Holt, Stephen B.; Vinopal, Katie

Working Paper

Response to Replication of Examining Inequality in the Time Cost of Waiting

I4R Discussion Paper Series, No. 244

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Holt, Stephen B.; Vinopal, Katie (2025) : Response to Replication of Examining Inequality in the Time Cost of Waiting, I4R Discussion Paper Series, No. 244, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/321371

INSTITUTE for REPLICATION

No. 244
DISCUSSION PAPER SERIES

Response to Replication of Examining Inequality in the Time Cost of Waiting

Stephen B. Holt

Katie Vinopal

This paper responds to:

Hall, J.D., and D. Thiele. 2025. A comment on “Examining Inequality in the Time Cost of Waiting”. *I4R Discussion Paper Series No.* 243. Institute for Replication.

July 2025

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 244

Response to Replication of Examining Inequality in the Time Cost of Waiting

Stephen B. Holt[1], Katie Vinopal

[1]University at Albany/USA

JULY 2025

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Response to Replication of Examining Inequality in the Time Cost of Waiting

Stephen B. Holt* Katie Vinopal

July 6, 2025

Abstract

Holt and Vinopal (2023) provides evidence of an income-based gap in time spent waiting for services on the typical day in the United States. The gap was estimated at the extensive margin, unconditional intensive margin, and intensive margin conditional on some waiting time. The analysis was complemented with a heterogeneity assessment that found the income-based gap in the length of waiting spells conditional on some waiting was observed across race categories except Black Americans. Unfortunately, Hall and Thiele identified an error in the code regarding missing household income data that led to a larger analytic sample than was warranted. In addition, Hall and Thiele claim some sensitivity in the hypothesis tests of some results to an expanded definition of the dependent variable and standard error estimation choice. However, correcting our original coding errors regarding missing income, adding weights when omitted, and adding controls when omitted does not meaningfully alter the estimated coefficients or their precision. Hall and Thiele's claims that our race results are not robust is misleading, as they seem driven by interpretation of simplified and space-constrained writing in the abstract, rather than lack of reproduction or sensitivity of the empirical results. To the extent we observe some sensitivity, it involves Hall and Thiele's move to an expansive definition of waiting time to include more leisure and entertainment services, a change that substantially alters the interpretation of results and goes beyond a simple test of robustness.

1 Introduction

We write to respond to "A comment on 'Examining Inequality in the Time Cost of Waiting." We have carefully reviewed Hall and Thiele's (henceforth "HT") thorough replication and robustness checks and thank them for their time. HT tests claims adopted from our abstract and some expansions from our full manuscript. They argue that our claim that the income-based gap in waiting time remains after conditioning on observables is only robust in some specifications and our claim regarding race-based gaps in waiting time is not robust. Their argument for the former seems to rely primarily on the use of a measure of income that is not comparable to ours and an alternative definition of waiting time that introduces additional activities, as corrections for other coding errors HT document do not alter the statistical significance of the results in the published manuscript.

Regarding our claim on race differences in time spent waiting when waiting occurs, part of HT's argument seems to hinge entirely on imprecision in our language. On page 13 of their comment, HT lay out the claims

*Corresponding author. Email: sbholt@albany.edu

they are assessing and assess our analyses against broader claims than were intended. In the first instance, they seem to interpret our use of "longer average wait times" as "waiting times" when "wait times" was intended as a short hand for "time spent waiting when waiting" or "length of waiting spells." We admit to the imprecision of the language; however, we were under strict character limits in communicating in the abstract imposed by the journal's editing standards and rules.

In the second pull quote they use in their evaluation, they do indeed find us referencing "waiting times" when our true intended reference was "wait times" or "time spent waiting when waiting occurs." This editing error arose, again, out of the necessity of fitting within the strict table, figure, and word limits of the journal.[1] Indeed, our focus on waiting time conditional on waiting in the race analysis was due to space constraints. Our original draft and public working paper written prior to review present results assessing racial differences in all four outcomes and provides a lengthier and more complete discussion and interpretation of the results. We also explored gender gaps and gaps between parents and non-parents. During review, the anonymous reviewers and editing team suggested focusing on wait times (time spent waiting when waiting occurs) given these space constraints and moving the remainder of the results to the online appendix.

We discuss this as a winding way to note that we find the conclusion that the claims regarding race gaps as not being robust a misleading conclusion. After correcting for errors in the code, the results regarding racial gaps in conditional waiting time replicate and are not sensitive to standard error specification, waiting time definition, or the inclusion of all controls in the full specification of the main model. While HT acknowledge this in their text, their conclusion in the abstract and throughout regarding this analysis seems driven by a comparison to a claim we did not intend (and from a single typo in the manuscript).

