Dear Author

Please use this PDF proof to check the layout of your article. If you would like any changes to be made to the layout, you can leave instructions in the online proofing interface. First, return to the online proofing interface by clicking "Edit" at the top page, then insert a Comment in the relevant location. Making your changes directly in the online proofing interface is the quickest, easiest way to correct and submit your proof.

Please note that changes made to the article in the online proofing interface will be added to the article before publication, but are not reflected in this PDF proof.

If you would prefer to submit your corrections by annotating the PDF proof, please download and submit an annotatable PDF proof by clicking the link below.

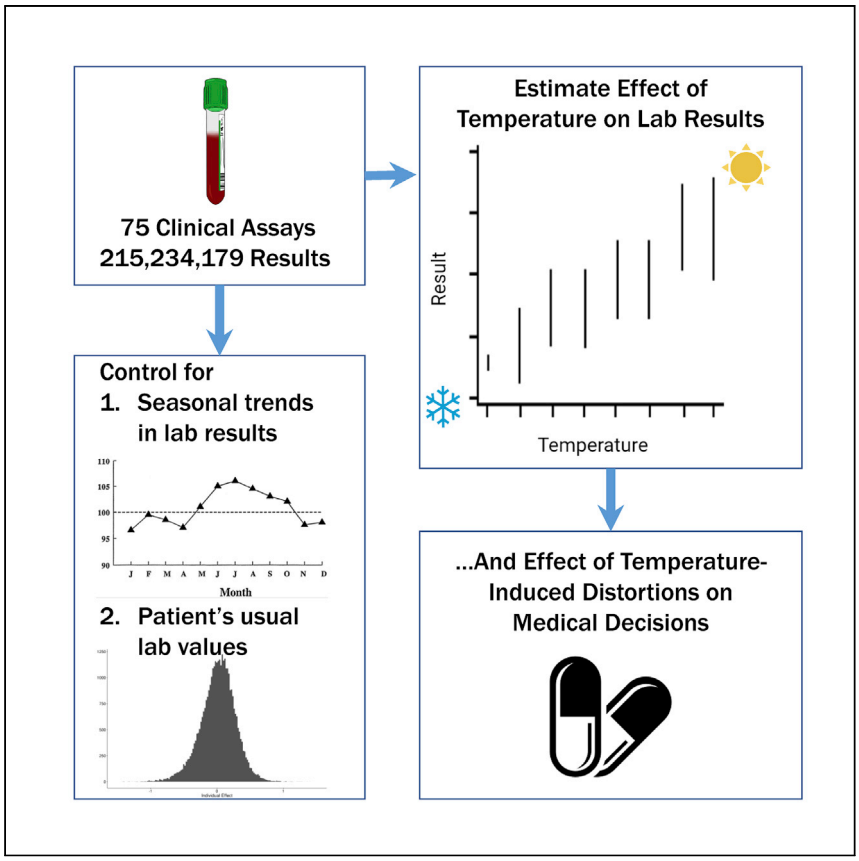
Annotate PDF

Med

CellPress

Clinical and Translational Report

Variation in common laboratory test results Because of ambient temperature



Obermeyer and Pope show that ambient temperature on the day an individual's blood is drawn can affect laboratory test results. Analyzing 215,234,179 results, they find that many core clinical assays are affected. Medical decision-making changes as a result; doctors prescribe fewer statins to individuals whose lipids are checked on colder days.

Ziad Obermeyer, Devin Pope

zobermeyer@berkeley.edu

Highlights

Analysis of a dataset of nearly 5 million and over 200 million laboratory test results

Ambient temperature on the day an individual's blood was drawn affected the test result

Affected assays included renal function tests, complete blood count, and lipid panels

Doctors prescribed fewer statins to individuals whose lipids were checked on colder days.

Translation to Population Health

Obermeyer & Pope, Med 2, 1–13
December 10, 2021
https://doi.org/10.1016/j.medj.2021.11.003

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

Med

CellPress

Clinical and Translational Report

Variation in common laboratory test results Because of ambient temperature

Ziad Obermeyer[1,3,4,*] and Devin Pope[2,3]

SUMMARY

Background: Laboratory tests measure important aspects of physiology, but their results also vary for idiosyncratic reasons. We explore an underappreciated source of variation: ambient temperature on the day blood is drawn.

Methods: In a sample of 4,877,039 individuals between 2009–2015, we model 215,234,179 test results as a function of temperature, controlling for individual and city-week fixed effects. This measures how day-to-day temperature fluctuations affect results over and above the individual's mean values, and seasonal variation.

Findings: 51 of 75 assays are significantly affected by temperature, including measures of kidney function (increased creatinine, urea nitrogen, and urine specific gravity), cellular blood components (decreased neutrophils, erythrocytes, and platelets), and lipids (increased high-density lipoprotein [HDL] and decreased total cholesterol, triglycerides, and low-density lipoprotein [LDL]). These small, day-to-day fluctuations are unlikely to correlate with long-term physiological trends; for example, lipid panels checked on cooler days look lower risk, but these short-term changes probably do not reflect stable changes in cardiovascular risk. Nonetheless, doctors appear to treat these individuals differently. We observe 9.7% fewer statin prescriptions for individuals checked on the coolest versus the warmest days (–0.42% versus baseline of 4.34%, $p < 0.001$).

Conclusions: Ambient temperature affects the results of many laboratory tests. These distortions, in turn, affect medical decision-making. Statistical adjustment in reporting is feasible and could limit undesired temperature-driven variability.

Funding: None.

Context and significance

Doctors use laboratory tests to measure many important aspects of physiology. But test results can also vary for arbitrary reasons, like ambient temperature on the day of the test. In a United States-based sample of nearly 5 million people and 215 million test results, day-to-day temperature fluctuations affected the results of some of the most commonly used laboratory tests in medicine, including lipids and red and white blood cells. Although the changes were small, they did influence doctors' medical decisions; individuals whose lipid panels were checked on colder days appeared to be at lower risk for cardiovascular disease, leading to a 10% lower likelihood of being prescribed a statin. Understanding these changes means that laboratories could correct some results based on ambient temperature.

INTRODUCTION

Every year, 13 billion laboratory tests are performed in the United States,[1] nearly 800 million in the United Kingdom,[2] and many more worldwide. Test results provide critical data on clinically important changes in patient physiology—including acute variations in plasma volume, body temperature, circadian rhythms, etc.[3–6]—but can also vary for more idiosyncratic reasons from one draw to the next, like differences in technique or sample processing.[7–12]

Here we explore an underappreciated source of variation in laboratory tests: the ambient temperature on the day blood is drawn. A growing body of literature in social science uses short-term temperature fluctuations as a "natural experiment" to study the effect of temperature on important outcomes.[13,14] We adapt these techniques to our setting, where we suggest that the particular day a test was done—and, specifically, whether that day was hotter or colder than usual for that particular

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

week in that particular city—is as good as random. This mimics random assignment to hotter or colder temperature and allows estimation of causal effects, not just correlations.

Until recently, the large-scale laboratory data needed to deploy these techniques were scarce, meaning that this issue has been explored rarely and only for specific tests in small, homogeneous samples.[15,16] Here we use a large dataset of test results spanning several climate zones to detect these changes. We then explore their implications by linking them to doctors' treatment decisions that depend on test results.

