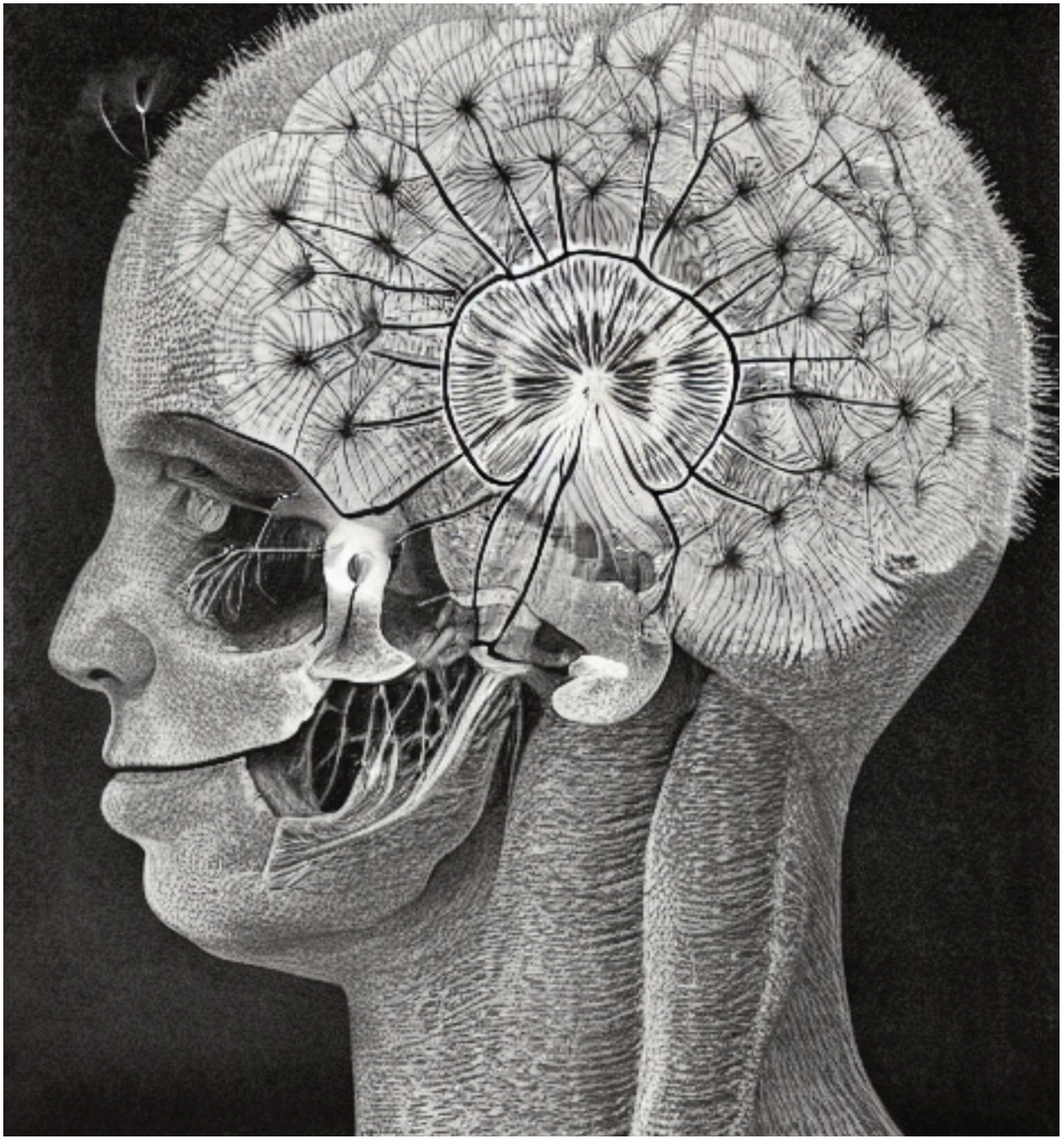
MSRJ™
Medical Student Research Journal

Spring, 2023



MICHIGAN STATE UNIVERSITY | College of Human Medicine

MSRJ

Michigan State University College of Human Medicine

Vol: 10, Issue: Spring, 2023

The Medical Student Research Journal (MSRJ) is the longest-running international academic journal in the United States authored, reviewed, edited, and published by medical students for medical students. It is dedicated to promoting the scientific achievements of medical students, teaching principles of peer and article review, and providing editorial, publishing and leadership learning experiences. Medical students worldwide are invited to submit manuscripts and serve as trained reviewers. The MSRJ publishes original research, case studies, editorials, research letters, reviews, and reflections that meet required standards, are authored by a medical student, and advance science. It is sponsored by the Michigan State University College of Human Medicine. Visit www.msrj.org for more information.

COVER ARTIST

The origin of consciousness as an emergent property: sketch of reticular activating system projecting into cerebral cortex taking the form of dandelion pappi [Latin *pappus*; old man].

The wind will scatter our thoughts over the years, seed by seed, until we've declined beyond the mantle of introspect. Here, entropy is our friend and vessel. Those seeds of your knowledge - of your essence and being - will blossom elsewhere, and your consciousness will find new life.

Moustafa Hadi, Class of 2024 College of Human Medicine, Michigan State University

SPECIAL THANKS

Aron Sousa, M.D., Dean, College of Human Medicine
Andrea Wendling, M.D., Senior Associate Dean for Academic Affairs
Walter Esselman, Ph.D., Senior Associate Dean for Research
Michigan State University, College of Human Medicine, Office of Research
Datapage.ie

MICHIGAN STATE UNIVERSITY | College of Human Medicine

MSRJ

Michigan State University College of Human Medicine

Vol: 10 Issue: Spring, 2023 **Pages 01–56**

Spring, 2023

ORIGINAL RESEARCH

CASE STUDIES

REVIEWS

An Assessment of Knee Flexion in Lateral Knee X-rays

Christopher W. Keyes[1*], Lodewijk van Holsbeeck[2], Jean M Kerver[3], Chi Chang[3,4]

[1]Michigan State University College of Human Medicine, East Lansing, Michigan, USA
[2]Ascension St. Mary's Hospital, Saginaw, Michigan, USA
[3]Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, USA
[4]Office of Medical Education Research and Development, College of Human Medicine, Michigan State University, USA

Corresponding Author: Christopher W. Keyes; keyesch2@msu.edu

Purpose: *Patient positioning plays a crucial role in the field of radiology. Lateral knee X-rays are a type of image that often has incorrect positioning of the angle of knee flexion. The ideal range is between 20 and 30 degrees. The goal of this study was to assess the angle of knee flexion at two different locations in a single hospital system while determining if several variables influence the angle.*

Method: *This study is a retrospective chart review that assessed the angle of knee flexion in patients 18 years or older that underwent a lateral-mediolateral knee X-ray taken at an urgent care center and a general diagnostic center of a hospital within the same system between March 1 and December 1, 2021. Variables including age, sex, BMI, technologist, and location were collected from these patients' charts and evaluated. MRI information was gathered for patients who underwent an MRI within 30 days of a lateral knee X-ray. The research team assessed effusions reported on X-ray compared to effusions reported on MRI for these patients.*

Results: *Among patients included in the study (n = 665) the average angle of knee flexion was 51.28 degrees. Age, sex, BMI, and location were not significantly associated with the mean angle of knee flexion with p-values of 0.63, 0.13, 0.55, and 0.15 respectively. The radiology technologist taking the image did have an association with the angle of knee flexion with a p-value of 0.001. Differences in the mean angle of knee flexion between the groups of X-rays with effusions reported compared to the groups of X-rays where effusions were not reported but found on MRI resulted in a p-value of 0.83.*

Conclusions: *The technologist taking the image was the only variable of this study that had a significant difference in mean angle of knee flexion. Additional studies are needed to determine what technologist factors are most important in determining the angle of knee flexion. Using MRI information to evaluate if effusions were not reported due to the angle of knee flexion was limited in this study due to small sample size.*

Keywords: *radiological positioning; angle of knee flexion; radiology; X-ray interpretation*

INTRODUCTION

Radiology plays an important role in the care of patients. Physicians often turn to imaging for assistance with diagnosis and treatment. There are many factors that go into the production and interpretation of a radiological study. The radiology technologists have a crucial part in the production of these images. These individuals are tasked with providing radiologists with the images to diagnose patients. Factors that need to be considered by the radiology technologist include type of image, body part being imaged, flexion, extension, rotation, comfort, ability of the patient, and much more. Some of the images, such as the lateral-mediolateral knee X-ray, can be difficult to obtain by the technologists. The difficulty with this particular image is positioning. The knee needs to be flexed, rotated, and aligned properly.[1]

When evaluating a lateral-mediolateral knee X-ray, the quality of the image is of high importance. Specifically, the angle of knee flexion should be between 20 and 30 degrees.[2,3] According to studies assessing different knee anatomy,[4,5] lateral knee imaging rotated as little as 5 degrees off-axis from true lateral can have a significant effect including misreading important pathology. In addition to rotation, when the knee is hyperflexed, it can decrease the view of the suprapatellar fat pads causing decreased visualization of effusions and possible missed pathology.[6]

This project aimed to evaluate the lateral-mediolateral knee X-rays at a healthcare center. The research team looked to assess the angles of knee flexion as well as the possible effects of age, sex, BMI, technologist taking the image, and location had on the angle of flexion. The team also attempted to compare X-ray and MRI

reports to assess whether knee effusions could be underreported due to the angle of knee flexion.

METHODS

This study retrospectively collected data from patients' charts who were 18 or older and underwent a lateral-mediolateral knee X-ray taken at Munson Medical Center and Foster Family Community Health Center in Traverse City, Michigan, between March 1, 2021 and December 1, 2021. Patients with both knees imaged had information collected on both images and were then randomized with a random number generator to only include either their left or right knee. This prevented any 'double counting' of any patient or variable. Only patients with BMI included in their chart were included for the BMI analysis. Only BMI measurements at the time of X-ray were included. The BMI variable was filtered and divided into categories BMI ≤ 18.5, 18.6–24.9, 25–29.9, 30–34.9, and ≥35. These ranges were chosen because they are the standard BMI categories indicating underweight, normal weight, overweight, obese, and morbidly (extremely) obese. The age variable was filtered and divided into categories 18–29, 30–39, 40–49, 50–59, 60–69, 70–79, and 80+. Information on the angle of knee flexion, age, sex, BMI, technologist, and location was collected. F-tests were used to determine if equal variances could be assumed for sex and location. Bartlett's tests of equal variances were used to determine if equal variances could be assumed for age, BMI, and technologists. T-tests were used to assess differences of means for sex, location, and effusions reported. ANOVA tests were used to determine mean differences for age and BMI. A Brown-Forsythe test was used to determine mean differences for technologists. The angle of knee flexion was measured by a single medical student on the team who was instructed on proper technique by a radiologist in order to assure consistency in measurement. A radiology reading application, Inteleconnect, was used to measure the angle of knee flexion as seen in Fig. (1). The angle shown in Fig. (1) was recorded and its supplementary angle was calculated by taking the measured angle and subtracting it from 180 degrees. The supplementary angle was used for the purposes of this study. To decrease chance of bias in measurement, the document used to collect the angle of knee flexion data was separate from the document used to collect variable information. Any patient identifying information was kept on a separate, password-protected spreadsheet, and each patient was assigned a unique patient identifying number.

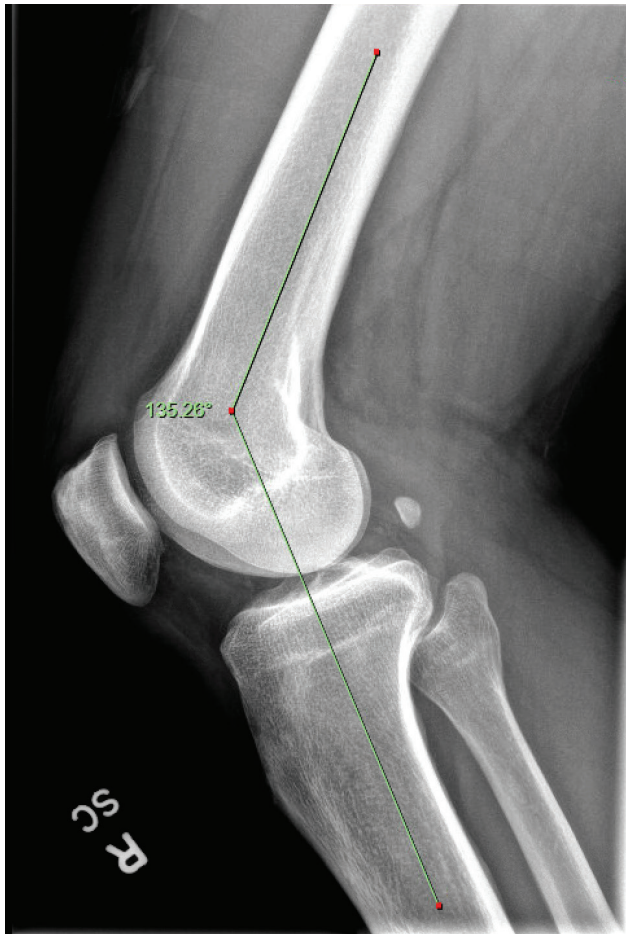


Figure 1. Measuring the angle of knee flexion. Example of the radiology application measurement on a lateral knee X-ray. The green line indicates the angle of knee flexion with the value calculated through the application.

MRI information that followed a lateral-mediolateral knee X-ray within 30 days or less was also collected. The research team attempted to evaluate if effusions were underreported on X-rays due to improper angle of knee flexion by comparing them to the reports of MRIs on the same patient. The patients who had effusions reported on MRI but not X-ray were considered to have effusions 'missed' on X-ray. The patients who had effusions reported on X-ray and MRI were considered to have the effusion properly reported. The patients who did not have effusions reported on either X-ray or MRI were considered to be effusion-free and excluded from further analysis. An F-test was used to determine if equal variances could be assumed for reported effusions. A T-test was used to assess differences of means between the

categories of 'missed' effusions to the properly reported effusions.

RESULTS

For this study, 665 patients were found to fit the inclusion criteria. Of these 665 patients, 105 had both knees imaged. For these patients, only one knee was selected to be included by a random number generator as mentioned in the methods section. Summaries for the variable's demographics can be seen in the appendix Tables (A1–A3).

Looking at the angle of knee flexion for all patients measured, there was a range of 12.54–89.79 degrees with a mean angle of 51.28 degrees and a median of 51.27 degrees. The summary statistics for the angle of knee flexion can be seen in the appendix in Table A1. See Fig. (2) for the distribution of the angle of knee flexion.

For age, variance testing showed a *p*-value of 0.68 meaning equal variance could be assumed. Statistical testing revealed that there was not a statistically significant difference in the mean angle of knee flexion between age groups $p = 0.63$. Refer to Table 1 for age group information.

For male sex versus female sex, variance testing showed a *p*-value of 0.05 meaning equal variances could not be assumed. Statistical testing resulted in a *p*-value of 0.13 meaning there was no statistically significant difference between sex and mean angle of knee flexion. Refer to Table 2 for sex group information.

For BMI, 296 patients were included in the statistical analysis. There were 396 patients who did not have BMI included in their charts. Variance testing showed a *p*-value of 0.51 meaning equal variance could be assumed. Statistical testing revealed that there was not a statistically significant difference in the mean angle of knee flexion between BMI groups $p = 0.76$. Refer to Table 3 for BMI group information.

For analysis of technologists taking the image, there were 43 technologists that took images. For the purpose of this variable assessment, only the technologists with 10 or more measurements were considered, resulting in analysis of 12 technologists. Variance testing showed a *p*-value of 0.00024 meaning equal variance could not be assumed. Statistical testing resulted in a *p*-value of 0.001 meaning that there was a statistically significant difference in the mean angle of knee flexion between technologists. Refer to Table 4 for technologist measurement information.

With respect to the location that the images were taken at, variance testing showed a *p*-value of 0.33 meaning that equal variances could be assumed.

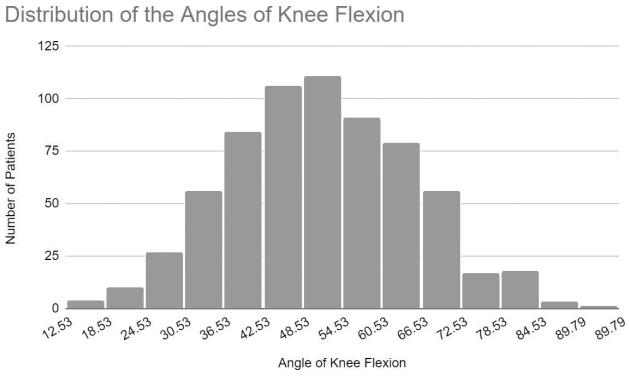


Figure 2. Distribution of the angles of knee flexion. Histogram of the angle of knee flexion of all measured patients.

Table 1. Age analysis. Variable analysis of age group on mean angle of knee flexion. Differences in mean $p = 0.63$

Age group	Mean angle of knee flexion (in degrees)	Standard deviation
18–29	48.11	16.28
30–39	49.02	15.14
40–49	50.47	12.80
50–59	51.52	13.05
60–69	51.79	13.49
70–79	51.34	13.85
80+	53.64	14.33

Table 2. Sex analysis. Variable analysis of sex on mean angle of knee flexion. Differences in mean $p = 0.13$.

Sex group	Mean angle of knee flexion (in degrees)	Standard deviation
Male	50.36	14.55
Female	52.02	13.09

Table 3. BMI analysis. Variable analysis of BMI on mean angle of knee flexion. Differences in mean $p = 0.55$.

BMI group	Mean angle of knee flexion (in degrees)	Standard deviation
<18.5	59.10	22.21
18.6–24.9	49.87	14.74
25–29.9	52.91	15.28
30–34.9	50.53	13.56
>35	51.90	13.45

Table 4. Technologist analysis. Analysis of technologists on mean angle of knee flexion. Differences in mean $p = 0.001$.

Technologist	Mean angle of knee flexion (in degrees)	Standard deviation
Tech 1	60.43	11.90
Tech 2	50.47	13.08
Tech 3	62.50	10.96
Tech 5	38.49	10.74
Tech 7	53.84	15.55
Tech 9	54.29	14.72
Tech 11	51.27	9.63
Tech 13	49.79	13.05
Tech 14	41.30	9.90
Tech 15	51.67	15.91
Tech 18	46.32	8.07
Tech 19	51.39	10.19

Table 5. Location analysis. Variable analysis of location on mean angle of knee flexion. Differences in mean $p = 0.15$.

Location	Mean angle of knee flexion (in degrees)	Standard deviation
Hospital	52.20	13.31
Urgent care	50.63	14.04

Statistical testing revealed a *p*-value of 0.15 meaning there was no statistically significant difference between locations and mean angle of knee flexion. Refer to Table 5 for location information.

When comparing effusion reporting with MRI and X-ray, there were 13 patients who fit the criteria of 'missed' effusions and 21 patients who fit the criteria of reported effusions. Comparing the angle of knee flexion in groups of effusions reported on both X-ray and MRI (reported effusions) to effusions reported on MRI but not X-ray ('missed' effusions), variance testing showed a *p*-value of 0.29 meaning that equal variances could be assumed. Statistical testing resulted in a *p*-value of 0.83 meaning there was no statistically significant difference between reported effusion groups and mean angle of knee flexion. Refer to Table 6 for reported effusion group information.

Table 6. Effusion analysis. Effusions reported on MRI but not X-ray considered 'missed' while effusions reported on both X-ray and MRI were reported appropriately. Differences in mean $p = 0.83$.

Effusion report	Mean angle of knee flexion (in degrees)	Standard deviation
Effusion 'Missed' on X-ray	52.00	15.41
Effusion reported on X-ray	51.00	11.88

DISCUSSION

Overall, the angle of knee flexion in these patients showed a distribution along a standard curve with the mean falling well above the accepted angle 20–30 degrees. This could be due to several factors, only some of which were measured in this study. Since this study found the radiology technologist measurements to have a statistically significant difference between mean angles measured, the variables involving the technologists are likely to have the largest impact. Some variables that were not included in this study that could be researched further include technologist age, training level, amount of work experience, and areas of imaging experience.

None of the possible patient-centered variables (age, sex, and BMI) measured in this study were statistically significant. One possible explanation for this is patient willingness to participate in their care. Most patients will position themselves however the technologist asks them to do if they are able. Some of these variables were considering that the patient might not be able to flex their knee to the proper angle due to BMI or age, for example, but these did not seem to be a factor to a significant degree.

The final variable assessed in this study was location. This did not appear to have a significant association to the angle of knee flexion either. Possible explanations for this would be that both the urgent care and hospital are in the same hospital network within the same city. They use the same training systems for technologists. In addition to this, some technologists had recorded images at both locations. Though this may introduce some bias, this variable was mainly intended to assess differences in equipment and procedure at each location. It would be useful for future studies to assess these variables at different facilities in several different hospital systems and cities. It is likely that the results will be similar to this study with the technologist having the largest impact on angle of knee flexion with other variables having minimal, if any, significance. This study is generalizable because it includes a large sample size, has a variety of ages and BMIs with a ratio of sexes being 1.33/1. Though ethnicity was not included in this study, it is unlikely that it would impact measurements.

There are several limitations to consider for this study. Firstly, errors could have been made in measurement of the angles of knee flexion. This error was limited by having a single researcher measure all of the angles and to have this researcher trained on measurements of this angle by a radiologist. Secondly, BMI was not recorded in all the patient's charts, so some of the information regarding this variable could be biased to patients who were hospitalized at one point in time resulting in this measurement being in the chart. There are two primary limitations to the evaluation of effusions between X-ray and MRI. There was a small number of patients who had an MRI follow-up within 30 days with reported effusions – a total of 21 in the group where effusions were reported on both X-ray and MRI and 13 in the group where it was not reported on X-ray but was reported on MRI. This is a very small sample size, and it does not have the power to determine significance. In addition to this, effusions reported on X-ray can be very subjective to the radiologist dictating the report and may not be reported if the radiologist does not determine the effusion to be of significance.

Some considerations to take away from this study are that the angle of knee flexion varied from technologist to technologist, but overall, the knee joint was hyperflexed. A question that needs to be researched further is 'Does the angle of knee flexion matter clinically?' Theoretically, it has been taught that the angle of flexion is important to fully visualize certain anatomy and pathologies. However, further research needs to be conducted as to if knee flexion influences a radiologist's interpretation of these images, and if so, to what degree. Other future research can be directed toward correcting the discrepancies in knee flexion angles. Creating a tool that technologists can use to quickly and effectively position the patient's knee to the appropriate angle could be implemented. In addition to this, refresher courses could be implemented into continuing education training for radiologic technologists to remind them on important positioning as well as common positioning errors.

CONCLUSION

According to prior research and radiology literature, patient positioning plays a crucial part of producing quality images for radiologists to read. Errors in positioning can lead to images of poor quality and possibly missed pathologies. In lateral knee X-rays, rotation and flexion of the knee are two critical components in producing the image. The research team set out to assess the angle of knee flexion in images produced at two locations in a large rural health system. After analysis, it was found that the angle of knee flexion was significantly different from the ideal range of 20–30 degrees. Several variables were measured to evaluate possible correlations for this discrepancy in angles. Of these variables, the radiology technologist taking the image was the only variable that had a significant difference in the mean angle of knee flexion. Further research looking into whether the angle of knee flexion influences the report generated by the radiologist, possibly with 'missed diagnoses', such as effusions, needs to be performed. In addition to this, a tool to allow radiology technologists quick and accurate measurements of knee flexion angle could be explored along with a refresher course on positioning.

Institutional Review Board Statement

Our Institutional Review Board approved the review of the patients' charts in this study and waived the need for consent.

Financial Support

None. The authors did not receive grant or outside funding in support of their research or preparation of this manuscript. They did not receive payment or any benefits from commercial entities.

Conflicts of Interest

None. The authors were not compensated or funded in any way for the preparation of this manuscript. This study has not been submitted elsewhere. We understand and agree that if the manuscript is accepted for publication, copyright in the article, including the right to reproduce the article in all forms and media, shall be assigned to the publisher.

REFERENCES

1. Martensen KM. A consistent method to produce lateral knee radiographs. Radiol Technol 1990; 62(1):24–7.
2. Tai AW, Alparslan HL, Townsend BA, Oei TN, Govindarajulu US, Aliabadi P, et al. Accuracy of cross-table lateral knee radiography for evaluation of joint effusions. Am J Roentgenol 2009; 193(4):W339–44. doi: 10.2214/AJR.09.2562
3. Thakur AJ. Knee. Strategic Radiographic positioning. Philadelphia: Wolters Kluwer Health; 2010.
4. Ziegler CG, Fulkerson JP. & Edgar C. Radiographic reference points are inaccurate with and without a true lateral

radiograph: The importance of anatomy in medial patellofemoral ligament reconstruction. Am J Sports Med 2016; 44(1):133–42. doi: 10.1177/0363546515611652
5. Huddleston HP, Redondo ML, Cregar WM, Christian DR, Hannon CP, Yanke AB. The Effect of Aberrant Rotation on Radiographic Patellar Height Measurement Using Canton-Deschamps Index: A Cadaveric Analysis. J Knee Surg 2023;36(3):254–260. doi: 10.1055/s-0041-1731720.
6. Martensen KM. Radiographic image analysis. 2nd ed. Philadelphia: Elsevier Saunders; 2006, pp. 317–24.

APPENDIX

Table A1. Study demographics. Highlighting the summary statistics of angle of knee flexion and variables patient age and BMI.

Statistical measurment	Angle of knee flexion (in degrees)	Patient age	BMI
Mean	51.28	61.06	30.84
Median	51.27	64	29.7
Standard deviation	13.73	14.82	7.17
Range	12.53–89.79	18–95	15.2–67.6

Table A2. Patient sex demographics. Study demographics of patient population regarding the variable of sex.

Sex	Number of patients
Male	378
Female	285
Male:Female ratio	1.33/1

Table A3. Location demographics. Study demographics regarding the variable of location.

Location	Number of patients
Hospital	286
Urgent care	377
Hospital: Urgent care ratio	0.76/1

The Impact of Specific Teaching Methods on Communication and History Taking in Second-Year Medical Students

Aadrika Kashyap, Jignesh Gandhi, Pravin Shinde*

Department of General Surgery, Seth GSMC and KEMH, Mumbai, India

*Corresponding Author: Pravin Shinde; pravinshinde@kem.edu

***Introduction*:** *This study aims to assess the impact of various teaching methods including role play, didactic lectures, and case studies on the history taking and communication skills of second-year Bachelor of Medicine and Surgery (MBBS) students. The goal is to help students become better doctors by arriving at diagnoses quicker by asking relevant questions in their history taking. A secondary goal is to improve the doctor–patient relationship through better communication skills.*

***Methods*:** *The students were assessed on their history taking and communication skills before and after the application of specific teaching methods. The teaching methods were chosen according to efficacy and impact as shown by other research articles, in addition to the convenience of applying them to our study and the curriculum of similar schools. The improvement was scored by the faculty at KEM Hospital in Mumbai, India, where the study was conducted, using a checklist that includes the main aspects of communication and general history taking. We tested the students on their communication skills, completeness of their history taking with regard to the history of the presenting illness, history of past illnesses, personal history, family history, and mental status report. The results of the pre- and post-intervention scores were analyzed using paired t-tests.*

***Results*:** *Fifteen students were assessed in this study. The results showed improvement in their mean scores after the teaching methods were applied. Using the student t-test, we statistically analyzed the students pre- and post-intervention. The p-value was found to be statistically significant (<0.05) in communication skills, completeness of their history taking with regard to the history of the presenting illness, history of past illnesses, family history, and mental status report. It was found to be non-significant with regard to personal history taking.*

***Conclusions*:** *The students benefited from the teaching sessions conducted during their surgical rotations. Applying these teaching tools helped students come to diagnoses better through history taking alone. Their communication skills were also found to be significantly improved, which has shown to positively impact physician–patient rapport and treatment compliance. We have concluded that it would be meaningful to incorporate these teaching tools in the curriculum of second-year undergraduate students with the goal of making them better physicians in the future.*

Keywords: *undergraduate; teaching; history taking; communication; medical education; medical student*

INTRODUCTION

History taking involves listening to patients' concerns and asking patients relevant questions with the purpose of coming to a diagnosis. Many diseases can be diagnosed solely through history taking and clinical examination alone.[1,2] Hence, it is vital to hone students' history taking skills before they become practicing doctors. This can avoid unnecessary tests, which are often costly and sometimes risky.[3,4] Taking a good history is an important skill for all medical students to learn; however, teaching methods are sometimes lacking due to the sparsity of time or resources. In this research study, we aim to quantify student improvement in history taking skills after the implementation of specific teaching methods chosen through a literature review.

