GOOD EVIDENCE MATTERS

GEMs of the Week

SPOTLIGHT

Are You Smarter Than a Computer? Comparing Doctors' Diagnostic Reasoning to AI

Under Pressure

Does Exertion Change Concussion Assessments?

Speeding Birth, Slowing Recovery?

Oxytocin and Postpartum Depression

Cognitive Training, Exercise, or Both for Cognitive Outcomes in Multiple Sclerosis

To Have or Not to Have

Timing of Smartphone Exposure and Adolescent Health

Botox and Babies

Reassuring Safety Data From a 29-Year Review

FPIN
Family Physicians Inquiries Network

Volume 6 Issue 33

Are You Smarter Than a Computer? Comparing Doctors' Diagnostic Reasoning to AI

Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial

Goh E, Gallo R, Hom J, et al. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. *JAMA Netw Open*. 2024;7(10):e2440969. Published 2024 Oct 1. doi:10.1001/jamanetworkopen.2024.40969

KEY TAKEAWAY: Physician use of large language models (LLMs) as a diagnostic aid did not improve clinical reasoning compared with conventional resources. However, LLM alone diagnostic accuracy compared to both physician groups was statistically significant.

STUDY DESIGN: Single-blind randomized clinical trial

LEVEL OF EVIDENCE: STEP 2

BRIEF BACKGROUND INFORMATION: Diagnostic errors are a significant source of patient harm. LLMs alone have demonstrated strong performance in medical reasoning exams, but their impact as a physician tool for diagnostic reasoning was previously unknown. The study aimed to evaluate whether access to LLMs improves physicians' diagnostic reasoning.

PATIENTS: Physicians

INTERVENTION: Use of LLM (Open AI GPT-4)

CONTROL: Conventional online resources

PRIMARY OUTCOME: Diagnostic reasoning, accuracy, and speed

METHODS (BRIEF DESCRIPTION):

- 50 attending and resident physicians were randomized to use either ChatGPT-4 or conventional resources such as UpToDate and Google to answer case questions for six clinical vignettes in 60 minutes.
- The only criteria for enrollment were any practicing physician with training in a general medicine specialty (Internal Medicine, Family Medicine, or Emergency Medicine).
- Diagnostic reasoning was measured on a validated rubric developed for the study and rated by single blind raters. Scores were out of 100%.
- LLM alone was prompted to answer the six diagnostic cases three separate times, and the responses were entered for blind rating along with physician responses.

INTERVENTION (# IN THE GROUP): 25

COMPARISON (# IN THE GROUP): 25

FOLLOW-UP PERIOD: Not available

RESULTS:

Primary Outcome –

- There was no difference in the median score per case between the LLM group and the control group (76% vs 74%, respectively; adjusted difference 2 percentage points; 95% CI, −4 to 8).
- In addition, there was no statistical difference in the accuracy of diagnosis between LLM + physician group and the control (odds ratio [OR] 1.4; 95% CI, 0.7−2.8).
- The LLM alone, compared to the control group, scored higher with a median score per case of 92% (adjusted difference 16 percentage points; 95% CI, 2–30).
- There was no difference in time spent per case between the LLM group and the control group (adjusted difference, –82 seconds; 95% CI, –195 to 31).

LIMITATIONS:

- The study used pre-studied vignettes which are not real world.
- Although all physicians in the LLM group used the tool at least once, there was no standardized or enforced method of use. This variability may have diluted the observed effect.
- While the structured reflection tool provided a richer measure of diagnostic reasoning, the study did not evaluate whether improved scores translated into better patient outcomes.
- Participants were recruited from academic institutions and may not represent the broader physician population in terms of experience or familiarity with AI tools.
- A single LLM was used, and other resources were not tested.
- This was primarily tested on Internal Medicine physicians which limits generalizability.

Pooja Kanthawar, MD
Advocate Health Care Illinois Masonic FMRP
Chicago, IL

Under Pressure: Does Exertion Change Concussion Assessments?

