UR Can Become a Self-Driving Car

With Humans at the Helm and AI in the Loop

By Tammy Gavin, Munzoor Shaikh and Christopher Steel

Vol 2. NO. 9 | September 16, 2025

The Rise of the Algorithmic UR Leader

Key Learnings

- **AI without human oversight in Utilization Review (UR) can lead to greater losses in human and financial terms.** Algorithms that ignore human factors from patient and providers or ignore fast-changing payer rules have already led to legal blowback, financial losses and damage to patients. For example, UnitedHealthcare's NaviHealth algorithm predicted number of days in a post-acute setting, but insufficient human oversight led to inappropriate discharges affecting thousands of patients in 21 states. As another example, UCHealth, paid $23 million after an automated ER algorithm, that auto-assigned CPT 99285 when vital signs were checked frequently, led to a federal False Claims Act settlement.
- **The best results come from human-centered design and proper organizational structure where humans lead, and AI enhances.** Systems that build human-centered structures with clear roles, governance and user adoption protocols reduced denials, saved millions (e.g., $1 million per month at a Midwestern IDN), and gave clinicians back their time, all by letting people handle the hard calls while algorithms did the preparatory work.
- **Winning health systems will bring forward a new type of clinician-operator – "the algorithmic leader."** Health systems that succeed will not be the ones with the "best" technical solution, but ones who foster algorithmic leaders. These individuals know how to create the right structure, pair people with algorithms, design human-centered workflows that prioritize empathy and judgment, and measure success in both trust and dollars. Simply put, they deliver better decisions at scale (e.g., Sharp HealthCare paired an AI status engine with daily MD–RN huddles and a change-management boot camp, cutting review time by 80% while keeping human judgment in the driver's seat).

Why Aren't Health Systems in UR Self-Driving Mode Yet?

Health systems face a tight labor market challenge today and UR remains a key area of administrative pain points. So why hasn't UR gone into a full self-driving mode of AI solutions yet? To understand this, it helps to examine the systemic challenges the industry faces today. In theory, UR automation promises speed and scale. In practice, it delivers confusion, callbacks, and class-action lawsuits.

Most of us have heard by now about Cigna's PXDX algorithm, which allegedly enabled one medical director to deny 300,000 claims in two months without reviewing a single patient record. Or UnitedHealth's naviHealth model, which, according to internal whistleblowers and patient lawsuits, was wrong nearly 90% of the time, cutting off elderly rehab stays prematurely. A 2024 Senate investigation found that similar AI tools at United, Aetna, and Humana led to post-acute denials being tripled (in some cases multiplied 16X) compared to historical norms.

Not only is the current trend potentially hurtful to patients, but it also creates significant abrasion with providers, who often mistrust the payer's decisions and act as an advocate for the patient. Without a clean structure and clean process, no amount of technology or automation can solve the UR pain points for health systems.

Algorithms Can't Read What's Not Documented

Algorithms judge only what the record hands them, yet the record is often missing critical parts of a patient's story. Some key information may exist outside of the EMR documentation system such as patient emotions, family barriers, frailty cues, social determinants, payer specific documentation needs etc. NIH researchers estimate that 80% of clinically relevant detail exists in free-text notes or nowhere at all. AI can certainly read free-text notes, but notes may not all be collated in one place in a well-organized fashion in the EMR. The social pressures of keeping a patient in bed, the family barrier delaying discharge, the payer-specific phrase that flips a case from "observation" to "inpatient", all too often reside in the clinical team's heads, not in the EMR.

Clinical-documentation-improvement (CDI) programs try to close the gap, but they cover just 55% of discharges on average (AHIMA, 2024). Even the latest ambient voice tools have limits, a multi-site pilot found that it missed critical nuance in one out of four encounters unless a clinician edited the transcript. Especially for social determinants, which can be strong factors in risk mitigation, data can be missing or misleading. A 2024 JAMA Network Open study showed a sepsis-risk algorithm systematically under-scored Black patients when social-determinant fields were blank. Food insecurity remains another critical example. When it comes to food insecurity and glycemic control, the later days of the month are riskier for hypoglycemia admissions among low-income patients EVEN though the Z59.4 social determinant code for food insecurity may not indicate so due to a patient having intake in the early parts of the month (see appendix for more details). These shortfalls quickly morph into denial risk for a health system's UR team.