RESULTS

4,877,039 individuals have laboratory result data. Daily temperature ranges from a low of −28°C (Fairbanks, Alaska; February 16, 2011) to 49°C (Yuma, Arizona; July 23, 2014). Table 1 lists, for the 75 tests we studied, sample size, mean values, and the effect of temperature (Table S1 provides additional results for all 75 tests). Overall, 51 of 75 assays are significantly affected by temperature, accounting for 196,635,337 of 215,234,253 results (91.4%).

Does temperature affect laboratory test results?

We begin by considering several measures of kidney function: creatinine, blood urea nitrogen, and urine specific gravity. These tests may be particularly susceptible to temperature-related changes, given the known links between temperature and plasma volume (see further discussion of these mechanisms below; additional results for the 75 most common tests in our sample can be found in Data S1). Figure 1 shows that higher temperature produces increases in creatinine (linear coefficient: 0.0002 mg/dL/°C, $p < 0.001$; urine creatinine: 0.2640 mg/dL/°C, $p < 0.001$; Figure S24; Table S3), blood urea nitrogen (BUN: 0.0048 mg/dL/°C, $p < 0.001$), and urine specific gravity (0.00002 u/°C, $p < 0.001$). Distributed lags (Figures S1–S75, panel 4) show that temperatures on the day before and the day of the test have the largest influence. As a sensitivity analysis, we create an "apparent temperature" measure incorporating humidity and wind speed.[22] Results are substantively unchanged (Figures S1–S75, panel 6). As a falsification test, we also show that vitamin D has clear seasonal trends, as expected because of sun exposure, but no relationship to temperature (Figure S75). Overall, the size of these fluctuations is small: a 1 SD change in temperature produced a less than 1% change in all assays in this sample (Table 1, column 7).

It is plausible that temperature-induced variation in these tests reflects real physiological changes. Indeed, these tests are often used clinically to measure exactly such short-term changes in plasma volume and kidney function. In contrast, other laboratory tests in our sample are used clinically to measure longer-term physiological trends unrelated to the particular temperature on the day of measurement. For example, lipid testing is meant to capture long-term cardiovascular risk to target lipid-lowering therapy. Thus, it would be surprising—and perhaps disconcerting—to discover temperature-induced changes in these tests.

Nonetheless, higher temperature makes cardiovascular risk appear lower based on the results of lipid panels (Figure 2): increases in HDL (0.0035 mg/dL/°C, $p < 0.001$) and decreases in total cholesterol (−0.0613 mg/dL/°C, $p < 0.001$) and triglycerides (−0.0680 mg/dL/°C, $p < 0.001$). Low-density lipoprotein (LDL), measured by the usual Friedewald method[23] (total cholesterol less triglycerides and high-density lipoprotein [HDL]), decreases (−0.0502 mg/dL/°C, $p < 0.001$), as expected, given the aforementioned changes in cholesterol and HDL. In contrast, the less frequently

[1]University of California, Berkeley, School of Public Health, 2121 Berkeley Way, Berkeley, CA 94704, USA

[2]University of Chicago Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, USA

[3]These authors contributed equally

[4]Lead contact

*Correspondence: zobermeyer@berkeley.edu

https://doi.org/10.1016/j.medj.2021.11.003

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

CellPress

Table 1. Summary statistics and temperature effects

	Unique individuals (1)	Mean age (2)	Female (%) (3)	Mean result (4)	Result units (5)	Reference range (6)	Change (%) 1 SD temp ↑ (7)	p Value (8)	CV_{temp}/ CV_{max} (%) (9)
Basic metabolic panel									
Sodium	3,067,503	47.1	56.2	139.6	mmol/L	135–145	0.001	0.7626	0.2
Potassium	3,169,956	47.3	56.3	4.3	mmol/L	3.5–5.0	−0.12	<0.0001	5.0
Chloride	2,719,013	47.0	56.3	103.6	mmol/L	96–106	0.09	<0.0001	14.6
Carbon dioxide	2,576,957	47.1	56.3	24.8	mmol/L	22–29	−0.44	<0.0001	21.8
Urea nitrogen	3,170,312	47.3	56.2	14.8	mg/dL	7–20	0.32	<0.0001	5.4
Creatinine	3,254,730	47.6	56.8	0.9	mg/dL	0.6–1.2	0.23	<0.0001	7.7
Glucose	2,909,001	47.1	56.8	98.6	mg/dL	70–110	0.05	0.0498	2.0
Complete blood count									
Leukocytes	2,727,716	44.1	60.2	6.78	k/μL	4.5–11.0	−0.32	<0.0001	5.7
Erythrocytes	2,659,982	44.0	61.4	4.51	m/μL	4.7–6.1 M, 4.2–5.4 F	−0.29	<0.0001	18.0
Lymphocytes	2,415,443	44.2	60.4	1.98	k/μL	1.0–4.8	0.15	<0.0001	2.9
Neutrophils	2,421,142	44.1	60.2	4.05	k/μL	1.5–8.0	−0.55	<0.0001	6.4
Monocytes	2,426,381	44.0	60.6	0.47	k/μL	0.2–1.0	−0.32	<0.0001	3.6
Basophils	1,752,303	44.0	59.6	0.03	k/μL	0–0.19	−0.54	<0.0001	3.8
Eosinophils	2,349,009	44.1	60.2	0.16	k/μL	0.03–0.35	−0.24	0.0008	2.3
Hematocrit	3,062,695	43.9	62.0	40.2	%	40–54 M, 37–47 F	−0.21	<0.0001	15.6
Hemoglobin	2,901,558	43.9	62.1	13.5	g/dL	14–18 M, 12–16 F	−0.38	<0.0001	26.9
Platelets	2,650,676	44.1	60.1	242.3	k/μL	150–400	−0.10	0.0001	2.2
EDW	2,596,008	44.0	60.0	14.0	%	11.9–15.5	0.35	<0.0001	19.4
MCV	2,615,807	43.8	60.4	89.7	fL	80–96	0.08	<0.0001	11.3
Liver function tests									
ALT	2,954,560	47.0	56.0	23.4	u/L	7–56	−0.54	<0.0001	7.3
AST	2,887,995	46.9	55.9	21.8	u/L	10–40	−0.19	<0.0001	3.1
Alkaline phosphatase	2,724,156	47.4	56.2	72.3	u/L	20–140	−0.45	<0.0001	13.8
Bilirubin, total	2,869,760	46.9	56.0	0.6	mg/dL	0.2–1.3	0.36	<0.0001	3.3
Albumin	2,856,008	47.1	56.4	4.4	g/dL	3.5–5.0	−0.32	<0.0001	20.0
Globulin	2,483,073	46.7	56.0	2.7	g/dL	2.0–3.5	−0.29	<0.0001	10.4
Prothrombin time	216,671	50.5	51.0	17.0	s	9–12	−0.55	0.0003	27.4
Lipid panel									
Cholesterol, total	3,078,321	47.9	53.3	185.8	mg/dL	< 200	−0.33	<0.0001	11.1
HDL	2,978,509	48.2	52.7	52.1	mg/dL	> 35	0.07	0.0009	1.7
LDL	2,832,248	48.2	53.6	106.5	mg/dL	< 100	−0.47	<0.0001	11.8
Triglycerides	2,940,361	48.3	53.6	129.0	mg/dL	40–150	−0.53	<0.0001	5.3
Other									
Urine specific gravity	872,433	44.7	59.7	1.02	SI u	1.01–1.03	0.03	<0.0001	n/a
Creatinine in urine	629,030	50.1	49.0	139.7	mg/dL	40–300 M, 37–250 F	0.32	<0.0001	10.4
Vitamin D	812,795	47.5	69.2	29.9	ng/mL	20–100	0.09	0.4818	0.6

