München

Universität Augsburg
Lehrstuhl für Allgemeinmedizin
Herr Zekeriya Aktürk
Gutenbergstraße 7
86355 Neusäß

Vorsitzender:
Prof. Dr. R. M. Huber
Telefon+49 (0)89 440055191
Telefax+49 (0)89 440055192
Ethikkommission@
med.uni-muenchen.de
www.ethikkommission.med.uni-muenchen.de

Anschrift:
Pettenkoferstr. 8a
D-80336 München

München, 22.04.2025/Vb/ce

Projekt Nr.: **25-0205** (bitte bei Schriftwechsel angeben)

Beratung nach Deklaration von Helsinki / § 15 Berufsordnung der Ärzte Bayerns / Beratung nach geltendem Fakultätsrecht

Studientitel	Durchführung und Evaluation eines Forschungsmethoden- und Biostatistikkurses für Allgemeinmediziner
Antragsteller:	Universität Augsburg, Lehrstuhl für Allgemeinmedizin, Herr Zekeriya Aktürk, Gutenbergstr. 7, 86356 Neusäß
Untersucher:	Marco Roos, Institut für Allgemeinmedizin, Gutenbergstr. 7, 86356 Neusäß

Sehr geehrter Herr Aktürk,

besten Dank für Ihr Schreiben vom 09.04.2025 mit der Beantwortung unserer Fragen bzw. Erfüllung der Auflagen und den noch ausstehenden bzw. überarbeiteten Unterlagen.

Die Ethikkommission (EK) kann Ihrer Studie nun die ethisch-rechtliche Unbedenklichkeit zuerkennen.

Das Votum ist für die Dauer der Studie gültig, längstens jedoch bis zum 21.04.2030 (5 Jahre).

Sofern ein Votum über diesen Zeitraum hinaus benötigt wird, bitten wir, der EK unaufgefordert einen Antrag auf Verlängerung des Votums mindestens 3 Monate vor Gültigkeitsende des Votums vorzulegen.

Vorsorglich möchte ich darauf hinweisen, dass auch bei einer positiven Beurteilung des Vorhabens durch die EK die ärztliche und juristische Verantwortung für die Durchführung des Projektes uneingeschränkt bei Ihnen und Ihren Mitarbeitern verbleibt.

Allgemeine Hinweise:

- Änderungen im Verlauf der Studie sind der EK zur erneuten Prüfung vorzulegen.
- Schwerwiegende unerwartete studienabhängige Ereignisse sind der EK mitzuteilen (trifft nur für interventionelle Projekte zu).
- Das Ende der Studie ist anzuzeigen und das Ergebnis vorzulegen.
- Die ärztliche und juristische Verantwortung bei der Durchführung der Studie verbleibt uneingeschränkt bei Ihnen und Ihren Mitarbeitern.
- Bitte berücksichtigen Sie, dass diese Bewertung die ggf. erforderliche Konsultation des behördlichen Datenschutzbeauftragten nach Art. 30 DSGVO nicht ersetzt und ein Eintrag in das Verfahrensverzeichnis der Klinik/des Instituts zu erfolgen hat.
- Die Ethikkommission erklärt, dass an der Bewertung des vorliegenden Antrags niemand beteiligt war, der gemäß Bayerischem Verwaltungsverfahrensgesetz (BayVwVfG) Art. 20 als befangen anzusehen ist.

Für Ihre Studie wünsche ich Ihnen viel Erfolg.

Mit freundlichen Grüßen

Huber

Prof. Dr. R. M. Huber
Vorsitzender der Ethikkommission

Vorgelegte Unterlagen:

1) Antragsformular.pdf vom 26.02.2025
2) Anschreiben.pdf vom 26.02.2025
3) Anhang1Fragebogen.pdf vom 26.02.2025
4) Anhang2Pretest.pdf vom 26.02.2025
5) Anschreiben.pdf vom 26.02.2025
6) Anschreiben.pdf vom 26.02.2025
7) Antragsbestätigung.pdf vom 26.02.2025
8) Studienprotokoll4ALLVersion2_02042025.pdf vom 08.04.2025
9) Anhang4Einwilligungserklaerung.pdf vom 08.04.2025
10) Anhang3COLLES.pdf vom 09.04.2025