Communication skills are also very important and are often overlooked.[5] Connecting with the patient and establishing a rapport is vital in eliciting key history elements.[6] Studies show that good communication can positively impact the patient's outcome, adherence to treatment, and overall satisfaction.[7] The learning curve of these soft skills can be steep, and some doctors remain poor communicators throughout their career. This can lead to poor patient outcomes that may otherwise be preventable.[8] We hypothesise that this can be avoided by incorporating certain teaching tools in the medical curriculum of undergraduate students.

In KEM Hospital, Mumbai, where we have conducted our study, second-year students are posted in various departments including Surgery, Medicine, Pediatrics

and Obstetrics and Gynecology. During this time, the role of the student is to observe patients while attending rounds with the senior doctor and residents, which is the practice in most hospitals in Mumbai. Students are encouraged to take patient histories, perform clinical examinations, and ask the residents any questions they might have. However, they are not required to gather information by rote or to demonstrate hypothesis generation and testing. In this study, we plan to implement specific teaching tools during the surgical postings of second-year students to aim to improve student history taking and communication skills.

METHODS

Study Design

The study was a prospective cohort study in a group of 15 students who were attending their second MBBS surgical clinical postings at KEM Hospital, Mumbai, from September 2019 to November 2019. The students were assessed on their history taking and communication skills by the faculty members after 2 weeks and 8 weeks of their surgical posting. The 2-week observation served as the 'before' measure. During the 2 weeks before the intervention, the students attended normal ward rounds, a practice that is followed during postings in all schools in Mumbai. At the 2-week mark, students were tested according to a checklist to evaluate their communication and history taking skills. After this, the teaching methods selected based on research and the convenience of incorporating them into the curriculum were used, and the students were assessed at the 8-week mark of their postings. The checklist used to assess students, shown in Table 1, was derived using sources including the 'Simplified Checklist of Calgary Cambridge Guide' for assessment of communication, and guidebooks like the 'PCM Guidebook for History taking and Physical Exams' for history taking.

Ten faculty members assessed the 15 students. To assure the faculty members assessed the students in a similar way, there were different assessors for each student. No student was assessed by the same assessor twice. The assessment was carried out by asking the student to take the history of a patient with faculty present. This patient would have already been examined, and the history would have already been taken by the doctor assessing the student, so the examiner knew what history to expect out of the student. The patient complaints and presentations were not similar pre- and post-intervention and were not similar between the students. As students were not assessed on their physical exam, problem solving or clinical decision-making skills,

Table 1. Checklist against which the students were graded.

Communication skills	Skills expected to be demonstrated	Total marks
Introducing yourself and the purpose of examination	Stating your name, title, and purpose fluently and not missing details.	1
Building a rapport with patient	Putting the patient at ease, smiling, respectful conduct.	1
Eye contact	Ensuring eye contact with the patient and not only the history taking sheet or the body part being examined.	1
Listening to the patient's complaints	Not interrupting the patient, letting him speak without directing him in a certain way	1
Asking correct open-ended questions (question style)	Not asking specific questions that will lead to a definite diagnosis and letting the patient speak for himself.	1
Empathy	Not acting too proud or dealing roughly with the patients.	1
Communicating in patient's language	Establishing a fluent conversation in the language that the patient is most comfortable in, that is, the regional language	1
Confidence, fluency, comfort to the patient	Being confident in the procedures and executing them with ease and providing comfort to the patient in letting him know that the examiner is aware of exactly what needs to be done.	1

Continued

Table 1. Checklist against which the students were graded.

Communication skills	Skills expected to be demonstrated	Total marks
Patient's particulars (name, age, occupation, religion, residence)	Asking about all these necessary personal details and taking note of them.	2
Chief complaints		
1. Interpretation of patient's complaints	Knowing what is important to highlight in the chief complaints and asking the correct open-ended questions regarding them	4
2. Chronology in reporting	Following a specific order of questioning (particulars, chief complaints, history of presenting illness, associated diseases, history, drug history, history of allergy, personal history, family history, history of immunisation)	1
History of present illness		
1. Onset	Ask when the symptoms started to occur and the mode of onset – sudden or gradual – and if there was any causative factor involved.	2
2. Duration	Ask for how long did they occur for and if they subsided and why did they subside	2
3. Progress	Ask about the evolution of symptoms and the exact order in which they occurred.	2
4. Negative history	Ask about possible other symptoms and receive a proper negative history regarding them.	2
History of past illness		
1. Previous occurrence of the disease	Ask whether the disease occurred earlier or if this is the first time	1
2. Previous operations/accidents	Ask about previous operations/accidents irrespective of whether they are related to the current disease or not.	1
3. Tuberculosis, diabetes, hypertension	Ask about the previous history of tuberculosis, diabetes and hypertension.	2
Personal history		
1. Addictions	Ask if he is addicted to any drugs or other things	2
2. Bowel habits	Ask about regularity, pain and consistency and if blood or any other abnormalities are present.	2
3. Bladder habits	Ask about regularity, pain and consistency and if blood or any other abnormalities are present.	2
4. Menstrual history (if patient is female) a) Length of cycle b) Length of menses c) Previous pregnancies		3
Family history		
1. History of similar illness	Ask if it has occurred and the relation of the patient to this family member.	1
2. Diabetes, hypertension	Ask if these chronic diseases are present in the family	1
Mental state and intelligence	Ask about the level of consciousness and grade him according to the five stages of consciousness if not fully conscious.	1

we assumed that communication and history taking would not differ significantly based on the presenting complaints of the patient.

Teaching Methods

We performed a literature review to investigate which teaching tools have been shown to be effective in prior studies.[5,9–19] Additionally, the teaching tools needed to be convenient to perform in the hospital setting and be in accordance with the guidelines of our hospital and college. The six teaching tools used by the faculty included:

1. Focus scripts and literature – Literature was provided by the investigators, which was to be read by the students before they attended the teaching sessions. The literature contained papers on how to improve one's communication skills and its importance and implications on patient–physician relationship. It also included literature on the basic steps of history taking and the important questions to be asked to achieve a well-rounded history of the disease. We included the 'Simplified Checklist of Calgary Cambridge Guide' for assessment of communication, and guidebooks like the 'PCM Guidebook for History taking and Physical Exams' for history taking. Reading before attending lectures has been found to be impactful in students' learning.[20] We gave the students their reading assignments every weekend before discussing that topic with them in the other five teaching methods.
2. Video demonstrations – We showed the students videos of experienced doctors taking the history of patients presenting with various illnesses. We introduced communication skills for eliciting history and then showed the students videos that showed the various components of a patient history and contrasted them with poor examples so they can learn which path to take. The better, more effective history was shown as more holistic and effective in arriving at the diagnosis. The various scenarios that were used in videos helped students identify the proper communication skills that they are required to use when eliciting history.
3. Online course – An online course to teach the history taking skills was emailed to the students as streaming videos on the skills they were supposed to demonstrate. This was prepared by the authors themselves. Streaming the online course provided the opportunity for the students to access this information repeatedly. An online discussion platform to discuss questions was also provided. This collaborative e-learning using streaming videos and discussion boards was found beneficial in previously done research. The authors addressed questions on the techniques, recognition of cues, how to recognise crucial features of a symptom, and stages and techniques of asking the history of presenting illness, among other queries that the students had.
4. Small group roleplay and feedback – Students acted as patients with various illnesses that were not known to the other students. Each student had to ask pertinent questions and correctly examine the acting patient to arrive at a correct diagnosis.
5. Didactic lectures – Faculty members conducted lectures on how to take histories and arrive at correct diagnoses. This focused on the importance of non-verbal skills like eye contact and confidence. We used mental rehearsal methods as they were found to be useful to improve the retention of information according to research.
6. Mock round with patients from the ward – We conducted a round of history taking with patients from the ward. During this mock round students were asked to take the history of the patient and were evaluated using an Objective Structured Clinical Examination (OSCE) examination pattern. We used OSCE, as it is a commonly used method of clinical skill evaluation and is an effective method of judging clinical skills and knowledge.[20,21]

Time devoted by students to learning methods is given in Table 2, where the hours per day required for each

Table 2. Time and frequency of teaching methods used on the students.

	Lit.	Videos	Course	Role play	Lectures	Mocks
Hours/day	2	1	2	2	2	2
Times/week	1	2	1	3	3	2
Total time (hours)	2×1×5=10	1×2×5=10	2×1×5=10	2×3×5=30	2×3×5=30	2×2×5=20

teaching method has been provided, and the frequency per week of practicing these tools during the interval between the pre- and post-intervention period has also been mentioned. The interval period between pre- and post-intervention was 5 weeks, and the total time devoted by each student (hours/day × times/week × 5 weeks) for each activity in total has also been calculated in Table 2.

Sample Size and Eligibility Criteria

The inclusion criteria were all second-year MBBS students who are attending surgery postings under the PI's unit at KEM Hospital. Out of 20 students posted, 15 had agreed to participate in our research and completed all the teaching methods required in this study.

Data Collection and Statistical Analysis

We compared the performance in each category of the checklist (e.g. Introduction of self, Eye Contact) of the student with his earlier performance (e.g. before the teaching tools were applied). We used the student *t*-test to analyze the students' pre- and post-intervention scores.

RESULTS

The scores of students before and after the intervention are shown in Table 3. The pre- and post-intervention comparison is shown in Table 4. This data is depicted in Figure 1 to serve as a visual comparison of the means of the test values pre- and post-intervention. Overall, the results showed improvement in the mean scores after

Table 3. Scores of the students before and after the intervention.

Before intervention	CS	CC	HOPI	HOaI	PH	FH	MS
1	5	3	5	3	5	2	0
2	4	4	4	3	7	3	0
3	4	4	4	4	4	2	0
4	0	3	5	2	6	2	0
5	1	4	4	2	6	1	1
6	2	0	3	2	6	1	0
7	9	1	2	2	6	0	1
8	3	4	4	3	7	3	0
9	9	4	0	1	1	0	0
10	8	0	4	4	4	2	0
11	5	0	2	4	4	2	0
12	5	0	3	2	6	2	0
13	5	3	5	2	6	2	1
14	0	4	4	2	6	3	0
15	9	2	1	1	1	1	1
After intervention							
1	8	4	6	3	5	2	0
2	9	5	7	7	9	4	1
3	4	4	5	4	5	2	0
4	5	3	6	3	4	3	1
5	5	4	5	3	6	1	1
6	5	4	5	3	6	1	1
7	10	5	5	3	6	1	1
8	5	3	6	3	4	4	0
9	10	5	1	6	6	1	0
10	5	3	6	3	4	4	0
11	7	3	6	3	4	4	1
12	6	3	6	3	4	4	0
13	7	4	6	3	7	1	1
14	9	4	5	3	4	4	1
15	10	5	6	4	6	3	1

the teaching methods were applied. Using the student *t*-test, we statistically analyzed the students pre- and post-intervention. The *p*-value was found to be statistically significant (<0.05) in communication skills, completeness of their history taking with regard to the history of the presenting illness, history of past illnesses, family history, and mental status report. It was found to be non-significant with regard to personal history taking. The completeness of asking the patient's personal history (patients' addictions, bowel and bladder habits and menstrual history) needs to be emphasised separately for an improvement in this aspect, which was not done in the teaching tools we had used.

Table 4. Comparison between pre- and post-intervention values using student *t*-test.

	Pre-intervention		Post-intervention		Change	*p*
	Mean	SD	Mean	SD		
CS	4.6	3.1	7.0	2.2	2.4	0.004
CC	2.4	1.7	3.9	0.8	1.5	0.003
HOPI	3.3	1.5	5.4	1.4	2.1	<0.001
HOaI	2.5	1.0	3.6	1.2	1.1	0.021
PH	5.0	1.9	5.3	1.4	0.3	0.587
FH	1.7	1.0	2.6	1.4	0.9	0.003
MS	0.3	0.5	0.6	0.5	0.3	0.019

DISCUSSION

The usual method of conducting postings for undergraduate students is to assign them to a patient and observe them throughout the course of their postings and have discussions with their teachers or the resident of the ward. The methodology of history taking and being able to arrive at a differential diagnosis is very often overlooked, and professors concentrate more on explaining the pathophysiology and treatment of diseases, rather than the method of coming at the particular diagnosis. As learning to diagnose is as important as knowing how to treat a disease, it should be included as a cornerstone of medical education.

According to our study, if the teaching tools that we have applied are used regularly on undergraduate students, it could help students improve in eliciting the chief complaints and the history of the disease. The students showed a significant improvement in skills required for eliciting history from the patient. The students' communication skills improved by a third, the manner of eliciting the history of presenting illness, past illness and the patient's personal and family history improved by more than half. The most improvement was shown in the manner of eliciting chief complaints and asking further relevant questions to arrive at a provisional diagnosis. Students in the preclinical stage will benefit from creative methods of teaching in the form of the above-mentioned tools, which would make them focus on history taking skills and its importance.

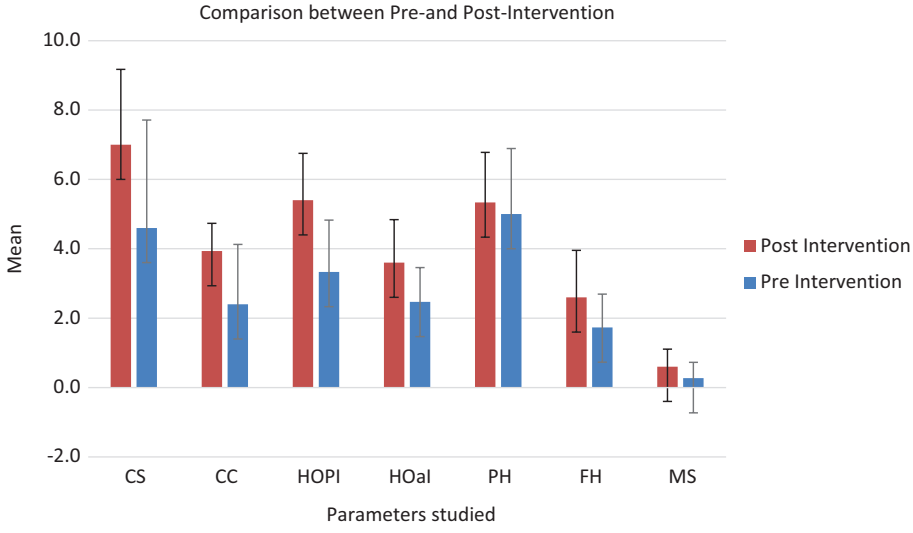


Figure 1. Box diagram of the comparison students' mean scores pre- and post-intervention.
CS = Communication Skills, CC = Chief Complaint, HOPI = History of presenting illness, HOaI = History of past illness, PH = Personal history, FH = Family history, MS = Mental status report

Future aims of our study include incorporating these teaching methods into the regular curriculum for second-year students. Learning a systematic and logical order to elicit a proper history from patients is imperative in coming to a correct diagnosis. Strong communication skills, which are normally not emphasised enough while teaching medical students, will also help them connect to the patient in a holistic way and put them at ease so they do not hold back information that could be crucial to the diagnosis.[22] The completeness of asking the patient's personal history (patients' addictions, bowel and bladder habits, and menstrual history) needs to be emphasised separately for an improvement in this aspect, which was not done in the teaching tools we had used.

As the sample size of this pilot study was quite small, we have intentions of doing a follow-up study with a larger sample size in the future, over a longer time period, to look at retention. As the teaching tools we used in our study are convenient to implement and have been found impactful, and there are no barriers to include the teaching tools we have used in our study in the medical school curriculum in Mumbai, we believe doing so will shorten the learning curve of students across similar schools.

LIMITATIONS

The first limitation is that this study was done only at one university. Secondly, as this study was limited to one batch of students posted under the PI, the sample size is small at 15 students. We hope our study can serve as a model for further studies to be conducted on a larger scale, which would further prove the benefits of adding the specified teaching tools to the curriculum at an earlier stage, so that students can acquire good history taking and clinical examination skills earlier than what is the current norm.

CONCLUSION

In this study, we have concluded that the introduction of teaching sessions focused on communication skills and history taking was beneficial to student learning. Students showed significant improvement in communication skills, completeness of their history taking with regard to the history of the presenting illness, history of past illnesses, family history, and mental status report. The student's completeness of asking the patient's personal history (patients' addictions, bowel and bladder habits, and menstrual history) did not improve drastically and hence needs to be emphasised using other teaching methods that we had not chosen for our study. We believe that it would be meaningful to include these teaching tools into the curriculum for undergraduate students.

Acknowledgments

None.

Conflict of interest and funding

The authors have not received any funding or benefits from industry or elsewhere to conduct this study.

Other disclosures

None.

Ethical approval

This study was approved by the Institutional Ethics Committee (IEC-II) relating to Biomedical and Health Research (BHR) (Project Number EC/OA-123/2019).

Disclaimer

None.

Previous presentations

None.

REFERENCES

1. Summerton N. The medical history as a diagnostic technology. Br J Gen Pract 2008; 58(549): 273–6. doi: 10.3399/bjgp08X279779

2. Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. JAMA 1997; 278(17): 1440–5. doi: 10.1001/jama.1997.03550170070035

3. Carroll AE. The high costs of unnecessary care. JAMA 2017; 318(18): 1748–9. doi: 10.1001/jama.2017.16193

4. Sox HC, Hickman DH, Marton KI, Moses L, Skeff KM, Sox CH, et al. Using the patient's history to estimate the probability of coronary artery disease: a comparison of primary care and referral practices. Am J Med 1990; 89(1): 7–14. doi: 10.1016/0002-9343(90)90090-Z

5. Kellie B, Zaza L. Communication skills in medical education: an integrated approach. Educ Res Perspect 2011; 38(2): 56–8.

6. Nichols LO, Mirvis DM. Physician-patient communication: does it matter? Tenn Med 1998; 91(3): 94–6.

7. Wendy L, Cara SL, Ronald ME. Developing physician communication skills for patient-centered care. Health Affairs 2010; 29(7): 1310–18. doi: 10.1377/hlthaff.2009.0450

8. Tiwary A, Rimal A, Paudyal B, Sigdel KR, Basnyat, B. Poor communication by health care professionals may lead to life-threatening complications: examples from two case

reports. Wellcome Open Res 2019; 4: 7. doi: 10.12688/wellcomeopenres.15042.1
9. Arun Kumar T, Sangeeta A, et al. Formative assessment and documentation of history taking in pediatrics: a comparative study. International Journal Of Contemporary Pediatrics 2018; 2(2): 621–628. doi: 10.18203/2349-3291.ijcp20180567
10. Katharina K, et al. Teaching history taking to medical students: a systematic review. BMC Med Educ 2015; 15: 159.
11. Wiecha J, Gramling R, et al. Collaborative e-learning using streaming video and asynchronous discussion boards to teach the cognitive foundation of medical interviewing: a case study. J Med Internet Res 2003; 5(2): e13. doi: 10.2196/jmir.5.2.e13
12. Evans BJ, Sweet B, Coman GJ. Behavioural assessment of the effectiveness of a communication programme for medical students. Med Educ 1993; 27(4): 344–50. doi: 10.1111/j.1365-2923.1993.tb00279.x
13. Windish DM, Price EG, Clever SL, Magaziner JL, Thomas PA. Teaching medical students the important connection between communication and clinical reasoning. J Gen Intern Med 2005; 20: 1108–13. doi: 10.1111/j.1525-1497.2005.0244.x
14. Thomas F, Jean-François C, Christina K, Michael K, Anne S, Hermann S, et al. 'Basic medical skills' – evaluation of a primary care oriented course concept within the new medical curriculum in Germany. GMS Zeitschrift für Medizinische Ausbildung 2005; 22(3): Doc59.
15. Novack DH, Dubé C, Goldstein MG. Teaching medical interviewing: a basic course on interviewing and the physician-patient relationship. Arch Intern Med 1992; 152(9): 1814–20. doi: 10.1001/archinte.1992.00400210046008
16. Al-Wardy NM. Assessment methods in undergraduate medical education. Sultan Qaboos Univ Med J 2010; 10(2): 203–9.
17. Fromme B, Karani R, Downing S. Direct observation in medical education: a review of the literature and evidence for validity. Mount Sinai J Med 2009; 76: 365–71. doi: 10.1002/msj.20123
18. Kuo A, Irby D, Loeser H. Does direct observation improve medical students' clerkship experiences? Med Educ 2005; 39: 518. doi: 10.1111/j.1365-2929.2005.02132.x
19. Miller N, Angstman KB, van Ryn M, Garrison GM, Fischer K, Phelan S. The association of direct observation of medical students with patient-centered care for obesity. Fam Med 2020; 52(4): 271–7. doi: 10.22454/FamMed.2020.314090
20. Moravec M, Williams A, Aguilar-Roca N, O'Dowd DK. Learn before lecture: a strategy that improves learning outcomes in a large introductory biology class. CBE Life Sci Educ 2010; 9(4): 473–81. doi: 10.1187/cbe.10-04-0063
21. Banerjee A. Disconnect between principles and practice of medical education technology. Med J DY Patil Univ 2014; 7: 691–2. doi: 10.4103/0975-2870.144827
22. Albaradie RS. Perception of students and teachers about didactic teaching: a cross-sectional study. Saudi J Health Sci 2018; 7: 107–15. doi: 10.4103/sjhs.sjhs_28_18

Just Like the Simulations: Improving Medical Student Confidence with Simulated Emergency Medicine Scenarios

Adam Pasquinelly[1], Alexander Luna[1], Alexander Belaia[2], Shaza Aouthmany[1]

[1]College of Medicine and Life Sciences, University of Toledo, Toledo, OH, USA
[2]New York-Presbyterian Brooklyn Methodist Hospital, Brooklyn, NY, USA

**Corresponding Author: Adam Pasquinelly; adam.pasquinelly@rockets.utoledo.edu*

Background: *Incoming emergency medicine (EM) residents may feel unsure of their ability to handle common emergency department scenarios, even if they are well educated on the proper steps to take in those scenarios. This may not stem from a lack of skill so much as a lack of confidence in their ability to perform with skills they have.*
Objective: *We look to establish a link between completion of simulation-based training in common EM scenarios and learner self-reported confidence in their ability to perform competently in those scenarios.*
Methods: *Fourth-year medical students who matched into an EM residency program participated in a Transitional Educational Program (TEP) at the Interprofessional Immersive Simulation Center at the University of Toledo in April 2021. Simulations of 16 procedural skills and clinical judgement cases were carried out using high-fidelity mannequins and real medical equipment in a hospital-based setting. Subjects were given pre- and post-TEP survey questionnaires assessing their self-reported confidence to competently perform in common EM clinical scenarios, using a 5-grade Likert scale. Data were analyzed using a one-tailed Wilcoxon signed-rank matched-pairs test.*
Results: *Of 19 participating subjects, 16 (84.2%) consented and responded to the pre-survey. Of those 16 subjects, 10 (62.5%) completed the surveys at the correct time and order. The pre- and post-surveys consisted of the same 14 questions. In 11 of 14 survey questions, there was a significant increase in subject self-reported confidence ($p < 0.05$) between pre- and post-survey.*
Conclusions: *Simulation-based training in the setting of high-fidelity equipment and faculty guidance improved the self-reported confidence of incoming EM residents to perform in common EM scenarios.*

Keywords: *Simulation; medical education; emergency medicine; technology; confidence*

INTRODUCTION

Real-world clinical situations can be daunting for both medical students and residents. While clinical experience is a vital part of medical school and residency curricula, beginning learners of both levels may not feel confident in their ability to manage the clinical scenarios that they will face on these rotations.[1] Most medical students entering clinical education will have just completed several years of preclinical work, in which they learned about the foundations of medicine mostly through reading and attending lectures. Though incoming first-year residents have completed their medical school clinical rotations, they are quickly transitioning from those curated, supervised clinical experiences to real-world clinical scenarios in which they will be the primary decision makers in patients' medical care. While the preclinical and pre-residency education undertaken by these learners provides the groundwork for clinical competency, it does not always directly translate to clinical ability. For example, even though a first-year resident might know the proper steps for many various procedures or exams indicated in a trauma patient, they may lack confidence in their ability to adequately perform these steps in a real clinical setting, and thus can still 'freeze up' and find themselves unable to perform when needed. One way that medical education has evolved to help learners better navigate the transition to real-world clinical practice is through the use of simulations.

Several studies cited here have shown proof of concept for the ability of simulation-based educational programs to increase the self-reported confidence of medical students and incoming residents in their ability to perform emergency medicine (EM) procedures and make clinical decisions. However, available literature did not provide a standardized version of simulation-based education; variations in simulation programs include the use

Note: Appendix A and Appendix B can be found at DOI: 10.15404/msrj/03.2023.237

of mannequins, standardized patients, computer-based scenarios, replica equipment, and more. Several studies identified advantages of using the most lifelike equipment available, while also eliminating any risk to real or standardized patients as learners complete the simulated procedures.[2–4] Therefore, our investigation explores specifically the use of a transitional education program (TEP) consisting of high-fidelity mannequin human patient simulators and real clinical equipment, taking place in a hospital-based simulation setting. The TEP was led by faculty in the Department of Emergency Medicine at the University of Toledo College of Medicine and Life Sciences (UTCOMLS) and was aided by other trained medical staff employed by the university. As the focus of this study is EM procedures, we restricted our subject population to graduating fourth-year medical students who had matched into an EM residency program that same year.

Our study investigates the use of simulation-based education for fourth-year medical students and assesses their self-reported confidence levels through scaled surveys both before and after they perform in the role of a first-year EM resident in several simulated clinical scenarios. Our primary outcome measure is the change in self-reported confidence measured by the Likert scale described here, with higher scores indicating higher confidence in each respective clinical scenario. We hypothesize that the use of simulation-based training will have a positive effect on self-reported confidence levels in this cohort's ability to perform the simulated procedures in real-world clinical practice. We hope to add to the available body of literature regarding simulation use in medical education, with an emphasis on self-reported confidence levels as an important factor alongside competence in evaluating medical learners' ability to perform in clinical scenarios.

METHODS

Settings and Participants

The TEP was held in April 2021 at the Interprofessional Immersive Simulation Center, a state-of-the-art facility at the University of Toledo Health Science Campus with capabilities including high-fidelity mannequins, advanced clinical simulations, virtual reality, and anatomy/surgical skills training. The TEP was designed by teaching and clinical faculty of both the Department of Graduate Medical Education and the Department of Emergency Medicine at the UTCOMLS. All designing and participating faculty had at least 5 years of experience, with most eclipsing 10 years. Participants in the TEP consisted of 19 fourth year medical students (M4s) who matched into an EM residency program and planned to enroll as a first-year resident upon graduation from medical school. The subjects volunteered to participate in the TEP by responding to an email sent to them several weeks prior to the event. The faculty and staff that conducted the TEP were all involved with the Department of Emergency Medicine or the Department of Graduate Medical Education at UTCOMLS. Each subject received the intervention (attended the TEP) and self-reported confidence in the ability to perform each procedure was assessed before and after the intervention.