Columns 1–6 show summary statistics, including normal ranges for males (M) and females (F). Column 7 shows the effect of a 1 SD (9.97°C) increase in temperature on the test result, in percent relative to its mean, and column 8 shows the p value of the linear effect on which this effect is based. Column 9 shows the fraction of maximum tolerated imprecision (measured by coefficient of variation [CV_{max}]) a 1 SD change in temperature would account for. Additional data on all 75 tests are available in Table S1.

measured LDL by direct assay is unaffected (Figure S48). Interestingly, other liver products were also affected (Figures S1–S75). Albumin (−0.0014 g/dL/°C, $p < 0.001$) and globulin (−0.0008 g/dL/°C, $p < 0.001$) also decreased, as did ALT, AST, alkaline phosphatase, and PT (−0.0127 u/L/°C, −0.0042 u/L/°C, −0.0323 u/L/°C, and −0.0093 s/°C; all $p < 0.001$). Bilirubin, in contrast, increases (0.0002 mg/dL/°C, $p < 0.001$). Distributed lag models (Figures S1–S75, panel 4) indicated that changes are most correlated with temperatures in the 1–3 days before tests.

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

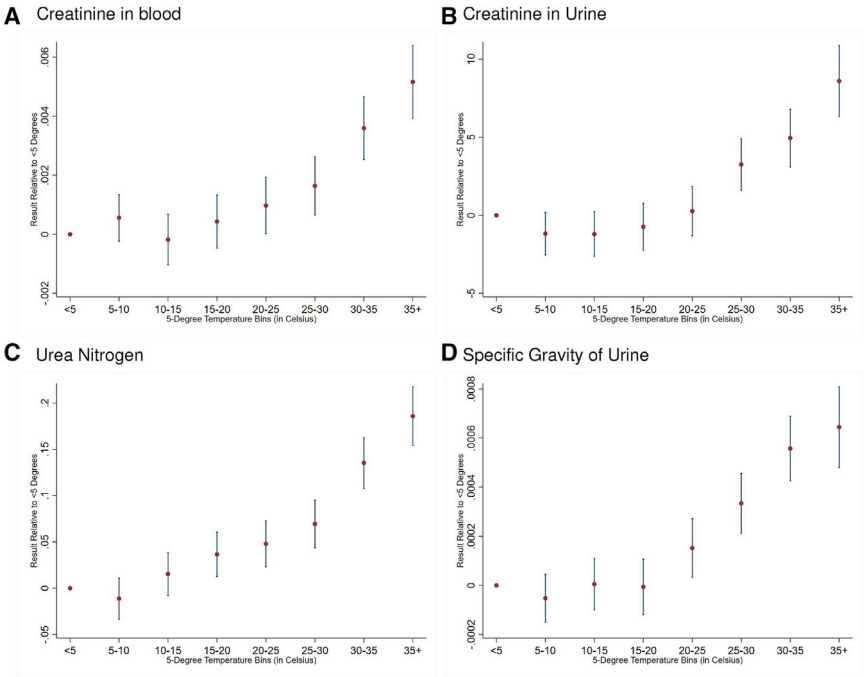


Figure 1. Temperature effects on measures of kidney function

(A–D) Estimates and 95% confidence intervals (CIs) for the effect of temperature on laboratory studies related to kidney function. The effect of temperature on the day of the test (e.g., temperature falling into the range of 15°C–20°C) is shown relative to the coldest days (<5°C, the omitted category, shown without 95% CI). Effects on the y axis are temperature-induced changes in units of the laboratory result (blood and urine creatinine and urea nitrogen, mg/dL; specific gravity of urine, SI units).

Higher temperature also affects complete blood count (Figure 3). Erythrocytes decrease (−0.0013 million cells/μL/°C, $p < 0.001$), as do hematocrit and hemoglobin (−0.0085%/°C and −0.0052 g/dL/°C, both $p < 0.001$). Leukocytes also decrease (−0.0022 thousand cells/μL/°C, $p < 0.001$), driven by neutrophils (−2.2334 cells/μL/°C, $p < 0.001$), whereas lymphocytes, monocytes, and eosinophils are unaffected (Figures S25, S51, and S53). Platelet counts decreased (−0.0244 thousand cells/μL/°C, $p < 0.001$), as does platelet volume (−0.0055 fL/°C, $p < 0.001$).

When using laboratory tests to measure stable physiological states, temperature-induced changes present a dilemma. Temperature causes results to vary from one draw to the next, but the underlying quantities of interest are presumably more stable and unlikely to be affected by day-to-day temperature fluctuations. So of the many possible test result values at many possible temperatures, which one is "true"? Although we could choose an arbitrary "reference temperature," the reality is that we do not know which temperature best correlates with the true underlying quantity of interest (e.g., at which temperature does LDL best correlate with 10-year cardiovascular risk?). This question could, in principle, be answered using data from long-standing cohort studies of cardiovascular risk (e.g., Framingham); provided there is temperature variation on the day labs were drawn. Using methods similar to ours, long-term risk could be more or less correlated with results from hotter or cooler days.

Do temperature-induced changes matter for laboratory performance?

Temperature-induced changes appear to violate a key principle of laboratory performance: precision. The same sample tested on the same equipment should have the

