

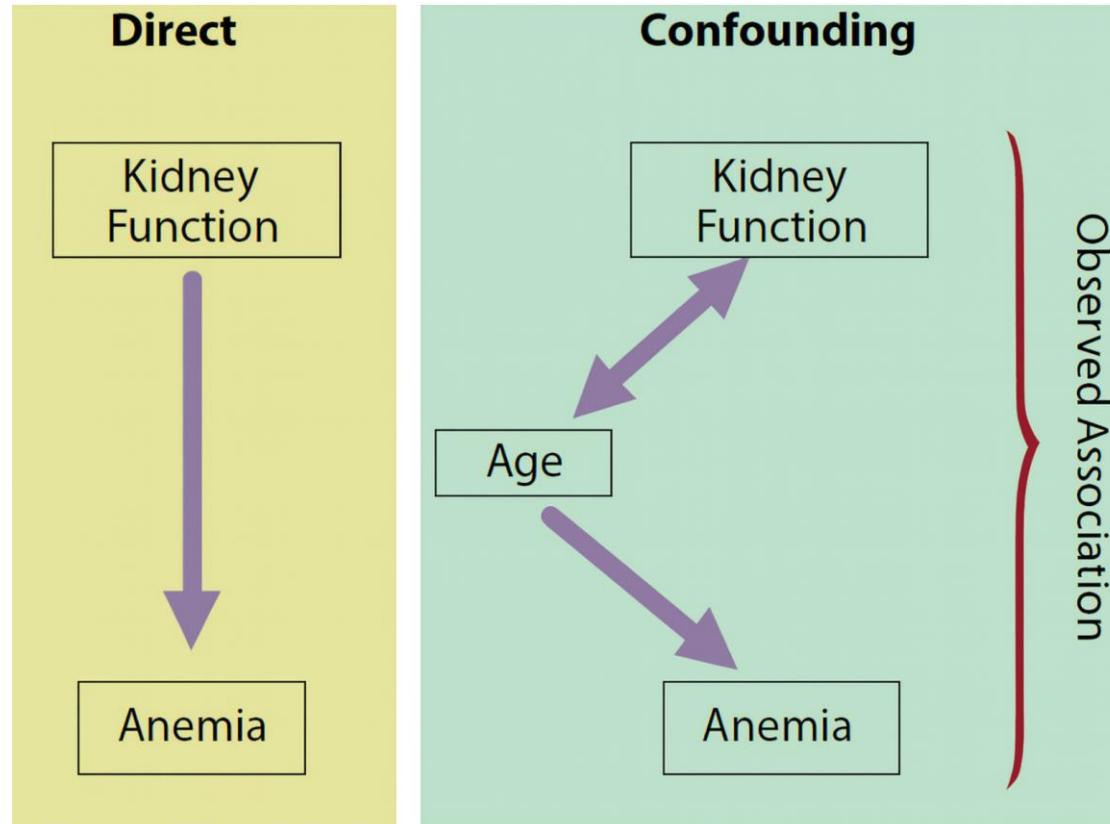


How kidney function and patient characteristics influence hemoglobin levels in CKD patients

Group 4: Ranxuan Li, Sinuo Lu, Yongxian Zhang

11/25/2025

Lead-in



Background & Research Questions

Chronic kidney disease is commonly accompanied by anemia and hemoglobin concentration is a clinically validated indicator of CKD-related anemia. Large epidemiologic studies have shown that declining kidney function is strongly associated with reduced hemoglobin levels, reflecting impaired erythropoietin production in CKD (Stauffer & Fan, 2014; Babitt et al., 2021). For this reason, hemoglobin can be used as a reliable proxy for anemia severity. Serum creatinine is a core biochemical marker of kidney dysfunction and is routinely used in clinical practice and research to define and stage CKD (Levey & Coresh, 2012). Thus creatinine is an appropriate surrogate for renal impairment in our analysis.