Effect of Mild and Moderate Exertion on the Sideline Assessment of Concussion: A Randomized Controlled Trial

Jain RK, Marshall KA, Leddy JJ, et al. Effect of Mild and Moderate Exertion on the Sideline Assessment of Concussion: A Randomized Controlled Trial. *Clin J Sport Med*. 2025;35(6):613-619. doi:10.1097/JSM.0000000000001356

KEY TAKEAWAY: Mild-to-moderate physical exertion does not improve sideline concussion assessment performance compared to no exertion in athletes in ≥1 sports without recent concussion.

STUDY DESIGN: Randomized controlled trial

LEVEL OF EVIDENCE: STEP 3 (downgraded due to small sample size)

BRIEF BACKGROUND INFORMATION: Sports-related concussions are mild traumatic brain injuries (TBIs) commonly resulting from sport and recreational activity participation. Immediately after an injury, athletes are brought to the sideline and screened for symptoms suggestive of a concussion injury. Typically, an evaluation of orientation, memory, balance, cognition, and vestibular-ocular function. Concussions are especially important to recognize immediately. Experiencing a second injury shortly after the first can result in Second Impact Syndrome, a rare but often fatal condition characterized by rapid brain swelling. In addition, athletes who experience multiple concussions are at higher risk for Persistent Post Concussive Syndrome which can result in long-term headaches, dizziness, and issues with cognitive and emotional function. This study aimed to investigate if physical exertion affects concussion assessment compared to no exertion.

PATIENTS: Athletes 13–18 years old in ≥1 sports without recent concussion

INTERVENTION: Mild and moderate physical exertion

CONTROL: No exertion

PRIMARY OUTCOME: Changes in sideline concussion assessment performance

METHODS (BRIEF DESCRIPTION):

- Athletes 13–18 years old in ≥1 sports from two high schools in Western New York were included.
- Participants were excluded if they had a diagnosed or suspected concussion in the past one month, were experiencing persistent post-concussion symptoms, had a history of moderate-to-severe brain injury, or any contraindications to exercise according to the American Heart Association (AHA) guidelines.
- Participants were randomized into three groups using a commercial treadmill: Mild physical exertion (10 min at 50–55% of age predicted max heart rate) and moderate physical exertion (10 min at 70–75% of max heart rate) which was compared to no exertion (15 min seated rest).
- Participants were assessed using two common sideline concussion assessment tools, the Sports Concussion Assessment Tool (SCAT5) and Vestibular Ocular Motor Screening (VOMS).
- SCAT5 symptoms consisted of 22 symptoms scored 0–6 by the participant, with a score of zero being none and six being the most severe, the maximum score being 132.
- SCAT5 cognition testing consisted of orientation, immediate memory given five words in three different orders, digits backwards, months in reverse order, and delayed recall of the five words, with a higher score meaning better cognition, and a maximum score of 29.
- VOMS tests smooth pursuits, saccades horizontal and vertical, convergence, vestibular-ocular reflex horizontal and vertical, and vision motor sensitivity, with four symptoms listed after testing then scored 0–10 by the participant with a score of zero being none and 10 being the most severe, the maximum score being 210.
- Group-wise comparisons were done at baseline, during exercise, and post-exertion to evaluate for differences.
- An analysis of variance (ANOVA) was used for continuous measures, and a x2 test was used for categorical measures. If significant differences were observed, further pair-wise comparison was done using Tukey correction.
- Within each group, a paired t test was used to evaluate for any change in assessments and exercise

and an effect size calculated (Cohen d). A P-value of <0.05 was considered statistically significant.

INTERVENTION (# IN THE GROUP):

- Mild exertion: 16
- Moderate exertion group: 15

COMPARISON (# IN THE GROUP): 16

FOLLOW-UP PERIOD: Immediately after intervention

RESULTS:

Primary Outcome –

- No exertion significantly improved total SCAT5 symptom scores between assessments compared to baseline (Cohen d 0.87; 95% CI, 0.28–1.4).
- No exertion did not significantly improve total SCAT5 cognition scores compared with baseline (Cohen d 0.14; 95% CI, –0.36 to 0.63).
- No exertion did not significantly improve VOMS total errors compared with baseline (Cohen d 0; 95% CI, –0.49 to 0.49).
- Mild exertion did not significantly improve total SCAT5 symptom scores compared with baseline (Cohen d 0.41; 95% CI, –0.10 to 0.92).
- Mild exertion did not significantly improve total SCAT5 cognition scores compared with baseline (Cohen d 0.16; 95% CI, –0.34 to 0.65).
- Mild exertion did not significantly improve VOMS total errors compared with baseline (Cohen d 0; 95% CI, –0.49 to 0.49).
- Moderate exertion did not significantly improve total SCAT5 symptom scores compared with baseline (Cohen d 0.33; 95% CI, –0.20 to 0.84).
- Moderate exertion did not significantly improve total SCAT5 cognition scores compared with baseline (Cohen d –0.07; 95% CI, –0.57 to 0.44).
- There was no change in VOMS score for mild or moderate exertion; therefore, an effect size could not be determined.

LIMITATIONS:

- There was a small sample size.
- Only healthy, uninjured athletes participated, limiting application to concussed athletes.
- Exertion intensity did not exceed moderate levels.
- Lack of blinding
- Short-term follow-up, only immediately post-exertion.

Nikolas Jen, DO
Northeast Georgia Medical Center FMRP
Gainesville, GA

Speeding Birth, Slowing Recovery? Oxytocin and Postpartum Depression

Intrapartum Synthetic Oxytocin and Its Effects on Maternal Well-Being at 2 Months Postpartum

Gu V, Feeley N, Gold I, et al. Intrapartum Synthetic Oxytocin and Its Effects on Maternal Well-Being at 2 Months Postpartum. *Birth*. 2016;43(1):28-35. doi:10.1111/birt.12198

KEY TAKEAWAY: Higher doses of intrapartum synthetic oxytocin are associated with slightly greater symptoms of depression, anxiety, and somatization at two months postpartum.

STUDY DESIGN: Prospective cohort study

LEVEL OF EVIDENCE: STEP 3

BRIEF BACKGROUND INFORMATION: In the United States, the use of synthetic oxytocin (synOT) for induction and augmentation of labor as well as prevention of postpartum hemorrhage has more than doubled since 1990. The influence of synOT on maternal psychological wellbeing has not been well studied. This study aimed to examine the relationship between synOT administration and maternal mental health at two months postpartum.

PATIENTS: Adult pregnant patients with a singleton pregnancy who delivered after 35 weeks' gestation

INTERVENTION: Combined total dose of synOT administered intrapartum and postpartum

CONTROL: No or low levels of intrapartum synOT

PRIMARY OUTCOME: Maternal psychological well-being at two months postpartum

METHODS (BRIEF DESCRIPTION):

- Women ≥18 years old with singleton pregnancies were recruited from a hospital and birthing center in Montreal, Quebec (n=386), as part of two separate studies: One with enrollment at a clinic at 12–14 weeks' gestation and the other with enrollment during the birth hospitalization.
  - This study was comprised of data pooled from the two separate studies.
- Participants received intravenous (IV) synOT for standard obstetric indications during labor or induction, and/or for use in prevention or management of postpartum hemorrhage as an intramuscular (IM) or IV administration per clinician.
  - Infusion rates were increased in increments of 3–6 mL/hr to achieve effective contractions at a maximal frequency of every 2–3 minutes, lasting 45–90 seconds each.
  - SynOT could also be administered by IM injection to prevent postpartum hemorrhage postnatally.
    - The total dose was calculated from the sum of each rate of infusion by minutes at that rate plus any IM injections.
- Women who did not receive any synOT were included in the study (n=29).
- Symptoms of depression, anxiety, post-traumatic stress disorder (PTSD), and somatization were assessed to examine maternal psychological well-being.
  - Depression was measured using the Edinburgh Postnatal Depression Scale (EPDS). Scores range from 0–30, a score of ≥12 indicating a risk of depression.
  - Anxiety was measured using the Generalized Anxiety Disorder 7 Item Scale (GAD-7). Seven symptoms of anxiety are rated on a scale of 0–3, with higher scores indicating greater anxiety.
  - PTSD was measured using the Perinatal Posttraumatic Stress Questionnaire (PPQ). Scores range from 0–56 with a score of 19 indicating a cutoff for clinical PTSD.
  - Somatiziation was measured use the somatization subscale of the Symptom Checklist 90-Revised (SCl-90-R). Scores range from 0–48 with higher scores indicating greater distress.
- The primary outcome was assessed using bivariate correlations (Pearson's correlation), analysis of covariance, and linear regression analyses.
  - The Pearson correlation coefficient (r) measures the strength and direction of a linear relationship between two variables where +1 indicates a perfect positive linear correlation, –1 indicates a perfect negative linear correlation, and 0 indicates no linear correlation.