Until this invisible data is accounted for, every denial engine operates with a built-in bias. The CDI nurse who reviews the chart, the hospitalist who clarifies frailty in real time, the coder who knows one payer's magic words, these people don't just polish the EMR record, they provide a human perspective to the Algorithmic process. Without them, the metaphorical UR car can self-drive but may not reach the destination the driver has in mind or may get there with a "bumpy ride".

The Terrain Keeps Shifting

The best algorithm can steer only based on the road, the map and traffic conditions it knows. The problem is that the entire information set changes frequently. UnitedHealthcare issued more than 400 discrete policy edits in the first four months of 2024, each bulletin a small phone book of new criteria, exclusions, and caveats. Many UR vendors refresh rule engines quarterly and by the time their patches deploy, half the terrain has already shifted.

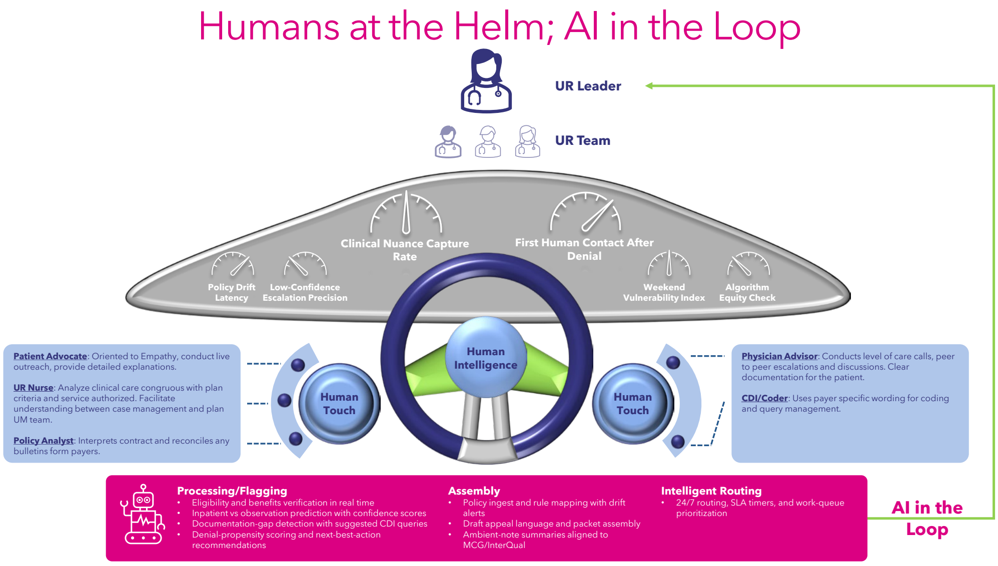
Health systems that copy payer auto-denial logic inherit the same liability and risk piloting outdated rules into new terrain. Regulators have noticed the mismatch between payer decisions and the terrain shifting. As a result, CMS warned Medicare Advantage plans in plain language: "Do not deny care solely on an algorithm." California's SB 1120 goes further, requiring a licensed clinician to sign off on every AI-generated denial. The Office of Inspector General has opened audits, and Humana's recent $12 million settlement over automated post-acute denials shows that legal traction is real, not theoretical.

In this environment, a human navigator is not optional, but essential. Policy analysts must reconcile bulletins in real time, physician advisors must vet algorithmic outliers, and revenue leaders must renegotiate when the solutions no longer match the contract. Until models can update at the speed of all relevant changes, UR will need people riding shotgun, calling out the detours and supervising the UR self-driving program.

Humans at the Helm, AI in the Loop

The most successful health systems have discovered a simple recipe where seasoned clinicians and savvy analysts in the lead can outperform a warehouse of computers acting alone. At a Midwestern IDN, embedding physician advisors inside the UR workflow delivered a 398% ROI and kept roughly $1 million a month from leaking out in preventable denials. Sharp HealthCare's case tells a similar story where 80% cut in review time arrived only after daily MD–RN huddles and a change-management boot camp taught staff how, when, and why to override the algorithm.

Why does a human-led model win? Because revenue integrity takes both a smart GPS and a seasoned driver's human instincts. For example, any clinician knows that a mother denied postpartum coverage doesn't want an F-score, she wants a voice with the authority to escalate, advocate or educate her when the legitimate answer is "no". Cross-functional pods such as UR nurse, physician advisor, payer rep all provide that assurance in real time on *human* terms. They translate clinical nuance into *human* language, imbue the language with empathy, and decide when a "no" deserves a second look or a *human* explanation.