Appendix 5: R-Script

R-Script

```
# R-Code for data analysis 4ALL study - Zekeriya Aktürk

rm(list = ls()) #remove all variables and datasets from the global environment

# Install (if needed) and load necessary packages

if (!require(readxl)) install.packages("readxl"); library(readxl)

if (!require(janitor)) install.packages("janitor"); library(janitor)

if (!require(dplyr)) install.packages("dplyr"); library(dplyr)

if (!require(tidyr)) install.packages("tidyr"); library(tidyr)

if (!require(ggplot2)) install.packages("ggplot2"); library(ggplot2)

if (!require(psych)) install.packages("psych"); library(psych)

setwd("C:/Users/ZekeriyaAktuerk/Downloads")

data <- read_excel("0_4ALLData_.xlsx")

# Define a named vector for the full variable names

var_labels <- c(

  Q01_F1Advice = "Heard from friends",

  Q01_F1EMail = "Heard via IAM email/newsletter",

  Q01_F1Socialmedia = "Heard via social media",

  Q01_F1Moodle = "Heard via Moodle",

  Q01_F1Other = "Heard via other means",

  Q05_F3Age = "Age",

  Q06_F4Sex0F1M = "Sex (0=F, 1=M)",

  Q07_F5Occupation = "Occupation (1=Doctor, 2=Resident, 3=Student)",

  Q09_F6JobExperience = "Job experience (years)",

  Q10_F7REsidencyStatus = "Residency status (1=In training, 2=Specialist, etc.)",

  Q12_F8Improve = "Motivation: Improve knowledge",

  Q12_F8Projects = "Motivation: Scientific projects",

  Q12_F8Interpret = "Motivation: Better interpretation",

  Q12_F8Preparation = "Motivation: Prepare for research",

  Q12_F8GeneralInterest = "Motivation: General interest",

  Q12_F8Other = "Motivation: Other",

  Q14_F9SelfRating = "Self-rated competence",

  Q15_F10Experience = "Experience in research",

  Q16_F11NumberPublications = "Number of publications",

  COLLES1_1 = "COLLES Module 1.1",

  COLLES1_2 = "COLLES Module 1.2",

  COLLES1_3 = "COLLES Module 1.3",

  COLLES1_4 = "COLLES Module 1.4",

  COLLES2_1 = "COLLES Module 2.1",

  COLLES2_2 = "COLLES Module 2.2",

  COLLES2_3 = "COLLES Module 2.3",

  PretestScore = "Pretest score",

  PosttestScore = "Posttest score"

)

# Convert variable types

data <- data %>%

  mutate(

    # Scale (numeric) variables

    Q05_F3Age = as.numeric(Q05_F3Age),

    Q09_F6JobExperience = as.numeric(Q09_F6JobExperience),

    Q16_F11NumberPublications = as.numeric(Q16_F11NumberPublications),

    COLLES1_1 = as.numeric(COLLES1_1),

    COLLES1_2 = as.numeric(COLLES1_2),

    COLLES1_3 = as.numeric(COLLES1_3),

    COLLES1_4 = as.numeric(COLLES1_4),

    COLLES2_1 = as.numeric(COLLES2_1),

    COLLES2_2 = as.numeric(COLLES2_2),

    COLLES2_3 = as.numeric(COLLES2_3),

    PretestScore = as.numeric(PretestScore),

    PosttestScore = as.numeric(PosttestScore),


    # Likert-scale variable (treated as ordinal, but numeric for simplicity)

    Q14_F9SelfRating = as.numeric(Q14_F9SelfRating),


    # Nominal (factor) variables

    Q07_F5Occupation = factor(Q07_F5Occupation,

                              levels = c(1, 2, 3),
```

```
                    labels = c("Doctor", "Resident", "Medical student")),
  Q10_F7REsidencyStatus = factor(Q10_F7REsidencyStatus,
                                 levels = c(1, 2, 3, 4),
                                 labels = c("Still in training", "Specialist", "No training planned", "Other")),