Interventions

To better prepare graduating medical students for their intern year in the EM setting, the TEP was designed with 10 simulations of common EM skills, six simulations of common patient cases requiring rapid clinical judgment, and several structured and unstructured discussions with EM faculty and residents (Fig. 1) (Appendix A). The schedule of events was assembled after discussions with many medical students, residents, and faculty at UTCOMLS, who collectively identified the listed topics as areas in which first-year residents had a high potential to feel underprepared or unsure of their abilities.

Before the program started at 8:00 am, subjects were asked to complete the pre-survey individually. Subjects were then put into small groups of two to four and were randomly placed into one of the six morning session activities (three skill simulations and three case simulations). Subject groups spent 30 min at each station, including a briefing and debriefing before and after each simulation. After the station time had elapsed, subject groups rotated to the next station. This process repeated

Morning Sessions (8 AM - 12 PM)	Afternoon Sessions (1 PM - 5 PM)
Cases:	**Cases:**
Cardiac Pacing	GI Bleed
Airway: CHF or Asthma	AMS: Hypoglycemia
Trauma: ACLS	Stroke: Hemorrhagic vs. Ischemic
Skills:	**Skills:**
Placement:	Urgent Airway: Intubation
IO	Emergent Airway: Cricothyroidotomy
Splint	Sick vs. Non-sick: Identification
Chest Tube	Picking the Right Medication
Pigtail	
Suturing	
Lumbar Puncture	
3rd Year Clinical Horror Stories	**Q&A with Residents & Faculty**

Figure 1. TEP Schedule.

until all morning session stations were completed. There was a 1-h lunch break between the morning sessions and the afternoon sessions. The afternoon session was completed in the same manner as the morning session. After the TEP was finished, subjects who completed the pre-survey were asked to complete the post-survey.

Outcomes Measured

To assess the impact of the TEP on M4 self-reported confidence in the EM setting, this study used a descriptive pre/post survey design, with the pre-survey given immediately prior to the simulations program and the post-survey given immediately after the program was completed. The study survey consisted of 14 statements asserting confidence in various EM procedures (all of which would later be topics covered in a simulation during the TEP), and subjects were asked to indicate their level of agreement with each statement on a standard 5-grade Likert scale (1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, 5 = strongly agree) (Appendix B). These survey questions were designed according to guidelines proposed by Nemoto and Beglar.[5] The pre-survey included the informed consent agreement, three demographic questions, the study survey, and a field for the entry of a unique 4-digit PIN that would be used to link pre- and post-survey responses to an individual without the need to collect any other identifiers. The post-survey included the PIN field, the study survey, and another 8-item survey (the secondary survey). The secondary survey asked the subjects to indicate their level of agreement, on the same 5-grade Likert scale used in the primary survey, with statements about the usefulness of simulations as an education tool, their perception of how closely the simulations matched real clinical scenarios that they had experienced, and their satisfaction with the TEP.

The pre-survey was made available at 9:30 pm the night before the start of the TEP, and submissions were accepted until the program started at 8:05 am. The post-survey was made available at completion of the TEP, and submissions were accepted for 10 days; this longer window was used to allow subjects adequate time for survey completion during a very busy time at the end of their medical school careers. Both surveys were administered with Microsoft Office Forms and sent to the subjects' personal email addresses. Likert scale data from each survey was exported from Office Forms, linked by PIN, and analyzed using Microsoft Excel. Data were only accepted for analysis if: (1) the pre-survey was completed before the start of the TEP (8:05 am on 4/6/2021) and (2) there was both a pre- and post-survey linked to the same PIN.

The primary outcome of this study is the measured change in self-reported confidence in subjects' ability to adequately perform in each situation that was covered in the TEP simulations, with a secondary outcome of the measure of participants' opinions on the effectiveness of simulations in medical education.

Analysis of Outcomes

Pre-TEP and post-TEP survey responses (n = 10) were analyzed using a one-tailed Wilcoxon signed-rank matched-pairs test.[6] Statistical significance was determined with a p-value of 0.05 and a critical value of 10, as per the critical value table computed by McCornack.[7] The secondary survey regarding simulation effectiveness in medical education was assessed for general positive or negative responses to each question by assigning numerical values to the Likert scale responses (−2 = strongly disagree, −1 = disagree, 0 = undecided, 1 = agree, 2 = strongly agree) and calculating average values for each question and each subject response.

IRB Statement

This study was approved as an institutional review board (IRB)-exempt study by the University of Toledo Institutional Review Board.

RESULTS

The study population was composed of fourth-year medical students at the University of Toledo who had matched into an EM residency program beginning after graduation. The TEP was advertised to all EM residency-matched students via email, and 19 signed up to attend the program. Of these 19 participants, 16 responded to the pre-survey and informed consent (84.2%). Of those 16 consenting subjects, 10 (62.5%) completed surveys at the correct times and used a PIN that linked a pre-survey response to a post-survey response, for a final study sample of 52.6% (10/19) of the available population. These 10 primary survey pairs comprised the analyzed data set (Tables 1 and 2).

Of the 14 questions on the primary survey, 11 showed a significant increase in subject self-reported confidence after the TEP (Fig. 2). Notably, the pre-survey had four total 'strongly agree' responses, and the post-survey had 29, an increase of over sevenfold. Similarly, the

Table 1. Demographic questions and responses given by the 10 study subjects.

Demographics					
What Is your level of training?	M4 (10)	Other (0)			
What gender do you Identify with?	Man (6)	Woman (4)	Other (0)		
What Is your age?	21–25 (3)	26–30 (6)	31–35 (0)	36–40 (0)	40+ (1)

Table 2. Primary survey questions and responses, grouped by question and timing (whether pre- or post-survey).

Survey Question	Q#	Survey	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
I feel confident in my ability to manage an emergent airway, e.g. performing a cricothyrotomy	**Q1**	Pre	1	4	3	2	0
		Post	0	1	0	7	2
I feel confident in my ability to manage a urgent airway, e.g. intubation	**Q2**	Pre	0	2	1	7	0
		Post	0	1	0	3	6
I feel confident in my ability to provide ACLS care	**Q3**	Pre	0	1	2	7	0
		Post	0	1	0	4	5
I feel confident in my ability to place a splint	**Q4**	Pre	0	4	3	3	0
		Post	0	2	1	7	0
I feel confident in my ability to place an intraosseus line	**Q5**	Pre	0	3	3	3	1
		Post	0	1	1	6	2
I feel confident in my ability to effectively manage patient situations involving opioids	**Q6**	Pre	0	5	3	2	0
		Post	0	1	2	7	0
I feel confident in my ability to place a chest tube	**Q7**	Pre	2	5	3	0	0
		Post	0	1	1	7	1
I feel confident in my ability to place a pigtail catheter	**Q8**	Pre	1	6	3	0	0
		Post	0	1	4	5	0
I feel confident in my ability to suture a wound	**Q9**	Pre	0	1	1	6	2
		Post	0	1	0	2	7
I feel confident in my ability to perform a lumbar puncture	**Q10**	Pre	1	8	1	0	0
		Post	0	1	2	7	0
I feel confident in my ability to effectively manage a patient with an altered mental status	**Q11**	Pre	0	2	6	2	0
		Post	0	1	0	7	2
I feel confident in my ability to effectively manage a patient with a GI bleed	**Q12**	Pre	1	1	5	3	0
		Post	0	2	0	8	0
I feel confident in my ability to manage a stroke patient	**Q13**	Pre	1	1	4	4	0
		Post	0	2	0	7	1
I feel confident in my ability to identify patients that are truly sick in the ED setting	**Q14**	Pre	1	1	4	3	1
		Post	0	2	0	5	3

GI, gastrointestinal; ED, emergency department.

pre-survey had eight total 'strongly disagree' responses, where the post-survey did not have any.

The secondary survey yielded generally positive subject responses for each question (range 0.92–1.76) as well as generally positive responses by each subject across the survey (range 0–2), indicating the subjects generally supported the effectiveness of simulations as an educational tool (average overall score 1.52) (Table 3). Secondary survey responses for all 13 subjects who completed the post-survey were included, regardless of pre-survey completion status or PIN linking the post-survey to a completed pre-survey, because the secondary survey was only present on the post-survey and thus did not require comparison to pre-survey responses.

DISCUSSION

This study found a statistically significant increase in subject self-reported confidence to competently perform procedures and make clinical decisions common to EM after participating in simulated versions of those

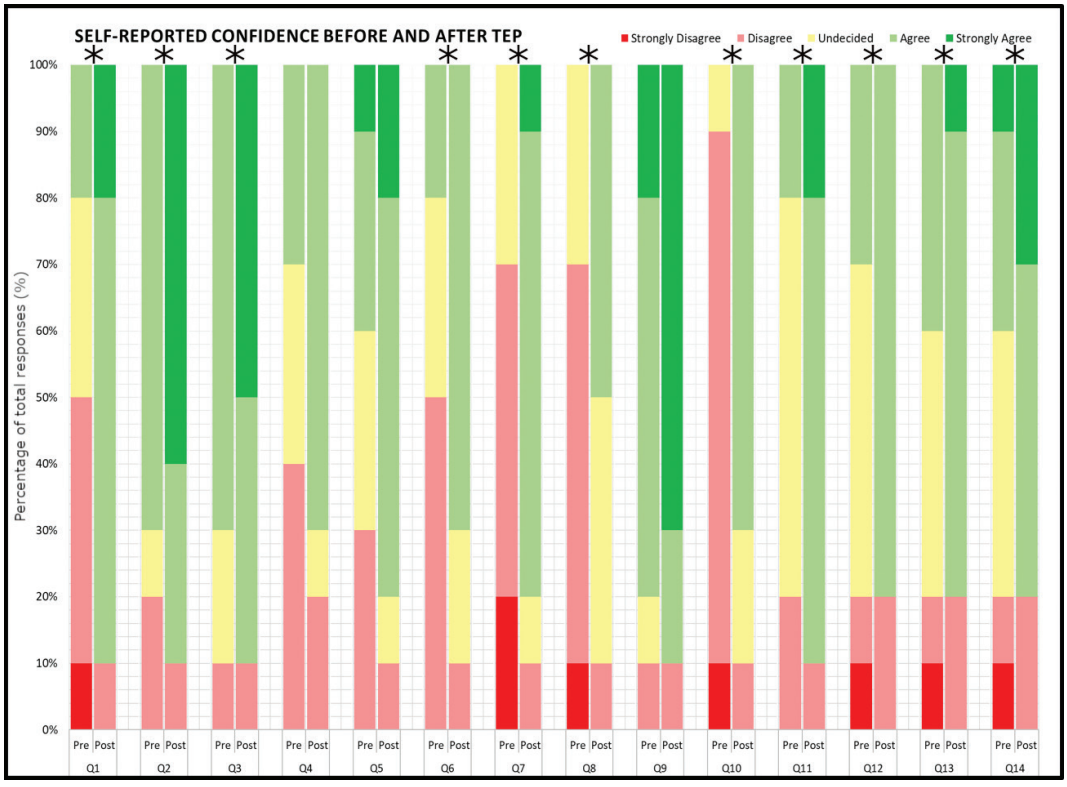


Figure 2. Results grouped by question. Pre-survey responses are on the left and post-survey responses are on the right; a star above a question column indicates a statistically significant increase in subject self-reported confidence from the pre-survey to the post-survey. The y-axis depicts the total responses, and each color band represents the percentage of total responses comprised by the corresponding response choice for each question. TEP, transitional education program. *: statistically significant, $p < 0.05$.

Table 3. Secondary survey questions and responses, grouped by question.

Secondary Survey Question	Q#	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Simulations are an effective method of training in general	1	0	0	1	2	10
Simulations are a productive way of learning how to respond to common patient presentations in emergency medicine	2	0	0	1	3	9
I was able to work with colleagues successfully during the simulations	3	0	0	1	1	11
I was able to practice the technique of each skill during the simulations	4	0	1	1	5	6
In my experience, these simulations were similar to responding to these patient presentations in a clinical setting	5	0	0	3	8	2
I was given ample opportunity for questions during the simulations	6	0	0	1	3	9
I felt that feedback given during the simulations was constructive and meaningful	7	0	0	1	3	9
Overall, I personally feel more confident in my ability to respond to these patient presentations after simulation training	8	0	0	1	2	10

procedures and cases. These simulations closely mirror real-world procedures and require many technical and management skills that first year residents are expected to perform. The results from this study support the growing trend found in medical education literature indicating some modality of simulation is beneficial in the training of medical students and residents – especially for those transitioning into real world clinical practice.[8–16] This study focused on those about to enter an EM residency program, but there are several other documented studies where the use of simulation has positively impacted the confidence of a broader range of students.[3,4,17–20]

Simulation has been established as a successful, evidence-based tool in medical education for several decades.[8,21] Simulation use in EM education – both in medical schools and residency programs – has become widespread in the last 20 years. Okuda et al[22] found that 91% of U.S. EM residency programs reported using some form of simulation in 2008, and other studies show even more 'boot camp'-style (defined as an early preparatory course or orientation sessions for learners undergoing a transition in medical education) simulation sessions are being developed at schools and residency programs across North America each year.[9–11,22,23]

The efficacy of simulation in medical education as a tool to facilitate *competency* in clinical skills is well supported in current literature.[8–11,21] However, studies assessing subject *confidence* in one's ability to perform procedures or make decisions in real-world clinical settings are substantially less prevalent in the current available literature. Okuda et al extensively aggregated evidence for the usefulness of simulations in several specific medical specialties, though their findings for EM focused more on competence in crew resource management and acute care team protocols, rather than self-reported confidence in the ability to independently perform procedures or make clinical decisions.[2] Gottleib thoroughly investigated the link between self-reported confidence and competence in learners and advocated for evaluation methods that consider both qualities simultaneously when assessing medical professionals' ability to perform tasks.[24]

Literature covering the correlation between subject confidence and simulation-based education in EM scenarios was severely lacking; however, several published studies supported a positive relationship between simulation-based education and subject confidence in specific medical skills or in other areas of healthcare. Bowers et al found that the comfort and confidence level of third and fourth year medical students increased after they participated in an advanced cardiac life support simulation training.[17] In a multi-specialty simulation-based course, Dermody et al observed that a majority of a cohort of 30 medical students had an increased confidence in their skills.[18] Biron et al observed that knowledge, skill, and confidence level increased between pre-session and post-session questionnaires and assessments in a cohort of 120 medical students participating in a multi-sensory cricothyrotomy educational experience.[12] Morgan and Cleave-Hogg demonstrated an increase in subject confidence in a population of 144 final-year medical students upon participation in several simulated anesthesia procedures.[25] Sattler et al found an increase in self-reported confidence in a population of internal medicine residents upon completion of a simulation course covering six invasive bedside procedures.[13]

Several studies, including this one, have used mainly self-reported perceived changes in their studied outcomes. Therefore, the role of simulations as an objective improvement tool in medical education is still unclear. A proposed solution to this limitation is a longitudinal approach assessing both competence and confidence, where students perform in these simulated scenarios across several sessions, studying both long-term retention and self-reported confidence in these skills, as well as objective measures of performance improvement in these skills. Other potential sources of bias in this study are the possibilities of central tendency bias in survey responses and observer bias during simulation sessions. The subjects were aware that they were part of a study investigating the effectiveness of simulations in education, creating some degree of bias in their responses. Central tendency bias is the tendency for subjects to avoid the extremes of a rating scale, for example, subjects avoiding the 'strongly agree' and 'strongly disagree' choices on the Likert scale in the study survey used here. Though this bias was likely present, its effect was unlikely to significantly sway results, and it has been argued to be closer to a data pattern than a bias.[26] Furthermore, simulations in general may be prone to some degree of inherent observer effect; participants are aware that there are likely no real consequences to failure, or at least none as severe as would be present in real-world scenarios, and therefore might not make the same decisions or actions in a simulation that they would in the real world. As the

technology underlying simulations advances, future studies might be able to simulate real-world scenarios even more accurately through the addition of virtual reality or related innovations, thus providing a more immersive experience and potentially an observer bias.

The secondary survey was included to gather general feedback about the TEP to identify subject perceived areas of strength and weakness in the TEP, which will serve to refine its implementation for future classes. Secondary survey responses indicated a generally positive perception of the use of simulation in medical education, which is consistent with prior studies.[2] The non-simulation sessions in the TEP, consisting of 3rd year clinical horror stories and Q&A sessions with residents and faculty, may have affected subject self-reported confidence as well. While these sessions were not focused on directly practicing a specific skill in a simulated scenario and thus were not addressed in the study surveys, they may have served to ease subject anxiety by establishing a sense of rapport and camaraderie between students and residents or faculty. The designers of the TEP wanted to include these sessions for student benefit but there is no good way to test the impact these sessions may have had on primary and secondary survey data, and thus we want to acknowledge these sessions as potential confounders.

One limitation in our study is the small sample size of 10 subjects – a result of sampling only UTCOMLS students who had already matched into EM residency programs. Additionally, not every student in the study population participated or followed study procedures properly, which invalidated several survey responses and reduced our study sample. One way to increase the statistical power of this study would be to replicate the TEP annually with students in each year's graduating class and aggregate data to generate a larger sample size. We plan to employ this plan for the next several years and incorporate more measurable outcomes in future manuscripts.

CONCLUSION

Despite the small number of subjects in this study, we found a significant correlation between participation in simulated EM procedures and cases and student self-reported confidence in performing competently in those scenarios. The study also found a strongly positive opinion held by subjects regarding the effectiveness of simulations as an educational tool. This study supports current literature citing the educational and self-reported confidence benefits of participating in simulated medical procedures, particularly when using advanced high-fidelity simulation technology. We advocate for the use of simulations in medical education and training, especially for graduating students transitioning into the first year of EM residency programs.

CONFLICT OF INTEREST AND FUNDING

None to report. The authors have not received any funding or benefits from industry or elsewhere to conduct this study.

REFERENCES

1. Smith HL, Craig SR, Yost WJ. Examination of entering residents' self-reported confidence and supervision needs performing AAMC entrustable professional activities. J Grad Med Educ 2018; 10(4): 474. doi: 10.4300/JGME-D-18-00520.1

2. Okuda Y, Bryson EO, DeMaria S, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence?: Utility of simulation in medical education. Mt Sinai J Med J Transl Pers Med 2009; 76(4): 330–43. doi: 10.1002/msj.20127

3. Gordon CJ, Buckley T. The effect of high-fidelity simulation training on medical-surgical graduate nurses' perceived ability to respond to patient clinical emergencies. J Contin Educ Nurs 2009; 40(11): 491–8. doi: 10.3928/00220124-20091023-06

4. Boling B, Hardin-Pierce M. The effect of high-fidelity simulation on knowledge and confidence in critical care training: an integrative review. Nurse Educ Pract 2016; 16(1): 287–93. doi: 10.1016/j.nepr.2015.10.004

5. Nemoto T, Beglar D. Developing Likert-Scale Questionnaires. Presented at: JALT 2013 Conference; October 25, 2013; Tokyo, Japan.

6. Woolson RF. Wilcoxon Signed-Rank Test. In: D'Agostino RB, Sullivan L, Massaro J, eds. Wiley encyclopedia of clinical trials. John Wiley & Sons, Inc., Hoboken, NJ; 2008: eoct979. doi: 10.1002/9780471462422.eoct979

7. McCornack RL. Extended tables of the Wilcoxon matched pair signed rank statistic. J Am Stat Assoc 1965; 60(311): 864–71. doi: 10.1080/01621459.1965.10480835

8. Chakravarthy B, Haar E, Bhat S, McCoy C, Denmark TK, Lotfipour S. Simulation in medical school education: review for emergency medicine. West J Emerg Med 2011; 12(4): 461–6. doi: 10.5811/westjem.2010.10.1909

9. Blackmore C, Austin J, Lopushinsky SR, Donnon T. Effects of postgraduate medical education 'boot camps' on clinical skills, knowledge, and confidence: a meta-analysis. J Grad Med Educ 2014; 6(4): 643–52. doi: 10.4300/JGME-D-13-00373.1

10. Fernandez GL, Page DW, Coe NP, Lee PC, Patterson LA, Skylizard L, et al. Boot camp: educational outcomes after 4 successive years of preparatory simulation-based training at

onset of internship. J Surg Educ 2012; 69(2): 242–8. doi: 10.1016/j.jsurg.2011.08.007
11. Malekzadeh S, Malloy KM, Chu EE, Tompkins J, Battista A, Deutsch ES. ORL emergencies boot camp: using simulation to onboard residents. ORL Emergencies Boot Camp. The Laryngoscope 2011; 121(10): 2114–21. doi: 10.1002/lary.22146
12. Biron VL, Harris M, Kurien G, Campbell C, Lemelin P, Livy D, et al. Teaching cricothyrotomy: a multisensory surgical education approach for final-year medical students. J Surg Educ 2013; 70(2): 248–53. doi: 10.1016/j.jsurg.2012.11.004
13. Sattler LA, Schuety C, Nau M, Foster DV, Hunninghake J, Sjulin T, et al. Simulation-based medical education improves procedural confidence in core invasive procedures for military internal medicine residents. Cureus 2020; 12(12): e11998. doi: 10.7759/cureus.11998
14. Hishikawa S, Kawano M, Tanaka H, Konno K, Yasuda Y, Kawano R, et al. Mannequin simulation improves the confidence of medical students performing tube thoracostomy: a prospective, controlled trial. Am Surg 2010; 76(1): 73–8. doi: 10.1177/000313481007600115
15. Hogg G, Miller D. The effects of an enhanced simulation programme on medical students' confidence responding to clinical deterioration. BMC Med Educ 2016; 16(1): 161. doi: 10.1186/s12909-016-0685-2
16. Turatsinze S, Willson A, Sessions H, Cartledge PT. Medical student satisfaction and confidence in simulation-based learning in Rwanda – pre and post-simulation survey research. Afr J Emerg Med 2020; 10(2): 84–9. doi: 10.1016/j.afjem.2020.01.007
17. Bowers KM, Smith J, Robinson M, Kalnow A, Latham R, Little A. The impact of advanced cardiac life support simulation training on medical student self-reported outcomes. Cureus. 2020; 12(3): e7190. doi: 10.7759/cureus.7190
18. Dermody SM, Masciello M, Malekzadeh S. A multispecialty critical airway simulation course for medical students. The Laryngoscope 2021; 131(7): 1482–6. doi: 10.1002/lary.29264
19. Nitschmann C, Bartz D, Johnson NR. Gynecologic simulation training increases medical student confidence and interest in women's health. Teach Learn Med 2014; 26(2): 160–3. doi: 10.1080/10401334.2014.883984
20. Megel ME, Black J, Clark L, Carstens P, Jenkins LD, Promes J, et al. Effect of high-fidelity simulation on pediatric nursing students' anxiety. Clin Simul Nurs 2012; 8(9): e419–28. doi: 10.1016/j.ecns.2011.03.006
21. Motola I, Devine LA, Chung HS, Sullivan JE, Issenberg SB. Simulation in healthcare education: a best evidence practical guide. AMEE Guide No. 82. Med Teach 2013; 35(10): e1511–30. doi: 10.3109/0142159X.2013.818632
22. Okuda Y, Bond W, Bonfante G, McLaughlin S, Spillane L, Wang E, et al. National growth in simulation training within emergency medicine residency programs, 2003–2008. Acad Emerg Med 2008; 15(11): 1113–6. doi: 10.1111/j.1553-2712.2008.00195.x
23. Patrick SP, Mastoras G, Krywenky A. University of Ottawa's Department of Emergency Medicine simulation boot camp: a descriptive review. BMJ Simul Technol Enhanc Learn 2020; 6(3): 184–5. doi: 10.1136/bmjstel-2019-000467
24. Gottlieb M, Chan TM, Zaver F, Ellaway R. Confidence-competence alignment and the role of self-confidence in medical education: a conceptual review. Med Educ 2022; 56(1): 37–47. doi: 10.1111/medu.14592
25. Morgan PJ, Cleave-Hogg D. Comparison between medical students' experience, confidence and competence: experience, confidence and competence. Med Educ 2002; 36(6): 534–9. doi: 10.1046/j.1365-2923.2002.01228
26. Douven I. A Bayesian perspective on Likert scales and central tendency. Psychon Bull Rev 2018; 25(3): 1203–11. doi: 10.3758/s13423-017-1344-2

Vulvovaginal Lacerations Following Consensual Versus Nonconsensual Sexual Intercourse

Erin Leach[1], Victoria Moaddel[1], Linda Rossman[2], Stephanie Solis[2], Lindsey Ouellette[3], Christian Kolacki[3], Jeffrey Jones[1,3*]

[1]College of Human Medicine, Michigan State University, Grand Rapids, Michigan, USA
[2]YWCA West Central Michigan Nurse Examiner Program, Grand Rapids, Michigan, USA
[3]Department of Emergency Medicine, Spectrum Health Hospitals, Grand Rapids, Michigan, USA

Corresponding Author: Jeffrey Jones; Jeffrey.Jones@spectrumhealth.org

Background**:** *The medical literature on vulvovaginal lacerations following consensual versus nonconsensual sexual intercourse is sparse and conflicting.*

Objectives**:** *To compare the predisposing factors, injury location and severity, as well as treatment of vulvovaginal lacerations sustained during consensual versus nonconsensual sexual intercourse, in adult women within a community-based cohort.*

Methods**:** *This is a retrospective comparative analysis of adult women presenting to the emergency departments of five hospitals and a free-standing nurse examiner clinic during a 7-year study period. All patients had documented vulvovaginal lacerations and reported vaginal penetration via consensual sexual intercourse (CSI) or nonconsensual sexual intercourse (NCSI) within 72 h of presentation.*

Results**:** *A total of 598 cases were identified: 81 (14%) reported CSI, and 517 (87%) reported NCSI. CSI patients were younger (21.3 vs. 25.7, $p < 0.001$) and reported a greater incidence of penile penetration (97.5% vs. 75.9%, $p < 0.001$). While NCSI subjects had a higher incidence of vulvovaginal lacerations overall (1.7 vs. 1.0, $p < 0.001$), their injuries were smaller (1.1 cm vs. 4.3 cm, $p < 0.001$) and more likely to be located on the posterior vulva (83% vs. 69%, $p = 0.003$) when compared with the CSI group. In addition, all the lacerations in the NCSI group were superficial. In contrast, 27 (33%) of CSI subjects had lacerations sutured in the ED; 6 (7%) required aggressive fluid resuscitation and 10 (12%) required surgical intervention.*

Conclusions**:** *In this community-based population, more severe vulvovaginal lacerations were noticed in women following CSI. The predisposing factors, injury location, and subsequent treatment in this group were significantly different when compared with women reporting NCSI.*

Keywords: *consensual; nonconsensual; sexual assault; sexual intercourse; treatment outcomes; vaginal penetration; vulvovaginal lacerations*

INTRODUCTION

In females, the spectrum of vulvovaginal injuries from vaginal sexual intercourse, also known as coitus, can range from superficial lacerations to life-threatening lacerations or perforation. Research has shown that most female vulvovaginal injuries resulting from coitus are minor and often occur during the first sexual experience.[1,2] These vulvovaginal injuries resolve spontaneously or with minor treatment.[2] However, coitus may result in more severe lacerations of the vulvovaginal area causing life-threatening hemorrhage, thus requiring immediate surgical repair.[3] It has been reported that vulvovaginal injuries occur more frequently and with more severity in sexually assaulted women.[2–4] However, data on vulvovaginal lacerations following consensual intercourse are sparse and conflicting.[4–7] In one of the few studies that compare injuries from CSI and NCSI, it was found that while injuries are more common in NCSI, there is no difference in the number of vulvovaginal lacerations and was limited by study number.[6] In another study, genital injuries were three times more likely in NCSI but there was no difference in location, size or type of injury from CSI.[5] A third shows a localized pattern in NCSI victims with injuries more likely in the posterior fourchette, labia minora, hymen and fossa navicularis. However, the study does not specify lacerations versus abrasions or whether a pattern was found in the CSI group.[4] In this study, we focus solely on vulvovaginal lacerations because of the possible need for emergency resuscitation and surgery. The purpose of our study is to compare the injury location, injury severity, predisposing factors, and treatment

of vulvovaginal lacerations sustained during vaginal penetration via CSI versus NCSI in adult women within a community-based cohort. By identifying differences or trends in presentations of vulvovaginal lacerations sustained by CSI and NCSI, clinicians can have a higher index of suspicion on presentation, ultimately leading to better patient outcomes.