Below is a scenario example of where the algorithm can be correct, but the human factors are ignored erasing the gains and hampering long term margins.

Scenario Example: Health System Self-Driving Algorithm Scenario Driving "Off the Cliff"

When the Algorithm Is Right but the Human Equation Goes Wrong

A health system installs an AI "status optimizer" designed to flag short-stay inpatients for automatic downgrades to observation. On the spreadsheet the logic is flawless where historical data showed CMS recouping $8 million a year in Two-Midnight audits; if the algorithm pre-emptively shifts those cases, finance projects a $7.2 million savings.

Month 1: Clinicians Revolt
Hospitalists discover the bot is downgrading stroke alerts and high-acuity heart-failure cases solely because length-of-stay probability was < 1.8 days. Physicians spend an extra 12 minutes per patient manually reversing decisions, and weekend coverage gaps mean many requests went unanswered until Monday, stretching actual LOS.

Month 3: Payers Push Back
Commercial insurers notice a spike in observation claims for DRGs that normally qualify as inpatient. They begin rejecting them outright, citing "provider error." Denials climb 22%, erasing half the projected savings and adding $200 000 in appeal labor.

Month 6: Regulators Knock
CMS reviewers flag the sudden pattern, opening a targeted probe. To avoid civil penalties, the health system self-reverses 900+ cases, refunding $2.3 million and issuing patient refunds for higher coinsurance bills.

Net Result after One Year
Projected Gain: $7.2 M
Actual Financial Outcome: (-$1.4 M) after refunds, overtime, and appeal costs)

Patient Experience Fallout: **11 pts dropped in HCAHPS "Communication of Care"**; 370 new complaints logged, a 3× increase, most citing "billing confusion" and "felt no one could fix mistake." Social media sentiment turned negative, and the system slid from 3.9 to 3.2 stars on Google reviews.

Other Intangible Costs: Surgeons deferred referrals to rival hospitals; staff satisfaction in Utilization Review dropped from 4.1 to 2.8.

The algorithm was right, but the human equation, patient trust, provider abrasion, clinical nuance, was wrong. The self-driving algorithm did its job, but without a human driver, the health system lost both reputationally and financially.

The Path Forward

Throwing only AI algorithms at UR treats the symptoms, not the disease. Algorithms alone amplify payer–provider mistrust, skip over the bedside subtleties, hurt referral patterns, and invite regulatory investigations. The cure is not to find the right self-driving UR AI technology; it is to re-commission humans as the primary drivers and recast algorithms as the powertrain.

What emerges is a new archetype, the algorithmic leader, a clinician-operator fluent in clinical process, payer policy and AI algorithms. This role orchestrates humans (patient, providers, payers) and algorithmic technologies rather than choosing one at the expense of the other. Their mandate looks as follows:

1. **Design human-centered workflows**. The workflow process is important, but the human impact of the process is even more critical (again, refer to the regional health system example shared earlier). Patients and staff should be trained in knowing how to foretell human behavior of patients, providers and other administrators. And of course, they should know how to reach a person who can overturn an errant "no" and do so with appropriate human finesse. As an example, UMC Health System rebuilt its operating model around an AI tool (Xsolis). It re-engaged staff, co-designed workflows with front-line teams, held weekly governance meetings, embedded the Care Level Score in its EMR, and formalized payer ties through Precision UM. Those changes drove a 5.7% drop in observation rate (Aug 2023–Jun 2024), 7,000+ nursing hours saved, over $800,000 in revenue captured, and a 20% shorter length of stay with 21% better patient flow. Using everyday humans to help design the process made adoption better and the algorithm in the loop performed better.
2. **Build the right structure to support the process**. No workflow survives without the right scaffolding. The difference between AI that works and AI that doesn't comes down to

structural elements such as roles, cadence, and governance. For example, clear ownership must be set for the level-of-care call, daily or weekly huddles must keep outliers in check, and payer collaboration must be wired into the workflow so policy shifts can be adopted quickly. For a real-life example, consider Sharp HealthCare. The team treated its utilization management overhaul as a change-management project first, technology second. Staff were trained on MCG, evidence-based guidelines were embedded, and adoption followed a disciplined plan. In Sharp Healthcare's case, review times dropped by 80% to under 3 minutes, and potential savings of $6.95 million came from correctly shifting observation cases to inpatient. Leaders were clear that the results flowed from process redesign and training, not the algorithm alone.