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

CellPress

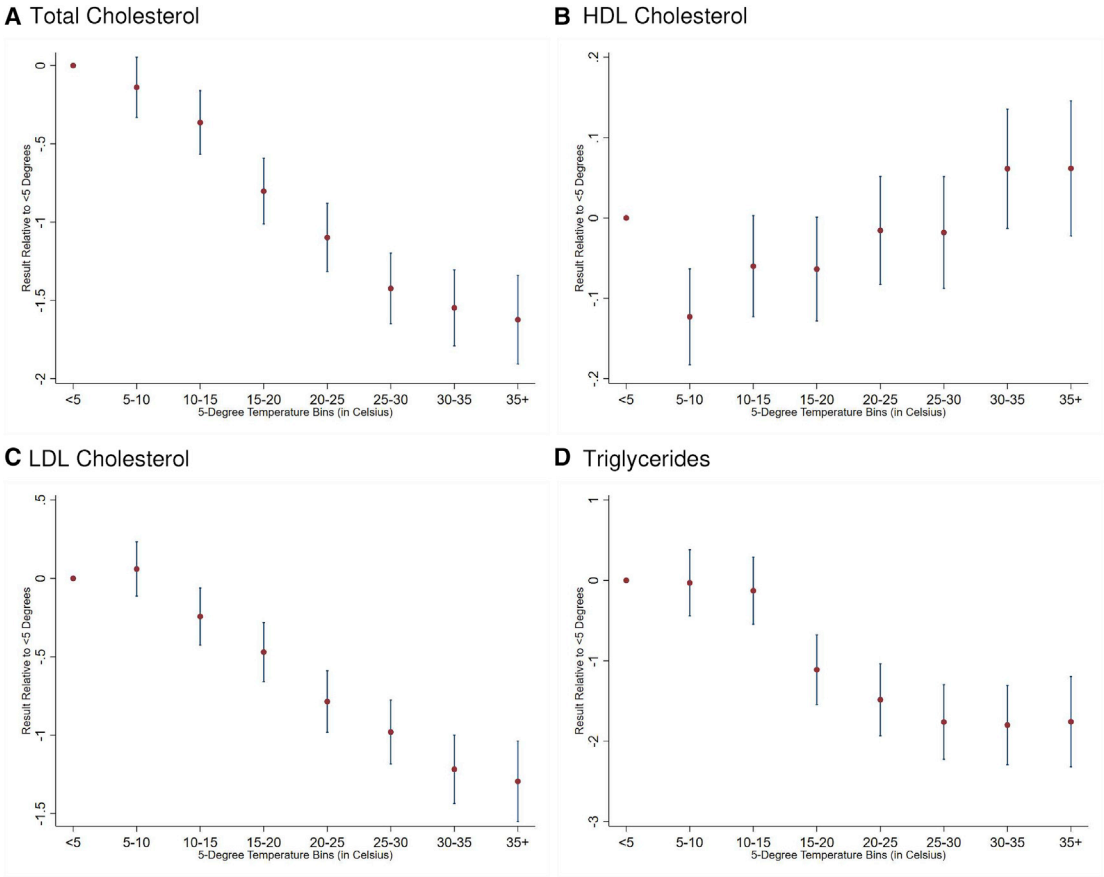


Figure 2. Temperature effects on measures of lipids and liver function

(A–D) Estimates and 95% CIs for the effect of temperature on laboratory studies related to lipids and other liver products. The effect of temperature on the day of the test (e.g., temperature falling into the range of 15°C–20°C) is shown relative to the coldest days (<5°C, the omitted category, shown without 95% CI). Effects on the y axis are temperature-induced changes in units of the laboratory result (mg/dL).

same result when measured repeatedly. But because temperature varies arbitrarily, the same individual tested with the same equipment would not have the same result when measured repeatedly on days with different temperatures.

From this point of view, temperature-induced distortions are analogous to imprecision. So we can quantify their magnitude by analogy to the metric laboratories use to measure imprecision more generally—the coefficient of variation (CV): dispersion in results, measured by their SD, divided by the mean. We estimate an analogous measure of imprecision generated by temperature (CV_{temp}): dispersion in results produced by a 1 SD temperature change, divided by the mean. We can then compare CV_{temp} with the maximum imprecision, CV_{max}, allowed for individual assays: the upper bound on the acceptable dispersion when testing and retesting similar samples.[23]

We use LDL as an illustrative example because it is clinically important and used widely to measure laboratory performance.[24] A 1 SD change in temperature (9.97°C, calculated within metropolitan areas) generates a CV_{temp} of 0.47% in LDL. This gives one estimate of the improvement in precision laboratories could attain by statistically adjusting for temperature. We view this method as giving a very rough sense of the order of magnitude of potential gains; it is far from a precise estimate. But it nonetheless lets us ask

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

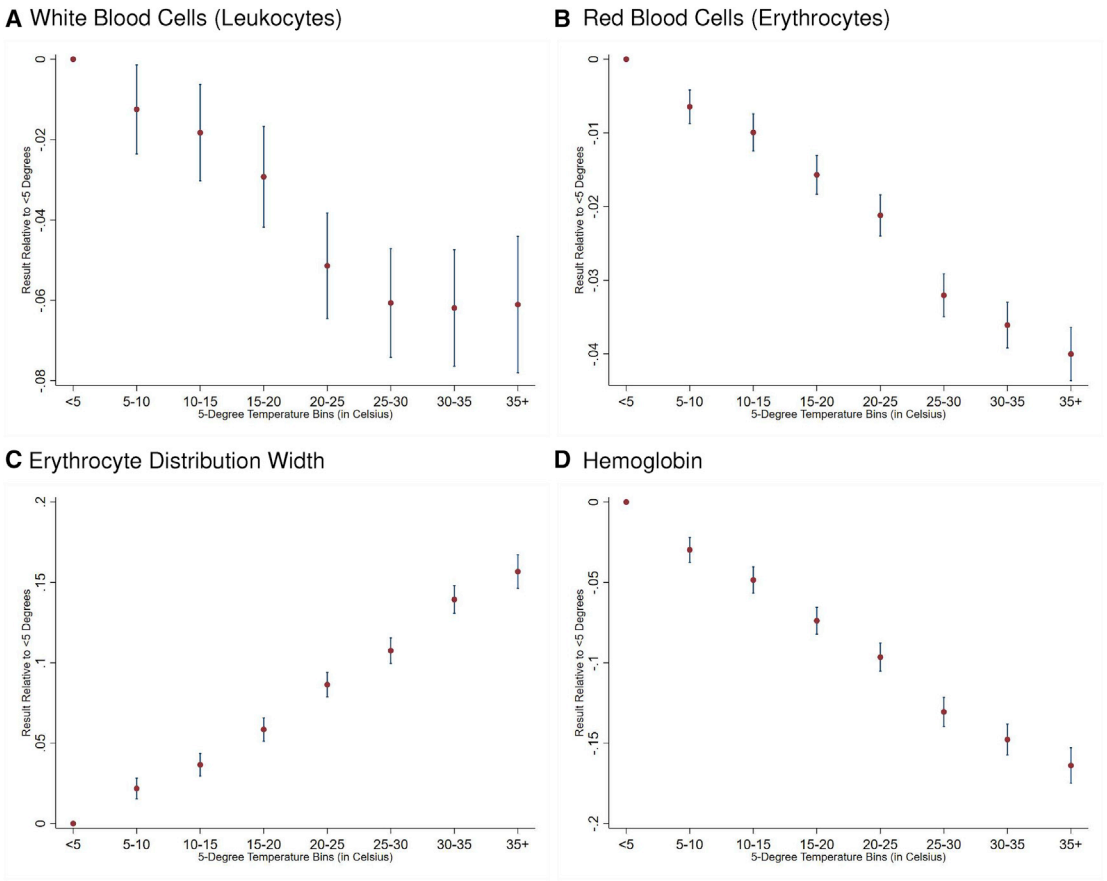


Figure 3. Temperature effects on measures of cellular components of the blood

(A–D) Estimates and 95% CIs for the effect of temperature on laboratory studies related to cellular components of the blood. The effect of temperature on the day of the test (e.g., temperature falling into the range of 15°C–20°C) is shown relative to the coldest days (<5°C, the omitted category, shown without 95% CI). Effects on the y axis are temperature-induced changes in units of the laboratory result (WBC, k/μL; RBC, m/μL; EDW, %; hemoglobin, g/dL).

whether such an improvement would be appreciable for laboratory performance, a question we attempt to answer in two ways:

(1) By comparison with improvements in measurement technology. Many innovations in LDL measurement have emerged to overcome the notorious imprecision of the Friedewald method. One study compared the precision of an early such method, second-generation immunoseparation (CV = 1.5%) versus Friedewald calculations (CV = 2.4%): an improvement of 0.9%.[25] Later, the best-performing third-generation "homogeneous" method achieved a CV of 0.6%, according to a literature review[26]—a further improvement of 0.9%.

