Daniel Casillas
MPH Capstone Final Report
Colorado School of Public Health, University of Colorado Anschutz Medical Campus
December 2025

Copeptin and Metabolic and Kidney Markers in Type 1 and Type 2 Diabetes: A Cross-Sectional Analysis

Abstract

Background:
Copeptin is the C-terminal portion of the arginine vasopressin precursor and is widely used as a stable surrogate marker of vasopressin activity [1-3]. Higher copeptin has been linked to insulin resistance, poor glycemic control, hypertension, chronic kidney disease, and cardiovascular events in adults [4-9]. People with type 1 diabetes and people with type 2 diabetes differ in pathophysiology, metabolic profiles, and kidney risk, but it is unclear whether copeptin relates to key metabolic and kidney outcomes in the same way across these groups.

Methods:
This capstone used cross sectional data from research cohorts of youth and young adults with type 1 or type 2 diabetes seen in a pediatric endocrinology setting. The analytic sample included 40 participants with copeptin, hemoglobin A1c (HbA1c), blood pressure, estimated glomerular filtration rate by CKD-EPI (eGFR), and diabetes type. Smaller subsets had glucose infusion rate during a hyper insulinemic clamp at 190 mg/dL (GIR 190) and baseline free fatty acids. Linear regression models treated copeptin as the outcome and modeled each predictor with a diabetes-type indicator and an interaction term. Models were fit with and without hematocrit. Simple slopes were estimated within type 1 and type 2 diabetes and reported with 95 percent confidence intervals and p values.

Results:
Across all models there was no strong evidence that copeptin was associated with HbA1c, systolic or diastolic blood pressure, insulin sensitivity, free fatty acids, or eGFR at the alpha level of 0.05. In models without hematocrit, the slope for HbA1c was larger in type 2 diabetes than in type 1 diabetes (1.60 versus 0.54 units of copeptin per one-unit higher HbA1c, p for the type 2 slope = 0.0835), but confidence intervals were wide and crossed the null. Results were similar after adjustment for hematocrit, and interaction terms between each predictor and diabetes type were not statistically significant.

Conclusions:
In this small cross-sectional sample, copeptin was not strongly associated with glycemic control, blood pressure, insulin sensitivity, lipids, or eGFR, and there was no clear evidence that associations differed by diabetes type. Any real differences in copeptin biology between type 1 and type 2 diabetes will likely require larger, adequately powered studies with repeated measures of copeptin and metabolic outcomes.

Background

Copeptin is a 39-amino-acid glycopeptide derived from the C-terminal part of prepro-vasopressin. It is secreted in equimolar amounts with arginine vasopressin (AVP) and is more stable in blood, which has led to its use as a surrogate marker of vasopressin activity [1-3]. Elevated copeptin has been associated with higher blood pressure, increased risk of chronic kidney disease, and cardiovascular events in adult cohorts [4-9]. Vasopressin pathways are involved in water balance, osmoregulation, vascular tone, endocrine stress responses, and metabolic stress and inflammation [1-3]. These features make copeptin an appealing biomarker in diabetes.

In people with diabetes, higher copeptin has been related to poorer glycemic control, insulin resistance, microalbuminuria, and higher risk of kidney outcomes, especially in type 2 diabetes [4-7,9-11]. Many people with type 2 diabetes have obesity, dyslipidemia, and hypertension, while people with type 1 diabetes typically develop hyperglycemia due to autoimmune beta-cell destruction. These differences could affect vasopressin activity and its downstream effects on the kidney and cardiovascular system [4,7,10]. Copeptin has also been linked to diabetic kidney disease and vascular complications in both types of diabetes [5,10,11].

There is limited work that directly compares copeptin associations across type 1 and type 2 diabetes in the same setting, especially in younger patients with detailed phenotyping. Understanding whether copeptin relates similarly to glycemic control, blood pressure, insulin sensitivity, lipids, and kidney function in both groups is important for judging its value as a risk marker. If copeptin behaves differently by diabetes type, this could shape how it is used for risk stratification or as a potential intervention target. If associations look similar, copeptin may reflect more general vasopressin-mediated pathways that operate regardless of diabetes etiology.

The goal of this project was to assess whether copeptin is associated with HbA1c, blood pressure, insulin sensitivity, free fatty acids, and eGFR, and to test whether these associations differ between people with type 1 and type 2 diabetes. Hematocrit was included in secondary models to explore possible volume effects.

Methods

Study design and data source

This capstone is a secondary analysis of data from pediatric endocrinology research cohorts of youth and young adults with type 1 or type 2 diabetes. Participants were recruited from specialty clinics and research studies focused on diabetes and kidney disease, where standardized protocols were used for clinical assessments and laboratory measurements.

For this project, the analysis is cross sectional and uses data from a single study visit for each participant, where copeptin and the other key measures were collected. All data were de identified before analysis. The aim of this project is exploratory and methodological rather than to provide definitive clinical recommendations. All parent studies were approved by the institutional review board, and only de-identified data were used for this secondary analysis. No

formal a priori power calculation was conducted because the sample size was fixed by available data.

Study population

The analytic sample included participants with type 1 or type 2 diabetes who had measured copeptin and at least one of the main predictors. For the primary set of models, 40 participants (20 with type 1 diabetes and 20 with type 2 diabetes) had complete data on copeptin, HbA1c, systolic and diastolic blood pressure, eGFR by CKD-EPI, diabetes type, and hematocrit.

For insulin sensitivity, 32 participants (16 in each diabetes group) had glucose infusion rate at 190 mg/dL from a hyper insulinemic clamp. For baseline free fatty acids, 30 participants (15 in each group) had complete data.

Participants were excluded from a specific model if they were missing copeptin, the relevant predictor, or diabetes type. As a result, the analytic sample size varies slightly across predictors and is reported in each table. No other exclusion criteria were applied.