INTERVENTION (# IN THE GROUP): 386

COMPARISON (# IN THE GROUP): Not available

FOLLOW-UP PERIOD: Two months postpartum

RESULTS:

Primary Outcome –

- SynOT use was significantly positively correlated with depressive symptoms (r=0.15, *p*<.01).
- SynOT use was significantly positively correlated with anxiety symptoms (r=0.11, *p*<.05).
- SynOT use was weakly positively correlated with PTSD symptoms (r=0.08, *p*>.05).
- SynOT use was significantly positively correlated with somatization symptoms (r=0.18, *p*<.01).

LIMITATIONS:

- Psychological well-being outcomes were assessed via self-reported symptomatology as opposed to diagnostic criteria.
- Prenatal mental health symptoms and labor associated variables were not controlled for.
- Correlational design identifies associations but cannot determine causation due to lack of variable manipulation and unclear directionality.

Kailey Flint, DO
Northwest Washington FMRP
Bremerton, WA

Cognitive Training, Exercise, or Both for Cognitive Outcomes in Multiple Sclerosis

Cognitive Training, Exercise Training or Combined Training? A Comparative Effectiveness Research Study on Subjective and Objective Cognitive Outcomes in Multiple Sclerosis

Gyger N, Monschein T, Filser M, et al. Cognitive Training, Exercise Training or Combined Training? A Comparative Effectiveness Research Study on Subjective and Objective Cognitive Outcomes in Multiple Sclerosis. *J Neurol*. 2026;273(2):82. Published 2026 Jan 16. doi:10.1007/s00415-025-13535-w

KEY TAKEAWAY: Treadmill exercise, alone or combined with cognitive training, improves information processing speed and self-perceived cognitive function in patients with multiple sclerosis

STUDY DESIGN: Single-center, randomized, superiority trial with three parallel arms

LEVEL OF EVIDENCE: STEP 3 (downgraded due to lack of blinding, small sample size, and high dropout rate)

BRIEF BACKGROUND INFORMATION: Cognitive impairment, particularly in information processing speed, affects about half of people with multiple sclerosis (pwMS) and significantly impacts their quality of life. While pharmacological options are limited, non-pharmacological interventions like cognitive rehabilitation and exercise are promising. This study aimed to evaluate the effectiveness of cognitive training, treadmill walking, or the combination of both in managing cognitive deficits in pwMS.

PATIENTS: Adults 18–60 years old with relapsing-remitting or secondary progressive multiple sclerosis

INTERVENTION: Computerized cognitive training (BS)

CONTROL: Treadmill walking (TW) or a combination of both (BS+TW)

PRIMARY OUTCOME: Change in self-perceived cognitive function and information processing speed

METHODS (BRIEF DESCRIPTION):

- Patients with relapsing-remitting or secondary progressive multiple sclerosis (MS) were recruited from an outpatient neurology clinic in Germany.
- Included participants were adults, 18–60 years old, with mild to moderate disability, defined as an Expanded Disability Status Scale (EDSS) score ≤6.0 (range 0–10 with 0 indicating no impairment and 10 indicating death due to MS) and cognitive impairment risk scores, defined as a Symbol Digit Modalities Test (SDMT) z-score between –0.2 and –3.0 (≤–1.5 indicates impaired cognitive performance).
- Participants were excluded if they had cardiovascular disease, another acute neurological or psychological disorder other than MS, a recent relapse, changes in MS treatment in the last one month prior to enrollment, or recent steroid therapy.
- Baseline participant characteristics included a mean age of 47 and 80% female.
- Participants were randomized 1:1:1 to computerized BS, TW, or BS+TW, with each intervention consisting of two 45-min sessions per week for 12 weeks (BS+TW group completed 4 sessions weekly).
- The primary outcomes were measured as follows:
  - Perceived cognitive deficits (subjective) - measured using the Perceived Deficits Questionnaire (PDQ-20), assessing four domains (attention/concentration, retrospective memory, prospective memory, organizing/planning) with scores ranging from 0–80. Scores of ≤34 indicated average or expected levels of cognitive function, while scores of ≥35 indicated higher risk of cognitive difficulties.
  - Information processing speed (objective) – measured using the SDMT z-score with higher scores indicating better performance and a change of ≥8 points considered clinically meaningful.
  - All participants completed neuropsychological testing and questionnaires at baseline, 12 weeks, and six months.