(2) By comparison with national guidelines. The National Cholesterol Education Program recommends a maximum imprecision, CV_{max}, of less than 4.0% for LDL.[27] Thus, temperature-driven imprecision (CV_{temp}/CV_{max}) was 11.8% of the maximum tolerance. Because similar guidelines for precision have been compiled for many laboratory studies,[23] we present this statistic (where available) in Table 1. For some tests, temperature induced imprecision amounts to over 25% of tolerated limits (CV_{max}).

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

Table 2. Temperature effects on statin prescriptions

Outcome	N without outcome	N with outcome	Regression coefficient (95% CI)	p Value
Abnormal LDL (>100 mg/dL)	749,164	1,047,442	−0.00074 (−0.00094 to −0.00054)	<0.01
Fill statin within 90 days	1,718,544	78,062	−0.00014 (−0.00022 to −0.000061)	<0.01
Fill statin within 180 days	1,701,619	94,987	−0.00013 (−0.00022 to −0.000043)	<0.01

Shown are results of linear probability models of an abnormal LDL and statin prescription, measured via prescription claims data after LDL measurement, on temperature (linear) as the independent variable.

Statistical adjustment could produce improvements in precision that are large relative to technological improvements in assays and by comparison with national guidelines.

Are temperature-induced changes clinically important?

The changes we measure are small; they are only detectable because of the very large dataset we use. So a natural question is, do these changes matter for clinical practice? Because laboratory tests affect physician decision-making in complex ways, it is difficult to estimate in general terms how much temperature-induced distortions affect decision-making. For some tests, however, there is a clear mapping from results to decisions; one such setting is the decision to prescribe cholesterol-lowering drugs on the basis of lipid testing, one of the most consequential decisions to result from lab tests in everyday practice.

Table 2 shows how temperature changes (1) the probability of exceeding LDL cutoffs relevant to statin initiation using current guidelines[28] and (2) the probability of filling a statin prescription after lipid testing (defined as LDL and HDL testing on the same day). Individuals with LDL measured on a 30°C day are 3.2% less likely overall to have LDL measured as greater than 100 mg/dL and ultimately 9.7% less likely to fill a statin prescription in the next 3 months (–0.42% versus baseline of 4.34%, $p < 0.001$) versus those with labs drawn on a 0°C day (–1.86% versus baseline of 58.3%, $p < 0.001$). These models omit individual fixed effects because we consider only individuals' first lipid test and use linear probability models rather than logit; logit models did not converge after several days because of the size of the dataset.)

What are the potential mechanisms for temperature-induced changes?

Understanding the nature of these temperature-induced changes is an interesting and important area for future research. The design of this study, which relies on an observational database, limits the extent of exploration we can do, and this means that any discussion of mechanisms is highly speculative. Nonetheless, some of our results may provide some insights into mechanisms, which we discuss using Fraser's distinction of analytical versus biological sources of variation in test results (omitting pre-analytic factors like posture, short-term food intake, or sample collection technique, regarding which we have little to say).[7]

Analytical variation encompasses dynamic systematic error *in vitro* as well as variability because of sample temperature and volume. To explore whether temperature induces analytical variation affecting a particular modality of testing (e.g., spectrophotometry, chemiluminescence, flow cytometry), we grouped all 75 tests by assay methodology, as shown in Table S1. Detailed comparison with the sign of the linear coefficients in Table S1 shows that temperature effects are not concentrated in a particular measurement technique, nor is the direction of change uniform within a technique, arguing against a particular artifact driving the changes.

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

Another analytical factor is prolonged time in transport between sample collection and processing, resulting in longer exposure to temperature, or correlations between temperature and transport time (e.g., because of traffic changes on colder days). This could cause a number of changes to samples, significantly altering results (e.g., via hemolysis or interference from delays in separating cells from serum[29]). To explore this source of variation, we compare our results with prior work that experimentally varied temperature and storage time to simulate the effect of transport under different conditions.[11] Comparing our estimates with these results shows that, for some labs, the effect we measure is quite close to experimentally simulated conditions reflecting transport: ALT (but not AST), albumin (we could not identify data on globulin), and erythrocytes (but not other components of the CBC; Table S1). However, for other laboratory studies, we find disagreement regarding sign (i.e., does the quantitative result increase or decrease) and magnitude of effect, suggesting that transport cannot explain many of our findings. Finally, it is important to note that the experimental literature on the effect of temperature exposure is not unanimous regarding the sign or magnitude of effects for the labs we study,[30] perhaps partially because temperature affects different assays very differently.[31]

Biological variation encompasses physiological changes *in vivo*, whether over the lifespan because of predictable cyclical rhythms or arising from short-term variability in an individual's biology. Changes in plasma volume are one obvious source of such biological variation. Higher temperature has several known and opposing effects on plasma volume. On one hand, there is a physiological expansion of plasma volume, likely because of venodilation-mediated reduction of capillary hydrostatic pressure, Q12 causing an influx of interstitial fluid.[16,20] On the other hand, exposure to higher temperature can also lead to reductions in plasma volume (and total body water) via insensible losses. Recall that, because our regression model adjusts for city-week trends, we only detect changes in these test results at the timescale of days (i.e., seasonal variation, which generally shows higher plasma volume in the summer Q13 months,[21] has been controlled for.)

Classical models of assay variability because of plasma volume imply that volume expansion should lower the concentration of some tests (i.e., we would see a negative effect of increased temperature in Table 1, column 7), whereas volume loss should increase their concentration (i.e., a positive effect of temperature). For example, in studies of positional changes (e.g., supine to sitting to standing), which decrease plasma volume,[32] many test results from CBC components[11] to lipids[33] Q14 were increased. A further implication of these models is that changes should be proportional across tests (i.e., in Table 1, column 7 should be the same magnitude for all Q15 such tests).[7] Inspecting our results on kidney function, as noted above, we see that creatinine, BUN, and urine specific gravity increase with temperature (i.e., the data in Q16 Table 1, column 7 are large and positive). In contrast, Table 1 shows that most of the other coefficients in column 7 7 are negative; if these results were due to plasma volume, they would have to be driven by volume expansion, which is in contradiction to the results for creatinine, BUN, and urine specific gravity.