Outcome: copeptin

Copeptin was the primary outcome for all models. It was measured in plasma using an immunoassay and recorded in pmol/L. Copeptin was treated as a continuous variable on its original scale. Given the modest sample size and the exploratory focus, no transformations or categorization were applied.

Predictors and covariates

The main continuous predictors, modeled one at a time, were:

- Hemoglobin A1c (HbA1c, percent)
- Systolic blood pressure (SBP, mmHg)
- Diastolic blood pressure (DBP, mmHg)
- Glucose infusion rate during a hyper insulinemic clamp at 190 mg/dL (GIR 190, mg/kg/min), as a measure of insulin sensitivity
- Baseline free fatty acids (FFA, mmol/L)
- Estimated glomerular filtration rate by CKD-EPI (eGFR, mL/min/1.73 m^2), as a measure of kidney function [12,13]

Diabetes type was included as a binary indicator, with type 1 diabetes as the reference group and type 2 diabetes coded as 1. In all models, an interaction term between the continuous predictor and diabetes type was included so that slopes could differ between groups.

Hematocrit was added as a continuous covariate in secondary models, based on the idea that volume status and red blood cell concentration might influence copeptin levels.

No additional covariates were included in the primary models to preserve precision and avoid overfitting in this small dataset.

Statistical analysis

Copeptin was modeled as a continuous outcome. For each predictor X listed above, two linear regression models were fit:

1. A base interaction model with X, diabetes type, and an X by diabetes type interaction term.
2. A second model that added hematocrit while keeping the interaction term.

The regression equation was

$$\text{copeptin} = \beta_0 + \beta_1 X + \beta_2 I(\text{T2D}) + \beta_3 X \cdot I(\text{T2D})[+ \beta_4 \text{hematocrit}],$$

where $I(\text{T2D})$ is an indicator for type 2 diabetes. Under this setup, β_1 is the slope of copeptin on X in type 1 diabetes, and $\beta_1 + \beta_3$ is the slope in type 2 diabetes. Simple slopes and 95 percent confidence intervals for each diabetes type were obtained from these models.

For each model, parameter estimates, standard errors, t statistics, p values, and 95 percent confidence intervals were extracted. Residual standard deviations were recorded as a summary of overall variability. The focus was on the simple slopes within each diabetes type and on whether the interaction term suggested a meaningful difference in associations between type 1 and type 2 diabetes.

Analyses used a two-sided alpha of 0.05. Because the project is exploratory and the sample is small, no formal multiple comparison adjustment was applied. All analyses were conducted in R, and model summaries were exported for tabular presentation.

Results

Study population

The analytic sample included 40 participants with complete data on copeptin, HbA1c, blood pressure, eGFR, diabetes type, and hematocrit, with 20 individuals in the type 1 diabetes group and 20 in the type 2 diabetes group. For the GIR models, 16 participants with type 1 diabetes and 16 with type 2 diabetes had clamp data. For baseline free fatty acids, 15 participants per group contributed data. These counts are summarized in Table 1.

The parent studies include a range of ages and clinical characteristics, but this capstone focuses on patterns of association between copeptin and the main metabolic and kidney measures by diabetes type, rather than on a full description of baseline characteristics. Table 1 highlights how many participants contributed to each set of models.

Associations between copeptin and metabolic markers without hematocrit

Table 2 shows the simple slopes of copeptin on each predictor within type 1 and type 2 diabetes for models that included only the predictor, diabetes type, and their interaction. None of the slopes reached the 0.05 significance level, but a few patterns are worth noting.

For HbA1c, the estimated slope in type 1 diabetes was 0.54 units of copeptin per one-unit higher HbA1c (95 percent CI −1.93 to 3.00, p = 0.6509), while the slope in type 2 diabetes was 1.60 (95 percent CI −0.24 to 3.44, p = 0.0835). Both confidence intervals include zero, but the point estimate is larger in type 2 diabetes, and the p value is closer to 0.05, suggesting a possible trend toward higher copeptin with worse glycemic control in type 2 diabetes. This pattern is illustrated in **Figure 1**, which plots copeptin versus HbA1c with separate fitted lines for each diabetes type.

For systolic and diastolic blood pressure, slopes were small and imprecise in both groups. Systolic blood pressure had near-null slopes in type 1 diabetes (−0.07, 95 percent CI −0.41 to 0.28) and type 2 diabetes (0.20, 95 percent CI −0.27 to 0.67). Diastolic blood pressure showed similarly weak associations, with wide intervals and p values well above 0.05.

The association between copeptin and insulin sensitivity, measured by GIR 190, was also not statistically significant. In type 1 diabetes, the slope was 0.53 (95 percent CI −1.36 to 2.43), and in type 2 diabetes the slope was −0.70 (95 percent CI −2.18 to 0.77). The point estimates differ in direction, but both confidence intervals are compatible with no association.

For baseline free fatty acids and eGFR, slopes were very close to zero in both groups. For example, the slope for eGFR was 0.08 in type 1 diabetes (95 percent CI −0.17 to 0.32) and 0.02 in type 2 diabetes (95 percent CI −0.29 to 0.32), with p values near 0.9. These results suggest little evidence that copeptin is strongly related to eGFR within the relatively preserved kidney function range of this sample.

Detailed regression coefficients, including interaction terms, are shown in Table 4 for the models without hematocrit.