INTERVENTION (# IN THE GROUP): 15

COMPARISON (# IN THE GROUP):

- TW: 16
- BS+TW: 15

FOLLOW-UP PERIOD: Six months

RESULTS:

Primary Outcome –

- At 12 weeks, all three groups showed a significant improvement in their perception of cognitive deficits through the PDQ-20 score.
  - BS (mean change –4.8, *p*=.018; Cohen's d 0.43)
  - TW (mean change –3.9, *p*=.028; Cohen's d 0.39)
  - BS+TW (mean change –5.0, *p*=.004; Cohen's d 0.39)
- At six months, only the TW and BS+TW groups maintained subjective cognitive improvements.
  - TW (mean change –3.7, *p*=.040, Cohen's d 0.32)
  - BS+TW (mean change –7.4, *p*=.012; Cohen's d 0.52)
- At 12 weeks and six months, the TW and BS+TW groups noted significant improvement in their SDMT scores.
  - TW at 12 weeks: (mean change 4.8, *p*=.001; Cohen's d 0.47)
  - TW at six months: (mean change 6.4, *p*=0.0; Cohen's d 0.66)
  - BS+TW at 12 weeks: (mean change 5.5, *p*=.026; Cohen's d 0.71)
  - BS+TW at six months: (mean change 5.0, *p*=.012; Cohen's d 0.62)
- There was no significant improvement in the SDMT scores in the BS group.

LIMITATIONS:

- The study's small sample size and high dropout rate (25%) limit its statistical power and may overestimate treatment effects.
- The lack of a true no-treatment control group makes it difficult to distinguish intervention effects from practice effects or natural disease progression.
- The nature of the interventions made it impossible to blind participants or investigators, increasing the risk of performance and detection bias.
- The study population consisted mostly of women with relapsing-remitting MS, which may limit the generalizability of the findings to other populations.

Sarah Prelaz, DO
Camp Lejeune Family Medicine Residency
Camp Lejeune, NC

To Have or Not to Have: Timing of Smartphone Exposure and Adolescent Health

Smartphone Ownership, Age of Smartphone Acquisition, and Health Outcomes in Early Adolescence

Barzilay R, Pimentel SD, Tran KT, Visoki E, Pagliaccio D, Auerbach RP. Smartphone Ownership, Age of Smartphone Acquisition, and Health Outcomes in Early Adolescence. *Pediatrics*. 2026;157(1):e2025072941. doi:10.1542/peds.2025-072941

KEY TAKEAWAY: Earlier and current smartphone ownership in early adolescence is associated with increased risks of obesity, insufficient sleep, and depression.

STUDY DESIGN: Prospective cohort study

LEVEL OF EVIDENCE: STEP 3

BRIEF BACKGROUND INFORMATION: Smartphone ownership has become nearly universal among youth, yet clinicians often lack objective data to guide conversations with families about potential health consequences. Although smartphones play an important role in daily functioning, including communication, entertainment, and social connectivity, the rapid rise of smartphone use among preteens has raised concerns about its impact on mental and physical well-being. Prior research has linked excessive screen time and problematic smartphone use to negative outcomes, but far less is known about the effects of simply owning a smartphone or the age at which it is first acquired. This study aimed to address that gap by examining whether smartphone ownership and the timing of its acquisition, is associated with depression, insufficient sleep, and obesity in early adolescence.