  # Dichotomous variables (logical or factor with yes/no)
  across(
    starts_with("Q01_F1->"),
    ~ factor(., levels = c(0, 1), labels = c("No", "Yes"))
  ),
  across(
    starts_with("Q12_F8->"),
    ~ factor(., levels = c(0, 1), labels = c("No", "Yes"))
  ),
  Q15_F10Experience = factor(Q15_F10Experience, levels = c(0, 1), labels = c("No", "Yes"))
 )

# Optional: Check structure
str(data)
summary(data)

library(readxl)
library(dplyr)

# Define analysis group: participants who have both Pretest and Posttest scores
data <- data %>%
 mutate(Analyzed = ifelse(!is.na(PretestScore) & !is.na(PosttestScore), "Analyzed", "Not analyzed"))

# Convert relevant variables
data$Q06_F4Sex0F1M <- factor(data$Q06_F4Sex0F1M, levels = c(0,1), labels = c("Female", "Male"))
data$Analyzed <- factor(data$Analyzed, levels = c("Analyzed", "Not analyzed"))

# Descriptive statistics
table(data$Analyzed)
summary(data$Q05_F3Age[data$Analyzed == "Analyzed"])
summary(data$Q05_F3Age[data$Analyzed == "Not analyzed"])

# Compare Age between groups (use Wilcoxon if not normal)
wilcox.test(Q05_F3Age ~ Analyzed, data = data)

# Compare Sex between groups
# Create the contingency table (Sex by Analyzed group)
table_sex <- table(data$Q06_F4Sex0F1M, data$Analyzed)
# Print the table with counts
print("Contingency Table (Counts):")
print(table_sex)
print("Contingency Table (Column Percentages):")
prop.table(table_sex, margin = 2)
# Perform Chi-squared test
print("Chi-squared Test Results:")
chisq.test(table_sex)
fisher.test(table_sex)

# Participant characteristics
if (!require(gtsummary)) install.packages("gtsummary"); library(gtsummary)
if (!require(gt)) install.packages("gt"); library(gt)

# Filter data to only include participants with Pretest and Posttest
data_analyzed <- data %>% filter(Analyzed == "Analyzed")

# Select relevant variables
vars_to_include <- c("Q05_F3Age", "Q06_F4Sex0F1M", "Q07_F5Occupation", "Q09_F6JobExperience", "Q15_F10Experience", "Q10_F7REsidencyStatus")

# Create a summary table using gtsummary
summary_table <- data_analyzed %>%
```

```
  select(all_of(vars_to_include)) %>%
  tbl_summary(
    statistic = list(all_continuous() ~ "{mean} ± {sd}", all_categorical() ~ "{n} ({p}%)"),
    digits = all_continuous() ~ 1,
    label = list(
      Q05_F3Age = "Age",
      Q06_F4Sex0F1M = "Sex",
      Q07_F5Occupation = "Occupation",
      Q09_F6JobExperience = "Job experience (years)",
      Q15_F10Experience = "Experience in research",
      Q10_F7REsidencyStatus = "Residency status"
    ),
    missing = "no"
  ) %>%
  bold_labels()

# Print the table to viewer pane and ready to export
summary_table %>% as_gt()

# Hearing about the course
data_analyzed <- data %>% filter(Analyzed == "Analyzed")

# Define variables related to how they heard about the course
hearing_vars <- c("Q01_F1Advice", "Q01_F1EMail", "Q01_F1Socialmedia", "Q01_F1Moodle", "Q01_F1Other")

# Create a formatted table
hearing_table <- data_analyzed %>%
  select(all_of(hearing_vars)) %>%
  tbl_summary(
    statistic = all_categorical() ~ "{n} ({p}%)",
    label = list(
      Q01_F1Advice = "Heard from friends",
      Q01_F1EMail = "Heard via IAM email/newsletter",
      Q01_F1Socialmedia = "Heard via social media",
      Q01_F1Moodle = "Heard via Moodle platform",
      Q01_F1Other = "Heard via other sources"