Figure 1. Flow from the parent research cohorts to the final analytic samples

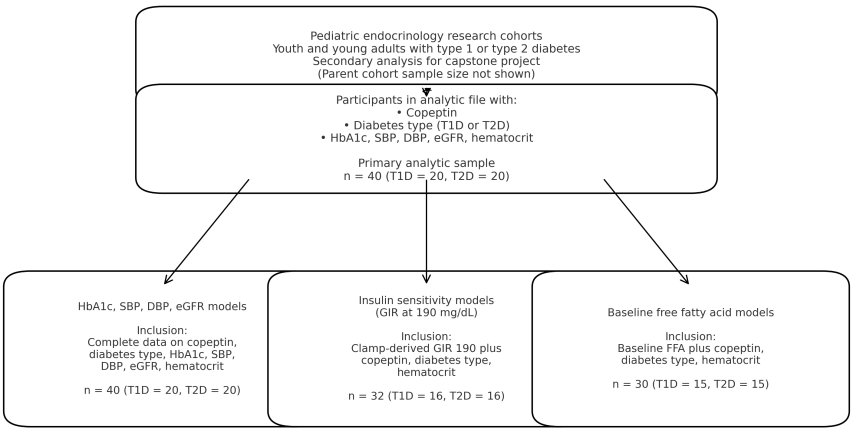
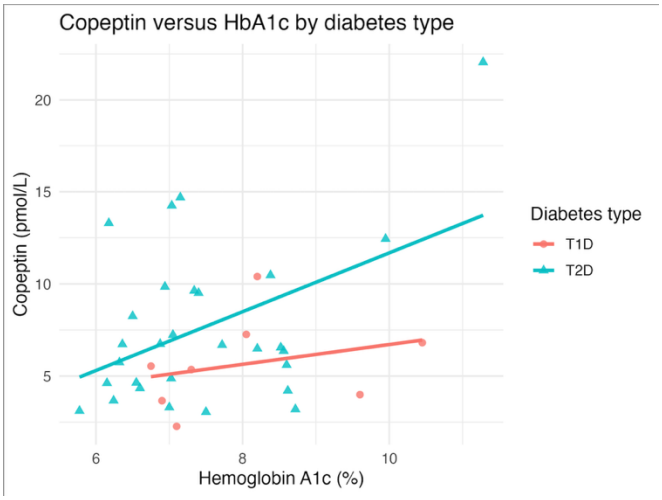


Figure 2. Copeptin versus hemoglobin A1c by diabetes type



Associations between copeptin and metabolic markers with hematocrit adjustment

Table 3 presents simple slopes from models that additionally adjusted for hematocrit. Overall, results were very similar to the unadjusted models, and adding hematocrit did not materially change the estimated associations.

For HbA1c, the slope in type 1 diabetes was 0.52 (95 percent CI −2.04 to 3.09, p = 0.6705), and the slope in type 2 diabetes was 1.59 (95 percent CI −0.32 to 3.50, p = 0.0969). These values are almost identical to the slopes from the models without hematocrit. Systolic and diastolic blood pressure slopes remained near zero in both types, with wide confidence intervals that included the null.

The slopes for GIR, baseline free fatty acids, and eGFR also remained small and non-significant after hematocrit adjustment. For GIR, the estimated slope was 0.69 in type 1 diabetes (95 percent CI −1.42 to 2.79) and −0.57 in type 2 diabetes (95 percent CI −2.23 to 1.09). For eGFR, slopes were 0.09 in type 1 diabetes (95 percent CI −0.17 to 0.34) and −0.01 in type 2 diabetes (95 percent CI −0.35 to 0.34).

Hematocrit itself was not significantly associated with copeptin in any of the models, and its inclusion did not change residual standard deviations or overall model fit in a meaningful way. Full coefficient tables for the hematocrit-adjusted models are also shown in Table 4.

Summary of interaction findings

Across all predictors, the interaction terms between the continuous predictors and diabetes type were not statistically significant. Some point estimates differed between type 1 and type 2 diabetes, but the data do not provide strong evidence that these differences are larger than what would be expected by chance in a small sample. This is especially true for HbA1c and GIR, where the point estimates differed in magnitude or direction, but interaction p values stayed above 0.05.

Overall, the results suggest that in this dataset there is no clear evidence of strong linear relationships between copeptin and the metabolic or kidney outcomes studied, and no convincing evidence that these relationships differ by diabetes type.

Discussion

This capstone examined whether copeptin is associated with glycemic control, blood pressure, insulin sensitivity, free fatty acids, and kidney function in a small sample of individuals with type 1 and type 2 diabetes, and whether those associations differ by diabetes type. Copeptin has been proposed as a biomarker of vasopressin activity that may be linked to adverse cardiometabolic and kidney outcomes in adults with and without diabetes [1-3,4-9,11]. In this analysis, however, all estimated associations were modest, imprecise, and not statistically significant.

The most suggestive pattern was a larger positive slope between copeptin and HbA1c in type 2 diabetes compared with type 1 diabetes. This lines up with previous work that has linked higher copeptin to insulin resistance, hyperglycemia, and incident type 2 diabetes in adults [4-7,9]. It is biologically plausible that vasopressin pathways and copeptin might be more closely tied to metabolic stress in type 2 diabetes, where obesity and insulin resistance are common. In this sample, though, the confidence intervals around the HbA1c slopes were wide, p values were above 0.05, and interaction tests did not provide strong evidence for a difference between diabetes types. For blood pressure, insulin sensitivity, free fatty acids, and eGFR, the data were even more consistent with no meaningful linear association.

Several factors could explain these largely null findings. The sample is small, especially for the GIR and free fatty acid models, which limits power to detect modest associations. Participants in these research cohorts are often relatively well managed, which can restrict variability in HbA1c,

blood pressure, and kidney function. If most participants fall in a narrow and healthy range, it becomes harder to see relationships with copeptin. The design is cross sectional, so a single copeptin value is being related to single time point outcomes. Previous prospective studies suggest that copeptin may be more informative for long term trajectories of kidney function and cardiovascular events rather than cross sectional measurements [4-9]. Relationships that are clear in large adult cohorts may not be detectable in a small pediatric and young adult sample [4-7,9-11].