PATIENTS: Early adolescents 12–13 years old

INTERVENTION: Smartphone ownership at 12 years old and the age at which the first smartphone was acquired

CONTROL: Youth who did not own a smartphone at 12 years old, youth who did not own a smartphone at 13 years old and those who acquired their first smartphone at an older age

PRIMARY OUTCOME: Depression, obesity, and insufficient sleep

Secondary Outcome: The effects of earlier age of smartphone acquisition and outcomes at 13 years old among youth who did not own a smartphone at 12 years old

METHODS (BRIEF DESCRIPTION):

- Children enrolled in the Adolescent Brain Cognitive Development (ABCD) Study, which recruited participants from 21 sites across the United States, creating a large and demographically diverse cohort (N=10,588).
- Participants were approximately 10 years old at enrollment, with analyses focused on outcomes at 12 and 13 years old.
- Demographic and developmental variables included socioeconomic status, race/ethnicity, caregiver-reported pubertal development, and parental monitoring behaviors.
- All participants were typically developing youth, and a diagnosed psychiatric disorder was not required for inclusion. As a result, the sample reflects a general population cohort rather than a clinical population.
- Exposure was caregiver reported smartphone ownership status at 12 years old.
- Additional exposure was age of first smartphone acquisition, reported in full years.
- No treatment or intervention was administered as this was a naturalistic observational study.
- Follow up exposure was among youth who did not own a phone at 12 years old, the study identified whether they acquired one between 12 and 13 years old.
- Primary comparison was youth who did not own a smartphone at 12 years old.
- Secondary comparison was among 12-year-old non-owners, those who acquired a smartphone between 12 and 13 years old, compared to those who remained non-owners.
- At the 2-year follow-up, depression was assessed using the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS).
- At the 3-year follow-up, depression was measured using the self-reported Brief Problem Monitor (BPM), with a clinical threshold defined as a T-score ≥65.
- Obesity was determined using body mass index (BMI) percentiles based on age and sex adjusted

CDC growth charts. Obesity was defined as a BMI >95th percentile.

- Insufficient sleep was defined as youth-reported sleep duration of less than nine hours per night.
- Outcomes were analyzed using mixed effects logistic regression models. Models adjusted for multiple potential confounders including socioeconomic status, developmental factors, other device ownership, and pre-exposure outcomes by 12 years old (psychopathology, obesity, insufficient sleep).

INTERVENTION (# IN THE GROUP):

- Smartphone owners at 12 years old: 6,739
- Owners who acquired smartphones between 12–13 years old: 1,546

COMPARISON (# IN THE GROUP):

- Did not own a smartphone at 12 years old: 3,849
- Non-smartphone owners between 12–13 years old: 1,940

FOLLOW-UP PERIOD:

- Primary outcome: Two years
- Secondary outcome: Three years

RESULTS:

Primary Outcome –

- Smartphone ownership at 12 years old was associated with a higher risk of depression, obesity and insufficient sleep compared to not owning a smartphone.
  - Depression (odds ratio [OR] 1.3; 95% CI, 1.1–1.6)
  - Obesity (OR 1.4; 95% CI, 1.2–1.6)
  - Insufficient sleep (OR 1.6; 95% CI, 1.5–1.8)

Secondary Outcome –

- Earlier smartphone acquisition was associated with an increased risk of obesity and insufficient sleep compared to later smartphone acquisition.
  - Obesity (OR 1.1; 95% CI, 1.02–1.2)
  - Insufficient sleep (OR 1.1; 95% CI, 1.02–1.1)
- There was no significant difference in the risk of depression for earlier smartphone acquisition compared to later smartphone acquisition.
- Acquiring a smartphone between 12–13 years old was associated with an increased risk of clinical-level psychopathology and insufficient sleep compared to those who had not.
  - Clinical level psychopathology (OR 1.6; 95% CI, 1.1–2.2)
  - Insufficient sleep (OR 1.5; 95% CI, 1.3–1.8)
- There was no significant difference in the risk of obesity for those acquiring a smartphone between 12–13 years old compared to those who had not.

LIMITATIONS:

- As this was an observational study, causal relationship between smartphone ownership and outcomes cannot be established.
- Some potential confounding variables were not measured, such as specific smartphone use (e.g., screen time, social media use or type of content).
- Exposure variables, including smartphone ownership and age of first acquisition, were reported by caregivers, which may introduce reporting bias.
- This study provided limited insight into how smartphones were used, which may influence the relationship between smartphone ownership and health outcomes.