The most narrow implication of this is that plasma volume cannot explain all of our results: the signs and magnitudes vary considerably, contrary to what we might expect. If there were heterogeneity in the population and in temperature effects on different labs, we could see such complex patterns, but we are unable to explore this directly in our analysis. However, overall, we cannot say with certainty whether plasma volume has expanded or contracted. Because creatinine, BUN, and urine specific gravity are so closely linked to plasma volume, we favor the interpretation

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

that volume decreases on average because of insensible losses on the day before and day of testing when distributed lags show that temperature matters. However, we cannot exclude plasma volume expansion. Creatinine could increase independent of plasma volume (e.g., in response to increased activity or muscle breakdown in the 24–48 h before testing), and urine specific gravity could reflect greater water retention to maintain increased plasma volume.

Another source of biological variation is aging. Panel 5 of Figures S1–S75 shows age-specific results, breaking out the effect of temperature for individuals over versus under 50 years of age (using an indicator term interacted with linear temperature in our regression models). One notable finding is that older individuals had larger increases in tests, directly reflecting volume depletion on hotter days. This may reflect the fact that elders are less able to maintain homeostasis in higher temperatures. It also suggests that, for these tests at least, biological rather than analytical factors are at play; factors operating after sample collection should affect all samples equally, meaning we would not observe differences by age. Another way in which age could affect our results is via confounding. For example, RDW is known to increase and hematocrit to decrease over the lifespan. Panel 5 of Figures S27 and S39 shows results by age for these two tests. The effect sizes estimated for younger versus older patients are statistically indistinguishable. We view this as reassuring; if age were a confounder (e.g., because tests were simply decreasing or increasing with age), then we would have expected to see larger differences between these two groups.

Temperature could also correlate with biological variability in red blood cell dynamics. Drawing on this literature, we can conjecture that increased temperature causes (smaller) older cells to be taken out of circulation.[34] This is supported by the observation that erythrocyte size increased (MCV: 0.0071 fL/°C, $p < 0.001$). If removal of smaller cells were the only mechanism, then we would expect variance of erythrocyte size (EDW) to decrease, but EDW increased (0.0049%/°C, $p < 0.001$), implying increased production of new cells alongside destruction of older cells. Higher temperature may lead to increased removal of red cells and increased production of new cells but not enough to offset removal, leading to a net decrease. With regard to platelet-related changes, higher temperature may be linked to decreased production as older (smaller) cells persist and fewer new (larger) ones are produced.

A final explanation for the variability we observe, which is neither analytical nor biological, is selection bias. Ambient temperature could affect the behavior of individuals; for example, if they are more likely to go outside on sunny days, undertake physical exercise, or cancel scheduled appointments for a blood draw. This hypothesis would imply a U-shaped (or inverse U-shaped) relationship between ambient temperature and labs because selection of individuals on very hot or cold days changes the relationship we observe on less extreme days. However, by and large, panel 3 of Figures S1–S75 shows an approximately linear, monotonic change as temperature increases, making this hypothesis less likely.

DISCUSSION

Using a large dataset of blood tests, we find that temperature significantly affects 51 of 75 assays and the vast majority (>90%) of individual tests in a large national sample. There are clear effects of temperature on tests reflecting short-term physiology and also distortions in a variety of other tests meant to measure stable physiological states, like cardiovascular risk. The variety of effects and magnitudes makes a single

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

unifying explanation unlikely, including plasma volume, specific assay performance, transport, or other known drivers of variability. Whatever their cause, temperature produces undesirable variability in at least some tests, which, in turn, leads to distortions in important medical decisions, notably statin prescription.

Our results may have practical implications for laboratory reporting. In some cases, laboratories may wish to consider statistically adjusting reported results for temperature. Doing so could reduce weather-related variability at low cost relative to the cost of new laboratory assay technology or investments in temperature control in transport vans, etc. There is some precedent for this idea; the "middleware" used by laboratory information systems adjusts raw assay results (e.g., rounding low results to "negative" based on reference ranges), although these are typically restricted to simple transformations. The idea of improving the "software" rather than the "hardware" of laboratory instruments is in some ways reminiscent of recent accounts of how Tesla improved braking performance via a software upgrade to cars' onboard computers as opposed to physical changes to the brakes.[35]

Naturally, the case for statistical adjustment is not black and white. For example, we would not normally want to adjust away temperature-induced changes in creatinine; but sometimes this lab is drawn to monitor progression of renal failure, which means that short-term changes are less clearly the quantity of interest. Thus, in practice, decisions regarding adjustment would need to be at the discretion of the laboratory staff and the treating physician, potentially on a case-by-case basis. There is some precedent for reporting adjusted and unadjusted values of certain labs and leaving the choice of which to use at the discretion of the treating physician.

This raises the broader question of how to identify the source of temperature-induced variations. Although our results indicate that much of the variation is biological as opposed to analytical, the nature of the dataset we used precludes a more fine-grained investigation of mechanisms. Follow-up work, using the detailed records on timing of sample collection, transport, and processing maintained by many health systems and laboratory testing companies, could be used to quantify the effect of post-collection factors. In addition, correlating temperature-induced changes to the results of certain clinical tests—e.g., sonographic venous compression studies, which provide measures of plasma volume—could also illuminate the source of variation. Another fascinating topic for future research is the correlation between the changes in results within individuals rather than the average effect across individuals. This could yield more fine-grained insights into mechanisms; for example, if plasma volume were driving results, then within-individual changes for several analytes should be highly correlated. This, in turn, raises a number of technical challenges; besides the proliferation of hypotheses to test and the need for individuals to have both tests done on at least 2 days, there are no established methods for computing such correlations. One potentially promising avenue is to calculate the observed within-individual correlation between two tests and compare it with what we would expect if the tests were independent (based on the average effect, as we document here), but such methods would need careful validation. We believe that all of these changes would be best studied in the outpatient, not inpatient, setting; because inpatients and their blood are kept in temperature-controlled hospitals, neither would be exposed to the full effect of ambient temperature fluctuations.

In the absence of such detailed testing, these temperature fluctuations raise fascinating new questions about human physiology. For example, the liver synthesizes

many circulating proteins as well as cholesterol and, as a result, is a major contributor to endogenous heat production. Exogenous heat could cause compensatory decreases in liver synthetic function. But instead of conjectures, we hope our findings will spur further research into these interesting and potentially poorly understood pathways.

Limitations of study

Limitations are primarily related to the observational nature of the study. Although we controlled for temporal and geographic averages as well as individuals' averages, we can only assume that the remaining short-term temperature fluctuations were as good as random; this assumption cannot be verified. That said, similar strategies are used widely in economics and social science to rigorously estimate causal effects of temperature. Our results are also valid for the prevailing conditions affecting measurement in the particular sample we study, which is influenced by country-specific patterns in laboratory testing, the way samples are collected and transported, the prevalence of air conditioning, and many other factors. If there were dramatic changes to these contextual factors, then the model would need to be re-estimated. In addition, although we control for time-invariant individual factors using fixed effects and overall temporal trends using year fixed effects, we observe only a limited amount of contextual information on each participant, meaning that we could not adjust for individual-specific trends in results. Finally, although temperature may have measurable effects on lab results and downstream clinical decisions, like statin prescription, most of the variation in test results is shaped by unmeasured factors related to individuals' physiology—after all, the reason why we get laboratory tests is that we cannot predict what they will show based on pre-test individual characteristics. As a result, even factors that are considered to be very important determinants of test results explain very small proportions of variation (for example, individuals have historically been encouraged to fast before testing to reduce unexplained variation in LDL from dietary factors, but fasting reduces variability in LDL by only small amounts; we can indirectly infer around 2-3 percentage points of R^2 from existing studies[26]).