METHODS

Study Design and Setting

We conducted a retrospective cohort analysis of adult women (> 16 years old) presenting to the emergency departments (ED) of three urban medical centers, two rural community hospitals, and a free-standing nurse examiner clinic during a 7-year study period (2013–2019). The study was reviewed and approved by the local institutional review board.

Patients

Patients were selected for inclusion in the study if they had documented vulvovaginal lacerations (ICD-9 codes 878 and 867; ICD-10 code S31.41) and reported vaginal penetration via consensual (CSI) or nonconsensual sexual intercourse (NCSI) within 72 h. For the purposes of this study, we defined CSI as vaginal penetration by a penis, fingers, or foreign object in a woman who was a willing and cooperative participant in the sexual activity. Nonconsensual sexual intercourse was defined as vaginal penetration involving force or the threat of force, incapacity, or no consent. Exclusion criteria included victims who declined forensic examination, had missing or incomplete documentation, vague or unclear patient history (e.g., intoxication), or prolonged time (greater than 72 h) following vaginal penetration.

Intervention

All patients were examined by board-certified emergency physicians or forensic nurses trained to perform medical forensic examinations. After each examination, clinicians documented the number and types of all vulvovaginal injuries visualized using a standardized classification system.[4] For the purposes of this study, we defined vulvovaginal laceration as any break in tissue (skin and mucous membranes) including fissures, tears, cuts, gashes, or rips.[4] The following nine anatomic sites were routinely evaluated for the presence and type of injury: the labia minora, labia majora, posterior fourchette, fossa navicularis, hymen, vagina, cervix, perineum, and perianal area. When labial traction was insufficient for visualization of the entire hymen, clinicians used the foley catheter technique to document tears to the hymen.[8]

Data Collection

Patient demographics, characteristics of the sexual encounter, predisposing factors, time to presentation and injury location(s) were recorded using a standardized abstraction form. Procedure or operative notes were reviewed to define the extent of vulvovaginal injuries and timing and type of definitive treatment. Derived from previous literature, the predisposing variables listed in Tables 3 and 4 are characteristics found to be associated with genital trauma and thus may predispose women to vulvovaginal lacerations.[9,10,11,12,13] These variables were chosen prospectively for analysis. All data were collected by four research associates who were blinded to the study objective. The research staff were trained in data abstraction using a set of mock case records. Another investigator supervised data abstraction and ensured that data variable definitions were uniformly applied.

Statistical Analyses

The primary outcome of interest was the location and severity of vulvovaginal lacerations in women following consensual versus nonconsensual sexual intercourse. Data were entered into an encrypted Microsoft Excel database (Microsoft Corp, Redmond, WA, USA). All analyses were performed using SAS statistical software (SAS Institute, Cary, NC, USA). One investigator performed a blinded critical review of a random sample of 10% of the medical records to determine inter-rater reliability in the identification, classification, and location of vulvovaginal injuries using the Kappa reliability test. Descriptive statistics (mean, standard deviation) were used to describe the frequency of vulvovaginal laceration, location, and severity of injury. Discrete variables were analyzed with the use of chi-squared tests; unpaired *t*-tests were used for comparisons of two means.

RESULTS

During the 7-year study period, 1545 women were found to have been evaluated for vulvovaginal trauma due to penetration by penis, fingers, or foreign object as pulled by ICD-9&10 codes listed in methods. Of these 1545, 108 (7.0%) had a documented history of CSI while 1437 (93.0%) reported NCSI. A total of 598 of these women sustained macroscopic vulvovaginal lacerations

and were entered into the study. Macroscopic lacerations were defined as those visualized directly without magnification. Eighty-one (13.5%) reported CSI and 517 (86.5%) reported NCSI. The mean age of the patients was 25.1 ± 9 years; the age range was 17–79 years. Seventy-one women (11.9%) were postmenopausal; 54 (9.0%) had no prior history of sexual intercourse. Consensual sexual intercourse and NCSI patients were comparable in terms of ethnicity, marital status, prior history of sexual intercourse, and the frequency of alcohol/drug use (Table 1). Consensual sexual intercourse patients were younger (21.3 vs. 25.7), had a shorter time interval from penetration to examination, and a greater incidence of penile penetration (97.5% vs. 75.9%). Overall, 66.6% (54) of CSI patients presented to the ED with marked vaginal bleeding, while 28.4% (23) reported perineal pain.

A total of 1034 macroscopic vulvovaginal lacerations were documented in the 598 patients included in the study (Table 2). Nonconsensual sexual intercourse victims had a greater mean number of lacerations (1.7 vs. 1.1, $p < 0.001$), injuries tended to be smaller (1.1 cm vs. 4.1 cm, $p < 0.001$), and more likely to be located on the fossa navicularis and posterior fourchette (49.3%). All lacerations in the NCSI population were superficial; none were severe enough to require surgical treatment or resuscitation (95% confidence interval, 0% to 0.6%). In contrast, CSI patients had more extensive lacerations, which were commonly located on the posterior vaginal wall and labia (56.7%). Overall, 27 (33.3%) of CSI subjects had vulvovaginal lacerations repaired in the ED; 10 (12.3%) were taken to the operating room (OR) for repair under anesthesia. A total of 37 CSI patients (45.7%) had lacerations requiring repair in the ED or the OR with a

Table 1. Patient demographics in adult women after consensual (CSI) and nonconsensual sexual intercourse (NCSI).

	CSI ($n = 81$)	NCSI ($n = 517$)	*p*-value
Age	21.3 ± 7.1	25.7 ± 9.4	< 0.001
Ethnicity (% white)	59 (72.8%)	361 (69.8%)	0.583
Marital status (% single)	63 (77.8%)	351 (67.9%)	0.073
No prior history of sexual intercourse	11 (13.6%)	46 (8.9%)	0.181
Alcohol or drug use < 24 h	37 (45.7%)	227 (43.9%)	0.762
Time interval to exam, mean hours (SD)	9.6 ± 5.5	17.2 ± 7.4	< 0.001
Time interval to exam < 24 h	62 (76.5%)	308 (68.1%)	0.128
Vaginal penetration			
Penile	79 (97.5%)	392 (75.9%)	< 0.001
Digital	11 (13.6%)	156 (30.2%)	0.002
Foreign body	3 (3.7%)	28 (5.4%)	0.521

Table 2. Injury characteristics (N = 598).

	CSI ($n = 81$)	NCSI ($n = 517$)	*p*-value
Mean no. of lacerations	1.1 ± 0.2	1.7 ± 1.3	< 0.001
Mean laceration length (cm)	4.1 ± 1.8	1.1 ± 0.9	< 0.001
Total lacerations	155	879	
Location of lacerations			
Vaginal wall	54 (34.8%)	71 (8.1%)	0.008
Labia	34 (21.9%)	179 (20.3%)	0.098
Hymen	32 (20.7%)	144 (16.4%)	0.032
Fossa navicularis	20 (12.9%)	239 (27.2%)	0.018
Posterior fourchette	13 (8.4%)	194 (22.1%)	0.013
Perineum	2 (1.3%)	52 (5.9%)	0.018
Lacerations requiring sutures	27 (33.3%)	0	< 0.001
Hemorrhagic shock	6 (7.4%)	0	< 0.001
OR repair	10 (12.3%)	0	< 0.001
Returned with rebleeding	3 (3.7%)	0	< 0.001

MSRJ © 2023 VOL: 10. Issue: Spring
epub March 2023; www.msrj.org

95% confidence interval of 34.6% to 57.1%. Lacerations that required repair were generally > 3 cm in length and located on the posterior vagina walls, vaginal vault and posterior fourchette. One patient had a laceration that extended into the peritoneal cavity. Hemorrhagic shock was present in six (7.4%) of the CSI patients. Variables that increased the need for operative intervention included uncontrolled bleeding despite gauze packing, a falling hematocrit, or a combined injury to the vagina and vulva. All patients had an uneventful postoperative course.

Known predisposing and etiologic factors for vulvovaginal lacerations were documented in 69 (85.2%) CSI patients (Table 3). This included alcohol intoxication, 'rough or aggressive' intercourse, awkward positioning, insertion of foreign bodies, and first coitus. Seventeen women (24.6%) had more than one predisposing factor documented. In comparison, the predisposing factors for vulvovaginal injury in NCSI victims are listed in Table 4. Overall, 410 (90.7%) patients had more than one predisposing factor documented; 239 (52.9%) had more than two factors documented; and 191 (42.3%) had more than three factors documented. Interrater reliability of the data abstraction was excellent with a median kappa statistic of 0.87.

DISCUSSION

Vulvovaginal trauma can span a continuum of severity from minor injuries to major lacerations. The actual prevalence of these injuries is difficult to discover, especially if the patient is reluctant to disclose the nature of the injury. In our community-based population, we found that women reporting consensual sexual intercourse had more severe vulvovaginal lacerations than victims of sexual assault, which is in alignment with previous studies.[5–7,14] Injuries from vaginal penetration may include lacerations, ecchymosis, abrasions, erythema, and edema. The vagina and vulva are vulnerable to hemorrhage due to their rich blood supply. The resulting bleeding from lacerations can be considerable and progress to hypovolemic shock and death if not promptly managed emergently and or operatively.[14] Further complications of severe vulvovaginal injuries include hemoperitoneum, pneumoperitoneum, retroperitoneal hematoma, and vaginal perforation.[7,14] Peritonitis from rupture of the posterior fornix of the vagina has also been reported, though very rare.[2]

Vulvovaginal lacerations may be isolated or multiple. Most lacerations will require only symptomatic therapy, but deeper wounds may require surgical care. One study reported that most women who presented to the emergency department with vulvovaginal lacerations required repair.[9] Women may delay presenting for emergent care resulting in significant blood loss. This delay may be due to fear, social stigma or simply embarrassment. The diagnosis of vulvovaginal laceration is often straightforward since most women present with significant bleeding and perineal pain. However, because

Table 3. Known predisposing factors for vulvovaginal laceration in consensual sexual intercourse patients (*n* = 69).

	No. of patients*
Alcohol intoxication	25 (36.2%)
'Rough or aggressive' intercourse	15 (21.7%)
Pre-existing vaginal infection	10 (14.4%)
Awkward positioning during intercourse	8 (11.6%)
Atrophic vagina in postmenopausal women	8 (11.6%)
First coitus	8 (11.6%)
Previous surgery	7 (10.1%)
Disproportion of male and female genitalia	3 (4.4%)
Insertion of foreign bodies	3 (4.4%)
Penile ornamentation	1 (1.5%)
Hx of pelvic radiation therapy	1 (1.5%)

*A total of17 patients (24.6%) had more than one predisposing factor documented.

Table 4. Known predisposing factors for genital trauma in nonconsensual sexual intercourse patients (N = 452).

	No. of patients*
Assailant known to victim	372 (82.3%)
Presence of nongenital injuries	239 (52.9%)
Alcohol or drug use by assailant	185 (40.9%)
Physical coercion	182 (40.3%)
Digital penetration	156 (30.2%)
Use of weapons	121 (26.8%)
Age between 17–19 years	104 (23.0%)
Postmenopausal women	63 (13.9%)
Multiple assailants	62 (13.7%)
Location of assault outdoors	53 (11.7%)
No prior sexual intercourse	46 (10.2%)
Pre-existing vaginal infection	37 (8.2%)
Insertion of foreign bodies	28 (6.2%)
Previous anogenital surgery	21 (4.6%)

*A total of 410 (90.7%) patients had more than one predisposing factor documented; 239 (52.9%) had more than two factors documented; and 191 (42.3%) had more than three factors documented.

severe lacerations are infrequently seen by clinicians, many centers do not have an organized treatment protocol.[14] Another important consideration is the poor visual estimation of blood loss, including the need for serial hemoglobin and hematocrit levels.[9] In our study, a falling hematocrit, uncontrolled bleeding, or a combined injury to the vulva and vagina increased the need for emergency surgery. Although rare, an expanding vulvar hematoma must be drained under anesthesia in order to prevent secondary infection and necrosis.[10,15]

In our initial population of adult NCSI victims, macroscopic lacerations were documented in 36.0% (517/1437). None of these lacerations in the NSCI group were severe enough to cause extensive bleeding or require surgical repair. However, the upper limits of the 95% confidence interval for this outcome were 0.6. This means that the true incidence of severe lacerations in our NCSI population could be as high as 0.6%. This is comparable to that reported by Geist who found vulvovaginal injuries in almost 50% of sexual assault cases, but only 1% needed surgical repair.[9] Similarly, in a retrospective study of more than 11 000 pediatric patients who were suspected of having been sexually abused or assaulted, only 11 cases requiring surgical repair were identified over a 20-year period at a tertiary care pediatric hospital.[16] One case-control study of 249 sexually assaulted women found that while 32% sustained anogenital injury, none required operative repair. Investigators concluded that the severity of the sexual assault was a poor predictor of injury.[17] A similar study of sexual assault in postmenopausal rape victims found that vulvovaginal lacerations occurred in 18.6%, with one in four severe enough to require suturing.[18] In summation of the literature on this subject, severe vulvovaginal lacerations following sexual assault are more uncommon than expected, but they can and do occur.[19] To postulate this may be due to CSI victims being more likely to present with more severe lacerations compared with NSCI due to the voluntary nature of CSI versus the added trauma and need for forensic examination associated with NCSI patients.

The location of vulvovaginal injuries in our NCSI patients were significantly different when compared with CSI patient (Table 2). Approximately half of the lacerations due to sexual assault were located on the fossa navicularis and posterior fourchette. This indicates that tears in the posterior fourchette, fossa, or along the long axis of the vagina may be more specific for forced vaginal intercourse.[11,20] In contrast, CSI patients had lacerations that were commonly located on the posterior vaginal wall, labia, and hymen. Injuries that required repair were generally located on the vagina walls, vaginal vault or posterior fourchette. Lacerations to the hymen are associated with younger sexual assault victims as well as those lacking prior sexual intercourse experience.[1] Both characteristics were more common among our CSI patients. Hymenal lacerations tend to be posterior (between the 5 o'clock and 7 o'clock positions) and to cause only minor bleeding and pain.[2] On rare occasions, hymenal lacerations may extend into the walls of the vagina, the perineal tendon or rectum, resulting in significant hemorrhage.[9]

Several known risk factors for vulvovaginal lacerations were present in our CSI population. In addition to those listed in Table 3, other studies have suggested previous anogenital surgery, congenital abnormalities of the vagina, friability of tissues, vaginal spasm, retroversion of the uterus, and clumsiness as etiologic factors.[1–3,6,14] Our study design prevents analysis of the contribution of each predisposing factor to the risk of developing vulvovaginal laceration. Details pertaining to the use of foreign bodies, coital positions, drug or alcohol use, and unusual activities during sexual intercourse should be elicited from the patient or from their partner. A patient might refer to a sexual act by its street name or use a euphemism to refer to a particular sex act. The clinician should clarify the meaning with the patient to accurately assess the possibility of injury. Even with a history of consensual intercourse, domestic abuse should be considered as a possible cause for anogenital injury and investigated.

Predisposing or etiologic factors specific for vulvovaginal lacerations following sexual assault in adult women are also not well defined in the literature. The variables listed in Tables 3 and 4 are characteristics associated with genital trauma in general and thus may predispose women to vulvovaginal lacerations.[9,10,11,12,13] The most consistently documented risk factors associated with genital injury are victim age (adolescent and postmenopausal), virginal status, the presence of a general body injury, foreign body, and multiple assailants. Preexisting vaginal infection and digital penetration were found to be associated with the presence of genital injury in one previously documented analysis.[6] Interestingly, sedative use or alcohol intoxication within hours of the sexual assault have been shown to be protective of anogenital injury.[21,22] It is prudent to notice that this broad assortment of risk factors is likely affected

by examiner training and experience, differences in injury definitions, patient population, as well as examination technique.

There is scant data in the current literature regarding risk factors for requiring surgical intervention after vulvovaginal laceration.[10] Women with lacerations less than 3 cm can be conservatively treated in the emergency department or clinic with local wound care, analgesia, suturing if needed, and antibiotics. In our study, clinical red flags that increased the need for operative intervention included uncontrolled bleeding, a falling hematocrit, or a combined injury to the vagina and vulva. All patients should undergo a rectal examination to ensure that the rectal mucosa is undamaged. A cystoscopic examination of the bladder and urethra may be necessary to rule out urinary tract injury. Sloin et al. created a guideline that provides a systemic approach to treating women with vulvovaginal lacerations.[14] According to these investigators, 'preparation for these uncommon emergencies may circumvent dangerous delays and inadequate examination and treatment.'[14]

LIMITATIONS

This study has several potential limitations that warrant consideration. Firstly, the sample population was drawn from emergency departments (EDs) of three urban medical centers, two rural community hospitals, and a free-standing nurse examiner clinic located in a single state of the Midwest United States. It is unknown how patient demographics and injury characteristics might differ in other settings or locations. Secondly, in this retrospective study we could not control for the differences in documentation or evaluations by different examiners. Five of the CSI patients had such profuse bleeding that a complete evaluation was not possible in the ED and the location and severity of vulvovaginal lacerations was determined in the operating room by gynecologic surgeons. Thirdly, although our examiners routinely use colposcopy with nuclear staining to document genital injuries including lacerations, we chose to count only macroscopic lacerations (those visualized directly without magnification). This made the clinical evaluation of the two groups (CSI vs. NCSI) more consistent. It is unlikely that that any microtrauma detected using the colposcope had any clinical significance.[22,23]

CONCLUSIONS

The assumption that genital injuries such as vulvovaginal lacerations are more severe in sexual assault cases is antiquated. It is important for the forensic clinician to recognize that physical findings in adult women who sustain vulvovaginal lacerations from consensual intercourse significantly differ from the findings in women who report sexual assault. In this community-based population, almost half of the women who presented to the ED and were found to have sustained a macroscopic vulvovaginal laceration following consensual intercourse required laceration repair in the ED or operating room. In contrast, while vulvovaginal lacerations were documented in over one-third of the women evaluated within 72 h after a sexual assault, none of these lacerations were severe enough to cause extensive bleeding or require repair. The predisposing factors and location of injury in victims of sexual assault were significantly different when compared with women presenting with lacerations due to consensual sexual intercourse. With an improved understanding of the types and severity of injuries sustained by NCSI and CSI, specifically vulvovaginal lacerations, clinicians can better anticipate clinical course, expedite diagnosis, and direct definitive management.

Conflict of interest and funding

The authors declare that they have no conflict of interest, and no funding was provided for this study.

Previous Presentation

This article was presented virtually at the Society of Academic Emergency Medicine Annual Meeting, May 2021, in Des Plaines IL.

REFERENCES

1. Jones JS, Rossman L, Hartmen M, Alexander C. Anogenital injuries in adolescents after consensual sexual intercourse. Acad Emerg Med 2003; 10(12): 1378–83. doi: 10.1197/S1069-6563(03)00555-4

2. Tchounzou R, Chichom-Mefire A. Retrospective analysis of clinical features, treatment and outcome of coital injuries of the female genital tract consecutive to consensual sexual intercourse in the Limbe Regional Hospital. Sex Med 2015; 3(4): 256–60. doi: 10.1002/sm2.94

3. Astrup BS, Ravn P, Lauritsen J, Thomsen JL. Nature, frequency and duration of genital lesions after consensual sexual intercourse – Implications for legal proceedings. Forensic Sci Int 2012; 219(1): 50–6. doi: 10.1016/j.forsciint.2011.11.028

4. Slaughter L, Brown CR, Crowley S, Peck R. Patterns of genital injury in female sexual assault victims. Am J Obstet Gynecol 1997; 176(3): 609–16. doi: 10.1016/S0002-9378(97)70556-8

5. McLean I, Roberts SA, White C, Paul S. Female genital injuries resulting from consensual and non-consensual vaginal intercourse. Forensic Sci Int 2011; 204(1): 27–33. doi: 10.1016/j.forsciint.2010.04.049
6. Lincoln C, Perera R, Jacobs I, Ward A. Macroscopically detected female genital injury after consensual and non-consensual vaginal penetration: a prospective comparison study. J Forensic Legal Med 2013; 20(7): 884–901. doi: 10.1016/j.jflm.2013.06.025
7. Muyshondt C, Monforte M, Droupy S. Sex-related injuries. Prog Urol 2013; 23(9):771–9. doi: 10.1016/j.jflm.2013.06.025
8. Jones JS, Dunnuck C, Rossman L, Wynn BN, Genco M. Adolescent Foley catheter technique for visualizing hymenal injuries in adolescent sexual assault. Acad Emerg Med 2003; 10(9): 1001–4. doi: 10.1197/S1069-6563(03)00321-X
9. Geist RF. Sexually related trauma. *Emerg Med Clin North Am 1998; 6(3):* 439–66. doi: 10.1016/S0733-8627(20)30540-X
10. Gambhir S, Grigorian A, Schubl S, Barrios C, Bernal B, Joe V, et al. Analysis on non-obstetric vaginal and vulvar trauma: risk factors for operative intervention. Updates Surg 2019; 71(4): 735–40. doi: 10.1007/s13304-019-00679-4
11. Jones ISC, O'Connor A. Non-obstetric vulval trauma. Emerg Med Australasia 2013; 25(1): 36–9. doi: 10.1111/1742-6723.12016
12. Sommers MS, Brown KM, Buschur C, Everett JS, Fargo JD, et al. Injuries from intimate partner and sexual violence: significance and classification systems. J Forensic Leg Med 2012; 19(5): 250–63. doi: 10.1016/j.jflm.2012.02.014
13. Sugar NF, Fine DN, Eckert LO. Physical injury after sexual assault: findings of a large case series. Am J Obstet Gynecol 2004; 190(1): 71–6. doi: 10.1016/S0002-9378(03)00912-8
14. Sloin MM, Karimian M, Ilbeigi P. Nonobstetric lacerations of the vagina. J Am Osteopath Assoc 2006; 106(5): 271–3. doi: 10.7556/jaoa.2006.106.5.271
15. Alcalde MVL, Hernández EH, Alfonso SB, Sánchez MJD. Non-obstetric traumatic vulvar hematoma: conservative or surgical approach? A case report. Case Rep Womens Health 2019; 22:e00109. doi: 10.1016/j.crwh.2019.e00109
16. Jones JG, Worthington T. Genital and anal injuries requiring surgical repair in females less than 21 years of age. J Pediatr Adolesc Gynecol 2008; 21(4): 207–11. doi: 10.1016/j.jpag.2007.10.010
17. Hilden M, Schei B, Sidenius K. Genitoanal injury in adult female victims of sexual assault. Forensic Sci Int 2005; 154(2–3): 200–5. doi: 10.1016/j.forsciint.2004.10.010
18. Ramin SM, Satin AJ, Stone IC Jr, Wendel GD. Sexual assault in postmenopausal women. Obstet Gynecol 1992; 80(5): 860–4.
19. Orellana-Campos C. Genital injuries: are they telling us something about sexual violence? Rev Bras Ginecol Obstet 2020; 42(2):106–13. doi: 10.1055/s-0040-1701465
20. Padoa A, Fishman NG, Tsviban A, Smorgick N. Vaginal postcoital injuries requiring surgical intervention: a case series and literature review. Int J Impot Res 2020; 33(1):110–117. doi: 10.1038/s41443-020-0234-8
21. Zilkens RR, Smith DA, Phillips MA, Mukhtar SA, Semmens JB, Kelly MC. Genital and anal injuries: a cross-sectional Australian study of 1266 women alleging recent sexual assault. Forensic Sci Int 2017; 275(1): 195–202. doi: 10.1016/j.forsciint.2017.03.013
22. Rogers A, McIntyre SL, Rossman L, Solis S, Bacon-Baguley T, Jones JS. The forensic rape examination: is colposcopy really necessary? Am J Emerg Med 2019; 37(5): 999–1000. doi: 10.1016/j.ajem.2018.10.015
23. Zink TM, Fargo JD, Baker RB, Buschur C, Fisher BS, Sommers MS. Comparison of methods for identifying ano-genital injury after consensual intercourse. J Emerg Med 2010; 39(1): 113–118. doi: 10.1016/j.jemermed.2008.08.024

Etiology and Clinical Features of Optic Neuritis in Two Children: A Case Report

Subah Nanda, BS[1], Amanda Schoonover, MPH[1], Rebecca Andrews-Dickert, MD[2], Jeffrey Jones, MD[3]*

[1]Spectrum Health Department of Emergency Medicine, College of Human Medicine, Michigan State University, Grand Rapids, MI, USA
[2]Spectrum Health Department of Emergency Medicine, College of Osteopathic Medicine, Sam Houston State University, Conroe, TX, USA
[3]Spectrum Health Department of Emergency Medicine, Grand Rapids, MI, USA

Corresponding Author: Jeffrey Jones; jones7@msu.edu

Background: *Optic neuritis (ON) is inflammation of the optic nerve that can occur in both adults and children. This disease is marked by a heterogeneous presentation in children and has clinical and epidemiologic characteristics that differ greatly from those found in adults. The purpose of this report is to illustrate the clinical features of ON that occur during childhood and to highlight the differences of ON in children versus adults. In doing so, we aim to add to the sparse current literature on this topic and help prevent the future misdiagnosis of ON in pediatric patients.*

Case Presentations: *An 11-year-old female presented with bilateral decreased visual acuity and significant ocular pain. The ophthalmic presentation and diagnostic workup led to the diagnosis of acute disseminated encephalomyelitis with ON. A second patient, a 12-year-old male, presented with decreased visual acuity and bilateral papilledema. Alongside a diagnosis of bilateral ON, a muscle biopsy confirmed mitochondrial cytopathy as the etiology of his presenting symptoms.*

Conclusions: *ON in children may be related to specific infections, autoimmune disorders, diseases of adjacent anatomical structures, or demyelinating disorders. Attacks may be acute or subacute with signs of reduced visual acuity, abnormal pupillary response, loss of color vision, impaired contrast sensitivity, and decreased peripheral vision. Awareness of this complex disease allows the clinician to initiate specific treatment and follow-up care that may reduce subsequent morbidity and the rate of recurrence.*

Keywords: *optic neuritis; diagnosis; clinical features; pediatric; case report*

INTRODUCTION

Optic neuritis (ON) is inflammation of the optic nerve that can affect patients at any age, including infancy. The disease is more common in adults with an incidence estimate of 1–2 per 100,000 people.[1] In contrast, the annual incidence of pediatric ON is only 0.15–0.57 per 100,000 people.[2–4] Although relatively rare in children, pediatricians and primary care clinicians should be aware of the specific features of pediatric ON in order to facilitate diagnostic testing and avoid misdiagnoses. Pediatric ON is usually associated with a good prognosis. However, a minority of children (22% in one study) will have persistent visual loss.[4] Knowledge regarding the clinical features, treatment, prognosis, and future neurologic implications of ON in children has grown significantly over the past decade.[5] These studies have shown that pediatric ON is a distinct clinical entity when compared to ON in adults.[3] Bilateral involvement, optic disc edema, and vision loss are more commonly seen in pediatric cases of ON when compared with adult cases of ON. Furthermore, orbital pain is less frequently reported in pediatric cases than in adult cases.[1] Since ON is a clinical diagnosis based on the history and physical examination findings, it is crucial for clinicians to be aware of these important clinical differences. This article presents two cases of ON in children with very different etiologies and highlights recent studies focused on the epidemiology and clinical features of pediatric ON.