Even with these limitations, the study has strengths. Copeptin was measured in a research environment with standardized protocols, alongside detailed measures of insulin sensitivity and kidney function. Including both type 1 and type 2 diabetes in the same analysis allowed a direct test of interaction by diabetes type. The modeling strategy was transparent, with clearly defined interaction terms, simple slopes, and secondary adjustment for hematocrit.

From a public health perspective, these findings do not support using copeptin as a stand-alone marker of glycemic control, blood pressure, or kidney function in youth and young adults with diabetes based on this dataset. They also do not rule out a useful role for copeptin in larger or more varied populations or as part of a broader risk panel. The main message is that future work on copeptin will need larger, longitudinal studies rather than that copeptin has no value.

Conclusions

In this cross-sectional analysis of 40 individuals with type 1 and type 2 diabetes, there was no strong evidence that copeptin was associated with HbA1c, systolic or diastolic blood pressure, insulin sensitivity, free fatty acids, or eGFR, and no clear indication that these relationships differed by diabetes type. Adjustment for hematocrit did not materially change the results. Any true associations between copeptin and these metabolic or kidney outcomes are likely modest and will require larger studies with better precision. Overall, this project suggests that copeptin is not a simple linear proxy for standard clinical measures in this setting and highlights the need for more comprehensive evaluation of copeptin as a biomarker in diabetes.

Limitations and future research

This project has several limitations. The sample size was small, which limited power to detect modest associations and made interaction tests by diabetes type especially underpowered. The analysis was cross sectional, so it cannot address temporal or causal relationships. Participants were enrolled from research cohorts in a pediatric endocrinology setting, which may not represent the broader population of people with diabetes. The models were intentionally simple and did not adjust for age, sex, body mass index, or diabetes duration to preserve precision.

Future research should evaluate copeptin in larger multicenter cohorts that include both type 1 and type 2 diabetes across a wider range of age, disease duration, and kidney function. Longitudinal studies with repeated measurements of copeptin, metabolic markers, and kidney outcomes would allow tests of whether copeptin predicts changes over time. It would also be useful to examine copeptin as part of multivariable risk prediction models and to test whether it adds information beyond standard clinical measures. Mechanistic studies could help clarify how

vasopressin pathways interact with glucose metabolism, blood pressure regulation, and kidney injury in different forms of diabetes.

Assessment of competencies

This capstone gave me a chance to bring together biostatistics, epidemiology, communication, and professional judgment in the context of a real clinical research question.

From a biostatistical perspective, I applied regression modeling to a small but complex dataset. I built analysis datasets from raw files, defined clear inclusion criteria, and chose linear regression with interaction terms to answer the question of whether associations between copeptin and metabolic or kidney outcomes differ by diabetes type. I summarized parameter estimates, confidence intervals, and p values in a way that can be interpreted both statistically and clinically. Working with a small sample pushed me to think carefully about model complexity, precision, and overfitting.

Epidemiologic reasoning guided how I framed and interpreted the work. I defined the exposure, outcomes, and covariates in line with the research question, considered confounding by factors such as hematocrit and diabetes type, and used interaction terms to test for effect modification. I interpreted findings within the limits of cross-sectional observational data and kept a clear distinction between association and causation. Thinking about generalizability also helped me remember that this sample reflects research participants, not all people living with diabetes.

This project also strengthened my communication skills. I translated statistical output into tables and one key figure that clinicians and collaborators can follow, and I practiced explaining results in plain language. Preparing both a slide presentation and this report required organizing methods and results into a clear story. Feedback from mentors on early versions of the slides and text helped me tighten the narrative and focus on what is most clinically relevant.

Professionalism and ethics were part of the process throughout. I worked with de identified data that were collected under approved protocols, and I was careful not to overstate what can be learned from a small exploratory study. I documented analytic decisions and maintained reproducible code. Being transparent about limitations and the exploratory nature of the work is an important part of responsible biostatistical practice.

Overall, this capstone deepened my sense of how biostatistics fits into the broader public health and clinical research system. Copeptin and related biomarkers sit at the interface of basic science, clinical care, and population health. Even though this project did not find strong associations, it highlighted the challenges of moving biomarkers from hypothesis to clinical application, especially in pediatric and young adult groups. The experience reinforced my interest in diabetes and kidney disease research and strengthened the skills I will bring to future multidisciplinary teams.

Tables

Table 1. Analytic sample by outcome and diabetes type

Outcome	Model set	Type 1 diabetes, n	Type 2 diabetes, n	Total n
HbA1c, SBP, DBP, eGFR	Base and hematocrit models	20	20	40
GIR at 190 mg/dL	Base and hematocrit models	16	16	32
Baseline free fatty acids	Base and hematocrit models	15	15	30

Table 2. Simple slopes of copeptin on metabolic and kidney measures by diabetes type, models without hematocrit

Predictor	Diabetes type	n	Slope (per unit increase in predictor)	95% CI	p value
HbA1c	T1D	20	0.536	−1.929, 3.002	0.6509
HbA1c	T2D	20	1.600	−0.237, 3.437	0.0835
SBP	T1D	20	−0.065	−0.408, 0.278	0.6931
SBP	T2D	20	0.200	−0.267, 0.667	0.3783
DBP	T1D	20	0.114	−0.568, 0.796	0.7277
DBP	T2D	20	−0.115	−0.638, 0.409	0.6488
GIR 190	T1D	16	0.532	−1.362, 2.427	0.5518
GIR 190	T2D	16	−0.703	−2.175, 0.770	0.3190
Baseline FFA	T1D	15	0.000	−0.004, 0.004	0.9034
Baseline FFA	T2D	15	0.001	−0.005, 0.008	0.6588
eGFR (CKD-EPI)	T1D	20	0.076	−0.168, 0.321	0.5166