Previous studies report that anemia is more common and more severe in older adults with CKD, but aging itself is also associated with physiological renal decline, chronic inflammation, and a higher prevalence of comorbidities such as hypertension (Santos et al., 2023; Wacka et al., 2024). These conditions independently reduce hemoglobin levels, suggesting that the crude association between age and hemoglobin may be confounded by kidney function or hypertension rather than representing a direct effect of aging. Systematic reviews of CKD-related anemia further emphasize the multifactorial nature of hemoglobin reduction, highlighting age, CKD severity, inflammation, and comorbidities as key determinants (Prapaiwong et al., 2023).

In our dataset, older patients have lower hemoglobin levels and markedly higher rates of hypertension. This pattern raises the possibility that age affects both the exposure and outcome variables, fulfilling the criteria for a confounding factor. Understanding whether age has an independent effect on hemoglobin after adjusting for kidney function and whether hypertension modifies this association forms the central motivation of our study.

Research questions:

- Is kidney function significantly associated with hemoglobin?
- Does age have an independent effect after adjusting for kidney function?
- Does hypertension modify this relationship and cause Simpson's paradox?

Dataset & Cleaning

Dataset Description:

- Approximately 400 patient records with 24 demographic, laboratory, and clinical variables related to kidney functions and hematology.
- The dataset contains multiple hematological and biochemical indicators, including such as *Blood Glucose (Random)*, *Blood Urea*, *Serum Creatinine*, *Hemoglobin Level* and etc.
- Variables include continuous biomarkers such as creatinine and urea, and categorical indicators such as hypertension.
- The dataset is applicable for hemoglobin-CKD relation analysis due to clinically meaningful variables and adequate sample size.

Cleaning Steps:

- Trim hidden spaces, convert blanks to true NA
- Reformatting fake character variables to numeric variables or factors as in need
- Impute missing numeric values with medians; categorical with modes
- Final dataset includes age, urine tests, serum creatinine, hemoglobin, PCV, hypertension, diabetes, anemia, CKD classification

Numerical & Categorical Comparisons (EDA)

Variable	p-value	Significant
Specific Gravity (sg)	< 2.2e-16	Yes
Albumin (al)	< 2.2e-16	Yes
Blood Glucose Random (bgr)	7.66e-08	Yes
Blood Urea (bu)	3.96e-13	Yes
Serum Creatinine (sc)	9.73e-08	Yes
Hemoglobin (hemo)	< 2.2e-16	Yes
Packed Cell Volume (pcv)	< 2.2e-16	Yes

Table 1. Wilcoxon Rank-Sum Test Results for Numerical Variables

The numerical variables were non-normally distributed; therefore, Wilcoxon rank-sum tests were used to compare CKD and non-CKD groups.

All kidney-related markers and anemia markers showed statistically significant differences between the two groups.



Variable	Category	Non-CKD	CKD
Hypertension	No	43.92	78.08
	Yes	28.08	49.92
Diabetes	No	46.8	83.2
	Yes	25.2	44.8
Anemia	No	60.48	107.52
	Yes	11.52	20.48

Table 2. Chi-Square Test Results for Categorical Variables

Hypertension, diabetes, and anemia show statistically significant differences between CKD and non-CKD groups based on Pearson's Chi-square tests.

Hypertension: $p < 2.2e-16$

Diabetes: $p = 7.1e-15$

Anemia: $p = 3.7e-06$

These findings indicate that CKD patients have a substantially higher prevalence of these conditions.

Baseline Characteristics

- CKD patients differ substantially from non-CKD individuals across most demographic, clinical, and laboratory measures.
- Kidney function markers including serum creatinine, blood urea, hemoglobin, and packed cell volume show large and statistically significant differences.
- CKD patients have markedly higher rates of hypertension, diabetes, proteinuria, glycosuria, urinary abnormalities, anemia, and poor appetite.
- These patterns collectively demonstrate the classical clinical and biochemical features of renal impairment.