Blood test results have small but systematic variation driven by short-term ambient temperature fluctuations. These differences can lead to meaningful changes in treatment decisions. Better understanding these perturbations could improve precision in measurement by accounting for ambient temperature in reporting lab results, and shed new light onto human physiology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.medj.2021.11.003.

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

ACKNOWLEDGMENTS

Q21 We are deeply grateful to Kristy Kim and Tashfeen Saeed for outstanding research assistance, Thomas C. Binns for invaluable contributions to our understanding of laboratory methods, Benjamin Mazer and Jeremy Sussman for extremely thoughtful comments, and three anonymous reviewers whose comments greatly improved the paper. There was no specific funding for this study, except related to the authors' faculty positions at the University of California, Berkeley and the University of Chicago. These institutions had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. We confirm our independence from funders and that we had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

AUTHOR CONTRIBUTIONS

Z.O. and D.P. performed and oversaw statistical analyses, had unrestricted access to all data, prepared the first draft of the manuscript, reviewed it, and edited it. Both authors agreed to submit the manuscript, read and approved the final draft, and take full responsibility for its content, including the accuracy of the data.

DECLARATION OF INTERESTS

The authors declare no competing interests. Unrelated to the submitted work, Z.O. has been supported by grants from Google, the Gordon and Betty Moore Foundation, the National Institutes of Health, and Schmidt Futures; has received speaking or consulting fees from AcademyHealth, Anthem, Blue Cross Blue Shield of Tennessee, Independence Blue Cross, Premier Inc, and The Academy; has equity interests in Dandelion Health and LookDeep Health; and is compensated as a staff physician at Tséhootsooí Medical Center.

Received: April 18, 2021
Revised: August 22, 2021
Accepted: November 9, 2021
Published: December 10, 2021

REFERENCES

1. Centers for Medicare & Medicaid Services (CMS), HHS; Centers for Disease Control and Prevention (CDC), HHS; Office for Civil Rights (OCR), HHS (2014). CLIA program and HIPAA privacy rule; patients' access to test reports. Final rule. Fed. Regist. *79*, 7289–7316.

2. Cole, J. (2006). Report of the Review of NHS Pathology Services in England. https://authorzilla.com/10De7/report-of-the-review-of-nhs-pathology-services-in-england.html.

3. Myers, G.L., Miller, W.G., Coresh, J., Fleming, J., Greenberg, N., Greene, T., Hostetter, T., Levey, A.S., Panteghini, M., Welch, M., and Eckfeldt, J.H.; National Kidney Disease Education Program Laboratory Working Group (2006). Recommendations for improving serum creatinine measurement: a report from the Laboratory Working Group of the National Kidney Disease Education Program. Clin. Chem. *52*, 5–18.

4. Sebastián-Gámbaro, M.A., Lirón-Hernández, F.J., and Fuentes-Arderiu, X. (1997). Intra- and inter-individual biological variability data bank. Eur. J. Clin. Chem. Clin. Biochem. *35*, 845–852.

5. Fraser, C.G., Wilkinson, S.P., Neville, R.G., Knox, J.D., King, J.F., and MacWalter, R.S. (1989). Biologic variation of common hematologic laboratory quantities in the elderly. Am. J. Clin. Pathol. *92*, 465–470.

6. Cotlove, E., Harris, E.K., and Williams, G.Z. (1970). Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. 3. Physiological and medical implications. Clin. Chem. *16*, 1028–1032.

7. Fraser, C.G. (2001). Biological variation: from principles to practice (AACC Press).

8. Lippi, G., Salvagno, G.L., Montagnana, M., Franchini, M., and Guidi, G.C. (2006). Phlebotomy issues and quality improvement in results of laboratory testing. Clin. Lab. *52*, 217–230.

9. Bruns, D.E., and Knowler, W.C. (2009). Stabilization of glucose in blood samples: why it matters. Clin. Chem. *55*, 850–852.

10. Gren, B. (2009). Incorrect guidelines for venipuncture affect the analytical results. Scand. J. Clin. Lab. Invest. *69*, 815–816.

11. Lima-Oliveira, G., Guidi, G.C., Salvagno, G.L., Brocco, G., Danese, E., and Lippi, G. (2016). Estimation of the imprecision on clinical chemistry testing due to fist clenching and maintenance during venipuncture. Clin. Biochem. *49*, 1364–1367.

12. van Balveren, J.A., Huijskens, M.J., Gemen, E.F., Péquériaux, N.C., and Kusters, R. (2017). Effects of time and temperature on 48 routine chemistry, haematology and coagulation analytes in whole blood samples. Ann. Clin. Biochem. *54*, 448–462.

13. Goodman, J., Hurwitz, M., Park, R.J., and Smith, J. (2020). Heat and learning. Am. Econ. J. Econ. Policy *12*, 306–339.

14. Burke, M., Hsiang, S.M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. Nature *527*, 235–239.

15. Vedel-Krogh, S., Kobylecki, C.J., Nordestgaard, B.G., and Langsted, A. (2019). The Christmas holidays are immediately followed by a period of hypercholesterolemia. Atherosclerosis *281*, 121–127.

16. Ockene, I.S., Chiriboga, D.E., Stanek, E.J., 3rd, Harmatz, M.G., Nicolosi, R., Saperia, G., Well, A.D., Freedson, P., Merriam, P.A., Reed, G., et al. (2004). Seasonal variation in serum cholesterol levels: treatment implications and possible mechanisms. Arch. Intern. Med. *164*, 863–870.

17. Pincus, R. (1995). Barnett, V., and Lewis T.: Outliers in Statistical Data. 3rd edition. J. Wiley & Sons 1994, XVII. 582 pp., £49.95. Biometrical Journal *37*, 256.

18. Wilcox, R. (2005). Trimming and winsorization. Q17 In Encyclopedia of Biostatistics (Wiley), p. 8.

19. Gasparrini, A., Armstrong, B., and Kenward, M.G. (2010). Distributed lag non-linear models. Stat. Med. *29*, 2224–2234.

20. Maw, G.J., Mackenzie, I.L., and Taylor, N.A. (2000). Can skin temperature manipulation, with minimal core temperature change, influence plasma volume in resting humans? Eur. J. Appl. Physiol. *81*, 159–162.

21. Sawka, M.N., Convertino, V.A., Eichner, E.R., Schnieder, S.M., and Young, A.J. (2000). Blood volume: importance and adaptations to exercise training, environmental stresses and trauma sickness. Med. Sci. Sports Exerc. *32*, 332–348.

22. Steadman, R.G. (1984). A universal scale of apparent temperature. J. Clim. Appl. Meteorol. *23*, 1674–1687.

23. Ricós, C., Alvarez, V., Cava, F., García-Lario, J.V., Hernández, A., Jiménez, C.V., Minchinela, J., Perich, C., and Simón, M. (1999). Current databases on biological variation: pros, cons and progress. Scand. J. Clin. Lab. Invest. *59*, 491–500.