Predictor	Diabetes type	n	Slope (per unit increase in predictor)	95% CI	p value
eGFR (CKD-EPI)	T2D	20	0.019	−0.286, 0.323	0.8976

Table 3. Simple slopes of copeptin on metabolic and kidney measures by diabetes type, models with hematocrit

Predictor	Diabetes type	n	Slope (per unit increase in predictor)	95% CI	p value
HbA1c	T1D	20	0.523	−2.044, 3.089	0.6705
HbA1c	T2D	20	1.589	−0.324, 3.503	0.0969
SBP	T1D	20	−0.067	−0.423, 0.290	0.6950
SBP	T2D	20	0.192	−0.299, 0.684	0.4173
DBP	T1D	20	0.113	−0.593, 0.820	0.7371
DBP	T2D	20	−0.104	−0.657, 0.449	0.6939
GIR 190	T1D	16	0.689	−1.415, 2.794	0.4859
GIR 190	T2D	16	−0.566	−2.225, 1.093	0.4683
Baseline FFA	T1D	15	0.000	−0.005, 0.005	0.9522
Baseline FFA	T2D	15	0.001	−0.006, 0.008	0.6643
eGFR (CKD-EPI)	T1D	20	0.085	−0.171, 0.342	0.4899
eGFR (CKD-EPI)	T2D	20	−0.006	−0.346, 0.335	0.9728

Table 4. Regression coefficients for copeptin on selected predictors, by model specification

(a) Models without hematocrit

Predictor	Term	Beta	SE	p value	95% CI low	95% CI high	Residual SD	n
HbA1c	HbA1c (main effect, T1D slope)	0.5364	1.1630	0.6509	−1.9290	3.0017	4.1383	20
HbA1c	Diabetes: T2D vs T1D	−4.4331	11.7014	0.7098	−29.2389	20.3728	4.1383	20
HbA1c	Interaction: HbA1c × T2D	1.0636	1.4504	0.4740	−2.0112	4.1384	4.1383	20
SBP	SBP (main effect, T1D slope)	−0.0650	0.1618	0.6931	−0.4081	0.2780	4.4481	20
SBP	Diabetes: T2D vs T1D	−28.0174	32.9838	0.4082	−97.9399	41.9051	4.4481	20
SBP	Interaction: SBP × T2D	0.2646	0.2733	0.3474	−0.3148	0.8440	4.4481	20
DBP	DBP (main effect, T1D slope)	0.1139	0.3216	0.7277	−0.5678	0.7957	4.5345	20
DBP	Diabetes: T2D vs T1D	21.0140	31.0126	0.5077	−44.7297	86.7578	4.5345	20
DBP	Interaction: DBP × T2D	−0.2286	0.4055	0.5808	−1.0881	0.6310	4.5345	20
GIR 190	GIR 190 (main effect, T1D slope)	0.5324	0.8697	0.5518	−1.3624	2.4272	3.3808	16
GIR 190	Diabetes: T2D vs T1D	12.4312	9.0147	0.1931	−7.2102	32.0726	3.3808	16
GIR 190	Interaction: GIR 190 × T2D	−1.2349	1.1013	0.2841	−3.6346	1.1647	3.3808	16
Baseline FFA	Baseline FFA (main effect, T1D slope)	−0.0002	0.0019	0.9034	−0.0043	0.0039	3.6800	15
Baseline FFA	Diabetes: T2D vs T1D	0.8337	4.1343	0.8439	−8.2660	9.9333	3.6800	15
Baseline FFA	Interaction: Baseline FFA × T2D	0.0016	0.0035	0.6611	−0.0061	0.0093	3.6800	15
eGFR CKD-EPI	eGFR (main effect, T1D slope)	0.0764	0.1152	0.5166	−0.1679	0.3207	4.5185	20
eGFR CKD-EPI	Diabetes: T2D vs T1D	10.1576	22.1552	0.6528	−36.8093	57.1246	4.5185	20
eGFR CKD-EPI	Interaction: eGFR × T2D	−0.0576	0.1841	0.7583	−0.4480	0.3327	4.5185	20

(b) Models with hematocrit

Predictor	Term	Beta	SE	p value	95% CI low	95% CI high	Residual SD	n
HbA1c	HbA1c (main effect, T1D slope)	0.5225	1.2041	0.6705	−2.0440	3.0890	4.2710	20
HbA1c	Diabetes: T2D vs T1D	−4.4460	12.0771	0.7179	−30.1877	21.2957	4.2710	20
HbA1c	Hematocrit	0.0306	0.2121	0.8871	−0.4214	0.4826	4.2710	20
HbA1c	Interaction: HbA1c × T2D	1.0668	1.4971	0.4870	−2.1242	4.2579	4.2710	20
SBP	SBP (main effect, T1D slope)	−0.0669	0.1673	0.6950	−0.4234	0.2897	4.5891	20
SBP	Diabetes: T2D vs T1D	−27.3737	34.2233	0.4363	−100.3190	45.5716	4.5891	20
SBP	Hematocrit	0.0408	0.2303	0.8617	−0.4501	0.5317	4.5891	20
SBP	Interaction: SBP × T2D	0.2593	0.2836	0.3749	−0.3451	0.8637	4.5891	20
DBP	DBP (main effect, T1D slope)	0.1134	0.3316	0.7371	−0.5935	0.8203	4.6762	20
DBP	Diabetes: T2D vs T1D	20.2318	32.1917	0.5391	−48.3832	88.8468	4.6762	20
DBP	Hematocrit	0.0500	0.2350	0.8343	−0.4508	0.5509	4.6762	20
DBP	Interaction: DBP × T2D	−0.2175	0.4213	0.6132	−1.1156	0.6806	4.6762	20
GIR 190	GIR 190 (main effect, T1D slope)	0.6894	0.9562	0.4859	−1.4151	2.7940	3.4945	16
GIR 190	Diabetes: T2D vs T1D	12.7981	9.3489	0.1983	−7.7787	33.3748	3.4945	16
GIR 190	Hematocrit	0.1167	0.2423	0.6395	−0.4166	0.6501	3.4945	16
GIR 190	Interaction: GIR 190 × T2D	−1.2556	1.1392	0.2939	−3.7629	1.2517	3.4945	16
Baseline FFA	Baseline FFA (main effect, T1D slope)	0.0001	0.0021	0.9522	−0.0046	0.0049	3.8297	15
Baseline FFA	Diabetes: T2D vs T1D	1.7104	4.8384	0.7310	−9.0703	12.4911	3.8297	15
Baseline FFA	Hematocrit	0.1245	0.3144	0.7003	−0.5760	0.8250	3.8297	15