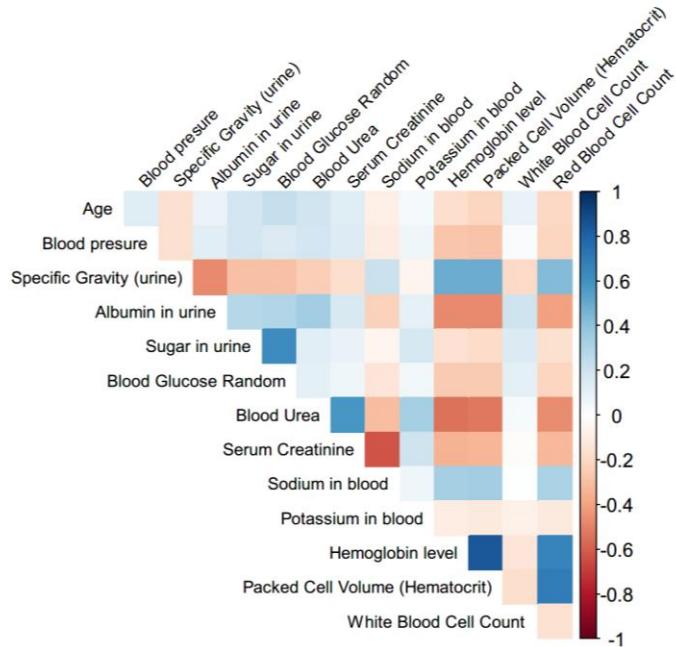
Variable	Overall N = 400¹	ckd N = 250¹	notckd N = 150¹	p- value²
age	51 (17)	55 (17)	47 (16)	<0.001
sg				
1.005	7 (2.0%)	7 (3.4%)	0 (0%)	
1.01	84 (24%)	84 (40%)	0 (0%)	
1.015	75 (21%)	75 (36%)	0 (0%)	
1.02	106 (30%)	31 (15%)	75 (52%)	
1.025	81 (23%)	11 (5.3%)	70 (48%)	
al				
0	199 (56%)	54 (26%)	145 (100%)	
1	44 (12%)	44 (21%)	0 (0%)	
2	43 (12%)	43 (21%)	0 (0%)	
3	43 (12%)	43 (21%)	0 (0%)	
4	24 (6.8%)	24 (11%)	0 (0%)	
5	1 (0.3%)	1 (0.5%)	0 (0%)	
bgr	148 (79)	175 (92)	108 (19)	<0.001
bu	57 (51)	72 (59)	33 (11)	<0.001
sc	3.1 (5.7)	4.4 (7.0)	0.9 (0.3)	<0.001
hemo	12.53 (2.91)	10.65 (2.19)	15.19 (1.28)	<0.001
pcv	39 (9)	33 (7)	46 (4)	<0.001
htn	147 (37%)	147 (59%)	0 (0%)	<0.001
dm	137 (34%)	137 (55%)	0 (0%)	<0.001
ane	60 (15%)	60 (24%)	0 (0%)	<0.001

¹Mean (SD); n (%)

²Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

Correlation Structure

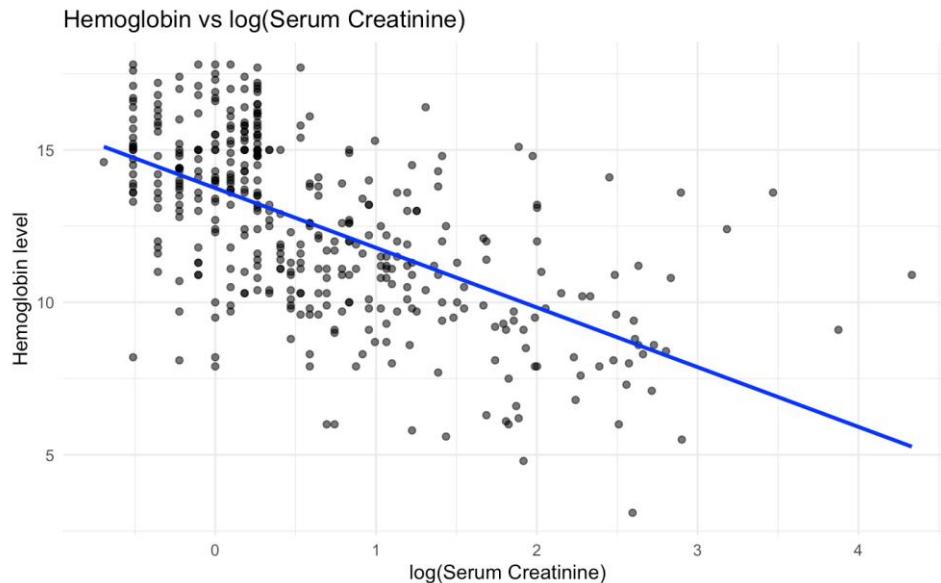
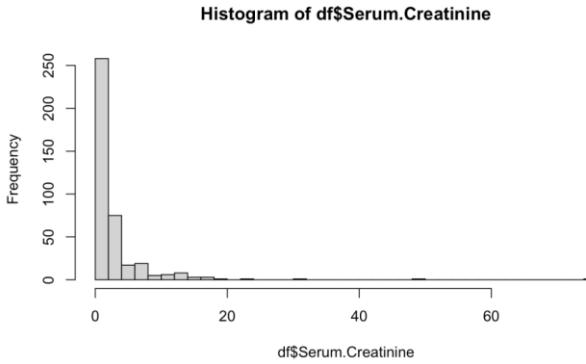
- The correlation heatmap shows clinically meaningful clusters among biochemical and clinical variables.
- Serum creatinine and blood urea are strongly correlated, reflecting shared impairment of renal filtration.
- Hemoglobin and packed cell volume also correlate closely, forming an anemia-related cluster.
- Blood glucose and urine sugar show moderate correlation, indicating a metabolic cluster related to diabetes.
- Electrolytes and white blood cell count exhibit weak correlations and contribute limited discriminatory value.
- Overall, the heatmap highlights three key domains: renal dysfunction, anemia, and diabetes.