3. **Delegate the rote to algorithms**. Let the system auto-flag missing authorizations and assemble appeal packets, freeing humans for complex judgment calls around medical necessity and unseen emotional and social factors.

4. **Govern by dual currencies**. Measure success in human trust AND financial gains because both metric collapses without the other. Here, setting the Key Performance Indicators (KPI) is critical and must be done as an inter-disciplinary team. Below is a sample of some KPI's that allow humans to oversee AI:

Sample KPI	What it Means	Definition
Policy Drift Latency (PDL)	How fast your UR program updates to payer policy changes.	Median hours from a payer bulletin release to the corresponding rule or workflow update in your UR platform and playbooks.
Low-Confidence Escalation Precision (LCEP)	Whether your "AI in the loop" is calibrated and handing off the right cases.	Among cases the model routes to humans due to confidence below threshold, the percentage where humans change the disposition or add documentation that changes the outcome.
Clinical Nuance Capture Rate (CNCR)	How often human documentation changes the trajectory of a case.	Percentage of admissions where added bedside nuance or SDoH documentation by UR or CDI leads to a different level of care, avoided denial, or successful preemptive P2P.
First Human Contact After Denial (FHCAD)	Patient access to a real person who can help.	Median hours from denial notification to first live contact with a UR nurse, physician advisor, or patient advocate who can escalate or explain next steps.
Weekend Vulnerability Index (WVI)	How well UR performs when staffing and decision support are thin.	Preventable denials per 100 admissions initiated Friday 3 p.m. to Monday 7 a.m., divided by the same rate for weekday admissions. Include status accuracy and P2P timeliness.
Algorithmic Equity Check (AEC)	Fairness of auto-denial patterns across patient groups.	Relative risk of denial or downgrade across cohorts after adjusting for severity, diagnosis, and payer. Example cohorts include age bands, race and ethnicity, primary language, disability status, and SDoH risk flags.

When algorithmic leaders steer the AI and nurses, coders, and analysts, the UR "vehicle" finally does what autonomous hype promises. It covers more ground, with fewer wrong turns, and everyone arrives intact.

The Long View

We're still in the high-gloss phase of the AI hype cycle. Health systems are chasing the next great AI vendor, hoping a smarter algorithm will finally fix UR. But speed without direction just leads to more wrong turns. In the long run, the winners won't be health systems with the best technical solution, they'll be the ones that create the appropriate structures to hard-wire human factors (workflow, bedside nuance, empathy, trust and incentives) into each algorithm they deploy. The health systems that pair structures and algorithmic leaders with UR self-driving technologies will quietly pull ahead with less detours and more results.

About the Authors

Tammy Gavin is a seasoned health system executive with 30+ years of experience navigating transformation elbow-to-elbow with front line clinical staff or leading from the front. She has a unique ability to develop and execute a vision and strategy that improves revenue generation and achieves operational efficiency while being a champion of the patient and maintaining strong business alliances among the medical staff, clinical staff and administration. She has proven results of changing culture and impacting outcomes across several health systems. For any questions, contact tgavin@longgame.com.

Munzoor Shaikh, MBA, is Founder and CEO of Long Game Health works with health plans, health systems and supporting organizations to optimize their performance and bring their missions to fruition. His expertise includes case management, population health, finance transformation, value-based care, artificial intelligence and healthcare private equity. For any questions, contact mshaikh@longgame.com.

Chris Steel, MD, CEO and CMO is a board-certified anesthesiologist and Interim Chief Executive and Chief Medical Officer at White River Health System, where he leads perioperative services, OPPE programs, and service-line growth. With 15+ years of experience spanning clinical care, operations, and regulatory strategy, he has also served as Chief of Staff, Chief Quality Officer, and directed anesthesia at four institutions. Nationally, he co-authored the ASA's Perioperative Surgical Home Implementation Guide and advises healthcare technology companies. For any questions, contact csteel@longgame.com.