Predictor	Term	Beta	SE	p value	95% CI low	95% CI high	Residual SD	n
Baseline FFA	Interaction: Baseline FFA × T2D	0.0013	0.0037	0.7449	−0.0071	0.0096	3.8297	15
eGFR CKD-EPI	eGFR (main effect, T1D slope)	0.0853	0.1205	0.4899	−0.1715	0.3420	4.6423	20
eGFR CKD-EPI	Diabetes: T2D vs T1D	14.2151	24.9457	0.5772	−38.9553	67.3856	4.6423	20
eGFR CKD-EPI	Hematocrit	0.1000	0.2516	0.6966	−0.4362	0.6363	4.6423	20
eGFR CKD-EPI	Interaction: eGFR × T2D	−0.0908	0.2068	0.6668	−0.5315	0.3499	4.6423	20

References

1. Khattri Bhandari S, et al. Copeptin as a marker for vasopressin dysregulation and related disorders. Journal of Applied Laboratory Medicine. 2025.
2. Mu D, et al. Copeptin as a diagnostic and prognostic biomarker in cardiovascular disease. Frontiers in Cardiovascular Medicine. 2022;
3. Choy KW, et al. Copeptin as a surrogate marker for arginine vasopressin. Clinical review. 2024.[
4. Wannamethee SG, et al. Copeptin, insulin resistance, and risk of incident diabetes in older men. Diabetes Care. 2015;38(3):422–429.
5. Roussel R, et al. Plasma copeptin, AVP gene variants, and incidence of type 2 diabetes. Journal of Clinical Endocrinology and Metabolism. 2016;101(6):2432–2440.
6. Velho G, et al. Plasma copeptin, kidney disease, and risk for cardiovascular morbidity in patients with type 2 diabetes. Diabetes Care. 2018;41(9):e149–e151.
7. Tasevska I, et al. Increased levels of copeptin predict decline in kidney function and risk of chronic kidney disease. American Journal of Nephrology. 2016;44(1):22–30.

8. Jensen T, et al. Copeptin and estimated insulin sensitivity in adults with and without type 1 diabetes. 2018.
9. Bjornstad P, et al. Copeptin and albuminuria in men with type 1 diabetes: a brief report from the T1D Exchange Biobank. 2017.
10. National Institute of Diabetes and Digestive and Kidney Diseases. eGFR Equations for Adults: CKD-EPI equations. NIDDK website. Accessed 2025.
11. Arroyo JP, et al. Peripheral insulin resistance is associated with copeptin in patients with CKD stage 3 to 4. Kidney360. 2021.