Relationship between kidney function and hemoglobin

Q1: Is kidney function associated with hemoglobin level in CKD patients?

$$\text{Hemoglobin} = \beta_0 + \beta_1 \cdot \text{Serum Creatinine}$$



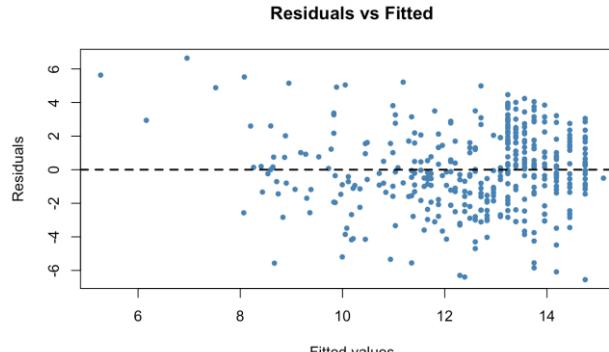
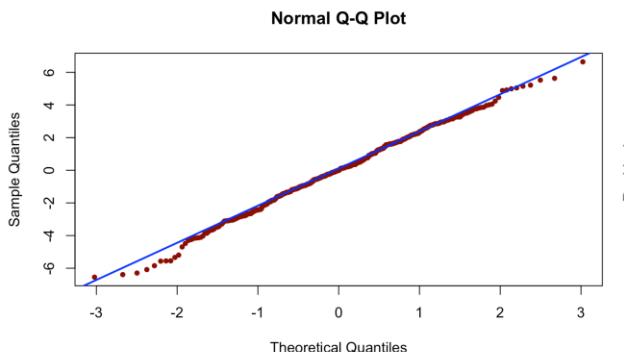
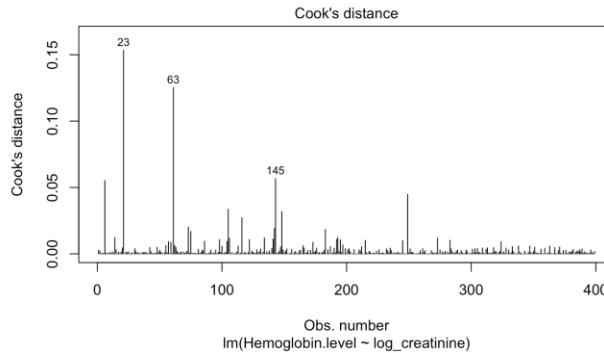
In CKD, Creatinine levels are inherently severely right-biased, I use the log approaches:
Taking the logarithm of Creatinine



Weill Cornell Medicine

Relationship between kidney function and hemoglobin

Q1: Is kidney function associated with hemoglobin level in CKD patients?