    ),
    missing = "no"
  ) %>%
  bold_labels()

hearing_table %>% as_gt()
motivation_vars <- c(
  "Q12_F8Improve",
  "Q12_F8Projects",
  "Q12_F8Interpret",
  "Q12_F8Preparation",
  "Q12_F8Generalnterest",
  "Q12_F8Other"
)
motivation_table <- data_analyzed %>%
  select(all_of(motivation_vars)) %>%
  tbl_summary(
    statistic = all_categorical() ~ "{n} ({p}%)",
    label = list(
      Q12_F8Improve = "Improve knowledge of statistics and research",
      Q12_F8Projects = "Support for projects or thesis",
      Q12_F8Interpret = "Better interpretation of studies/guidelines",
      Q12_F8Preparation = "Prepare for scientific activity",
      Q12_F8Generalnterest = "General interest in research",
      Q12_F8Other = "Other motivations"
    ),
    missing = "no"
  ) %>%
  bold_labels()
motivation_table %>% as_gt()
```

```
# Self rating and publications
data_analyzed <- data %>% filter(Analyzed == "Analyzed")
table_self_pub <- data_analyzed %>%
 select(Q14_F9SelfRating, Q16_F11NumberPublications) %>%
 tbl_summary(
  statistic = list(
   all_continuous() ~ "{mean} ± {sd}",
   all_categorical() ~ "{n} ({p}%)"
  ),
  digits = all_continuous() ~ 1,
  label = list(
   Q14_F9SelfRating = "Self-rated competence",
   Q16_F11NumberPublications = "Number of publications"
  ),
  missing = "no"
 ) %>%
 bold_labels()

# Show table in Viewer or export
table_self_pub %>% as_gt()

# COLLES scores
colles_vars <- c(
 "COLLES1_1", "COLLES1_2", "COLLES1_3",
"COLLES1_4",
 "COLLES2_1", "COLLES2_2", "COLLES2_3"
)

# Create labeled summary table with mean ± SD
colles_table <- data_analyzed %>%
 select(all_of(colles_vars)) %>%
 tbl_summary(
  statistic = all_continuous() ~ "{mean} ± {sd}",
  digits = all_continuous() ~ 2,
  label = list(
   COLLES1_1 = "COLLES Module 1.1",
   COLLES1_2 = "COLLES Module 1.2",
   COLLES1_3 = "COLLES Module 1.3",
   COLLES1_4 = "COLLES Module 1.4",
   COLLES2_1 = "COLLES Module 2.1",
   COLLES2_2 = "COLLES Module 2.2",
   COLLES2_3 = "COLLES Module 2.3"
  ),
  missing = "no"
 ) %>%
 bold_labels()

# Display table in Viewer
colles_table %>% as_gt()

# Pretest-posttest comparisons
# Reshape data to long format for paired comparison
prepost_long <- data_analyzed %>%
 select(ID, PretestScore, PosttestScore) %>%
 pivot_longer(cols = c(PretestScore, PosttestScore),
              names_to = "Time",
              values_to = "Score") %>%
 mutate(Time = recode(Time,
                      "PretestScore" = "Pretest",
                      "PosttestScore" = "Posttest"))
prepost_table <- prepost_long %>%
 tbl_summary(
  by = Time,
  statistic = all_continuous() ~ "{mean} ± {sd}",
  digits = all_continuous() ~ 2,
  label = list(Score = "Test Score"),
  missing = "no"
 ) %>%
 add_p(test = all_continuous() ~ "wilcox.test") %>%
 bold_labels()
prepost_table %>% as_gt()
```

```
# Bar graph of pretest vs. posttest scores
# if (!require(ggplot2)) install.packages("ggplot2"); library(ggplot2)
# if (!require(dplyr)) install.packages("dplyr"); library(dplyr)
# if (!require(tidyr)) install.packages("tidyr"); library(tidyr)

prepost_long <- data_analyzed %>%
  select(ID, PretestScore, PosttestScore) %>%
  pivot_longer(cols = c(PretestScore, PosttestScore),
               names_to = "Time",
               values_to = "Score") %>%
  mutate(Time = factor(recode(Time,