Code:

```
# ============================================================

# Copeptin ~ X * Diabetes (+/- hematocrit) - FINAL SCRIPT

# - Reads and cleans data (handles duplicate 'ethnicity')

# - Fits interaction models per predictor

# - Simple slopes via emtrends() with dynamic slope-column detection

# - Exports CSVs and a Word document with clean tables

# - Adds one key figure (copeptin vs HbA1c by diabetes type) to the Word file
```

```
# ============================================================

# 0) Install and load -------------------------------------------------------
need <- c(
  "readxl","dplyr","stringr","janitor","broom","purrr","emmeans",
  "tidyr","readr","rlang","flextable","officer","ggplot2"
)
to_install <- need[!need %in% rownames(installed.packages())]
if (length(to_install)) install.packages(to_install, quiet = TRUE)
invisible(lapply(need, library, character.only = TRUE))

# 1) Read -------------------------------------------------------------------
# Data were already imported into the Environment as CapstoneDataCasillas
# so we just assign it to 'raw' and continue.
raw <- CapstoneDataCasillas

# 2) Clean names and resolve duplicate ethnicity ----------------------------
master <- raw |>
  janitor::clean_names()

eth_candidates <- intersect(c("ethnicity_3","ethnicity_24","ethnicity"), names(master))
if (length(eth_candidates) == 0) {
  master$ethnicity <- NA_character_
} else if (length(eth_candidates) == 1) {
  master$ethnicity <- master[[eth_candidates[1]]]
} else {
```

```
  master <- master |>
    dplyr::mutate(ethnicity = dplyr::coalesce(!!!rlang::syms(eth_candidates)))
}

# 3) Select/rename variables, build diabetes factor ----------------------------
egfr_name <- dplyr::case_when(
  "e_gfr_ckd_epi" %in% names(master) ~ "e_gfr_ckd_epi",
  "egfr_ckd_epi"  %in% names(master) ~ "egfr_ckd_epi",
  "egfr"          %in% names(master) ~ "egfr",
  TRUE ~ NA_character_
)
if (is.na(egfr_name)) stop("Could not find eGFR column. Update 'egfr_name' lookup.")

df <- master |>
  dplyr::select(dplyr::any_of(c(
    "record_id","group","copeptin","hba1c","sbp","dbp","gir_190",
    "baseline_ffa", egfr_name, "hematocrit"
  ))) |>
  dplyr::rename(egfr_ckd_epi = !!egfr_name) |>
  dplyr::mutate(
    dplyr::across(
      c(copeptin,hba1c,sbp,dbp,gir_190,baseline_ffa,egfr_ckd_epi,hematocrit),
      ~ suppressWarnings(as.numeric(.))
    ),
    group_str = ifelse(is.na(group), "", as.character(group)),
    diabetes = dplyr::case_when(
```

```
    stringr::str_detect(group_str, "Type\\s*1") ~ "T1D",
    stringr::str_detect(group_str, "Type\\s*2") ~ "T2D",
    TRUE ~ NA_character_
   ),
   diabetes = factor(diabetes, levels = c("T1D","T2D"))
  ) |>
  dplyr::select(-group_str)

# Helper: find slope column in emtrends summary ------------------------------
.get_slope_col <- function(trs, x_var) {
  nm <- names(trs)
  cand <- nm[grepl("\\.trend$", nm, perl = TRUE)]  # e.g., "hba1c.trend"
  if (length(cand) >= 1) return(cand[1])
  if ("trend" %in% nm) return("trend")          # older emmeans
  if (x_var %in% nm) return(x_var)              # rare fall-back
  stop("Could not find slope column in emtrends summary for predictor: ", x_var)
}

# 4) Helpers -----------------------------------------------------------------
fit_one_slopes <- function(x_var, adjust_hematocrit = FALSE) {
  rhs <- paste0("`", x_var, "` * diabetes",
                if (adjust_hematocrit) " + hematocrit" else "")
  fml <- as.formula(paste("copeptin ~", rhs))

  dat <- df |>
    dplyr::select(copeptin, diabetes, hematocrit, dplyr::all_of(x_var)) |>
```

```
  tidyr::drop_na()

if (nrow(dat) < 10 || all(is.na(dat[[x_var]])) ||
    is.na(stats::var(dat[[x_var]], na.rm = TRUE)) ||
    stats::var(dat[[x_var]], na.rm = TRUE) == 0) {
  return(tibble::tibble(
    predictor = x_var,
    model = ifelse(adjust_hematocrit, "X * diabetes + hematocrit", "X * diabetes"),
    diabetes = c("T1D","T2D"),
    slope = NA_real_, se = NA_real_, df = NA_real_, t = NA_real_, p = NA_real_,
    ci_low = NA_real_, ci_high = NA_real_, n = nrow(dat)
  ))
}

mod <- stats::lm(fml, data = dat)

tr  <- suppressWarnings(emmeans::emtrends(mod, specs = ~ diabetes, var = x_var))
trs <- suppressWarnings(summary(tr, infer = TRUE))  # adds lower.CL / upper.CL

slope_col <- .get_slope_col(trs, x_var)

out <- trs |>
  dplyr::transmute(
    predictor = x_var,
    model = ifelse(adjust_hematocrit, "X * diabetes + hematocrit", "X * diabetes"),
    diabetes = as.character(.data$diabetes),
```

```
    slope = .data[[slope_col]],
    se    = .data$SE,
    df    = .data$df,
    t     = .data$`t.ratio`,
    p     = .data$`p.value`,
    ci_low  = .data$lower.CL,
    ci_high = .data$upper.CL,
    n = nrow(dat)
   )
  out
}

coef_table_one <- function(x_var, adjust_hematocrit = FALSE) {
  rhs <- paste0("`", x_var, "` * diabetes",
                if (adjust_hematocrit) " + hematocrit" else "")
  fml <- as.formula(paste("copeptin ~", rhs))

  dat <- df |>
    dplyr::select(copeptin, diabetes, hematocrit, dplyr::all_of(x_var)) |>
    tidyr::drop_na()

  if (nrow(dat) < 10 || all(is.na(dat[[x_var]])) ||
      is.na(stats::var(dat[[x_var]], na.rm = TRUE)) ||
      stats::var(dat[[x_var]], na.rm = TRUE) == 0) {
    return(tibble::tibble(
      predictor = x_var,
```

```
    model = ifelse(adjust_hematocrit, "X * diabetes + hematocrit", "X * diabetes"),
    term = NA_character_, beta = NA_real_, se = NA_real_, t = NA_real_, p = NA_real_,
    ci_low = NA_real_, ci_high = NA_real_, sigma_resid = NA_real_, n = nrow(dat)
  ))