CKD patients show:

- Higher serum creatinine & urea
- Higher random glucose
- Lower hemoglobin & PCV

These patterns reflect classical CKD pathology: impaired filtration, waste accumulation, and reduced erythropoietin production.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.7482	0.1432	95.97	<2e-16 ***
log_creatinine	-1.9576	0.1329	-14.73	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

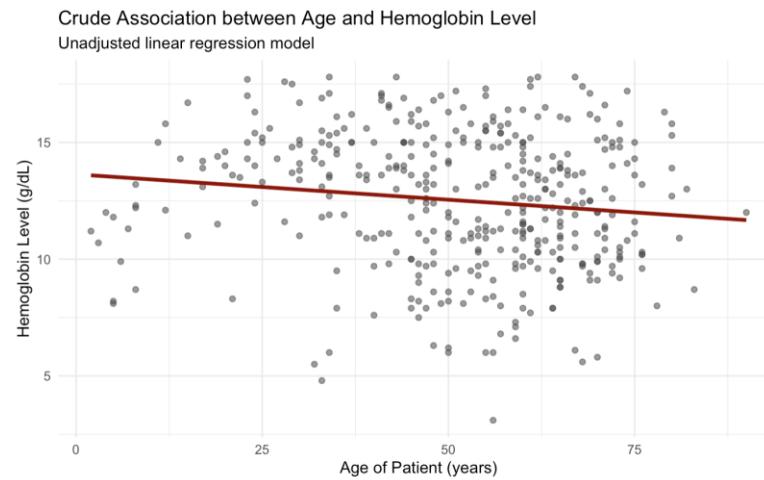
Residual standard error: 2.324 on 397 degrees of freedom
Multiple R-squared: 0.3533, Adjusted R-squared: 0.3517
F-statistic: 216.9 on 1 and 397 DF, p-value: < 2.2e-16

(Intercept) log_creatinine
13.748206 -1.957625

Crude association between age and hemoglobin

Q2: Is age associated with hemoglobin after accounting for kidney disease severity?

Crude model with only age



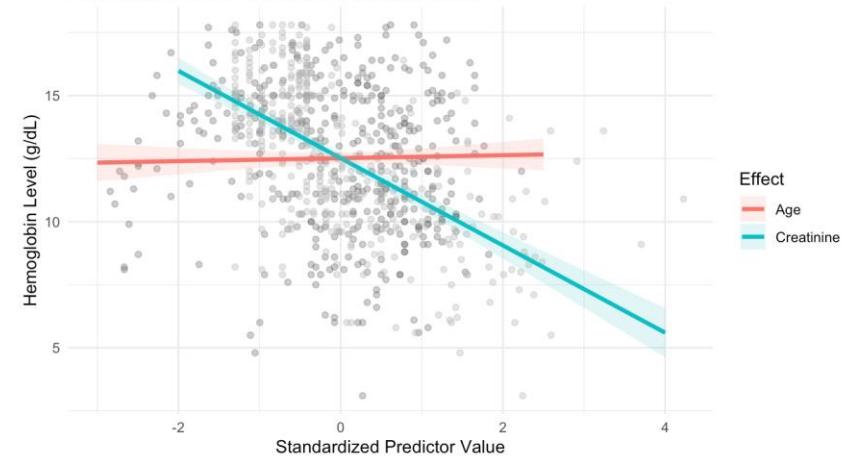
slope: -0.02178

p-value: 0.008766

Weill Cornell Medicine

Age effect after adjusting for renal function

Adjusted Effects of Age and Creatinine on Hemoglobin
Partial regression lines after standardizing predictors