                              "PretestScore" = "Pretest",
                              "PosttestScore" = "Posttest"),
                       levels = c("Pretest", "Posttest")))
prepost_summary <- prepost_long %>%
  group_by(Time) %>%
  summarise(
    mean_score = mean(Score, na.rm = TRUE),
    sd_score = sd(Score, na.rm = TRUE),
    .groups = "drop"
  )

ggplot(prepost_summary, aes(x = Time, y = mean_score, fill = Time)) +
  geom_bar(stat = "identity", width = 0.5, color = "black") +
  geom_errorbar(aes(ymin = mean_score - sd_score, ymax = mean_score + sd_score),
                width = 0.2, size = 0.7) +
  labs(
    title = "Comparison of Pretest and Posttest Scores",
    x = "",
    y = "Mean Score ± SD"
  ) +
  theme_minimal(base_size = 14) +
  theme(legend.position = "none")
```

```
# Univariate analysis of pretest
# Load necessary packages
if (!require(dplyr)) install.packages("dplyr"); library(dplyr)
if (!require(gt)) install.packages("gt"); library(gt)
if (!require(readxl)) install.packages("readxl"); library(readxl)

# Load your Excel file
data <- readxl::read_excel("0_4ALLData_.xlsx")

# Filter for analyzed participants
data_analyzed <- data %>% filter(Analyzed == "Analyzed")

# Create grouped variables
data_grouped <- data_analyzed %>%
  mutate(
    AgeGroup = ifelse(Q05_F3Age <= 40, "≤40", ">40"),
    Sex = factor(Q06_F4Sex0F1M, levels = c(0, 1), labels = c("Female", "Male")),
    Occupation = factor(ifelse(Q07_F5Occupation == 1, "Doctor", "Other")),
    SelfRatingGroup = factor(ifelse(Q14_F9SelfRating < 4, "<4", "≥4")),
    JobExperienceGroup = factor(ifelse(Q09_F6JobExperience <= 10, "≤10", ">10")),
    ResearchExperience = factor(Q15_F10Experience, levels = c(0, 1), labels = c("No", "Yes"))
  )

# Define labels
group_vars <- list(
  AgeGroup = "Age (≤40 vs >40)",
  Sex = "Sex (Female vs Male)",
  Occupation = "Occupation (Doctor vs Other)",
  SelfRatingGroup = "Self-rated competence (<4 vs ≥4)",
  JobExperienceGroup = "Job experience (≤10 vs >10)",
  ResearchExperience = "Research experience (No vs Yes)"
```

```
)

# Generate rows for the Pretest table
table_rows <- purrr::map_dfr(names(group_vars), function(var) {
  label <- group_vars[[var]]

  df <- data_grouped %>%
    select(group = !!sym(var), PretestScore) %>%
    filter(!is.na(group), !is.na(PretestScore))

  if (n_distinct(df$group) != 2) return(NULL)

  stats <- df %>%
    group_by(group) %>%
    summarise(
      `Pretest Score Mean ± SD` = sprintf("%.1f ± %.1f", mean(PretestScore, na.rm = TRUE), sd(PretestScore, na.rm = TRUE)),
      .groups = "drop"
    )

  p_val <- wilcox.test(PretestScore ~ group, data = df)$p.value
  p_val_fmt <- ifelse(p_val < 0.001, "<0.001", sprintf("%.3f", p_val))

  stats$`p-value` <- ""
  stats$`p-value`[1] <- p_val_fmt
  stats$Characteristic <- label

  stats %>%
    select(Characteristic, Subgroup = group, `Pretest Score Mean ± SD`, `p-value`)
})

# Create the gt table
pretest_gt <- table_rows %>%
  gt() %>%
  cols_label(
    Characteristic = "Characteristic",
    Subgroup = "Group",
    `Pretest Score Mean ± SD` = "Mean ± SD",
    `p-value` = "p-value"
  ) %>%
  tab_spanner(label = "Subgroup", columns = c(Subgroup), id = "spanner_subgroup") %>%
  tab_spanner(label = "Pretest Score", columns = c(`Pretest Score Mean ± SD`), id = "spanner_pretest") %>%