Appendix

A "food secure" chart that isn't so and why it drives end-of-month hypoglycemia and avoidable cost

What's documented (and why it misleads):
A 62-year-old man with insulin-treated type 2 diabetes is admitted on the 8th of the month. His intake screen (e.g., the two-item Hunger Vital Sign) is negative for food insecurity, so no Z59.41 Food insecurity code is applied. The chart now reads as if nutrition access is stable. Clinically, that's misleading rather than "incorrect," because food access for many low-income patients is time-varying across the month. A single, early-month negative screen overstates stability and masks a predictable end-of-month risk window when benefits and cash run out. The record's snapshot therefore misclassifies risk over the patient's discharge horizon, even though the data point itself was honestly collected. (ICD-10-CM SDOH Z-codes: Z59.41 "Food insecurity"; see CMS SDOH Z-code resource.

Clinical consequence (time-varying risk is real and large):
Multiple population studies show hypoglycemia spikes late in the month among lower-income patients as food budgets are exhausted. In a widely cited California analysis, hospital admissions for hypoglycemia were 27% higher in the last week vs the first week of the month in low-income groups, with no such pattern in higher-income patients—classic evidence of "month-end" food scarcity affecting glycemic control.

Financial consequence (order of magnitude):
End-of-month hypoglycemia isn't just more frequent; it's costly. A claims-based study estimated mean inpatient costs ≈ $11,632 per hypoglycemia admission (median ≈ $3,609). Costs were even higher in the last 7 days: mean ≈ $17,098 vs ≈ $5,278 earlier in the month. For ED visits, mean costs were ≈ $3,040 late-month vs ≈ $1,356 earlier. These figures quantify the avoidable spend tied to that misleadingly reassuring early-month "food secure" label.

Why this is misleading documentation (not just "wrong" data):
The intake field is truthful for that date yet fails to represent the discharge-relevant time horizon. It causes downstream

teams (inpatient, UR, discharge planning) to under-anticipate food-related risk during the next 3–4 weeks when the patient is home. Without a Z59.41 flag or a note capturing anticipated end-of-month scarcity, care plans omit meal supports or insulin adjustment counseling timed to that window, thus raising the odds of readmission and higher total cost.

What the record should capture to avoid the trap:
An emotionally intelligent human is needed to know how to solicit potentially vulnerable information such as food insecurity when conducting patient intake. In fact, most clinical nurses will share stories about how they notice signs when either talking to patietns or visiting their homes. Such a nurse should engage the patient in a HUMAN WAY and replace the static snapshot with a time-anchored SDOH note and code, for example: "Food access is adequate in weeks 1–2; risk of shortage in weeks 3–4 due to fixed income/SNAP cycle, apply Z59.41, arrange medically tailored groceries or pantry referral for weeks 3–4, and adjust insulin plan accordingly." Thus a HUMAN can make the SDOH field decision-useful for discharge timing, pharmacy teaching, and UR authorization narratives.

References

Key Learnings

1. Pierson, Brendan. "Lawsuit Claims UnitedHealth AI Wrongfully Denies Elderly Extended Care." Reuters, November 14, 2023. https://www.reuters.com/legal/lawsuit-claims-unitedhealth-ai-wrongfully-denies-elderly-extended-care-2023-11-14/.
2. U.S. Department of Justice, Office of Public Affairs. "UCHealth Agrees to Pay $23M to Resolve Allegations of Fraudulent Billing for Emergency Department Visits." Press release, November 12, 2024. https://www.justice.gov/archives/opa/pr/uchealth-agrees-pay-23m-resolve-allegations-fraudulent-billing-emergency-department-visits.

Why Aren't Health Systems in UR Self-Driving Mode Yet?

3. Allen, Lisa. "How Cigna's PXDX Tool Denied 300,000 Claims in Two Months." ProPublica, March 25, 2023. https://www.propublica.org/article/cigna-health-denials-pxdx.
4. Farber, Madeline. "Whistleblowers Say naviHealth's Algorithm Got It Wrong 90 Percent of the Time." STAT News, November 14, 2023. https://www.statnews.com/2023/11/14/navihealth-lawsuit-ai-denials.
5. Cohrs, Rachel. "Senate Report Slams MA Carriers for Post-Acute Denial Surge." Healthcare Dive, October 4, 2024. https://www.healthcaredive.com/news/ma-denials-senate-report.

Algorithms Can't Read What's Not Documented

6. National Institutes of Health. "Unstructured Clinical Data: Challenges and Opportunities." NIH Data Brief 14 (2024): 1-6.
7. Ranney, Andrea, et al. "Effect of Missing SDoH on Sepsis Risk Prediction." JAMA Network Open 7, no. 9 (2024): e243812. https://doi.org/10.1001/jamanetworkopen.2024.3812.
8. American Health Information Management Association. 2024 CDI Staffing Benchmark Report. Chicago: AHIMA Press, 2024.
9. Patel, Rohit. "Ambient Voice in the ICU: A Multisite Pilot." Journal of Hospital Informatics 12, no. 2 (2024): 77-84.