CASE PRESENTATIONS

Case 1: An 11-year-old female in good general health presented to the emergency department (ED) with a complaint of viral symptoms that began 1 month earlier. There was no significant family history, chronic illness, or recent vaccinations. Her symptoms included headache, body aches, decreased visual acuity, eye pain, and fatigue. Gradually, most of her symptoms resolved, except for the eye pain and decreased visual acuity. Her ocular examination displayed hippus, blurred optic discs

bilaterally, and enlarged retinal vessels. Visual acuity was 20/800 in her right eye and 20/200 in her left. A relative afferent pupillary defect (RAPD) was observed in the right eye. On hospital admission, color vision (ishihara) was severely impaired in both eyes. Ocular movements were full range but associated with pain. Visual fields were intact. No conjunctival congestion or lid swelling was noticed. The anterior segments of both eyes was normal. The remaining sections of systemic and neurological examination were normal. Magnetic resonance imaging (MRI) of the orbits showed normal fibers within the orbits and in the optic nerves, but it also revealed abnormal fibers outside the orbits. A lumbar puncture was performed, which had a normal opening pressure of 22 cm. Cerebrospinal fluid (CSF) studies were negative, including bacterial and viral cultures, mycoplasma and viral serologies, antistreptolysin O titer, antinuclear antibody panel, C-reactive protein, and rheumatoid factor. The diagnosis of bilateral ON was made. The patient's symptoms improved over 5 days of intravenous methylprednisolone (30 mg/kg per day), and she was discharged home with the diagnosis of acute disseminated encephalomyelitis with ON. At her three-month follow up, she had full visual recovery. One year later, the patient again developed symptoms of eye pain and decreased visual acuity. Repeat MRI indicated persistent areas of demyelination of the optic chiasm and pre-chiasmic optic nerves. This was consistent with recurrent ON, which responded to steroids again. Due to the recurrence of the ON, the patient was referred for evaluation for multiple sclerosis (MS).

Case 2: A 12-year-old male presented to the ED with a history of 1 week of decreased vision in his right eye followed by decreasing vision in his left eye. There were no similar complaints in his past medical history. He also had no history of fever, symptoms of upper respiratory tract infection, recent vaccinations, bleeding tendencies, or trauma. His physical examination was significant for pain with lateral eye movements. The patient was unable to visualize symbols that were greater than 3 inches away. Color vision and red desaturation were reduced in both eyes. Visual field assessment was not performed because the patient grew restless. RAPDs were documented in both eyes. Fundoscopic examination with dilation revealed a loss of the optic discs in both eyes with 1- 2+ edema bilaterally. No conjunctival congestion or lid swelling was noticed. The anterior segments of both eyes were normal. The remainder of the systemic and neurological examination was normal. A lumbar puncture was performed, and it showed a normal opening pressure of 18 cm. Inpatient CSF studies (IgE, glucose, lactic acid, and electrophoresis for oligoclonal bands), CSF cultures, and Complete Blood Count (CBC) were all found to be normal. An MRI was obtained and found to be unremarkable. Amino acid studies revealed elevations of alanine and proline, which was concerning for primary lactic acidemia. The patient was started on intravenous methylprednisolone for 5 days (30 mg/kg per day) for presumed ON. This was followed by an oral corticosteroid taper over 2 weeks. A subsequent muscle biopsy showed significantly reduced activity of complex I and mildly reduced activity of complexes III and IV of the electron transport chain. This was consistent with Leber Hereditary Optic Neuropathy with associated bilateral ON. The patient was discharged on a mitochondrial cocktail of coenzyme Q10, thiamine, riboflavin, and carnitine. At his one-year follow-up, the patient had maintained his vision and his neurological examination was normal.

DISCUSSION

ON is a relatively rare condition in children, which can cause mild to severe vison loss. As demonstrated in these two cases, early recognition is important for diagnosis and prompt treatment. The diagnosis includes a wide range of inflammatory and demyelinating conditions associated with optic neuropathy (Table 1). ON can present in isolation or be the first manifestation of a chronic demyelinating illness, such as MS or neuromyelitis optica.[5] Secondary causes of ON are protean and include infections, diseases of the adjacent sinuses or orbital structures, trauma, vascular insufficiency, metastases, toxins, or nutritional deficiencies.[6–11] The exact pathogenesis of ON is not well understood. It is likely due to a delayed type IV hypersensitivity reaction induced by cytokines and other inflammatory mediators released from activated peripheral T-cells, which cross the blood–brain barrier and cause an autoimmune reaction.[4–7,13] However, the specific mechanism and target antigen(s) remain unknown. In many cases, direct injury to the axon may also play a role in the pathophysiology.[13] Emerging case reports indicate that COVID-19 is a rare, but potential cause of ON in both children and adults.[11,14,15] While the exact link between COVID-19 and ON is still under study, proposed mechanisms behind its pathogenesis include direct viral invasion, blood–brain barrier disruption, cytokine storm, autoimmunity, and coagulopathy.[16–18]

Table 1. Noninfectious etiologies of optic neuritis.[6–12]

Demyelinating diseases	Autoimmune	Drugs and chemicals	Miscellaneous
Multiple sclerosis	Sarcoidosis	Lead	Systemic vasculitis
Idiopathic inflammatory demyelination	Systemic lupus erythematosus	Methanol	Diabetes
Neuromyelitis optica	Sjögren's syndrome	Quinine	Vitamin A, B12 deficiencies
Acute disseminated encephalomyelitis	Behçet's disease	Arsenic	Tumor metastasis
Myelin oligodendrocyte glycoprotein autoantibody disease	Graves ophthalmopathy	Ethambutol	Bee and wasp stings
Hereditary neuropathies		Antibiotics	Leukemia
			Vaccination
			Sinusitis
			Sickle cell
			Trauma

Rapid determination of the underlying cause of ON is vital for implementing both timely and appropriate treatment in a child with acute vision loss. As seen in both cases described in this report, a diagnostic workup may require laboratory investigations and neuroimaging that extend beyond a simple history and ophthalmologic examination. Differentiating between various causes of ON may require serologic testing, CSF analysis and cultures, visual field perimetry, muscle biopsy, optical coherence tomography, or gadolinium-enhanced MRI of the brain and orbits.[6] Although MRI is not required to diagnose ON in children, it is the best imaging technique to confirm the diagnosis of acute demyelinating ON. One investigator has recommended that 'all children with ON should undergo neuroimaging not only to evaluate for other signs of demyelinating disease, but also to exclude the possibility of an intracranial lesion'.[4] Recent research in biomarkers, such as aquaporin-4 and myelin oligodendrocyte glycoprotein, may also be helpful in differentiating between infectious and autoimmune disease.[9,19] This information can then be used in counseling patients and their families about the disease, prognosis, and risk of recurrence.

Much of what we know about ON in childhood is based upon limited case series and retrospective reviews.[19,20] To date, there are no prospective clinical trials or published guidelines for children. Additionally, most early reports of pediatric ON focused on children older than 12 years of age.[20] However, it is generally accepted that pediatric ON has very different clinical features when compared to those found in adults (Table 2). For example, while the most common cause of ON in adults is demyelination, post-infectious or post-vaccination inflammation represents most cases of pediatric ON.[12,20,21] This was evident in our first case where ON was preceded by a viral syndrome. In general, these children will not require expensive laboratory studies, such as a lumbar puncture or an MRI, to make a diagnosis of ON.[22] Instead, a careful history should be aimed at detecting recent infections, vaccinations, or presence of vasculitis. In contrast with the adult presentation of ON, children are more likely to have bilateral disease, anterior optic nerve involvement with papillitis, and more severe vision loss on initial presentation.[21] Young children may not notice unilateral vision loss and may casually accept bilateral vision loss until it becomes incapacitating.[8] Eye pain, which is associated with ocular movements and may precede or coincide with the visual symptoms, occurs in more than 90% of adults.[21] Observational studies in children have demonstrated the absence of periocular pain in over half of cases in their pediatric study population.[21] However, both cases presented here did have eye movements associated with pain.

Table 2. Infectious etiologies of optic neuritis.[6–12]

Herpes Zoster	COVID-19
Borrelia	Rubella
Syphilis	Cytomegalovirus
Tuberculosis	Toxocariasis and helminths
Toxoplasmosis	Flavivirus
Leptospirosis	Adenovirus
Mononucleosis	Coxsackievirus
Brucella	Bartonella
Pertussis	Streptococcus
Varicella	

Table 3. Optic neuritis in adults versus children.[1,4,7,8,11,12,14,15]

Adult	Pediatric
1–2 per 100,00 incidence	0.15–0.57 per 100,00 incidence
Mean age 31.8 years	Mean age 9.5 years
Most common cause is demyelination	Most often postinfectious or postimmunization
Pain with eye movements (90%)	Headache (53%)
Unilateral disease (70%)	Bilateral disease (72%)
Visual acuity <20/200 (36%)	Visual acuity <20/200 (90–95%)
Female preponderance 2:1	Female preponderance 1:1 prepuberty
Optic disc swelling or papillitis (35%)	Optic disc swelling or papillitis (64–87%)
Retrobulbar (65%)	Retrobulbar (13–36%)
Risk of multiple sclerosis 38%	Risk of multiple sclerosis 10–29%
Visual recovery (>20/40) 90–95%	Visual recovery (>20/40) 80–89%

A careful ophthalmologic examination may help to differentiate a typical presentation of ON from atypical cases (Table 3). Vision loss occurs over a period of hours to days, peaking within several weeks of symptom onset.[12] In our patients, dyschromatopsia or color desaturation, was found to be a sensitive sign of acute ON.[12] Ishihara color plates were used to assess red color desaturation. However, if these plates are unavailable, simply ask the patient to compare the color of a bright red object with each eye. Furthermore, physicians should assess contrast sensitivity by shining a light in each eye and asking the patients to compare the brightness. A useful technique is to ask, 'If I were to give you a dollar for this brightness' (shine light in normal eye), 'How much would you give me for this ...?' (shine light in affected eye). Lastly, physicians ought to perform a confrontational visual field test, specifically looking for central or paracentral scotomas.[23] It may be helpful to say to children, 'If what you see is like a television screen, then where is the part that is missing?' Scotomas typically occur over the course of a few hours to days, with maximum defects reached within several days.[23]

Both patients had dilated fundoscopic examinations performed to avoid missing other retinal diseases that could be mistaken for ON, such as retinal detachment. Before dilating the pupils, determine the presence or absence of a RAPD. This can be demonstrated with the swinging flashlight test, which is performed by moving a penlight back and forth between the eyes. The afferent pupillary defect becomes obvious when stimulation of the normal eye elicits a brisk constriction of both pupils, while stimulation of the diseased eye causes dilation of both pupils. Note that in bilateral involvement, the RAPD may not be apparent. One observational study of ON in children reported the presence of color vision defects in 50% of children, visual field defects in 58.5%, and RAPD in 67%.[24] The most important fundoscopic finding in ON is the presence of optic disc edema, known as papillitis. Most children with ON have optic disc edema, where as with only 35% of adults with ON present in this manner.[18] This was documented in both of our patients. However, optic disc edema can be difficult to visualize in a fidgeting child. In addition to blurred disc margins, look for optic disc pallor, filling-in of the physiological disc cup, elevation of the optic disc, distended retinal veins, and dilated disc capillaries. Of these features, evidence of thickening of the peripapillary nerve fiber layer was found to provide the highest level of accuracy as a single sign of optic disc edema.[25] Ocular point-of-care ultrasound can also enable pediatricians to detect optic disc elevation and abnormal optic nerve sheath diameter at the bedside, thereby expediting the diagnosis of ON.[26]

Fortunately, both patients had none of the signs and symptoms that might indicate a more serious pathology. Hemorrhages of the nerve fiber layer, detectable by fundoscopic examination, are rare in children with ON and should prompt an investigation to exclude other diagnoses. Additional clinical features that indicate a more serious pathology include insidious onset, progressive visual loss for more than 2 weeks, painless visual loss, severe optic nerve pallor at presentation, marked uveitis or retinal periphlebitis, slow visual recovery, ongoing neurologic symptoms, and any deterioration after withdrawal of steroids.[10,23] These clinical red flags require careful diagnostic assessment of other diseases that are associated with optic neuropathy (Table 4). However, subclinical or spontaneously resolving disease

Table 4. Clinical features of pediatric optic neuritis.[1–7,12,13,23]

- Preceding viral illness or febrile prodrome
- Acute or subacute visual loss in one or both eyes
- Progressive over days to several weeks
- Photopsias precipitated with eye movement
- Abnormal color vision (notably red color desaturation)
- Reduced contrast sensitivity
- Visual field defect (central or paracentral scotomas)
- Periocular pain and pain with eye movement
- Relative afferent pupillary defect (Marcus Gunn pupil)
- Normal (retrobulbar) or swollen optic disc (papillitis)
- Normal macula and peripheral retina
- Uveitis or retinal periphlebitis possible

states, atypical symptoms, recurrent isolated attacks, and a poor history provided by young children can make diagnosis of ON challenging. In response to this diagnostic complexity, Yeh and colleagues have created a comprehensive algorithm for the approach to a child presenting with acute ON.[5] This approach was utilized in our patients and consisted of an ophthalmologic examination, lumbar puncture, MRI of the brain, and a broad rheumatologic workup.

The first patient described here was referred for a MS workup after a recurrent episode of ON. MS is an immune-mediated demyelinating disorder that attacks myelinated axons in the central nervous system, which leads to significant physical disability.[10] The relationship between an initial episode of ON and the development of MS has been established by many previous studies, none of which are prospective. The reported risk varies greatly from 13 to 36% of children.[5] Lucchinetti et al studied childhood ON and estimated the risk of MS to be 13% by 10 years of age, 19% by 20 years of age, 22% by 30 years of age, and 26% by 40 years of age.[27] A meta-analysis by Waldman et al showed that for every one year increase in age, the odds of a child developing MS after an initial episode of ON increased by 32%.[28] In addition, the risk of progression to MS was markedly increased (27-fold) with the presence of demyelinating lesions on brain MRI scans. Other reported risk factors for MS include racial or geographic factors, female gender, and recurrent ON.[24,29] However, prospective studies are necessary to support these findings and address the conflicting data concerning pediatric ON.

Treatment of ON is aimed at identifying and treating the underlying cause of the condition. However, even without treatment, it was observed that 80% of children spontaneously recovered their vision within 2–3 weeks.[24] Even if a child's vision does recover, pediatric eyes may retain some functional defects in visual fields, low-contrast vision, and color perception.[30] ON treatment guidelines are based on large-scale studies in adult patients and a preferred protocol has been developed from the optic neuritis treatment trial (ONTT).[1] These treatments often include intravenous methylprednisolone (4–30 mg/kg per day) for 3–5 days followed by an oral corticosteroid tapered over 2 weeks.[24,29] Both patients presented here were initially treated with IV methylprednisolone (30 mg/kg per day), with differing subsequent treatment due to the varying pathologies and causes of their ON. Physicians will likely tailor their decision to treat based on age, gender, laterality, and level of visual acuity. If initial treatment with steroids is not successful, additional treatment options include a second round of intravenous steroids, intravenous immunoglobulin, or plasma exchange.[29] Factors that may predict a poor recovery in children include age over 10 years old, profound loss of visual acuity at presentation, optic atrophy, a diagnosis of MS, and bilateral involvement at presentation.[30–32] Both children presented here were over 10 years old and presented with severely decreased visual acuity bilaterally. Although they initially recovered, the first child was referred for evaluation for MS.

CONCLUSIONS

ON is a complex and challenging disease in children. It has specific clinical epidemiologic characteristics that differ from those found in adults. Blurred vision and headache may be the first and only presenting symptoms of a systemic disease, such as MS, serious infection, or an underlying mitochondrial abnormality. In the two cases presented here, early recognition of the disease was made after a thorough ophthalmologic examination, CSF analysis, MRI imaging, and rheumatologic studies. Prompt diagnosis of this complex disease allows clinicians to initiate corticosteroids and arrange follow-up care that may reduce subsequent morbidity in this vulnerable population.

Conflict of interest and funding

The authors declare that they have no conflict of interest and that no funding was provided for this case report.

REFERENCES

1. Lock JH, Newman NJ, Biousse V, Peragallo JH. Update on pediatric optic neuritis. Curr Opin Ophthalmol 2019 Nov; 30(6): 418–25. doi: 10.1097/ICU.0000000000000607

2. Toosy AT, Mason DF, Miller DH. Optic neuritis. Lancet Neurol 2014 Jan; 13(1): 83–99. doi: 10.1016/S1474-4422(13)70259-X
3. Brady KM, Brar AS, Lee AG, Coats DK, Paysse EA, Steinkuller PG. Optic neuritis in children: clinical features and visual outcome. J AAPOS 1999 Apr; 3(2): 98–103. doi: 10.1016/S1091-8531(99)70078-9
4. El-Dairi MA, Ghasia F, Bhatti MT. Pediatric optic neuritis. Int Ophthalmol Clin 2012 Summer; 52(3): 29–49. doi: 10.1097/IIO.0b013e318259dec6
5. Yeh EA, Graves JS, Benson LA, Wassmer E, Waldman A. Pediatric optic neuritis. Neurology 2016 Aug 30; 87(9 Suppl 2): S53–8. doi: 10.1212/WNL.0000000000002822
6. Bennett JL. Optic neuritis. Continuum (Minneap Minn) 2019 Oct; 25(5): 1236–64. doi: 10.1212/CON.0000000000000768
7. Günbey C, Konuşkan B. Optic neuropathies in childhood: a review of etiology and treatment. Turk J Pediatr 2019; 61(4): 471–6. doi: 10.24953/turkjped.2019.04.001
8. Hoorbakht H, Bagherkashi F. Optic neuritis, its differential diagnosis and management. Open Ophthalmol J 2012; 6: 65–72. doi: 10.2174/1874364101206010065
9. Collinge JE, Sprunger DT. Update in pediatric optic neuritis. Curr Opin Ophthalmol 2013 Sep; 24(5): 448–52. doi: 10.1097/ICU.0b013e3283641b86
10. Voss E, Raab P, Trebst C, Stangel M. Clinical approach to optic neuritis: pitfalls, red flags and differential diagnosis. Ther Adv Neurol Disord 2011; 4(2): 123–34. doi: 10.1177/1756285611398702
11. Parvez Y, AlZarooni F, Khan F. Optic neuritis in a child with COVID-19: a rare association. Cureus 2021 Mar 24; 13(3): e14094. doi: 10.7759/cureus.14094
12. Herce HH, Sathappan VS, Ing EB. Childhood optic neuritis, in Medscape. 2021. Available from: https://emedicine.medscape.com/article/1217290-overview [cited 17 January 2022].
13. Pau D, Al Zubidi N, Yalamanchili S, Plant GT, Lee AG. Optic neuritis. Eye (Lond) 2011 Jul; 25(7): 833–42. doi: 10.1038/eye.2011.81
14. Sinha A, Dwivedi D, Dwivedi A, Bajaj N. Optic neuritis as a presenting symptom of post-COVID-19 multisystem inflammatory syndrome in children (MIS-C). Indian J Pediatr 2021 Dec; 88(12): 1269. doi: 10.1007/s12098-021-03921-3
15. François J, Collery AS, Hayek G, Sot M, Zaidi M, Lhuillier L, et al. Coronavirus disease 2019-associated ocular neuropathy with panuveitis: a case report. JAMA Ophthalmol 2021 Feb 1; 139(2): 247–9. doi: 10.1001/jamaophthalmol.2020.5695
16. Baig AM, Sanders EC. Potential neuroinvasive pathways of SARS-CoV-2: deciphering the spectrum of neurological deficit seen in coronavirus disease-2019 (COVID-19). J Med Virol 2020 Oct; 92(10): 1845–57. doi: 10.1002/jmv.26105
17. Shah S, Danda D, Kavadichanda C, Das S, Adarsh MB, Negi VS. Autoimmune and rheumatic musculoskeletal diseases as a consequence of SARS-CoV-2 infection and its treatment. Rheumatol Int 2020 Oct; 40(10): 1539–54. doi: 10.1007/s00296-020-04639-9
18. Seah I, Agrawal R. Can the coronavirus disease 2019 (COVID-19) affect the eyes? A review of coronaviruses and ocular implications in humans and animals. Ocul Immunol Inflamm 2020 Apr 2; 28(3): 391–5. doi: 10.1080/09273948.2020.1738501
19. Lehman SS, Lavrich JB. Pediatric optic neuritis. Curr Opin Ophthalmol 2018 Sep; 29(5): 419–22. doi: 10.1097/ICU.0000000000000509
20. Pérez-Cambrodí RJ, Gómez-Hurtado Cubillana A, Merino-Suárez ML, Piñero-Llorens DP, Laria-Ochaita C. Optic neuritis in pediatric population: a review in current tendencies of diagnosis and management. J Optom 2014 Jul–Sep; 7(3): 125–30. doi: 10.1016/j.optom.2013.12.008
21. Boomer JA, Siatkowski RM. Optic neuritis in adults and children. Semin Ophthalmol 2003 Dec; 18(4): 174–80. doi: 10.1080/08820530390895172
22. Sachdeva V, Kekunnaya R. Working up a child with demyelinating optic neuritis: striking a balance! Oman J Ophthalmol 2021 Jun 28; 14(2): 74–7. doi: 10.4103/ojo.ojo_105_21
23. Gise RA, Heidary G. Update on pediatric optic neuritis. Curr Neurol Neurosci Rep 2020 Mar 3; 20(3): 4. doi: 10.1007/s11910-020-1024-x
24. Chang MY, Pineles SL. Pediatric optic neuritis. Semin Pediatr Neurol 2017 May; 24(2): 122–8. doi: 10.1016/j.spen.2017.04.004
25. Carta A, Favilla S, Prato M, Bianchi-Marzoli S, Sadun AA, Mora P. Accuracy of funduscopy to identify true edema versus pseudoedema of the optic disc. Invest Ophthalmol Vis Sci 2012; 53(1): 1–6. doi: 10.1167/iovs.11-8082
26. Pancer Z, Tessaro MO, Vo A. Adolescent male with eye pain. Ann Emerg Med 2021 Oct; 78(4): 567–76. doi: 10.1016/j.annemergmed.2021.04.003
27. Lucchinetti CF, Kiers L, O'Duffy A, Gomez MR, Cross S, Leavitt JA, et al. Risk factors for developing multiple sclerosis after childhood optic neuritis. Neurology 1997 Nov; 49(5): 1413–8. doi: 10.1212/WNL.49.5.1413
28. Waldman AT, Stull LB, Galetta SL, Balcer LJ, Liu GT. Pediatric optic neuritis and risk of multiple sclerosis: meta-analysis of observational studies. J AAPOS 2011 Oct; 15(5): 441–6. doi: 10.1016/j.jaapos.2011.05.020
29. Borchert M, Liu GT, Pineles S, Waldman AT. Pediatric optic neuritis: what is new. J Neuroophthalmol 2017 Sep; 37(Suppl 1): S14–22. doi: 10.1097/WNO.0000000000000551
30. Writing Committee for the Pediatric Eye Disease Investigator Group (PEDIG), Pineles SL, Repka MX, Liu GT, Waldman AT, Borchert MS, et al. Assessment of pediatric optic neuritis visual acuity outcomes at 6 months. JAMA Ophthalmol 2020 Dec 1; 138(12): 1253–61. doi: 10.1001/jamaophthalmol.2020.4231
31. Averseng-Peaureaux D, Mizzi M, Colineaux H, Mahieu L, Pera MC, Brassat D, et al. Paediatric optic neuritis: factors leading to unfavourable outcome and relapses. Br J Ophthalmol 2018 Jun; 102(6): 808–13. doi: 10.1136/bjophthalmol-2016-309978
32. Ambika S, Padmalakshmi K, Venkatraman V, Noronha OV. Visual outcomes and clinical manifestations of pediatric optic neuritis in Indian population: an institutional study. J Neuroophthalmol 2018 Dec; 38(4): 462–5. doi: 10.1097/WNO.0000000000000646

Do Probiotics Prevent Gestational Diabetes in Obese/ Overweight Patients? A Systematic Review

Christina Carfagnini*

Saba University School of Medicine, The Bottom, Saba, The Caribbean Netherlands

*Corresponding Author: Christina Carfagnini; c.carfagnini_037@saba.edu

Introduction: *While some studies suggest probiotic supplements may prevent Gestational Diabetes Mellitus (GDM), it is unclear if probiotics effectively prevent GDM among overweight and obese patients. This systematic review synthesizes recommendations for clinical practice and future research by evaluating the quality of evidence regarding* Lactobacillus *and* Bifidobacterium *containing probiotics to prevent GDM among obese and overweight patients.*

Methods: *PubMed, Embase, CINAHL, and Web of Science were searched using appropriate MeSH terms. Results were limited to randomized controlled trials published between 2011 and 2021. Titles and abstracts were screened for relevance after duplicates were removed. The inclusion criteria were as follows: studies that diagnosed GDM according to the International Association of Diabetes and Pregnancy Study Group criteria, suspended probiotic use prior to intervention, excluded participants with altered glucose metabolism, included participants with a body mass index ≥ 25 kg/m^2, and provided a specified dose of probiotic supplements. Articles without statistical analysis were excluded. Resulting articles were critically appraised using Version 2 of the Cochrane Risk of Bias tool.*

Results: *This search strategy resulted in 24 articles after duplicates were removed. Five double-blind randomized controlled trials found that the incidence of GDM during the third trimester was not significantly different between probiotic and control groups. There was wide variation in the bacterial species, dose, and duration of probiotic treatments used. All studies have a high risk of bias due to non-adherence to the treatment.*

Discussion: *This review used highly sensitive criteria for GDM diagnosis that may mask a preventative effect of probiotics. Noncompliance may bias results toward the null, given insufficient analysis of the effect of adhering to the intervention. No patterns between the length of probiotic intervention or probiotic species and improved glucose tolerance were noted.*

Conclusions: *Current evidence is not sufficient to recommend probiotic supplements to prevent GDM in overweight and obese patients. Future evidence should address the effect of adhering to probiotic interventions and develop consistent probiotic intervention protocols.*

Keywords: *dietary supplements; glucose metabolism disorders; obesity; probiotics; gestational diabetes mellitus*

INTRODUCTION

The rising prevalence of Gestational Diabetes Mellitus (GDM) represents a public health concern. During a healthy pregnancy, increased human placental lactogen and decreased insulin receptor substrate-1 (IRS-1) tyrosine phosphorylation decrease maternal insulin sensitivity to improve glucose availability for the fetus.[1,2] GDM refers to pathological insulin resistance that occurs when pancreatic β-cells cannot produce sufficient insulin to compensate for this increased demand.[3] GDM is associated with reduced tyrosine phosphorylation of IRS-1 and the insulin receptor's intracellular domain that attenuates insulin response and persists after delivery.[2] GDM affected 6% of pregnancies in the United States from 2012 to 2016,[4] which increased from 0.3% of pregnancies from 1979 to 1980.[4,5] Despite the development of more sensitive diagnostic criteria, the rise in GDM has been attributed to the increasing prevalence of obesity as GDM is more common among patients whose body mass index (BMI) classifies as being overweight (BMI: 25–29.9 kg/m^2) or obese (BMI ≥ 30 kg/m^2).[4]

The impact of GDM on healthcare spending and patient outcomes demands novel preventative measures. The total economic burden of GDM and complications was approximately $1.6 billion in the United States during 2017.[6] Neonatal complications from GDM include macrosomia, respiratory distress syndrome, and hyperbilirubinemia.[7–9] Neonates may also develop seizures, obesity, and metabolic syndrome later in their life.[7,8] Maternal long-term complications include type-2 diabetes mellitus and metabolic syndrome.[8] While diet and exercise can prevent GDM,[10] multiple reviews found no significant changes in the incidence of GDM when

patients of diverse BMI categories were assigned to lifestyle interventions.[11–13] These results were attributed to poor self-efficacy and low adherence to recommendations, especially for overweight and obese patient populations with a higher prevalence of GDM.[4]

Host–microbiome interactions that influence glucose metabolism suggest modulating the gut microbiome using probiotics could be a novel target to prevent GDM. Hasain and colleagues have previously reviewed the molecular pathway, by which gram-negative bacteria and lipopolysaccharides (LPS) in the colon can attenuate insulin signaling.[14] The mechanism by which *Bifidobacterium*[14] and *Lactobacillus*[15] modulate this pathway and improve glucose metabolism has also been described. This basic science research aligns with results from a randomized controlled trial (RCT) that found a statistically significant decreased risk of GDM among women who received *Lactobacillus rhamnosus* and *Bifidobacterium lactis* containing probiotics compared to those who did not, with a risk ratio of 0.37 (95% confidence interval [CI]: 0.15–0.89).[15] The majority of participants in this study had a healthy BMI, which limits the generalizability of this trial due to the high prevalence of obesity in the United States.[16] This discrepancy is especially important, as obesity is associated with changes in the gut microbiome and a greater risk of GDM due to increased insulin resistance and low-grade inflammation.[17]

It is unclear if probiotics effectively prevent GDM among overweight and obese patients. A meta-analysis-pooled results found probiotics did not significantly prevent GDM among obese and overweight patients, as diagnosed by International Association of Diabetes and Pregnancy Study Groups (IADPSG) criteria.[18] However, the substantial methodological diversity across RCTs limits the utility of this meta-analysis. This systematic review aims to synthesize recommendations for clinical practice and future research by evaluating the quality of evidence regarding *Lactobacillus* and *Bifidobacterium* containing probiotics to effectively prevent GDM, as defined by the IADPSG criteria, among obese and overweight patients.