  fmt_markdown(columns = everything())

# Print table
pretest_gt

#
# Univariate analysis of posttest
# Load necessary packages
if (!require(dplyr)) install.packages("dplyr"); library(dplyr)
if (!require(gt)) install.packages("gt"); library(gt)
if (!require(readxl)) install.packages("readxl"); library(readxl)

# Read the data (adjust path if needed)
data <- readxl::read_excel("0_4ALLData_.xlsx")

# Filter for analyzed participants
data_analyzed <- data %>% filter(Analyzed == "Analyzed")

# Create grouped variables
data_grouped <- data_analyzed %>%
  mutate(
    AgeGroup = ifelse(Q05_F3Age <= 40, "<=40", ">40"),
    SexGroup = factor(Q06_F4Sex0F1M, levels = c(0, 1), labels = c("Female", "Male")),
    OccupationGroup = factor(ifelse(Q07_F5Occupation == 1, "Doctor", "Other")),
    SelfRatingGroup = ifelse(Q14_F9SelfRating < 4, "<4", "≥4"),
```

```
  JobExperienceGroup = ifelse(Q09_F6JobExperience <= 10, "≤10", ">10"),
  ResearchExperienceGroup = factor(Q15_F10Experience, levels = c(0, 1), labels = c("No", "Yes"))
 )

# Define variable labels
group_vars <- list(
 AgeGroup = "Age (years)",
 SexGroup = "Sex",
 OccupationGroup = "Occupation",
 SelfRatingGroup = "Self-rated competence",
 JobExperienceGroup = "Job experience",
 ResearchExperienceGroup = "Research experience"
)

# Generate rows for the table
library(purrr)
table_rows <- purrr::map_dfr(names(group_vars), function(var) {
 label <- group_vars[[var]]

 df <- data_grouped %>%
  select(group = !!sym(var), PosttestScore = PosttestScore) %>%
  filter(!is.na(group), !is.na(PosttestScore))

 if (n_distinct(df$group) != 2) return(NULL)  # Only include binary comparisons

 # Compute mean ± SD
 stats <- df %>%
  group_by(group) %>%
  summarise(
   `Posttest Score Mean ± SD` = sprintf("%.1f ± %.1f", mean(PosttestScore), sd(PosttestScore)),
   .groups = "drop"
  )

 # Wilcoxon rank-sum test
 p_val <- wilcox.test(PosttestScore ~ group, data = df)$p.value
 p_val_fmt <- ifelse(p_val < 0.001, "<0.001", sprintf("%.3f", p_val))

 stats$`p-value` <- ""
 stats$`p-value`[1] <- p_val_fmt
 stats$Characteristic <- label

 stats %>% select(Characteristic, Subgroup = group, `Posttest Score Mean ± SD`, `p-value`)
})

# Format with gt
table_rows %>%
 gt() %>%
 cols_label(
  Characteristic = "Characteristic",
  Subgroup = "Subgroup",
  `Posttest Score Mean ± SD` = "Posttest Score Mean ± SD",
  `p-value` = "p-value"
 ) %>%
 tab_spanner(label = "Subgroup", columns = Subgroup, id = "spanner_subgroup") %>%
 tab_spanner(label = "Posttest Score", columns = `Posttest Score Mean ± SD`, id = "spanner_score") %>%
 fmt_markdown(columns = everything())


data_analyzed <- data %>% filter(Analyzed == "Analyzed")
summary(data_analyzed$Q14_F9SelfRating)
#
# Scatter graph of age and pretest/posttest scores
if (!require(ggplot2)) install.packages("ggplot2"); library(ggplot2)
if (!require(dplyr)) install.packages("dplyr"); library(dplyr)
if (!require(tidyr)) install.packages("tidyr"); library(tidyr)
```

```
# Filter analyzed participants
data_analyzed <- data %>% filter(Analyzed == "Analyzed")

# Prepare long-format data
score_long <- data_analyzed %>%
  select(ID, Q05_F3Age, PretestScore, PosttestScore) %>%
  pivot_longer(cols = c(PretestScore, PosttestScore),
               names_to = "TestTime",