24. Kidd, B.A., Hoffman, G., Zimmerman, N., Li, L., Morgan, J.W., Glowe, P.K., Botwin, G.J., Parekh, S., Babic, N., Doust, M.W., et al. (2016). Evaluation of direct-to-consumer low-volume lab tests in healthy adults. J. Clin. Invest. *126*, 1734–1744.

25. Schectman, G., Patsches, M., and Sasse, E.A. (1996). Variability in cholesterol measurements: comparison of calculated and direct LDL cholesterol determinations. Clin. Chem. *42*, 732–737.

26. Nauck, M., Warnick, G.R., and Rifai, N. (2002). Methods for measurement of LDL-cholesterol: a critical assessment of direct measurement by homogeneous assays versus calculation. Clin. Chem. *48*, 236–254.

27. Warnick, G.R., Kimberly, M.M., Waymack, P.P., Leary, E.T., and Myers, G.L. (2008). Standardization of measurements for cholesterol, triglycerides, and major lipoproteins. Lab. Med. *39*, 481–490.

28. Grundy, S.M., Stone, N.J., Bailey, A.L., Beam, C., Birtcher, K.K., Blumenthal, R.S., Braun, L.T., de Ferranti, S., Faiella-Tommasino, J., Forman, D.E., et al. (2019). 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J. Am. Coll. Cardiol. *73*, e285–e350.

29. Ford, L., and Berg, J. (2008). Delay in separating blood samples affects creatinine measurement using the Roche kinetic Jaffe method. Ann. Clin. Biochem. *45*, 83–87.

30. Felding, P., Petersen, P.H., and Hørder, M. (1981). The stability of blood, plasma and serum constituents during simulated transport. Scand. J. Clin. Lab. Invest. *41*, 35–40.

31. Fantz, C.R., and Greene, D.N. (2018). Where Are the Preanalytical Stability Standards? J. Appl. Lab. Med. *2*, 830–832.

32. Maw, G.J., Mackenzie, I.L., and Taylor, N.A. (1995). Redistribution of body fluids during postural manipulations. Acta Physiol. Scand. *155*, 157–163.

33. Rifai, N., Warnick, G.R., and Dominiczak, M.H. (2000). Handbook of Lipoprotein Testing (AACC Press).

34. Higgins, J.M., and Mahadevan, L. (2010). Physiological and pathological population dynamics of circulating human red blood cells. Proc. Natl. Acad. Sci. USA *107*, 20587–20592.

35. Olsen, P. (2018). Tesla Model 3 Gets CR Recommendation After Braking Update. Consumer Reports, May 30, 2018. https://www.consumerreports.org/car-safety/tesla-model-3-gets-cr-recommendation-after-braking-update/.

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Ziad Obermeyer (zobermeyer@berkeley.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

We cannot make data available under the terms of our data use agreement, but data can be independently licensed for research via IBM Analytics. Our results and code are publicly available at https://gitlab.com/labsysmed/weather-to-test.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We use IBM Analytics claims data on commercially-insured US workers under 65 from 2009-2015. The dataset contains private insurance claims from a set of major US payers, with a subset of records linked to electronic health record data from a specific (but unnamed) set of health systems. This contains outpatient laboratory test results, which we relied on for the study. We restrict to those with results from at least one of the 75 most common quantitative tests (excluding qualitative or binary tests). We delete duplicates, and winsorize the data at the top and bottom 2.5% for each test, as is customary in large administrative datasets with erratic outlier values.[17,18]

The resulting dataset contains 215,234,179 test results for 4,877,039 individuals. The most and least common tests were creatinine (n = 7,475,654) and choriogonadotropin (n = 126,325). Demographic information is limited by privacy concerns: we observe only age, gender, and home Metropolitan Statistical Area (which we abbreviate 'metro area').

We merge test date and participants' home metro area with weather data from the National Oceanic and Atmospheric Administration, the gold standard for climate data in the US. We average weather across zip codes by metro area to form metro area-level data. We use the maximum daily temperature, rather then the mean, to reduce the influence of nighttime conditions, when patients are less likely to be outside and affected by temperature, as is customary in studies of heat exposure.[13] While using metro-area maxima does not capture temperature variation within metro areas, the resulting mis-measurement of our independent variable would induce attenuation bias. In other words, the coefficients we estimate would be biased toward zero, making it less likely to find an effect of temperature. We thus view our estimates as a lower bound on temperature effects.

QUANTIFICATION AND STATISTICAL ANALYSIS

We study the effect of day-to-day temperature fluctuations (the exposure) on quantitative results of the 75 most widely-used laboratory tests, and likelihood of filling statin prescriptions after lipid testing (the outcomes). To do so, we perform regressions modeling individual test results as a function of ambient temperature (daily high, °C), controlling for:

Please cite this article in press as: Obermeyer and Pope, Variation in common laboratory test results Because of ambient temperature, Med (2021), https://doi.org/10.1016/j.medj.2021.11.003

1. Fine-grained temporal and geographic trends. We include metro area–week fixed effects (19,812 indicators, indexing 52 weeks in 381 metro areas; this controls for, e.g., the average creatinine results observed in Dallas in the second week of July) and calendar-year fixed effects. Together these account for seasonal and geographic differences, which have many influences besides the temperature on the day the lab was drawn (e.g., aging or other trends across years, comorbidity differences across cities, dietary patterns across seasons[15]).
2. Individual patients' average values. We also include patient fixed effects (i.e., indicators for each individual), that control for patients' average values (e.g., a patient's mean creatinine over multiple measurements). This restricts to patients with multiple labs (e.g., $\geq$ 2 creatinines) checked over the span of our data. By controlling for a patient's average value (e.g., the fact that a particular patient has mildly elevated creatinine at baseline due to mild renal insufficiency), we estimate the effect of temperature *within* patients: how the same test, for the same patient, changes when drawn on different days with different temperatures.

This strategy estimates—over and above *average values in a region and week*, and over and above *patients' usual values*—the effect of temperature on the day the test was conducted. We first measure temperature effects using a set of temperature-bin indicators (in 5°C bins: < 5°C, 5-10°C, …, > 35°C) to account for potential non-linearities. Then, since relationships appeared largely linear, we also re-estimate models with linear coefficients to summarize temperature effects concisely. As a sensitivity Q20 analysis, we also construct distributed lag model[19] (i.e., 31 lag/lead variables measuring temperatures $\pm$ 15 days around, and the day of, testing) to quantify time-varying effects, hypothesizing that the largest effects would be on the day of and before the test. Results are robust to the inclusion of metro-area–calendar year fixed effects as additional controls (not shown), which makes it less likely that the variation we observe is driven by secular increases in temperature (as opposed to quasi-random fluctuations). Standard errors are clustered at the metro area-week level. We also re-run all results clustered at the metro-area level, which increases the standard errors in most cases (between 50%–100%), but does not change the interpretation for the vast majority of effects.

Analyses are performed using STATA 14.0/15.0. Since data are deidentified, the University of Chicago Institutional Review Board judged this not to be human subjects research. Patients were not involved in the design of this study, but permitted their data to be used for this research; without their data this study would not be possible.