METHODS

PubMed, Embase, CINAHL, and Web of Science were searched for articles investigating the incidence of GDM among overweight or obese women receiving probiotics. PubMed was searched using the following MeSH terms: probiotic, *Lactobacillus*, *Bifidobacterium*, overweight, obesity, gestational diabetes, gestational, and gestational weight gain (GWG). CINAHL, EMBASE, and Web of Science were searched using the same terms. Results were limited to RCTs published within the last 10 years. Literature searches resulted in 24 articles after duplicates were removed, as illustrated in Fig. 1. Titles and abstracts were screened for relevance to the research question. Full texts of relevant articles were reviewed for inclusion and exclusion criteria by the author of this review. Studies were included if:

1. GDM was diagnosed according to IADPSG criteria.[19] It is important to use IADPSG criteria as it provides a consistent criterion to compare research produced in different countries. This is the most sensitive definition of GDM, which will lead to more conservative results for prevention studies.
2. Participants received a specified dose of probiotics with a noted species of *Lactobacillus* or *Bifidobacterium*, or a placebo/vehicle control.
3. All participants had a BMI ≥ 25 kg/m^2.
4. Probiotic use prior to the intervention was suspended.
5. Participants on medications or diagnosed medical conditions that alter glucose metabolism before the intervention period were excluded from the trial.

Articles were excluded if they did not provide results using statistical tests appropriate to the study design. Inclusion and exclusion criteria resulted in five selected articles, as shown in Fig. 1.

The first author assessed the quality of evidence using Version 2 of the Cochrane Risk of Bias tool for randomized controlled trials (RoB2) after studying the RoB2 guidance document.[20] A narrative synthesis focused on critical appraisal and methodological heterogeneity of probiotic interventions was completed. An institutional review board was not needed, given the nature of a systematic review.

Study Tabulations and Outcomes Measured

The incidence of GDM during the third trimester of pregnancy was the primary outcome in this review. Differences between intervention and control groups were determined by appropriate statistical tests for each study design, specifically odds ratios,[21,22] relative risks,[23] unpaired t-test,[24] and ANOVAs.[25] The included primary studies completed statistical analysis using SPSS Version 21,[21] SPSS Statistics 23,[22] SAS 9.4,[23,25] and R 3.2.3.[24]

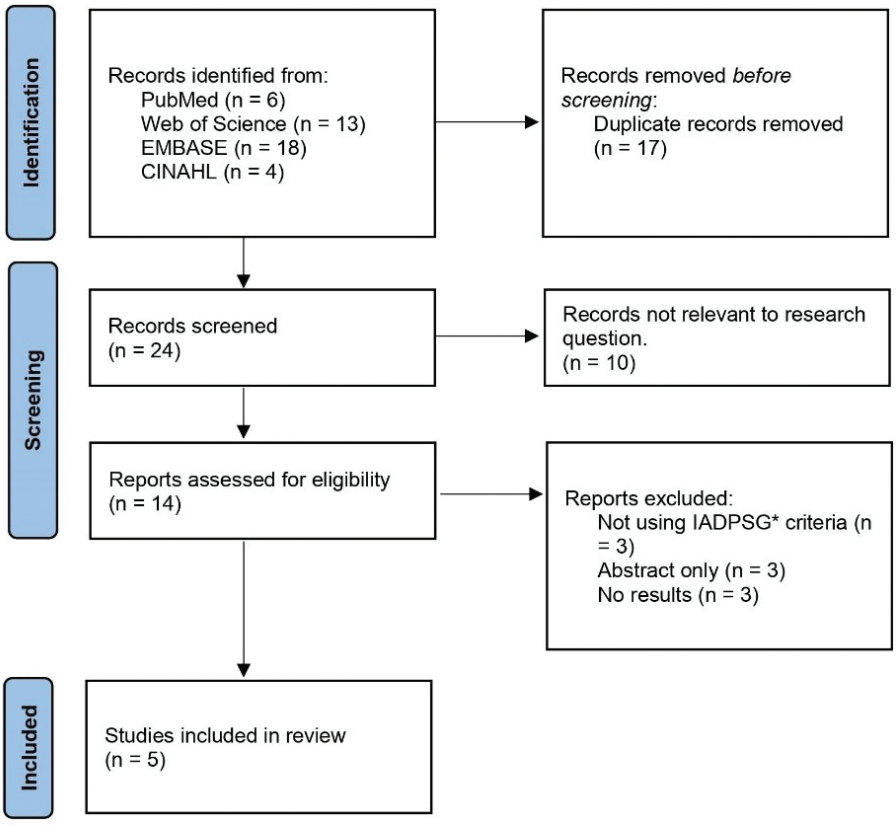


Figure 1. PRISMA flow diagram. Adapted from PRISMA 2020 Statement.[38] *International Association of Diabetes and Pregnancy Study Groups (IADPSG).

Statistical significance was defined by $p < 0.05$. The odds ratio of developing GM in the probiotic group compared to placebo/vehicle controls along with 95% CIs is listed in Table 1.

RESULTS

Probiotic Supplements on GDM

This literature search resulted in five RCTs summarized in Table 1. Sample sizes ranged from 49 to 460 participants. Three studies included overweight and obese patients.[21,22,25] Two studies included obese patients only.[23,24] All studies included patients in their second trimester ranging from 12 to 22 weeks of gestation as per the inclusion criteria. Three studies compared the incidence of GDM among patients who received probiotic supplements and placebo/vehicle controls.[21,22,24] One study compared the incidence of GDM among those treated with a combination of probiotics and fish oil, probiotics only, fish oil only, or a double placebo control.[25] Another study compared the incidence of GDM among patients treated with probiotic capsules, placebo capsules, dietary interventions, and routine dietary advice.[23] Three studies reported a decrease in the incidence of GDM among the intervention compared to the control group,[21,23,25] while two studies found an increase in GDM. Of the two studies that included obese women only, one study found decreased rates of GDM among the probiotic group compared to placebo controls,[23] while the other found an increase.[24] The incidence of GDM among those treated with probiotic supplements compared to placebo controls was not statistically significant in all five studies, as outlined in Table 1.

Probiotic Supplements on Maternal Outcomes

Although GDM results were not significant for all articles included in this review, there were several significant secondary outcomes. One study found fasting

Table 1. Evidence table of five randomized controlled trials that investigate if *Lactobacillus* or *Bifidobacterium* containing probiotics administered to overweight or obese patients can effectively prevent Gestational Diabetes Mellitus according to International Association of Diabetes and Pregnancy Study Group.

First author, date of publication	Study design	Study population			Therapy		Outcome
		Sample size (*n*)	Pre-pregnancy BMI (kg/m^2)	Gestational age (weeks)	Probiotic exposure	Comparison groups	Odds ratio of GDM in the probiotic group compared to placebo controls (95% confidence interval)
Asgharian, 2020	RCT	130	≥25	20–22	100 g probiotic yogurt	100 g conventional yogurt**	0.5 (0.2–1.5)
Halkjəer, 2020	RCT	49	≥30, ≤35	12–14	Probiotic capsules	Placebo capsules	2.1 (0.4–12.7)
Callaway, 2019	RCT	411	≥25	<20	Probiotic capsules	Placebo capsules	1.62 (0.9–2.8)
Okesene-Gafa, 2019	RCT	460	≥30	12–17	Probiotic capsules	Dietary intervention***, routine dietary advice, placebo capsule	0.95 (0.2–4.9)
Pellonperä, 2019	RCT	439	≥25	<18	Probiotic capsules	Fish oil**** and probiotic capsules, probiotic and placebo capsules, fish oil and placebo capsules, placebo capsules	0.84 (0.5–1.5)

BMI, body mass index; GDM, Gestational Diabetes Mellitus; RCT, randomized controlled trial.
**Both probiotic and conventional yogurt contained *Streptococcus* thermophilus and *Lactobacillus* delbrueckii subsp. bulgaricus in the dosage of 107 CFU/g (colony forming units per gram) used for biotransformation of milk (as starter).
***HUMBA handbook with information about healthy nutritious foods, recipes, unhealthy drinks, managing cravings, and ways to be more physically active. In addition, they received 4 home-based education sessions by a community health worker. Women in the dietary intervention also received motivational text messages 3 times weekly from randomization until birth.
****The fish oil capsules contained a total of 2.4 g of n-3 fatty acids (1.9 g docosahexaenoic acid [22:6n-3] [DHA] and 0.22 g eicosapentaenoicacid [20:5 n-3]).

plasma glucose (FPG) was 4 mg/dL lower (95% CI: – 6.9, – 1.1) at 28 weeks gestation among the probiotic yogurt group compared to placebo controls when adjusted for baseline FPG and BMI, which was significant according to Analysis of Covariance (ANCOVA) (p = 0.008).[21] The mean 2-h oral glucose tolerance of 103.9 ± 21.0 mg/dL in the probiotic group was significantly lower than the mean 2-h oral glucose tolerance of 115.5 ± 26.3 g/dL in the vehicle control group at 28 weeks of gestation according to ANCOVA tests (p = 0.002).[21] Oral glucose tolerance at 1 h was not significantly different between the two groups in this study.[21] Another study found that the mean FPG of 77.5 ± 8.1 mg/dL among women who received probiotic supplements was significantly higher than the mean FPG of 79.3 ± 9.0 mg/dL among the placebo group at 28 weeks gestation, according to a general linear model adjusted for clinical center and BMI category (25–29 kg/m^2, 30–39 kg/m, and ≥40 kg/m^2) (p = 0.049).[22] However, these changes are not clinically significant as changes in FPG do not cross the threshold for GDM diagnostic criteria. In addition, oral glucose tolerance at 1- and 2-h time points was not statistically significant between groups in this study.[22] A significant interaction between previous GDM and the probiotic group with respect to change in FPG was found in another study.[25] The change in blood glucose concentration was significantly different among the probiotic and fish-oil groups depending on the duration of the intervention or pre-pregnancy BMI after excluding women with GDM in early pregnancy.[25] There were no significant differences between FPG and oral glucose tolerance found in the two other studies.[23,24]

There were inconsistent results regarding the effect of probiotic supplements on GWG and preeclampsia. One study found that 33% of participants in the probiotic group experienced excessive GWG, which was significantly lower than 85% of participants receiving

placebos at 28 weeks of gestation according to a binary logistic regression adjusted for clinical center and BMI category.[22] Mean GWG was not significantly different between the two groups. No study found a significant change in GWG between the probiotic and placebo-control groups. One study reported 10% of participants with preeclampsia in the probiotic-treated group were greater than the 5% of controls according to a binary logistic regression adjusted for clinical center and BMI category ($p = 0.09$).[22] However, no other study found a significant difference between preeclampsia or gestational hypertension.[21,23–25]

Methodological Heterogeneity Between Study Designs

Methodological diversity between study designs included in this review is summarized in Table 2. *Lactobacillus rhamnosus* and *Bifidobacterium lactis* were the most common strains of bacteria contained in probiotic supplements. Two studies that used these species observed a non-significant decrease in the incidence of GDM among the probiotic group compared to placebo controls,[23,25] while one study observed an increased rate of GDM among the intervention group.[22] One study provided the intervention group with probiotics containing *Lactobacillus acidophilus* and *Bifidobacterium lactis* and observed a non-significant decrease in GDM rates compared to placebo controls.[21] This study provided probiotics via yoghurt compared to vehicle control created with the same starting bacteria. The other four studies provided probiotics using capsules. One study provided the intervention group with a multi-strain probiotic supplement containing strains of *Streptococcus*, *Bifidobacterium*, and *Lactobacillus*. This study found a non-significant increase in the incidence among GDM between groups.[24]

Table 2. Heterogeneity between study designs.

First author, date of publication	Bacterial strains in probiotics	Vehicle of administration	Daily dose (CFU)	Duration of probiotic therapy	Weeks of gestation at GDM diagnosis (weeks)	Self-reported compliance for probiotics definition
Asgharian, 2020	*Lactobacillus acidophilus La5, Bifidobacterium lactis* Bb12	Yogurt	5×10^8	24 weeks of gestation until delivery	28	Not specified
Halkjəer, 2020	Vivomixx*	Capsules	4.5×10^{11}	14–20 weeks of gestation until delivery	27–30	Self-reported >80% capsule intake
Callaway, 2019	*Lactobacillus rhamnosus (LGG), Bifidobacterium animalis subspecies lactis (BB-12)*	Capsules	$>1.0 \times 10^9$	Enrollment until delivery**	28	Self-reported capsule use, verified by the presence of bifidobacterium DNA tested by end-stage PCR in fecal sample taken at 28 weeks of gestation
Okesene-Gafa, 2019	*Lactobacillus rhamnosus GG, Bifidobacterium lactis (BB-12)*	Capsules	$>6.5 \times 10^9$	Enrollment until delivery***	26–28	Self-reported >75% capsule intake
Pellonperä, 2019	*Lactobacillus rhamnosus HN001, Bifidobacterium animalis ssp. lactis 420*	Capsules	1.0×10^{10}	First study visit**** until 6 months postpartum	24–28	Self-reported capsule intake ≥5 days/week

CFU, colony forming units; GDM, Gestational Diabetes Mellitus.
*Vivomixx contains *Streptococcus thermophilus* DSM 24,731, bifidobacteria (*Bifidobacterium breve* DSM 24,732, *Bifidobacterium longum* DSM 24,736, and *Bifidobacterium infantis* DSM 24,737), and lactobacilli (*Lactobacillus acidophilus* DSM 24,735, *Lactobacillus plantarum* DSM 24,730, *Lactobacillus paracasei* DSM 24,733, and *Lactobacillus delbrueckii* subsp. *bulgaricus* DSM 24,734).
**Average gestational age at enrollment was 15.95 weeks ± 1.45.
***Average gestational age at enrollment was 15.13 ± 1.8.
****Average week of gestation is 13.9.

There were differences in regard to the daily dose of probiotics and duration of the intervention. The daily dose of probiotic supplements ranged from 5×10^8 to 4.5×10^{11} colony forming units (CFUs). There did not appear to be a pattern with decreased GDM rates in the probiotic group compared to placebo controls reported by studies with higher doses of probiotics. The duration of probiotic supplementation ranged from 4 weeks total to 13 weeks of gestation to delivery. A decreased incidence of GDM in the intervention compared to the control group was not found with longer durations of treatment.

Each study defined adherence to probiotic interventions differently. One study did not define a threshold for compliance but reported a mean yogurt intake of 27.8/28 days and 27.6/28 days among the probiotic and conventional yogurt groups, respectively.[21] Four studies provided different thresholds to define adherence to the intervention. For example, Halkjær et al. defined compliance at ≥80% of capsule intake, which was met by 21/25 participants in the intervention and 17/24 participants in the placebo group.[24] Another study defined compliance as ≥75% of capsule intake, which was met by 76% of participants assigned to either probiotic or placebo capsules.[23] In comparison, 81% of participants complied with dietary intervention in the same study. Pellonpera et al. considered probiotic intake ≥5 days per week as a threshold for compliance, which was met by 89% of participants among all four study groups.[25] These figures were derived from self-reported data. Interestingly, one studied defined compliance at ≥75% self-reported capsule use, which was verified by fecal sample analysis.[22] While over 90% of participants in the probiotic group were compliant according to self-reported data, only 76% of compliance was found on fecal sample analysis.

Quality Assessment

Four included studies had a low risk of bias arising from the randomization process in the studies, as indicated in Table 3. All studies randomly allocated participants to study groups and performed blocked randomization stratified by pre-pregnancy BMI categories to ensure each study group had an equal number of overweight and obese participants. Intervention and control groups were not statistically significant for all baseline characteristics included in two studies.[21,24] Despite adequate randomization, a significantly higher prevalence of family history for diabetes among the fish-oil/placebo group raised some concerns for bias in one study.[25] Two studies did not statistically identify differences among study groups at baseline,[22,23] which was considered a low risk of bias according to the RoB2.

All studies had a low risk of bias due to deviations from the intended interventions with respect to the assignment to the intervention. These double-blind placebo-controlled trials used appropriate measures to conceal the intervention assignment from participants, investigators, and those involved in data analysis. All studies analyzed participants according to the intervention they have been assigned to as part of an intention to treat (ITT) analysis.

However, all studies had a high risk of bias due to deviations from the intended intervention with respect to adherence to the treatment. Despite adequate concealment of the intervention from participants and researchers, non-adherence to daily probiotic use in all studies was sufficient to raise concerns. Three studies did not complete an analysis to estimate the effect of adhering to the intervention.[21,22,25] Two studies performed a per-protocol sensitivity analysis, which excluded participants who did not adhere to the probiotic or placebo capsule directions from further analysis.[23,24] A second study performed a per-protocol analysis for GWG outcome only.[23] The RoB2 does not consider per-protocol analysis as an acceptable method for studying the effect of adherence.

Results from the included studies may be biased due to missing outcome data. Three studies included in this review have a low risk of bias due to missing data, as data were reported for 95% of randomized participants.[21,22,24] Two studies did not perform a sensitivity analysis to demonstrate outcomes were not biased by missing data.[23,25] However, it is unlikely that missing outcome data depended on the true value of measured outcomes because the proportions of missing data were approximately equal between the study groups, and reasons for missing data were consistent across study groups. For these reasons, the two studies raised some concern for bias due to missing data.

There was a low risk of bias due to the outcome measurements for all included studies. Four studies did not describe how blood glucose concentrations during the oral glucose tolerance test (OGTT) were measured.[21–24] Inappropriate methods may have been used, as portable blood glucose monitors have poor validity.[26] However, the non-differential bias on outcome measures and blinding of outcome assessors suggest a low

Table 3. Results from Version 2 of the Cochrane Risk of Bias Assessment (RoB2) for studies included in this review.

Signaling questions	Asgharian, 2020	Halkjæer, 2020	Callaway, 2019	Okesene-Gafa, 2019	Pellonperä, 2019
Domain 1: Risk of bias arising from the randomization process	Low	Low	Low	Low	Some concerns
Was the allocation sequence random?	Yes	Yes	Yes	Yes	Yes
Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Yes	Yes	Yes	Yes	Yes
Did baseline differences between intervention groups suggest a problem with the randomization process?	No	No	No information	No information	Yes
Domaine 2: Risk of bias due to deviations from the intended intervention					
Were participants aware of their assigned intervention during the trial?	No	No	No	No	No
Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	No	No	No	No	No
Domaine 2a: Risk of bias due to deviations from the intended intervention (*effect of assignment to intervention*)	Low	Low	Low	Low	Low
Was an appropriate analysis used to estimate the effect of assignment to intervention?	Yes	Yes	Yes	Yes	Yes
Domain 2b: Risk of bias due to deviations from the intended interventions (*effect of adhering to intervention*)	High	High	High	High	High
Was there non-adherence to the assigned intervention regimen that could have affected participants' outcomes?	No information	Yes	Yes	Yes	Yes
Was an appropriate analysis used to estimate the effect of adhering to the intervention?	No information	No	No	No	No
Domain 3: Missing outcome data	Low	Low	Low	Some concerns	Some concerns
Were data for this outcome available for all, or nearly all, participants randomized?	Yes	Yes	Yes	No	No
Is there evidence that the result was not biased by missing outcome data?	Not applicable	Not applicable	Not applicable	No	No
Could missingness in the outcome depend on its true value?	Not applicable	Not applicable	Not applicable	Yes	Yes
Is it likely that missingness in the outcome depended on its true value?	Not applicable	Not applicable	Not applicable	No	No
Domain 4: Risk of bias in measurement of the outcome	Low	Low	Low	Low	Low
Was the method of measuring the outcome inappropriate?	No information	No information	No information	No information	No
Could measurement or ascertainment of the outcome have differed between intervention groups?	No	No	No	No	No
Were outcome assessors aware of the intervention received by study participants?	No	No	No	No	No
Domain 5: Risk of bias in selection of the reported result	Low	Low	Low	Low	Low
Were the data that produced this result analyzed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	Yes	Yes	Yes	Yes	Yes
Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple eligible outcome measurements (e.g., scales, definitions, and time points) within the outcome domain?	No	No	No	No	No
Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple eligible analyses of the data?	No	No	No	No	No

Note: Acceptable answers to signaling questions included: yes, probably yes, probably no, no, and no information. Levels of bias for each domain included: low, high, and some concerns. Questions that were not applicable for all included studies were not included in this table.

risk of bias. One study specified that they measured blood glucose concentrations using an enzymatic hexokinase assay,[25] which is a valid method.[27] In addition, measuring blood glucose concentration to diagnose GDM is an automated test that requires no judgement from outcome assessors, which eliminates observer bias.

All studies had a low risk of bias due to the selection of the reported results. All studies analyzed results according to the pre-specified plan outlined in the methods section. There is only one way to report GDM according to IADPSG guidelines, which eliminates bias due to selecting outcomes from multiple measures. All of these analyses eligible for consideration, as statistical tests between probiotic and placebo control groups at comparable time points, were provided and appropriate for the study designs.

DISCUSSION

Although probiotic supplements did not significantly affect the development of GDM among overweight or obese patients, the high sensitivity of IAPDSG may not capture the preventative effect of probiotic supplements. A study from New Zealand that included participants in all weight categories found probiotics contain *Lactobacillus rhamnosus*, or placebo capsules resulted in comparable rates of GDM according to IAPDSG guidelines, but a significant decrease in the incidence of GDM among the probiotic group when using New Zealand's diagnostic criteria.[28] While IAPDSG enables research to compare the incidence of GDM across countries, the highly sensitive criteria for diagnosing GDM provide a high number of false positives that may mask the potential preventative effect in probiotics.[29] It would have been helpful if primary studies reported the incidence of GDM according to population-specific criteria in addition to IAPDSG, as primary studies included in this review included populations from Iran,[21] Australia,[22] Denmark,[24] New Zealand,[23] and Finland.[25] However, the incidence of GDM diagnosed by Carpenter and Coustan criteria was not significantly different among overweight or obese participants at 28 weeks gestation who received *Lactobacillus* and *Bifidobacterium* containing probiotics or placebo control according to a generalized linear model adjusted for pre-pregnancy BMI ($p = 0.561$).[30] Therefore, probiotics may not effectively prevent GDM among overweight or obese women who enter pregnancy in a state of increased insulin resistance regardless of the sensitivity of the diagnostic criteria.

One study reported significant differences in glucose tolerance. After excluding patients with a history of GDM, one study found a significantly decreased FPG and 2-h glucose tolerance test in the probiotic group compared to placebo controls.[21] Although this difference in FPG may not be clinically significant, this improvement in glucose metabolism aligns with a study of patients from diverse weight groups that found 6 weeks of *L. casei*, *L. acidophilus*, and *B. bifidum* administration containing probiotics significantly decreased FPG and insulin resistance compared to placebos.[31] However, an included study that did not exclude patients with a history of GDM or perform statistical tests to identify differences in sample characteristics at baseline found a significantly greater FPG among patients who received probiotics compared to placebos.[22] History of GDM may influence the effect of probiotics on glycemic control in obese and overweight patients, as another included article reported a significant interaction between GDM history and the intervention group with respect to changes in FPG.[25] This interaction was attributed to the association of previous GDM with decreased glucose values in the probiotic group when compared with the increase in fish oil and placebo group.[25] Patients with a history of GDM may have persistently altered cell signaling pathways, such as decreased IRS-1 tyrosine phosphorylation, attenuating a potentially preventative effect of probiotic supplements for subsequent pregnancies.[2] For this reason, it is recommended that future research excludes patients with a previous history of GDM among the study sample.

Outcomes related to GWG or preeclampsia may also be affected by the prevalence of previous GDM. One study reported excessive GWG was significantly lower among those receiving probiotics compared to placebos, while the mean GWG was comparable between both groups.[22] The higher proportion of participants with a history of GDM in the probiotic group may contribute to the significantly decreased excessive GWG and higher prevalence of preeclampsia found in this group.[22] The four other studies reported no significant differences in GWG among probiotic and placebo-treated groups.[21,23–25] These results align with a study that found GWG was not significantly different among pregnant participants of diverse weight categories who received conventional or probiotic yoghurt enriched in *L. acidophilus* and *B. animalis*.[32] Another study found that women of diverse weight categories who received

L. rhamnosus and *B. lactis* containing probiotics were at a significantly decreased risk of central adiposity than placebo controls at 6 months postpartum.[33] Future studies should investigate the effect of probiotics at longer postpartum time points using multiple body composition outcomes. Although central adiposity is defined as a waist circumference >80 cm, which is not a suitable measure of body composition during pregnancy, a decreased hydration constant may underestimate fat mass in BMI calculations during pregnancy, while this would result in a non-differential bias.

The systematic quality appraisal is a strength of this review, which suggests the results may be biased due to the effect of adhering to the intervention. All RCTs completed an ITT analysis. However, high rates of non-compliance in an ITT analysis may increase type 2 error, as low rates of compliance can substantially impact on the power of an equivalence trial.[34] The threshold of capsule uptake to define compliance was inconsistent across all included studies. In addition, self-reported data likely over-estimated the proportion of compliant participants. Although two studies completed a per-protocol analysis,[23,24] such results may be biased by factors that influence participants' willingness or ability to comply with protocol guidelines. Future studies should measure compliance using fecal sample analysis and complete inverse probability weighting (IPW) to estimate causal effects, as suggested by the RoB2. As part of IPW, outcomes of each study arm are weighted by the inverse of the probability for receiving the treatment they were assigned, which creates an average potential outcome based on a pseudo-population where every participant received the treatment value. Currently, there are no relevant studies that complete such an analysis.