               values_to = "Score") %>%
  mutate(TestTime = recode(TestTime,
                           "PretestScore" = "Pretest",
                           "PosttestScore" = "Posttest"))

# Create plot with points and interpolation (trend) lines
ggplot(score_long, aes(x = Q05_F3Age, y = Score, color = TestTime)) +
  geom_point(position = position_dodge(width = 0.6), size = 3, alpha = 0.8) +
  geom_smooth(method = "loess", se = TRUE, formula = y ~ x, size = 1.2, alpha = 0.2) +
  scale_color_manual(values = c("Pretest" = "blue", "Posttest" = "darkgreen")) +
  labs(
    title = "Pretest and Posttest Scores by Age",
    x = "Age",
    y = "Score",
    color = "Test"
  ) +
  theme_minimal(base_size = 14)


# Polynomial regression pretest scores
if (!require(gtsummary)) install.packages("gtsummary"); library(gtsummary)
if (!require(gt)) install.packages("gt"); library(gt)

# Polynomial regression model
model_pre <- lm(PretestScore ~ Q05_F3Age + I(Q05_F3Age^2), data = data_analyzed)

# Create elegant regression table
regression_table <- tbl_regression(
  model_pre,
  intercept = TRUE,
  estimate_fun = ~style_sigfig(.x, digits = 2),
  label = list(
    `Q05_F3Age` = "Age",
    `I(Q05_F3Age^2)` = "Age²"
  )
) %>%
  add_glance_table() %>%  # No 'statistic =' here
  bold_labels()

# Show as gt table
regression_table %>% as_gt()


# Polynomial regression posttest scores
if (!require(gtsummary)) install.packages("gtsummary"); library(gtsummary)
if (!require(gt)) install.packages("gt"); library(gt)

# Polynomial regression model
model_pre <- lm(PosttestScore ~ Q05_F3Age + I(Q05_F3Age^2), data = data_analyzed)

# Create elegant regression table
regression_table <- tbl_regression(
  model_pre,
  intercept = TRUE,
  estimate_fun = ~style_sigfig(.x, digits = 2),
  label = list(
    `Q05_F3Age` = "Age",
```

```
  `I(Q05_F3Age^2)` = "Age²"
 )
) %>%
 add_glance_table() %>%  # No 'statistic =' here
 bold_labels()

# Show as gt table
regression_table %>% as_gt()

#
#
# lmer
if (!require(lme4)) install.packages("lme4")
if (!require(tidyr)) install.packages("tidyr")
if (!require(dplyr)) install.packages("dplyr")

library(lme4)
library(tidyr)
library(dplyr)

# Prepare the data in long format
data_long <- data_analyzed %>%
 mutate(ParticipantID = row_number()) %>%  # create ID first
 select(ParticipantID, Q05_F3Age, Q06_F4Sex0F1M, Q07_F5Occupation,
        Q14_F9SelfRating, Q15_F10Experience,
        PretestScore, PosttestScore) %>%
 pivot_longer(cols = c(PretestScore, PosttestScore),
              names_to = "Time", values_to = "Score") %>%
 mutate(
   Time = factor(ifelse(Time == "PretestScore", "Pre", "Post"), levels = c("Pre", "Post")),
   Sex = factor(Q06_F4Sex0F1M, levels = c(0, 1), labels = c("Female", "Male")),
   Occupation = factor(ifelse(Q07_F5Occupation == 1, "Doctor", "Other")),
   SelfRatedGroup = factor(ifelse(Q14_F9SelfRating < 4, "<4", "≥4")),
   ResearchExperience = factor(Q15_F10Experience, levels = c(0, 1), labels = c("No", "Yes"))
 )

# Fit the parsimonious linear mixed effects model
model_lmm <- lmer(
 Score ~ Time * (Q05_F3Age + Sex + Occupation + SelfRatedGroup + ResearchExperience) +
   (1 | ParticipantID),
 data = data_long
)

# Output model summary
summary(model_lmm)
confint(model_lmm, level = 0.95, method = "profile")
## Ende
```