The Road, the Map, and Traffic Conditions Keep Shifting

10. UnitedHealthcare. "Medical Policy Update Bulletins: January–April 2024." Corporate PDF archive, 50-90 pp each. https://www.uhcprovider.com/policies.
11. Centers for Medicare & Medicaid Services. HPMS Memo: Algorithmic Utilization Management, December 22, 2023.
12. California. SB 1120: Health Care Coverage Algorithms, Statutes of 2024.
13. Boddie, Erin. "Humana Pays $12 M to Resolve AI Denial Probe." Modern Healthcare, July 9, 2024. https://www.modernhealthcare.com/legal/humana-ai-denials-settlement.

Humans at the Helm, AI in the Loop

14. Hudson, Paul. "Physician Advisors Deliver 398 Percent ROI in UR." HFMA Strategic Financial Planning, May 2023. https://www.hfma.org/UR-physician-advisors.
15. Sharp HealthCare. "AI + Daily Huddles Cut Review Time 80 Percent." News release, June 7, 2025. https://www.sharp.com/news/ai-utilization-review.
16. MCG Health. "Leaders from Sharp HealthCare and MCG Discuss AI-Driven Utilization Management Success at CMSA 2025." News item, June 5, 2025. https://www.mcg.com/client-resources/news-item/cmsa-2025-dallas/.
17. Fee, James, Vaughn Matacale, and Natalia Dorf-Biderman. "Enterprisewide Physician Advisor Programs Are Key to Improving Costs and Revenue Cycle Performance." *hfm* (Healthcare Financial Management Association), May 2023, 1–5. Reprint PDF. https://www.enjoincdi.com/wp-content/uploads/2024/03/HFM-Article-May-2023-Fee-Matacale-Dorf-Biderman.pdf.

Regional Health System 'Off-the-Cliff' Case

18. A projected case summary obtained from Long Game past experience with clients.

The Path Forward

19. MCG Health. “Leaders from Sharp HealthCare and MCG Discuss AI-Driven Utilization Management Success at CMSA 2025.” News item, June 5, 2025. https://www.mcg.com/client-resources/news-item/cmsa-2025-dallas/.
20. XSOLIS. “AI-Powered Workflows Deliver 20% Shorter Length of Stay and Operational Efficiencies.” Case study (UMC Health System). Accessed August 28, 2025. https://www.xsolis.com/case-study/umc-health-system/
21. RevSpring and YouGov. 2025 Patient Financial Engagement Survey. Minneapolis: RevSpring, February 17, 2025.
22. Pew Research Center. “Public Views of Artificial Intelligence in Health Care.” February 2023. https://www.pewresearch.org/health-ai-survey-2023.

Appendix

23. Centers for Medicare & Medicaid Services (CMS), Office of Minority Health. *Improving the Collection of Social Determinants of Health (SDOH) Data with ICD-10-CM Z Codes.* Baltimore, MD: CMS, 2023. PDF. [https://www.cms.gov/files/document/cms-2023-omh-z-code-resource.pdf] (https://www.cms.gov/files/document/cms-2023-omh-z-code-resource.pdf)
24. Seligman, Hilary K., Ann F. Bolger, David Guzman, Andrea López, and Kirsten Bibbins-Domingo. “Exhaustion of Food Budgets at Month's End and Hospital Admissions for Hypoglycemia.” *Health Affairs* 33, no. 1 (January 2014): 116–23. https://doi.org/10.1377/hlthaff.2013.0096
(Open-access version:
https://pmc.ncbi.nlm.nih.gov/articles/PMC4215698/)
25. Basu, Sanjay, Seth A. Berkowitz, and Hilary K. Seligman. “The Monthly Cycle of Hypoglycemia: An Observational Claims-Based Study of Emergency Room Visits, Hospital Admissions, and Costs in a Commercially Insured Population.” *Medical Care* 55, no. 7 (July 2017): 639–45.
https://doi.org/10.1097/MLR.0000000000000728
(Publisher version: https://journals.lww.com/lww-medicalcare/fulltext/2017/07000/the_monthly_cycle_of_hypoglycemia__an.1.aspx)