The heterogeneity among included interventions is a limitation of this review. Differences between study designs, as summarized in Table 2, make it difficult to draw meaningful conclusions from pooled data from individuals. Future studies should apply a more consistent probiotic supplement and a dosing schedule before the efficacy of probiotic supplements for preventing GDM among high-risk patients can be concluded.

While one might presume increased exposure to probiotics would have beneficial effects on glucose metabolism, this may not be the case for overweight and obese patients. An included study that provided 4 weeks of probiotic supplements starting at 24 weeks of gestation found a significant decrease in FPG but no change in the incidence of GDM among study groups.[21] Similarly, a study of patients from diverse weight categories who received probiotic supplements for 6 weeks starting at 22 weeks of gestation found a significant decrease in FPG.[31] However, GDM diagnosis was not a measured outcome in this study. This study design is comparable to the study design by Wickens et al., which found a significant decrease in FPG and GDM according to New Zealand guidelines compared to placebo controls.[28] Other studies included in this review provided probiotic supplements for a longer duration and began treatment earlier in gestation but found no significant differences in FPG or GDM.[22,24,25] While the optimal time to provide probiotic interventions during gestation is unclear, this comparison suggests that high risk patients may benefit from probiotic interventions later in gestation. Furthermore, the study that provided the lowest dose of probiotic supplements was the only included study to report an improvement in glycemic control among obese and overweight patients.[21] This dose was also less than that provided in studies that improved glycemic control among normal-weight patients.[28,31] Potential differences in doses and treatment time for probiotic interventions to prevent GDM may relate to the rise in insulin resistance starting at 19–22 weeks of gestation and peak at 33–36 weeks of gestation among normal-weight patients compared to overweight or obese individuals who begin pregnancy with a higher insulin requirement. However, this argument conflicts with evidence from a meta-analysis reporting a dose-dependent improvement in glucose metabolism among GDM patients of diverse BMI categories.[35]

Current literature contains variation in bacterial species contained in probiotic supplements without any trends regarding the most effective combination for improving glycemic control. *L. acidophilus* and *B. lactis* containing probiotics were found to significantly decrease FPG among overweight and obese women.[21] However, probiotics containing *B. lactis* combined with *L. rhamnosus* did not produce such results among this patient population.[25] Conflicting results are also found among studies with healthy weight participants.[32,36] These inconsistencies suggest the effectiveness of probiotic species at improving glycemic control depends on the host gut microbiome at baseline. This explanation is likely, as *Lactobacillus* and *Bifidobacterium* improve host metabolism by increasing the availability of substrates for enteric butyrate-producing bacteria.[37]

CONCLUSIONS

Current evidence is not sufficient to recommend probiotic supplements to prevent GDM in overweight and obese patients, as the incidence of GDM among intervention and control groups was not statistically significant in all included studies. The majority of articles included in this review did not report any adverse effects of probiotic supplements. Therefore, there is insufficient evidence to suggest probiotics should be avoided due to adverse effects or contraindications. Clinical recommendations for preventing GDM among this high-risk patient population using probiotics may become more conclusive as future evidence addresses the effect of adhering to probiotic interventions using fecal sample analysis and IPW. In addition, a more consistent regime, possibly consistent Ing of lower dose probiotics later in gestation, should be established. Future studies may benefit from excluding participants with a history of GDM in addition to using multiple GDM diagnostic criteria and body composition measurements.

Disclaimers

None.

Statement of Source Support

No material support was provided for this article.

Conflict of interest and funding

No conflicts of interest to disclose. The author has not received any funding or benefits from industry or elsewhere to conduct this study.

REFERENCES

1. Sonagra AD, Biradar SM, Dattatreya K, Jayaprakash M. Normal pregnancy – a state of insulin resistance. J Clin Diagnostic Res 2014; 8(11): 1–3. doi: 10.7860/jcdr/2014/10068.5081

2. Catalano PM. Trying to understand gestational diabetes. Diabet Med 2014; 31(3): 273–81. doi: 10.1111/dme.12381

3. Baz B, Riveline JP, Gautier JF. Gestational diabetes mellitus: definition, aetiological and clinical aspects. Eur J Endocrinol 2016; 174(2): R43–51. doi: 10.1530/EJE-15-0378

4. Deputy NP, Kim SY, Conrey EJ, Bullard KM. Prevalence and changes in preexisting diabetes and gestational diabetes among women who had a live birth – United States, 2012–2016. MMWR Morb Mortal Wkly Rep 2018; 67(43): 1201–7. doi: 10.15585/mmwr.mm6743a2

5. Lavery J, Friedman A, Keyes K, Wright J, Anath C. Gestational diabetes in the United States: temporal changes in prevalence rates between 1979 and 2010. An Int J Obstet Gynaecol 2017; 124(1): 804–13. doi: 10.1111/1471-0528.14236.Gestational

6. Dall TM, Yang W, Gillespie K, Mocarski M, Byrne E, Cintina I, et al. The economic burden of elevated blood glucose levels in 2017: diagnosed and undiagnosed diabetes, gestational diabetes mellitus, and prediabetes. Diabetes Care 2019; 42(9): 1661–8. doi: 10.2337/dc18-1226

7. Kc K, Shakya S, Zhang H. Gestational diabetes mellitus and macrosomia: a literature review. Ann Nutr Metab 2015; 66(2): 14–20. doi: 10.1159/000371628

8. Reece EA. The fetal and maternal consequences of gestational diabetes mellitus. J Matern Neonatal Med 2010; 23(3): 199–203. doi: 10.3109/14767050903550659

9. McGillick E V., Morrison JL, McMillen IC, Orgeig S. Intrafetal glucose infusion alters glucocorticoid signaling and reduces surfactant protein mRNA expression in the lung of the late-gestation sheep fetus. Am J Physiol – Regul Integr Comp Physiol 2014; 307(5): 538–45. doi: 10.1152/ajpregu.00053.2014

10. Bowers K, Tobias DK, Yeung E, Hu FB, Zhang C. A prospective study of prepregnancy dietary fat intake and risk of gestational diabetes. Am J Clin Nutr 2012; 95(2): 446–53. doi: 10.3945/ajcn.111.026294

11. Wu S, Jin J, Hu K, Wu Y, Zhang D. Prevention of gestational diabetes mellitus and gestational weight gain restriction in overweight / obese pregnant. Nutrients 2022; 14(2383): 1–19. doi: 10.3390/nu14122383

12. Tanentsapf I, Heitmann BL, Adegboye ARA. Systematic review of clinical trials on dietary interventions to prevent excessive weight gain during pregnancy among normal weight, overweight and obese women. BMC Pregnancy Childbirth 2011; 11(81): 1–12. doi: 10.1186/1471-2393-11-81

13. Oostdam N, van Poppel MNM, Wouters MGAJ, van Mechelen W. Interventions for preventing gestational diabetes mellitus: a systematic review and meta-analysis. J Women Health 2011; 20(10): 1551–63. doi: 10.1089/jwh.2010.2703

14. Hasain Z, Mokhtar NM, Kamaruddin NA, Mohamed Ismail NA, Razalli NH, Gnanou JV, et al. Gut microbiota and gestational diabetes mellitus: a review of host-gut microbiota interactions and their therapeutic potential. Front Cell Infect Microbiol 2020; 10(188): 1–19. doi:10.3389/fcimb.2020.00188

15. Barrett HL, Dekker Nitert M, Conwell LS, Callaway LK. Probiotics for preventing gestational diabetes. Cochrane Database Syst Rev 2014; 4(2): 1–26. doi: 10.1002/14651858.CD009951.pub2

16. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017–2018. NCHS Data Brief 2020; (360): 1–8.

17. Diamant M, Blaak EE, de Vos WM. Do nutrient-gut-microbiota interactions play a role in human obesity, insulin resistance and type 2 diabetes? Obes Rev 2011; 12(4): 272–81. doi: 10.1111/j.1467-789X.2010.00797.x

18. Chatzakis C, Goulis DG, Mareti E, Eleftheriades M, Zavlanos A, Dinas K, et al. Prevention of gestational diabetes mellitus in overweight or obese pregnant women: a network meta-analysis. Diabetes Res Clin Pract 2019; 158: 107924. doi: 10.1016/j.diabres.2019.107924

19. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International Association of Diabetes and Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010; 33(3): 676–82. doi: 10.2337/dc09-1848
20. Higgins JP, Savović J, Page MJ, Sterne JAC. RoB 2 guidance: parallel trial. Cochrane Collab 2019; (July): 1–24. Available from: https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials [cited 20 May 2022].
21. Asgharian H, Homayouni-Rad A, Mirghafourvand M, Mohammad-Alizadeh-Charndabi S. Effect of probiotic yoghurt on plasma glucose in overweight and obese pregnant women: a randomized controlled trial. Eur J Nutr 2020; 59(1): 205–15. doi: 10.1007/s00394-019-01900-1
22. Callaway LK, McIntyre HD, Barrett HL, et al. Probiotics for the prevention of gestational diabetes mellitus in overweight and obese women: findings from the SPRING double-blind randomized controlled trial. Diabetes Care 2019; 42(3): 364–71. doi: 10.2337/dc18-2248
23. Okesene-gafa KAM, Li M, Mckinlay C, Taylor R, Rush EC, Wall CR, et al. Effect of antenatal dietary interventions in maternal obesity on pregnancy weight-gain and birthweight: Healthy Mums and Babies (HUMBA) randomized trial. 2019; 22(2): 152.e1–152.e13. doi: 10.1016/j.ajog.2019.03.003
24. Halkjær SI, Knegt VE De, Lo B, et al. Multistrain probiotic increases the gut microbiota diversity in obese pregnant women : results from a randomized, double-blind placebo-controlled study. Curr Dev Nutr 2020; 4(7): nzaa095. doi: 10.1093/cdn/nzaa095
25. Pellonperä O, Vahlberg T, Mokkala K, et al. Weight gain and body composition during pregnancy : a randomised pilot trial with probiotics and / or fish oil. Br J Nutr 2020; 126(4): 541–51. doi: 10.1017/S0007114520004407
26. Salacinski AJ, Alford M, Drevets K, Hart S, Hunt BE. Validity and reliability of a glucometer against industry reference standards. J Diabetes Sci Technol 2014; 8(1): 95–9. doi: 10.1177/1932296813514315
27. Dickson LM, Buchmann EJ, Janse Van Rensburg C, Norris SA. The impact of differences in plasma glucose between glucose oxidase and hexokinase methods on estimated gestational diabetes mellitus prevalence. Sci Rep 2019; 9(1): 1–7. doi: 10.1038/s41598-019-43665-x
28. Wickens KL, Barthow CA, Murphy R, Abels PR, Maude RM, Stone PR, et al. Early pregnancy probiotic supplementation with Lactobacillus rhamnosus HN001 may reduce the prevalence of gestational diabetes mellitus: a randomised controlled trial. Br J Nutr 2017; 117(6): 804–13. doi: 10.1017/S0007114517000289
29. Agarwal MM, Dhatt GS, Othman Y. Gestational diabetes : differences between the current international diagnostic criteria and implications of switching to IADPSG. J Diabetes Complications 2015; 29(4): 544–9. doi: 10.1016/j.jdiacomp.2015.03.006
30. Lindsay KL, Kennelly M, Culliton M, Smith T, Maguire O, Shanahan F, et al. Probiotics in obese pregnancy do not reduce maternal fasting glucose: a double-blind, placebo-controlled, randomized trial (Probiotics in Pregnancy Study). Am J Clin Nutr 2014; 99(6):1432–9. doi: 10.3945/ajcn.113.079723
31. Karamali M, Nasiri N, Taghavi Shavazi N, Jamilian M, Bahmani F, Tajabadi-Ebrahimi M, et al. The effects of synbiotic supplementation on pregnancy outcomes in gestational diabetes. Probiotics Antimicrob Proteins 2018; 10(3): 496–503. doi: 10.1007/s12602-017-9313-7
32. Asemi Z, Samimi M, Tabassi Z, Naghibi Rad M, Rahimi Foroushani A, Khorammian H, et al. Effect of daily consumption of probiotic yoghurt on insulin resistance in pregnant women: a randomized controlled trial. Eur J Clin Nutr 2013; 67(1): 71–4. doi: 10.1038/ejcn.2012.189
33. Ilmonen J, Isolauri E, Poussa T, Laitinen K. Impact of dietary counselling and probiotic intervention on maternal anthropometric measurements during and after pregnancy : a randomized placebo-controlled trial. Clin Nutr 2011; 30(2): 156–64. doi: 10.1016/j.clnu.2010.09.009
34. Sheng D, Kim MY. The e ects of non-compliance on intent-to-treat analysis of equivalence trials. Stat Med 2006; 25(7): 1183–99. doi: 10.1002/sim.2230
35. Łagowska K, Malinowska AM, Zawieja B, Zawieja E. Improvement of glucose metabolism in pregnant women through probiotic supplementation depends on gestational diabetes status: meta-analysis. Sci Rep 2020; 10(1): 1–17. doi: 10.1038/s41598-020-74773-8
36. Karamali M, Dadkhah F, Sadrkhanlou M, Jamilian M, Ahmadi S, Tajabadi-Ebrahimi M, et al. Effects of probiotic supplementation on glycaemic control and lipid profiles in gestational diabetes: a randomized, double-blind, placebo-controlled trial. Diabetes Metab 2016; 42(4): 234–41. doi: 10.1016/j.diabet.2016.04.009
37. Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L. Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. Front Microbiol 2016; 7: 979. doi: 10.3389/fmicb.2016.00979
38. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372: n71. doi: 10.1136/bmj.n71

Diffusion Tensor Imaging: A Step-by-Step Guide for Radiology and Neurology Clerkship Students, Residents, and Graduate Students Using Clinical Research Examples

Max J. Goodman[1]*, Wesley T. Richerson[2], Dawn F. Wolfgram[1]

[1]Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA
[2]Department of Biomedical Engineering, Marquette University and Medical College of Wisconsin, Milwaukee, WI, USA

Corresponding Author: Max J. Goodman; mgoodman@mcw.edu

Diffusion-tensor imaging has become common practice in radiology and imaging research due to its many applications in brain connectivity and neurodevelopment as well as for pathologies including tumors, ischemia, trauma, and neurodegeneration. However, its novelty compared to other neuroimaging techniques has meant that graduate programs, particularly medical schools, have not included opportunities to learn how diffusion tensor imaging can visualize the brain and interpretation of the data clinically and in research. Diffusion tensor imaging can be a challenging utility to understand for newcomers and is subject to wide interpretation. We offer for medical and graduate students as well as residents a step-by-step guide in interpreting diffusion tensor imaging results for clinical analysis using simple explanations of physics and neuroscience and its application in clinical and translational research.

Keywords: *magnetic resonance imaging; radial diffusivity; axial diffusivity; mean diffusivity; fractional anisotropy; radiology*

INTRODUCTION

Diffusion-tensor imaging (DTI) is a non-invasive method that utilizes magnetic resonance imaging (MRI) technology to measure tissue water diffusion rates *in vivo* with high sensitivity and has useful neuroimaging applicability particularly in tumors, ischemia, trauma, neurodevelopment, and neurodegeneration. However, DTI can be challenging to understand for newcomers and is subject to wide interpretation. This paper will serve as a step-by-step way for medical and graduate students as well as residents to interpret DTI results for clinical analysis using simple explanations of physics and neuroscience as well as its application in clinical and translational research. However, note that DTI and MRI are utilized beyond neuroimaging and are applied to every subfield of radiology.

To understand DTI, it is important to first discuss what MRI is. MRI is a diagnostic tool used to noninvasively visualize soft tissue structures through their protonic orientation. Ordinarily in tissue, protons are spinning in a random orientation, but in an MRI the external magnetic field causes them to become parallel or antiparallel to each other.[1] Radiofrequency pulses are then applied which flip the protons and generate an image from the energy it takes for protons to fall out of parallelism with each other and reach equilibrium. It is the variation of molecules that different tissues have which allow for MRI to distinguish anatomical and pathological features.

There are different sequences or types of MRI beyond DTI. The most common are T1-weighted, T2-weighted, fluid attenuated inversion recovery (Flair), proton density-weighted (PD), and diffusion-weighted (DWI). T1 is based on the rate it takes for protons to return to equilibrium, and T2 is based on the rate for protons to fall out of parallelism with each other.[1,2] A short T1 means a fast recovery to equilibrium which produces a strong or bright signal whereas a short T2 means a rapid decay which produces a weak or dark signal.[3] For example, cerebrospinal fluid (CSF) contains protons relatively far apart from each other and hence fewer interactions compared to fat, which allows for a longer time until equilibrium and hence a dark T1 but also a slower decay from parallelism and therefore a bright T2. T1 is generally utilized to assess morphology which includes anatomy, which is why it is also referred to as an anatomical sequence, as well as edema and fibrosis; T2 is more helpful to visualize edema and inflammation.[4] Flair is T2 but fluid is attenuated, allowing regions of tissue close to fluid to be more

MSRJ © 2023 VOL: 10. Issue: Spring
epub March 2023; www.msrj.org

carefully visualized, such as the periventricular area.[5] PD is when T1 and T2 are attenuated such that signal is produced based on the density of protons in tissue and is most frequently utilized for meniscal tears.[6,7]

Diffusion-weighted is different from other MRI sequences in that it is designed to visualize water diffusion.[1,8,9] Water has a unique diffusion pattern and rate depending on the cellular and tissue architecture.[8,9] For example, intracellular water will interact with structures such as the cytoskeleton, enzyme complexes, and organelles whereas extracellular water will interact with proteins in the extracellular matrix as well as different orientations of cells. Since these cellular and tissue architectural properties can change with alterations in tissue integrity due to the injury-repair response, DWI can be utilized for evaluating damage to components of brain structure, even before abnormalities are illustrated on anatomical MRI.[1,8] DTI is DWI but includes the mathematical models applied to DWI which provide a quantitative interpretation that DWI does not provide.

We can estimate cellularity and white matter microstructural organization based on DTI properties such as the degree of restricted diffusion or fractional anisotropy (FA), molecular diffusion rate or mean diffusivity (MD or apparent diffusion coefficient [ADC]), the diffusion rate along the main axis of diffusion known as axial diffusivity (AD or $\lambda_{\parallel}$), and the diffusion rate perpendicular to the main axis known as radial diffusivity (RD or $\lambda_{\perp}$).[10,11] These metrics are most commonly used for relating diffusion to an underlying pathophysiology of white matter and are calculated based on eigenvalues in a DTI model as seen in Figure 1. Eigenvalues explain how spread-out data is in a given direction, and in this case, eigenvalues describe the diffusion properties of water on a 3D grid, or tensor, of tissue in each unit (known as a voxel) of an MRI image of the brain.[12] There are other variables including trace, relative anisotropy and more, though their applicability is limited currently. Understanding these main terms' definitions and impressions is crucial towards understanding DTI (Table 1).

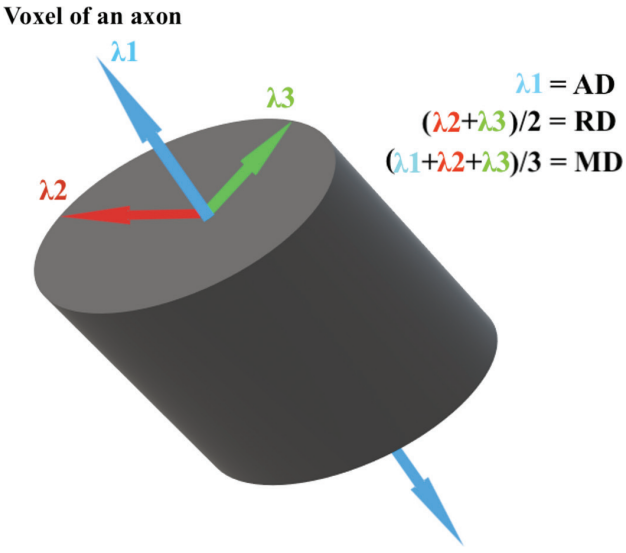


Figure 1. Eigenvalues and diffusion measures of a voxel of an axon. This schematic characterizes the eigenvalues of a tensor and how three of the most common diffusion measures are calculated.

DTI TERMINOLOGY

Fractional Anisotropy

Fractional anisotropy is measured from 0 to 1, with 0 being isotropic or the completely unrestricted diffusion of water that moves in any random direction, and 1 being anisotropic or the completely restricted diffusion of water in roughly one direction.[11] CSF is likely to have an FA closer to 0 (or isotropy) than grey or white matter of the brain for this reason, with white matter having a higher FA than grey matter.[13–15] White matter is mainly composed of axons and its myelin from oligodendrocytes that form tracts such as the cortico-spinal tract or the corpus callosum (CC). Axons are assumed to be tubular and unidirectional structures of neurons encased by myelin that restrict the movement of molecules into or out of the axon which therefore establishes anisotropy. Grey matter, on the other hand, consists of neuronal cell bodies and their dendrites, as well as unmyelinated axonal terminals that interact with one of the 1,000 synapses every neuron has on average, which allow for a greater possibility of directions for diffusion of water molecules.[16] This makes the grey matter FA less than that of white matter. However, not all white matter regions are created equal; the genu of the CC has a higher FA than other white matter tracts due to the highly organized fiber tracts within the CC.[14] A decrease in FA often is interpreted to mean either axonal or myelin degeneration; how to distinguish between these two possibilities requires further data.[15,17] This is because a reason for decreased FA may be due to an increase in RD or a decrease in AD and will be discussed later.[18]

Mean Diffusivity

Mean diffusivity is an average of three eigenvalues (λ_1, λ_2, and λ_3), or the magnitude of diffusion across the

3D-axes.[11] The sum of the magnitude of the three eigenvalues is the trace, so other papers may describe MD as trace/3. It can be generalized then that a low MD corresponds to low diffusivity and high MD high diffusivity. White matter and grey matter generally have similar MD, and CSF has higher MD than either type of brain matter.[13] While MD cannot be used to differentiate brain matters, it is sensitive to changes to the diffusion of water. Typically, if MD is higher than normal then it may reflect damage such as cytolysis as diffusion is less restricted.[15,19] Such damage may be from the loss of integrity of fiber tracts or myelin damage.[13,20] Cytotoxic edema, which can be from a redistribution of water extracellularly to intracellularly, would reduce cell-mediated water transport and therefore reduce MD.[21,22] Changes such as greater myelination or axonal packing would reduce MD, while demyelination, axonal degeneration, and increases in CSF or vasogenic edema increase MD.[13,15,20,22] Like FA, additional information is needed to interpret MD.

Axial Diffusivity

Axial diffusivity (AD or λ_1) is important for delineating between grey and white matter and for estimating changes to white matter specifically.[11] Because AD is assumed to be the parallel diffusion of water relative to a tract, AD is very low in grey matter and much higher in white matter.[23] However, AD is more sensitive to white matter maturation and similarly axonal degeneration than it is to the tissue's myelination; in fact, having elevated or lower myelination does not change AD.[24] Barriers to the diffusion of water down the axon such as cellular debris and protein aggregation (ex. amyloidosis) would decrease AD.[25] On the contrary, having more neurofilaments which increase the axon's diameter and therefore the axonal caliber would enhance AD.[24,26,27] Axonal packing (increased neurofilaments) is complicated by the changes in axonal diameters as well as myelination, so an absolute change in axonal packing may or may not be reflected in AD.[15,26] Additionally, AD is positively correlated with CSF and cytotoxic edema which would increase the axon's diameter and mimic neurofilament assembly, and so changes to AD may reflect extracellular and intracellular alterations.[15] A specific utilization of AD is in Wallerian degeneration; that is depicted by a decrease in FA and AD with an increase in RD (described subsequently) without a change in MD.[28] Generally, the interpretation for AD then is a marker of axonal injury, independent of myelination, but interpretation of changes in AD needs to be in the context of other conditions related to the intracellular and extracellular environments.

Table 1. This schematic depicts how changes in tissue integrity and architecture influence DTI measurements.

	FA	MD $(I_1+I_2+I_3)/3$	AD I_1	RD $(I_2+I_3)/2$
	FA is a summary measure of microstructural integrity. While FA is highly sensitive to microstructural changes, it is less specific to the type of change.	MD is an inverse measure of the membrane density, is very similar for both GM and WM and higher for CSF. MD is sensitive to cellularity, edema, and necrosis.	AD tends to be variable in WM changes and pathology. In axonal injury AD decreases. The ADs of WM tracts have been reported to increase with brain maturation.	RD increases in WM with de- or dysmyelination. Changes in the axonal diameters or density may also influence RD.
Gray Matter	↓	-	↓	↑
White Matter	↑	-	↑	↓
CSF	↓	↑	↑	↑
High myelination	↑	↓	-	↓
Dense axonal packing	↑	↓	-	↓
WM Maturation	↑	↓	↑	↓
Axonal degeneration	↓	↑	↓	↑
Demyelination	↓	↑	-	↑
Low SNR	↓	↑	↓	-

Source: From Tromp.[15]

Radial Diffusivity

Radial diffusivity [RD or $(\lambda_2+\lambda_3)/2$] can also separate grey from white matter and is useful for estimating changes to white matter though more specifically in regards to myelin.[11] Opposite of AD, RD is very low in white matter and higher in grey matter because it measures diffusion of water perpendicular to λ_1 rather than parallel.[23,27] One useful item, however, is that RD is also positively related to CSF and edema like AD, thus it is within reason that RD and AD both simultaneously increase if cytotoxic edema is occurring.[29] However, note that RD and AD can be age-related and increase simultaneously because of both myelin and axon loss.[30] One way to distinguish age-related white matter changes from edema is that RD is more sensitive to aging than AD, so a relatively equal change in both AD and RD is likely edema.[31] Additionally, RD is more closely related to myelination than it is to axonal injury. Studies found that dysmyelinated shiverer mutant mice had increased RD but AD remained unchanged; demyelination also increases RD.[32,33] Note that large enough axonal edema and inflammation may reduce water diffusivity after a period of time which would cause RD values to remain unchanged.[23] Therefore, analyzing RD either too early or too late may miss important data for the interpretation of myelin and oligodendrocyte cellularity.

APPLICATIONS OF DTI IN EVALUATING NEUROLOGICAL DISORDERS

While DTI has enormous potential for interpreting a macroscopic level of cellularity within white matter pathways, it is evident that other information is currently needed to make valid interpretations of its results that include but are not limited to pathophysiology, histology, anatomic neuroimaging, neuroanatomy, and careful consideration of the subject's clinical history. For example, important but often unappreciated clinical factors including elevated blood pressure and obesity appear to affect FA and AD respectively.[34] Basic neuroanatomy and neuroscience must also be utilized. A region of the hippocampus such as the dentate gyrus may entirely differ from the primary cornu ammonis in terms of cell density and cell type and therefore influences DTI data.[35]

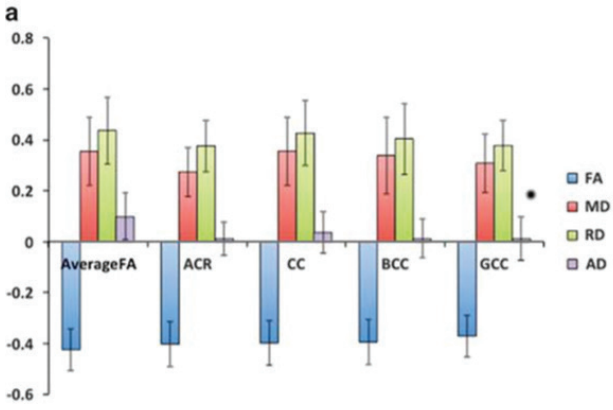
When understanding DTI and neuroimaging as it relates to aging, it is also important to analyze the changes in FA from posterior to anterior. Reports have observed a 'last in, first out', meaning the last parts of the brain to mature, such as the prefrontal lobe, are the first to degenerate in normal aging.[36,37] FA is often higher in the posterior relative to the anterior. This concept is important because a lower FA in the posterior portion of the brain in an older adult relative to the prefrontal cortex of the brain, for example, may underlie a pathology such as Alzheimer's disease or Lewy body dementia.[36]

Different regions also decline differently in normal aging,[38] though an overall global perspective should be taken as well to label either axonal injury or demyelination. In fact, one way of distinguishing between axonal injury from demyelination in either an acute or gradual pathophysiological process is to analyze whether or not the FA values are uniformly decreased globally within a tract or focally. Entire tracts of abnormally low FA signify axonal injury as the whole axon is dysfunctional, with focal differences owing to demyelination as each myelin sheath covers only a fraction of the axon. Readers of DTI literature should be cautioned in cases where researchers or clinicians are studying specific regions of interest (ROIs) and attempting to attribute it to either axonal injury or demyelination, as the results may only convey a small portion of information.