Appendix 6: Directed Acyclic Graph Model Code

DAG Model Code

The model code for dagitty.net (https://www.dagitty.net/dags.html) is given below:

```
dag {
"Knowledge Gain" [outcome,pos="0.184,0.251"]
"Research experience" [pos="-0.015,-0.885"]
"Self-rated Competence" [pos="-1.231,-1.110"]
"Time (pre/post)" [exposure,pos="-1.241,0.220"]
Age [pos="-0.292,-0.490"]
Occupation [pos="0.149,-1.094"]
Sex [pos="-0.162,-0.691"]
"Research experience" -> "Knowledge Gain"
"Research experience" -> "Self-rated Competence"
"Self-rated Competence" -> "Knowledge Gain"
"Time (pre/post)" -> "Knowledge Gain"
Age -> "Knowledge Gain"
Age -> "Self-rated Competence"
Occupation -> "Knowledge Gain"
Occupation -> "Self-rated Competence"
Sex -> "Knowledge Gain"
Sex -> "Self-rated Competence"
}
```

Appendix 7: Specification of individual contributions

Contributions

Effort	**Individual Contribution (%)**	**Contributions by Others (%)**
Conception	Zekeriya Aktürk (95%)	Raphael Kunisch (5%)
Course videos	Zekeriya Aktürk (90%)	Michaela Trompke (5%) Katherina Zeiser (5%)
Data evaluation	Zekeriya Aktürk (100%)	--
Dissertation	Zekeriya Aktürk (100%)	--

Appendix 8: Declaration of independent work

Declaration of independent work

Declaration of Independent Work

I hereby declare that I have written the present master's thesis entitled "Implementation and Assessment of a Research Methods and Biostatistics Course for General Practitioners" independently and without unauthorized assistance.

All sources and resources used in the preparation of this thesis have been properly cited and acknowledged. I have not submitted this thesis in the same or a similar form for assessment in any other examination or academic context.

Dinkelscherben, 24.07.2025

Zekeriya Aktürk

Appendix 9: Curriculum Vitae

Curriculum Vitae

Personal Information

First name:	Zekeriya
Last name:	Aktürk
Date of birth:	28.06.1968
Nationality:	Turkish
Marital status:	Married, two children
E-mail:	zekeriya.akturk@gmail.com
Web:	http://www.aile.net
YouTube:	https://www.youtube.com/zekeriyaakturk
Twitter:	https://x.com/zekeriyaakturk
Postal address:	Kohlstattstr. 16, 86424 Dinkelscherben, Germany

Education

- June 2000-June 2001
  Academic Development Fellowship Program, Virginia Commonwealth University Department of Family Practice, Richmond, USA
- April 1994-August 1998
  Postgraduate education in family medicine, Haydarpasa State Hospital, Istanbul, Turkey
- October 1985-March 1992
  Undergraduate student in Marmara University Faculty of Medicine, Istanbul, Turkey
- 1984-1985
  High school education, Ankara Dikmen Lisesi, Ankara, Turkey
- 1980-1984
  Middle school education, Hauptschule Uellendahl, 6500 Wuppertal, Germany
- 1975-1980
  Primary school education, Oymali 100. Yıl Ilkokulu, Trabzon, Turkey

Work Experience

From	To	Details
May 1992	August 1993	General Practitioner, Kömürlü Primary Care Health Clinic, Erzurum, Turkey
August 1993	March 1994	Research Assistant in Histology and Embryology, Atatürk University, Erzurum, Turkey
April 1994	January 1999	Resident trainee in Family Medicine, Haydarpasa Training Hospital, Istanbul, Turkey
February 1999	September 1999	Specialist in General Practice, Kandira Mother-Child Health Center, Kocaeli, Turkey
October 1999	June 2005	Assistant Professor, Department of Family Medicine, Trakya University, Turkey
June 2005	June 2006	Associate Professor, Department of Family Medicine, Trakya University, Turkey
June 2006	August 2009	Consultant, Primary Care Education Center, Ministry of Health, Riyadh, Saudi Arabia
August 2009	January 2010	Associate Professor, Department of Family Medicine, Atatürk University, Erzurum, Turkey
January 2010	August 2015	Professor, Department of Family Medicine, Atatürk University, Erzurum, Turkey
August 2015	July 2016	Professor, Department of Family Medicine, Sifa University, Izmir, Turkey
September 2016	November 2017	Political prisoner, Turkey
December 2016	November 2020	Independent Academic Consultant (www.aile.net)
November 2020	June 2023	Research Associate, Institute of General Practice and Health Services Research, Technical University of Munich
January 2023	Present	Research Associate, Institute of General Practice, University of Augsburg
July 2023	Present	Specialist in General Practice, Dr. Assefa Practice, 86368 Gersthofen

Hatice Dolduran, Hatice'nin Düşleri, https://www.instagram.com/p/BeTd0S6BZg_/?igsh=MXZ4azJkdzlqenZx