Risper Kirui, MD
Washington State University College of Medicine
Pullman, WA

Botox and Babies: Reassuring Safety Data From a 29-Year Review

Pregnancy Outcomes in Patients Exposed to OnabotulinumtoxinA Treatment: A Cumulative 29-Year Safety Update

Brin MF, Kirby RS, Slavotinek A, et al. Pregnancy Outcomes in Patients Exposed to OnabotulinumtoxinA Treatment: A Cumulative 29-Year Safety Update. *Neurology*. 2023;101(2):e103-e113. doi:10.1212/WNL.0000000000207375

KEY TAKEAWAY: Exposure to onabotulinumtoxinA before conception or during early pregnancy is associated with rates of major congenital malformations similar to those reported in the general population, supporting its relative safety when inadvertent exposure occurs, though data remains limited.

STUDY DESIGN: Retrospective observational study

LEVEL OF EVIDENCE: STEP 4 (downgraded due to retrospective design, lack of an internal comparison group, and reliance on voluntary postmarketing safety reports)

BRIEF BACKGROUND INFORMATION:

OnabotulinumtoxinA is commonly used to treat chronic migraine and other neurologic conditions in women of reproductive age. Due to limited pregnancy safety data, use is generally avoided despite frequent inadvertent exposure before pregnancy recognition. This study aimed to evaluate pregnancy outcomes following exposure to onabotulinumtoxinA using long-term postmarketing safety data.

PATIENTS: Women exposed to onabotulinumtoxinA before conception or during pregnancy

INTERVENTION: OnabotulinumtoxinA exposure before or during pregnancy

CONTROL: General population pregnancy outcomes

PRIMARY OUTCOME: Major congenital malformations and pregnancy outcomes

METHODS (BRIEF DESCRIPTION):

- Pregnancies reported to the Allergan Global Safety Database from January 1, 1990, to December 31, 2018, were reviewed.
- Maternal characteristics collected included age (<35 years old vs ≥35 years old), indication for treatment, timing of exposure relative to pregnancy, and onabotulinumtoxinA dose.
- Women were included if they were exposed to onabotulinumtoxinA during pregnancy or within three months prior to conception and had a known outcome.
- Exposure consisted of treatment with onabotulinumtoxinA for aesthetic or therapeutic indications. Dose information was reported in 50-unit increments, with most exposures (<200 units) consistent with typical clinical dosing.
- There was no internal unexposed comparison group.
- Pregnancy outcomes of live birth, fetal loss, and congenital malformations were compared with established general population rates.
- The primary outcome was the rate of major congenital malformations among live births from prospective pregnancies, as there is bias associated from retrospective reporting.

INTERVENTION (# IN THE GROUP): 195 pregnancies

COMPARISON (# IN THE GROUP): Not available

FOLLOW-UP PERIOD: From reported exposure through pregnancy outcome

RESULTS:

Primary Outcome –

- Of 195 prospective pregnancies with 197 fetuses, there were 152 (77%) live births and 45 (23%) fetal losses (32 spontaneous, 13 elective).
- Of 152 live births, 148 (97%) had normal outcomes and 4 had abnormal outcomes.
  - Among the four abnormal outcomes, there were one major birth defect, two minor fetal defects, and one birth complication.
- The prevalence rate for overall fetal defects and major fetal defects were low.
  - Overall fetal defects (2.6%; 95% CI, 1.0–6.6)
  - Major fetal defects (0.7%; 95% CI, 0.1–3.6)

LIMITATIONS:

- Retrospective design with reliance on voluntary postmarketing safety reporting.
- No internal comparison group.
- Incomplete reporting of dose, timing of exposure, and maternal characteristics in some cases.
- Few second- and third-trimester exposures so cannot reliably assess safety.

- Potential underreporting of adverse outcomes, which impacts internal validity of the results.
- Potential reporting bias as the authors sponsored the clinical trials.

Judy Trang-Mangonon, MD
Northside Hospital Gwinnett FMRP
Lawrenceville, GA