Clinical Research Examples: Schizophrenia

Our first example is from a 2018 study by Kelly et al,[39] where researchers from the Schizophrenia Working Group of the Enhancing Neuroimaging Genetics through Meta-Analysis consortium (ENIGMA-Schizophrenia) attempted to investigate white matter differences between 1963 adult subjects with schizophrenia and 2,359 healthy adult controls across the globe using DTI in a meta-analysis. As shown in Figure 2, the magnitude of the differences between the two groups in regards to FA, MD, AD, and RD through a post hoc analysis is described using Cohen's *d* effect size estimates and covariates including age, sex, and their interactions were accounted for. Across the average DTI measurement values and specifically in the anterior corona radiata (ACR), CC, body of corpus callosum (BCC), and genu of corpus callosum (GCC), FA was found to be lower in schizophrenia subjects than healthy controls, though MD and RD were higher ($p < 0.002$). AD was no different between schizophrenia subjects and healthy controls.

Numerous studies have suggested that a dysfunction in myelination due to an altered oligodendrocyte genotype contributes to schizophrenia.[40–43] Myelin



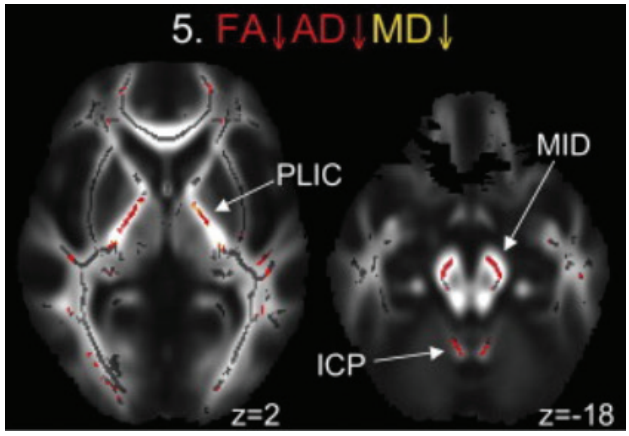
Source: Figure from Kelly et al[39]

Figure 2. (a) Cohen's *d* effect sizes, after meta-analysis, for FA, MD and RD differences in schizophrenia patients versus healthy controls, after including age, sex, age × sex, age2 and age2 × sex as covariates for the top four ROIs showing the largest FA effects (average FA, BCC, CC, ACR and GCC). Error bars represent the 95% confidence intervals. The CC, ACR, and FA across the whole brain are among the measures that show most robust effects in cohorts worldwide.

dysfunction is reported in subcortical white matter tracts of the prefrontal cortex in schizophrenia patients post-mortem.[40] In a DTI study, schizophrenia patients had low FA, high MD, high RD, and no difference in AD values in the ACR of the prefrontal cortex as well as the CC compared to controls that may be related to altered motor behavior.[39,44,45] So at first glance, the low FA and high MD from 2018 Kelly et al can signify axonal damage, demyelination, or edema. AD is unchanged, so axonal injury or edema is unlikely, which only leaves demyelination which is consistent with the RD value and histological data. This exact relationship has been found in patients with white matter hyperintensities experiencing cognitive impairment, where researchers also concluded it was consistent with demyelination.[46]

Clinical Research Examples: White Tract Variations

The second example is a study by Burzynska et al, in 2010 who compared DTI data from 63 healthy adults in their seventh decade of life from 80 other healthy adults in their third decade to discover whether or not there exists age-related DTI measurement variations using tract-based spatial statistics.[38] Figure 3 illustrates lower FA, AD, and MD values in older patients compared to younger patients in the posterior limb of the internal capsule, inferior cerebellar peduncles, and midbrain white matter ($p < 0.01$, 2-tailed).



Source: Figure from Burzynska et al[38]

Figure 3. Patterns of age-related differences in diffusivity parameters. Within voxels showing a significant age-related reduction in FA (in blue, $p < 0.01$, 2-tailed), we distinguished five patterns (in red, $p < 0.01$, 2-tailed) … (5) decreases in AD… showed additional increase in AD and a decrease in MD, respectively, are overlaid in orange. PLIC: posterior limb of the internal capsule; MID: midbrain WM; ICP: inferior cerebellar peduncle. The numbers indicate the *z*-axis coordinate in MNI152 space (mm).

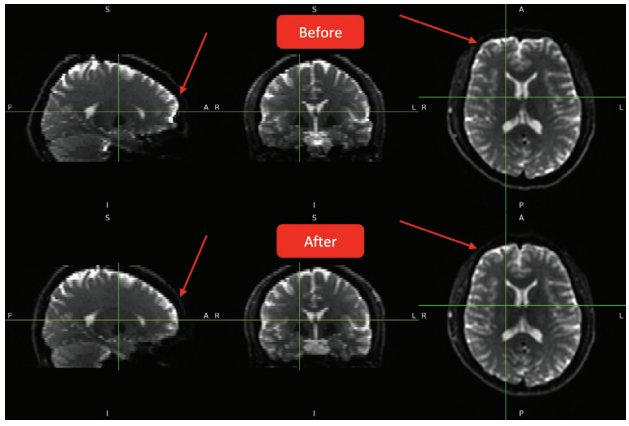
While this was one of only five patterns found in this investigation, we will be focusing on the aforementioned pattern for purposes of education. A low FA can mean either axonal or myelin degeneration, but that is generally accompanied by a high MD, which is not found in this case. Wallerian degeneration does not fit this pattern as described earlier. If gradual and chronic enough, cytotoxic edema or inflammation could be why RD was insignificantly impacted, but AD would have been higher. A low MD signifies low diffusivity, so while there may be axonal or myelin injury, it is possible that gliosis, a creation of more or larger glial cells, is occurring as a reaction to injury. Gliosis and resulting glial scars would also inhibit axon regeneration which may explain the low AD. RD is unchanged since gliosis is largely related to a hypertrophy of astrocytes and microglia, with little net change in oligodendrocytes and hence myelin. The discrepancy between the FA and MD in the study by Burzynska et al was thought to be due to gliosis.[38] One study found that gliosis could not be concluded based on AD and RD, though that investigation was done with a small sample size ex vivo in subjects with cerebral amyloid angiopathy without known postmortem interval and was largely exploratory and subject to false-positives.[47] Though this example leads us to conclude

gliosis, aging is subject to decreased white matter, demyelination, and axonal loss that can appear as unique DTI patterns regionally, which Burzynska et al describe in their investigation.[38]

Clinical Research Examples: Infarcts

In our final example, an investigation by Zhao et al, in 2019 investigated differences with a Student's *t*-test between subjects with leukoaraiosis (LA) and lacunar infarct subjects from two hospitals in China using subjects' DTI scans.[48] Lacunar infarct subjects had DTI values of (FA: 0.35 ± 0.03, MD: 0.40 ± 0.05) and LA subjects corresponded to (FA: 0.32 ± 0.02, MD: 1.08 ± 0.03) in normal appearing white matter, with a statistically significant difference between the two groups ($p < 0.05$).

This example is a nuanced example as it requires the attempt to distinguish a white matter hyperintensity as either LA or an acute lacunar infarction. Not only do they have a similar pathological basis owing to small-vessel disease and being age related, but also patients with acute lacunar infarcts are often predisposed to LA.[49–51] Symptoms of the stroke can assist physicians in identifying the specific white matter abnormality as a lacunar infarct from the other lesions associated with LA, but a lacunar infarct can be silent as well.[50,52] Additionally, a more sudden, local symptom such as dysarthria is more attributable to a lacunar infarct than LA, which has been related to gradual cognitive decline.[52,53] In this final DTI example study, it was found that patients with acute lacunar infarcts had higher FA and lower MD than LA patients.[48] The reason regarding the lower MD in the infarct was that the cytotoxic edema that occurred acutely decreased the ability for water diffusion, but over time MD would be increased once the edema is resolved. While ischemic injury may be thought to increase cellular damage and thus increase MD, this is dependent on the phase of injury and whether edema is present or not. A lower FA as seen in LA subjects may be two reasons: (1) because damage to white matter tracts typically reduces FA and (2) because LA lesions are in close proximity to the lateral ventricles, which contain elevated quantities of CSF. These FA and MD relative values are also subject to change depending on the condition of the stroke. A higher FA and lower MD like in this example are noted in the acute phase but may normalize during the subacute phase before MD increases and FA decreases once chronic, as indicative of structural disorganization.[48] A chronic lesion would be the same intensity to CSF, as Gore, Bansal, and Asuncion described.[53] This example should promote a well-rounded view for how to make an impression of DTI results.



Source: Images courtesy of Tugan Muftuler, PhD.

Figure 4. Before and after distortion correction caused by tissue susceptibility variations. Effect is most notable in the frontal lobe.

Limitations

There are a number of limitations both to the interpretation of DTI and to DTI itself. Data may be distorted if not properly processed, an example of which is illustrated in Figure 4. This can be said for motion correction, as it is natural for the patient to shift during the image acquisition.[11,54] Beyond background noise, which can be controlled via smoothing the dataset, eddy currents are natural opposing diffusion gradients and artifacts that can skew the voxels, known as shearing, or change the size of the brain volumes, which is called scaling.[11,55] To view motion and eddy currents, they can only be illustrated in video and are available in the citation.[56] Any DTI software should be researched to ensure that the eddy currents are corrected for; specific MR hardware is also capable of correcting for eddy currents. Ghosting is also a type of noise that makes repeated versions of parts of the brain in the image and reduces proper visualization; if the percent signal ghosting is greater than 3% then it may obstruct the image.[57] Ghosting can be due to patient motion, blood flow, and other respiratory and cardiac effects that if understood may be helpful for not only controlling for those factors but in interpreting the data afterwards. Fortunately, ghosting is now corrected by a majority of modern scanners and is rarely seen anymore, though an example can be found in the citation.[58]

In research and clinical interpretation, a comprehensive understanding of the patient cohort data and pathology is challenging but necessary as the influences of the DTI output are multifactorial. Partial volume effects from studying in an area of both white matter and grey matter can make the interpretation of results very uncertain and very challenging.[11,59] A heterogeneity of different fiber types within the same voxel can be minimized though using multi-tensor models that Soares, Marques, Alves, Sousa explain in their 2013 article.[19] Because partial volume effects can create a confounding bias in DTI measures, not taking partial volume effects into account can be devastating for valid impressions. Ozturk et al in 2008 cited partial volume effects as a potential reason for the high variability in DTI measurements found in deep white matter structures.[60] As for tissue architecture, crossing fibers with different directions would be expected to decrease FA, not due to either myelin or axonal cellularity. With 45–90% of white matter voxels containing a crossing fiber, it further reinforces the idea that other tools or modifications are needed to interpret DTI results.[61,62] For example, utilizing high-angular resolution diffusion imaging can measure an orientation distribution function and create a version of FA known as generalized FA that is more geared towards fiber crossings.[63] Finally, the most challenging dynamic to overcome is that complex diseases may demonstrate demyelination, axonal damage, and inflammation simultaneously. Currently, the solutions again are using DTI in conjunction with other medical tools along with a robust clinical understanding.

CONCLUSION

DTI is an essential tool towards diagnosis and understanding pathology in white matter. Its ability to estimate minute structural properties has made for a vast application both in research and clinically, though interpretation may be complicated and limited in some instances. This literature review has offered through basic scientific concepts guidance for not only medical students but also members across educational backgrounds to better enable them to make fair DTI inferences in clinical research. Because DTI is now an established tool in medicine and research, medical students, graduate students, and residents may need to become more familiar with DTI as its use widens.

Conflict of interest and funding

The authors have not received any funding or benefits from industry or elsewhere to conduct this study.

REFERENCES

1. Serai SD. Basics of magnetic resonance imaging and quantitative parameters T1, T2, T2*, T1rho and diffusion-weighted imaging. Pediatr Radiol 2022; 52(2): 217–27. doi: 10.1007/s00247-021-05042-7

2. Elster AD. An index system for comparative parameter weighting in MR imaging. J Comput Assist Tomogr 1988; 12(1): 130–134. doi:10.1097/00004728-198801000-00025

3. Young IR, Szeverenyi NM, Du J, Bydder GM. Pulse sequences as tissue property filters (TP-filters): a way of understanding the signal, contrast and weighting of magnetic resonance images. Quant Imaging Med Surg 2020; 10(5): 1080–120. doi: 10.21037/qims.2020.04.07

4. Dekkers IA, Lamb HJ. Clinical application and technical considerations of T1 & T2(*) mapping in cardiac, liver, and renal imaging. Br J Radiol 2018; 91(1092): 20170825. doi: 10.1259/bjr.20170825

5. Kates R, Atkinson D, Brant-Zawadzki M. Fluid-attenuated inversion recovery (FLAIR): clinical prospectus of current and future applications. Top Magn Reson Imaging 1996; 8(6): 389–96.

6. Farshad-Amacker NA, Jain Palrecha S, Farshad M. The primer for sports medicine professionals on imaging. Sports Health 2013; 5(1): 50–77. doi: 10.1177/1941738112468265

7. Sano T, Widmalm SE, Yamamoto M, Sakuma K, Araki K, Matsuda Y, et al. Usefulness of proton density and T2-weighted vs. T1-weighted MRI in diagnoses of TMJ disk status. Cranio. 2003; 21(4): 253–8. doi: 10.1080/08869634.2003.11746259

8. Huisman TAGM. Diffusion-weighted and diffusion tensor imaging of the brain, made easy. Cancer Imaging 2010; 10(1A): S163–71. doi: 10.1102/1470-7330.2010.9023

9. Winston GP. The physical and biological basis of quantitative parameters derived from diffusion MRI. Quant Imaging Med Surg 2012; 2(4): 254–65. doi: 10.3978/j.issn.2223-4292.2012.12.05

10. Alexander AL, Hurley SA, Samsonov AA, Adluru N, Hosseinbor AP, Mossahebi P, et al. Characterization of cerebral white matter properties using quantitative magnetic resonance imaging stains. Brain Connect 2011; 1(6): 423–46. doi: 10.1089/brain.2011.0071

11. Rajagopalan V, Jiang Z, Stojanovic-Radic J, Yue GH, Pioro EP, Wylie GR, et al. A basic introduction to diffusion tensor imaging mathematics and image processing steps. Brain Disord Ther 2017; 6: 2. doi: 10.4172/2168-975X.1000229

12. Basser PJ. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. NMR Biomed 1995; 8(7–8): 333–44. doi: 10.1002/nbm.1940080707

13. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. Radiology 1996; 201(3): 637–48. doi: 10.1148/radiology.201.3.8939209
14. Mädler B, Drabycz SA, Kolind SH, Whittall KP, MacKay AL. Is diffusion anisotropy an accurate monitor of myelination? Correlation of multicomponent T2 relaxation and diffusion tensor anisotropy in human brain. Magn Reson Imaging 2008; 26(7): 874–88. doi: 10.1016/j.mri.2008.01.047
15. Tromp D. DTI scalars (FA, MD, AD, RD) – how do they relate to brain structure? The Winnower; 2016.
16. Hawkins J, Ahmad S. Why neurons have thousands of synapses, a theory of sequence memory in Neocortex. Front Neural Circuits 2016; 10: 23. doi: 10.3389/fncir.2016.00023
17. Beaulieu C, Allen PS. Determinants of anisotropic water diffusion in nerves. Magn Reson Med 1994; 31(4): 394–400. doi: 10.1002/mrm.1910310408
18. Boretius S, Escher A, Dallenga T, Wrzos C, Tammer R, Brück W, et al. Assessment of lesion pathology in a new animal model of MS by multiparametric MRI and DTI. Neuroimage 2012; 59(3): 2678–88. doi: 10.1016/j.neuroimage.2011.08.051
19. Soares JM, Marques P, Alves V, Sousa N. A hitchhiker's guide to diffusion tensor imaging. Front Neurosci 2013; 7: 31. doi: 10.3389/fnins.2013.00031
20. Song SK, Sun SW, Ju WK, Lin SJ, Cross AH, Neufeld AH. Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. Neuroimage 2003; 20(3): 1714–22. doi: 10.1016/j.neuroimage.2003.07.005
21. Field AS, Hasan K, Jellison BJ, Arfanakis K, Alexander AL. Diffusion tensor imaging in an infant with traumatic brain swelling. AJNR Am J Neuroradiol 2003; 24(7): 1461–64.
22. Schaefer PW, Buonanno FS, Gonzalez RG, Schwamm LH. Diffusion-weighted imaging discriminates between cytotoxic and vasogenic edema in a patient with eclampsia. Stroke 1997; 28(5): 1082–5. doi: 10.1161/01.str.28.5.1082
23. Winklewski PJ, Sabisz A, Naumczyk P, Jodzio K, Szurowska E, Szarmach A. Understanding the physiopathology behind axial and radial diffusivity changes-what do we know? Front Neurol 2018; 9: 92. doi: 10.3389/fneur.2018.00092
24. Budde MD, Xie M, Cross AH, Song SK. Axial diffusivity is the primary correlate of axonal injury in the experimental autoimmune encephalomyelitis spinal cord: a quantitative pixelwise analysis. J Neurosci 2009; 29(9): 2805–13. doi: 10.1523/JNEUROSCI.4605-08.2009
25. Dijkhuizen RM, de Graaf RA, Tulleken KA, Nicolay K. Changes in the diffusion of water and intracellular metabolites after excitotoxic injury and global ischemia in neonatal rat brain. J Cereb Blood Flow Metab 1999; 19(3): 341–9. doi: 10.1097/00004647-199903000-00012
26. Feldman HM, Yeatman JD, Lee ES, Barde LHF, Gaman-Bean S. Diffusion tensor imaging: a review for pediatric researchers and clinicians. J Dev Behav Pediatr 2010; 31(4): 346–56. doi: 10.1097/DBP.0b013e3181dcaa8b
27. Aung WY, Mar S, Benzinger TL. Diffusion tensor MRI as a biomarker in axonal and myelin damage. Imaging Med 2013; 5(5): 427–40. doi: 10.2217/iim.13.49
28. O'Dwyer L, Lamberton F, Bokde ALW, Ewers M, Faluyi YO, Tanner C, et al. Multiple indices of diffusion identifies white matter damage in mild cognitive impairment and Alzheimer's disease. PLoS One 2011; 6(6): e21745. doi: 10.1371/journal.pone.0021745
29. Kim JH, Loy DN, Liang HF, Trinkaus K, Schmidt RE, Song SK. Noninvasive diffusion tensor imaging of evolving white matter pathology in a mouse model of acute spinal cord injury. Magn Reson Med 2007; 58(2): 253–60. doi: 10.1002/mrm.21316
30. Madden DJ, Bennett IJ, Burzynska A, Potter GG, Chen NK, Song AW. Diffusion tensor imaging of cerebral white matter integrity in cognitive aging. Biochim Biophys Acta 2012; 1822(3): 386–400. doi: 10.1016/j.bbadis.2011.08.003
31. Davis SW, Dennis NA, Buchler NG, White LE, Madden DJ, Cabeza R. Assessing the effects of age on long white matter tracts using diffusion tensor tractography. Neuroimage 2009; 46(2): 530–41. doi: 10.1016/j.neuroimage.2009.01.068
32. Song SK, Sun SW, Ramsbottom MJ, Chang C, Russell J, Cross AH. Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. Neuroimage 2002; 17(3): 1429–36. doi: 10.1006/nimg.2002.1267
33. Song SK, Yoshino J, Le TQ, Lin SJ, Sun SW, Cross AH, et al. Demyelination increases radial diffusivity in corpus callosum of mouse brain. Neuroimage 2005; 26(1): 132–40. doi: 10.1016/j.neuroimage.2005.01.028
34. Allen B, Muldoon MF, Gianaros PJ, Jennings JR. Higher blood pressure partially links greater adiposity to reduced brain white matter integrity. Am J Hypertens 2016; 29(9): 1029–37. doi: 10.1093/ajh/hpw026
35. Stolp HB, Ball G, So PW, Tournier J-D, Jones M, Thornton C, et al. Voxel-wise comparisons of cellular microstructure and diffusion-MRI in mouse hippocampus using 3D Bridging of Optically-clear histology with Neuroimaging Data (3D-BOND). Sci Rep 2018; 8(1): 4011. doi: 10.1038/s41598-018-22295-9
36. Douaud G, Groves AR, Tamnes CK, Westlye LT, Duff EP, Engvig A, et al. A common brain network links development, aging, and vulnerability to disease. Proc Natl Acad Sci 2014; 111(49): 17648–53. doi: 10.1073/pnas.1410378111
37. Bender AR, Völkle MC, Raz N. Differential aging of cerebral white matter in middle-aged and older adults: a seven-year follow-up. Neuroimage 2016; 125: 74–83. doi: 10.1016/j.neuroimage.2015.10.030
38. Burzynska AZ, Preuschhof C, Bäckman L, Nyberg L, Li SC, Lindenberger U, et al. Age-related differences in white matter microstructure: region-specific patterns of diffusivity. Neuroimage 2010; 49(3): 2104–12. doi: 10.1016/j.neuroimage.2009.09.041
39. Kelly S, Jahanshad N, Zalesky A, Kochunov P, Agartz I, Alloza C, et al. Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. Mol Psychiatry 2018; 23(5): 1261–9. doi: 10.1038/mp.2017.170
40. Flynn SW, Lang DJ, Mackay AL, Goghari V, Vavasour IM, Whittall KP, et al. Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem

with analysis of oligodendrocyte proteins. Mol Psychiatry 2003; 8(9): 811–20. doi: 10.1038/sj.mp.4001337
41. Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI. Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. Schizophr Res 2004; 67(2–3): 269–75. doi: 10.1016/S0920-9964(03)00181-6
42. Mitkus SN, Hyde TM, Vakkalanka R, Kolachana B, Weinberger DR, Kleinman JE, et al. Expression of oligodendrocyte-associated genes in dorsolateral prefrontal cortex of patients with schizophrenia. Schizophr Res 2008; 98(1–3): 129–38. doi: 10.1016/j.schres.2007.09.032
43. Alba-Ferrara LM, de Erausquin GA. What does anisotropy measure? Insights from increased and decreased anisotropy in selective fiber tracts in schizophrenia. Front Integr Neurosci 2013; 7: 9. doi: 10.3389/fnint.2013.00009
44. Clark KA, Nuechterlein KH, Asarnow RF, Hamilton LS, Phillips OR, Hageman NS, et al. Mean diffusivity and fractional anisotropy as indicators of disease and genetic liability to schizophrenia. J Psychiatr Res 2011; 45(7): 980–8. doi: 10.1016/j.jpsychires.2011.01.006
45. Walther S, Federspiel A, Horn H, Razavi N, Wiest R, Dierks T, et al. Alterations of white matter integrity related to motor activity in schizophrenia. Neurobiol Dis 2011; 42(3): 276–83. doi: 10.1016/j.nbd.2011.01.017
46. Min ZG, Shan HR, Xu L, Yuan DH, Sheng XX, Xie WC, et al. Diffusion tensor imaging revealed different pathological processes of white matter hyperintensities. BMC Neurol 2021; 21(1): 128. doi: 10.1186/s12883-021-02140-9
47. van Veluw SJ, Reijmer YD, van der Kouwe AJ, Charidimou A, Riley GA, Leemans A, et al. Histopathology of diffusion imaging abnormalities in cerebral amyloid angiopathy. Neurology 2019; 92(9): e933–43. doi: 10.1212/WNL.0000000000007005
48. Zhao D, Wang Z, Cheng Y, Yuan Z, Rene F, Liu H, et al. A DTI study of leukoaraiosis and the differential diagnosis between leukoaraiosis and acute lacunar infarction. CNS Neurosci Ther 2019; 25(9): 1064–7. doi: 10.1111/cns.13191
49. Miyao S, Takano A, Teramoto J, Takahashi A. Leukoaraiosis in relation to prognosis for patients with lacunar infarction. Stroke 1992; 23(10): 1434–8. doi: 10.1161/01.str.23.10.1434
50. Wardlaw JM. What causes lacunar stroke? J Neurol Neurosurg Psychiatry 2005; 76(5): 617–9. doi: 10.1136/jnnp.2004.039982
51. Mitchell GF, van Buchem MA, Sigurdsson S, Gotal JD, Jonsdottir MK, Kjartansson Ó, et al. Arterial stiffness, pressure and flow pulsatility and brain structure and function: the Age, Gene/Environment Susceptibility – Reykjavik study. Brain 2011; 134(Pt 11): 3398–407. doi: 10.1093/brain/awr253
52. Te M, Zhao E, Zheng X, Sun Q, Qu C. Leukoaraiosis with mild cognitive impairment. Neurol Res 2015; 37(5): 410–14. doi: 10.1179/1743132815Y.0000000028
53. Gore M, Bansal K, Khan Suheb MZ, Asuncion RMD. Lacunar stroke. StatPearls Publishing; 2022. Available from: http://www.ncbi.nlm.nih.gov/books/NBK563216/ [cited 27 February 2023].
54. Baum GL, Roalf DR, Cook PA, Ciric R, Rosen AFG, Xia C, et al. The impact of in-scanner head motion on structural connectivity derived from diffusion MRI. Neuroimage 2018; 173: 275–86. doi: 10.1016/j.neuroimage.2018.02.041
55. Leemans A, Jones DK. The B-matrix must be rotated when correcting for subject motion in DTI data. Magn Reson Med 2009; 61(6): 1336–49. doi: 10.1002/mrm.21890
56. eddy – a tool for correcting eddy currents and movements in diffusion data. FSL. Available from: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy [cited 4 March 2023].
57. Phantom test guidance for use of the large MRI phantom for the ACR accreditation program. 2018. Available from: https://www.acraccreditation.org/-/media/acraccreditation/documents/mri/largephantomguidance.pdf [cited 4 March 2023].
58. Nyquist (N/2) ghosts. Questions and answers in MRI. Available from: http://mriquestions.com/nyquist-n2-ghosts.html [cited 4 March 2023].
59. Alexander AL, Hasan KM, Lazar M, Tsuruda JS, Parker DL. Analysis of partial volume effects in diffusion-tensor MRI. Magn Reson Med 2001; 45(5): 770–80. doi: 10.1002/mrm.1105
60. Ozturk A, Sasson AD, Farrell JA, Landman BA, da Motta AC, Aralasmak A, et al. Regional differences in diffusion tensor imaging measurements: assessment of intrarater and interrater variability. AJNR Am J Neuroradiol 2008; 29(6): 1124–7. doi: 10.3174/ajnr.A0998
61. Jeurissen B, Leemans A, Tournier JD, Jones DK, Sijbers J. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Hum Brain Mapp 2013; 34(11): 2747–66. doi: 10.1002/hbm.22099
62. Sarwar T, Ramamohanarao K, Zalesky A. Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography? Magn Reson Med 2019; 81(2): 1368–84. doi: 10.1002/mrm.27471
63. Tuch DS. Q-ball imaging. Magn Reson Med 2004; 52(6): 1358–72. doi: 10.1002/mrm.20279