}

mod <- stats::lm(fml, data = dat)

tt <- broom::tidy(mod, conf.int = TRUE) |>
  dplyr::transmute(
    predictor = x_var,
    model = ifelse(adjust_hematocrit, "X * diabetes + hematocrit", "X * diabetes"),
    term = .data$term,
    beta = .data$estimate,
    se   = .data$std.error,
    t    = .data$statistic,
    p    = .data$p.value,
    ci_low  = .data$conf.low,
    ci_high = .data$conf.high
  )

gg <- broom::glance(mod) |>
  dplyr::transmute(
    sigma_resid = dplyr::coalesce(.data$sigma, NA_real_),
    n = dplyr::coalesce(.data$nobs, NA_real_)
  )
```

```
  dplyr::bind_cols(tt, gg[rep(1, nrow(tt)), ])
}

pretty_results <- function(slopes_tbl) {
  slopes_tbl |>
    dplyr::mutate(
      slope_ci = dplyr::if_else(
        is.na(.data$slope),
        NA_character_,
        sprintf("%.3f (%.3f, %.3f)", .data$slope, .data$ci_low, .data$ci_high)
      ),
      p_slope = dplyr::if_else(
        is.na(.data$p),
        NA_character_,
        sprintf("%.4f", .data$p)
      )
    ) |>
    dplyr::select(predictor, model, diabetes, n, slope_ci, p_slope)
}

pretty_term <- function(x, x_var) {
  x |>
    dplyr::mutate(
      term_pretty = dplyr::case_when(
        term == "(Intercept)" ~ "(Intercept)",
```

```
      term == x_var ~ paste0(x_var, " (main effect)"),
      term == "diabetesT2D" ~ "Diabetes: T2D vs T1D",
      term %in% c(paste0("`", x_var, "`:diabetesT2D"),
                  paste0(x_var, ":diabetesT2D")) ~ paste0("Interaction: ", x_var, " × T2D"),
      term == "hematocrit" ~ "Hematocrit",
      TRUE ~ term
    )
  )
}

# 5) Run models across predictors ------------------------------------------

predictors <- c("hba1c","sbp","dbp","gir_190","baseline_ffa","egfr_ckd_epi")

slopes_no_hema   <- purrr::map_dfr(predictors, fit_one_slopes, adjust_hematocrit = FALSE)
slopes_with_hema <- purrr::map_dfr(predictors, fit_one_slopes, adjust_hematocrit = TRUE)

models_no_hema   <- pretty_results(slopes_no_hema)
models_with_hema <- pretty_results(slopes_with_hema)

coef_no_hema   <- purrr::map_dfr(predictors, coef_table_one, adjust_hematocrit = FALSE)
coef_with_hema <- purrr::map_dfr(predictors, coef_table_one, adjust_hematocrit = TRUE)

coef_no_hema <- purrr::map2_dfr(
  split(coef_no_hema, coef_no_hema$predictor),
  names(split(coef_no_hema, coef_no_hema$predictor)),
```

```
  ~ pretty_term(.x, .y)
) |>
  dplyr::select(predictor, model, term_pretty, term, beta, se, t, p, ci_low, ci_high,
sigma_resid, n)

coef_with_hema <- purrr::map2_dfr(
  split(coef_with_hema, coef_with_hema$predictor),
  names(split(coef_with_hema, coef_with_hema$predictor)),
  ~ pretty_term(.x, .y)
) |>
  dplyr::select(predictor, model, term_pretty, term, beta, se, t, p, ci_low, ci_high,
sigma_resid, n)

# 6) Write CSVs ------------------------------------------------------------
readr::write_csv(models_no_hema,   "copeptin_models_no_hematocrit.csv")
readr::write_csv(models_with_hema, "copeptin_models_with_hematocrit.csv")
readr::write_csv(coef_no_hema,     "copeptin_coef_table_no_hematocrit.csv")
readr::write_csv(coef_with_hema,   "copeptin_coef_table_with_hematocrit.csv")

cat(
  "Wrote:\n",
  " - copeptin_models_no_hematocrit.csv\n",
  " - copeptin_models_with_hematocrit.csv\n",
  " - copeptin_coef_table_no_hematocrit.csv\n",
  " - copeptin_coef_table_with_hematocrit.csv\n"
)
```

```
# 7) Key figure: copeptin vs HbA1c by diabetes type ---------------------------
fig2_dat <- df |>
  tidyr::drop_na(copeptin, hba1c, diabetes)

fig2 <- ggplot(fig2_dat,
               aes(x = hba1c, y = copeptin,
                   color = diabetes, shape = diabetes)) +
  geom_point(size = 2, alpha = 0.8) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Hemoglobin A1c (%)",
    y = "Copeptin (pmol/L)",
    color = "Diabetes type",
    shape = "Diabetes type",
    title = "Copeptin versus HbA1c by diabetes type"
  ) +
  theme_minimal(base_size = 12)

# Optional file export if you want a standalone PNG
ggsave("figure_copeptin_hba1c_by_diabetes.png", fig2,
       width = 6, height = 4.5, dpi = 300)

# 8) Word document ------------------------------------------------------------
as_flex <- function(x, caption = NULL) {
  ft <- flextable::flextable(x)
  ft <- flextable::autofit(ft)
```

```
 if (!is.null(caption)) ft <- flextable::add_header_lines(ft, values = caption)
 ft
}

doc <- officer::read_docx()
doc <- officer::body_add_par(doc, "Copeptin ~ X * Diabetes (+/- hematocrit)", style = "heading 1")

doc <- officer::body_add_par(doc, "Model tables (simple slopes by diabetes group)", style = "heading 2")
doc <- flextable::body_add_flextable(doc, as_flex(models_no_hema,   "No hematocrit"))
doc <- officer::body_add_par(doc, "")
doc <- flextable::body_add_flextable(doc, as_flex(models_with_hema, "With hematocrit"))

doc <- officer::body_add_par(doc, "Coefficient tables (betas, SEs, p, 95% CI)", style = "heading 2")
doc <- officer::body_add_par(doc, "No hematocrit", style = "heading 3")
doc <- flextable::body_add_flextable(doc, as_flex(coef_no_hema))
doc <- officer::body_add_par(doc, "")
doc <- officer::body_add_par(doc, "With hematocrit", style = "heading 3")
doc <- flextable::body_add_flextable(doc, as_flex(coef_with_hema))

# Add key figure into the Word document
doc <- officer::body_add_par(
 doc,
 "Figure 1. Copeptin versus HbA1c by diabetes type",
 style = "heading 2"
```

```
)
doc <- officer::body_add_gg(
  doc, fig2,
  width = 6, height = 4.5
)

out_doc <- "copeptin_interaction_results_with_figure.docx"
print(doc, target = out_doc)
cat("\nWord document written:", out_doc, "\n")
```