Age:

slope: 0.003381
p-value: 0.626

log-serum

slope: -1.974547
p-value: <2e-16

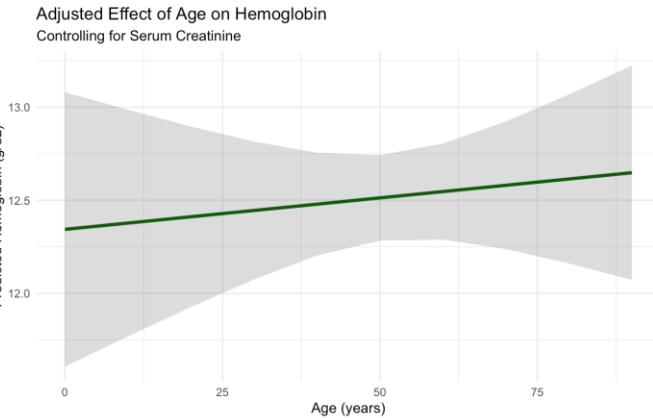
Adjusted Effect of Age on Hemoglobin

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.585324	0.363748	37.348	<2e-16 ***
log_creatinine	-1.974547	0.137518	-14.358	<2e-16 ***
Age.of.the.patient	0.003381	0.006940	0.487	0.626

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error: 2.326 on 396 degrees of freedom
Multiple R-squared: 0.3537, Adjusted R-squared: 0.3504
F-statistic: 108.3 on 2 and 396 DF, p-value: < 2.2e-16



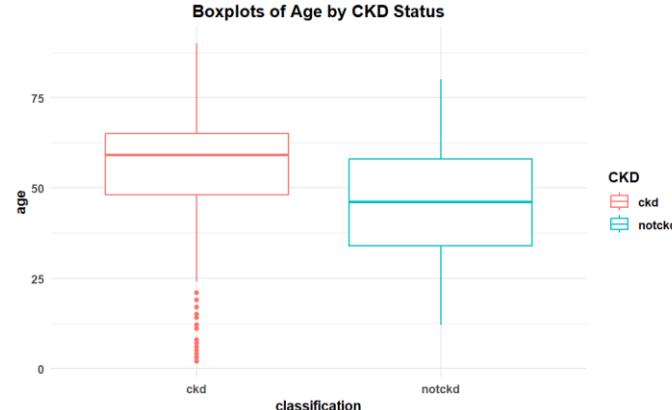
Statistical Results:

- Adjusting for log creatinine, the effect of age on hemoglobin becomes non-significant.
- Creatinine remains strongly associated with hemoglobin, confirming kidney dysfunction as the dominant predictor.
- The age coefficient is essentially zero, and the model's explanatory power is driven almost entirely by kidney function.

Clinical Interpretation:

- The adjusted regression line is nearly flat, and the predicted change in hemoglobin across age is clinically trivial.
- The wide confidence intervals at very young and very old ages indicate unstable estimation.
- The negative age effect observed in the crude model was explained by renal dysfunction rather than age.

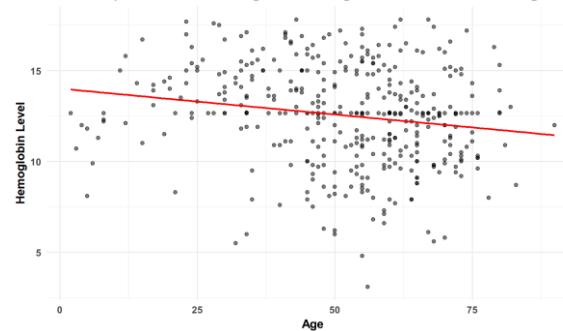
Age as a Confounder



Wilcoxon rank sum test with continuity correction

```
data: age by classification  
W = 24941, p-value = 3.17e-08  
alternative hypothesis: true location shift is not equal to 0
```

Relationship Between Patient Age and Hemoglobin Level with Linear Regression



Call:
lm(formula = age ~ sc, data = data)

Residuals:

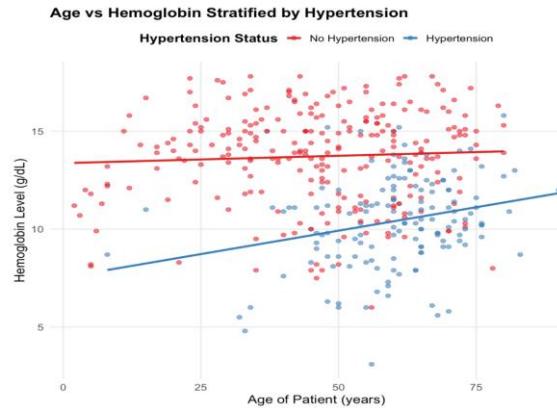
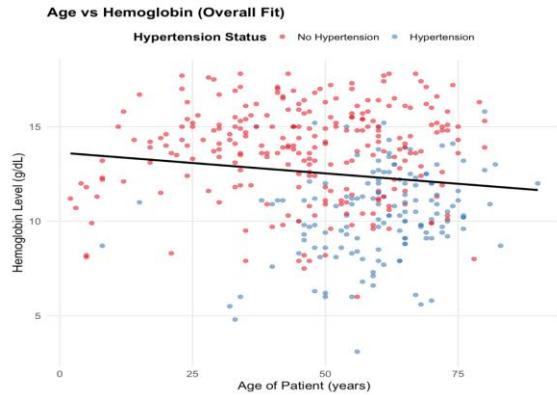
Min	1Q	Median	3Q	Max
-48.876	-8.856	2.733	12.012	38.436

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	50.3509	0.9548	52.734	< 2e-16 ***
sc	0.4042	0.1499	2.697	0.00729 **

- Hemoglobin tends to be slightly lower at older ages, indicating a mild negative association between age and hemoglobin.
- CKD patients are substantially older than non-CKD patients, as seen in most chronic conditions.
- Age is related to both hemoglobin levels and CKD prevalence but neither a cause or result to both variables and therefore acts as a confounding factor in evaluating the associations between kidney function and hemoglobin level.

Simpson's Paradox



We fitted linear regression models using:

$$\text{Hemoglobin} = \beta_0 + \beta_1 \cdot \text{Age}$$

- Model A: Overall (ignoring hypertension)
- Models B1/B2: Stratified by hypertension status

3. Direction Reversal Summary

Model	Slope of Age	Direction
Overall	-0.02197	Negative
Hypertension	+0.04785	Positive
No Hypertension	+0.00752	Positive

- In the crude analysis, older age appears associated with lower hemoglobin.
- After stratifying by hypertension status, both hypertensive and non-hypertensive groups show a positive association.
- This reversal demonstrates Simpson's paradox, driven by the higher prevalence of hypertension among older adults and its independent association with lower hemoglobin.
- Hypertension therefore modifies the age–hemoglobin relationship and must be considered in evaluating age-related anemia in CKD.

Conclusions & Limitations

Conclusions:

- Kidney function is the primary determinant of hemoglobin.
- Age has minimal independent effect after adjusting for kidney function and hypertension.
- Hypertension introduces Simpson's paradox; stratification is essential.

Limitations:

- Small sample sizes at extreme ages produced wide confidence intervals and unstable adjusted estimates.
- Nonlinear associations between age, kidney function, and hemoglobin may exist but were not modeled.
- Important clinical determinants of anemia such as inflammation markers, iron indices, and erythropoietin levels were not available, leaving residual confounding.
- The dataset originated from a single clinical source, which limits generalizability of findings to broader CKD populations.

Reference

- [1] Levey AS, Coresh J. Chronic kidney disease. *Lancet*. 2012;379(9811):165–180. doi:10.1016/S0140-6736(11)60178-5
- [2] Stauffer ME, Fan T. Prevalence of anemia in chronic kidney disease in the United States. *PLoS ONE*. 2014;9(1):e84943. doi:10.1371/journal.pone.0084943
- [3] Babitt JL, Eisenga MF, Haase VH. Anemia in chronic kidney disease: From pathophysiology and current treatments to future agents. *Front Med*. 2021;8:642296. doi:10.3389/fmed.2021.642296
- [4] Santos S, et al. Anemia in elderly patients: Contribution of renal aging and chronic kidney disease. *Geriatrics*. 2023;8(2):43. doi:10.3390/geriatrics9020043
- [5] Prapaiwong P, et al. Determinants of anemia among patients with chronic kidney disease: A systematic review. *J Multidiscip Healthc*. 2023;16:1441–1456. doi:10.2147/JMDH.S535423
- [6] Wacka E, Nicikowski J, Jarmuzek P, Zembron-Lacny A. Anemia and its connections to inflammation in older adults: A review. *J Clin Med*. 2024;13(7):2049. doi:10.3390/jcm13072049

Thank you



Weill Cornell Medicine