Our response proceeds in three parts. First, we acknowledge minor coding errors (and thank HT for finding them), correct our code and tables, and show that these mistakes do not substantially change our results. Second, we discuss the two major changes HT made that drive differences between our paper and HT's robustness checks. Third, we address the claim that our race results are not robust, which we believe is incorrect. Fourth, we address HT's assessment that our Claim 2 – that the income gap in waiting times cannot be explained by observable characteristics–is not robust in all specifications.

We find that their claims rest entirely on a change to the dependent variable that fundamentally differs from our focus on necessary services by including more discretionary and leisure-oriented activities in the analysis. Given our primary focus on socioeconomic differences in time autonomy - discussed plainly throughout the text and motivating our examination of necessary services - the inclusion of services for which consumption is highly discretionary in the outcome variable represents a substantive change in the definition and interpretation of the measure. As such, we do not believe they present any meaningful evidence to justify a claim that our results are not robust. Instead, they present a meaningfully different measure as a robustness check that should instead be viewed as raising a different research question regarding socioeconomic gaps in waiting for services in leisure activities.

[1] We note that the unconventional requirement of the journal involving data and methods description at the end of the manuscript also involved a great deal of editing for compliance on our part, adding to the possibility that some of the errors HT document arose in our revisions.

2 Responses

2.1 Coding Errors

First, HT helpfully finds a potential serious coding error wherein a number of households who report missing income were incorrectly categorized as low income in some years. This was simply an honest mistake; a result of our overlooking how missings were coded on one of the variables used to construct our measure of household income. We have corrected this error and reproduced our results, finding that they hold with this correction with only modest impact on the coefficients.

Second, we have fixed our code for Table 1, Column 6, Panel A such that it correctly matches what is described in the text of the paper. Third, we corrected the trivial error that prevents full reproducibility in Line 79 of 03_tables.do, which, as HT note, does not affect the results reported in the paper. Fourth, we apply weights to column 6 of Table 1 and Supplementary Tables 3-7, to match the text, and find only modest changes to coefficient magnitudes.

HT also point out some inconsistency between the text and the code in the definition of the dependent variable; specifically, what activities are included in our waiting time measures. Some of this was due to edits made with word and character constraints in mind. We describe our waiting time variable as: "shopping; medical care for the respondent, household children and adults, and non-household children and adults; household services (for example, maintenance, lawn services, etc.); legal services; financial services; educational services for the respondent or household and non-household children; government services; and personal care services (for example, salons, barbers, etc.)." In the definition, we thought of "purchasing childcare services" as a form of educational services and "civic obligations" as falling under the heading of "government services." We recognize now, after an external team undertook a replication effort, that activities like "purchasing childcare services" and "civic obligations" should have also been explicitly stated in our list.

Further, we acknowledge the errant inclusion of "waiting associated with helping household adults" and "waiting associated with helping non-household adults." We are glad to confirm in our response that the correction of this error does not meaningfully change our results.

For posterity, we have opted to simply correct our code to match the text definition, which includes waiting associated with purchasing childcare services, lawn services, and civic obligations as intended and removes waiting associated with helping household and non-household adults. Everything we do in this report referring to "textual definition" or "text DV", then, matches the textual definition described here. As we will show, this change makes little difference in the robustness of our results. As mentioned above, HT re-runs all analyses using an expanded definition of waiting time. We do not adopt this definition, for reasons we make clear below.

Other minor issues HT correctly note involve some mismatches between text description and table demarcation of controls, likely attributable to errors we made when converting table formats from working paper and submission formats to *Nature* journal formatting standards and text limitations, as this involved a great deal of re-arrangement of the presentation of results and the discussion of the data and methods of analysis. The mistakes are, of course, ours. We also take the point that our use of “unemployed” rather than “not employed” was imprecise.

3 Drivers of Inconsistencies Between Manuscript and Robustness Checks

The areas in which HT document the most sensitivity of the results involve models in which HT use replicate weights to estimate the standard errors of the regressions, usually in conjunction with their new definition of the dependent variable. That these changes co-occur in their presentation obscures the individual impact that each have individually on our results. In this section, we discuss each in turn.

3.0.1 Standard Errors

HT note that we use clustered standard errors in our regressions. Clustered standard errors at the highest level of selection into the sample when estimating fixed-effects models has been a conventional approach in applied econometrics for some time to account for correlated errors within sampling strata. See, for instance, the discussion in Abadie et al. (2023) regarding the roots and rationale of this common practice. However, Abadie et al. (2023) note that cluster adjustments are often too conservative, particularly in the context of fixed-effects models that, by design, absorb cluster-specific variation. We call attention to this more recent changing literature on the use of cluster adjustment in standard errors for a few reasons. First, we simply want to acknowledge that our approach was consistent with common practice in using both the ATUS (and related data, such as the CPS) to estimate fixed-effects regression models at the time of drafting, submitting, and revising our manuscript. See, for instance, Aguiar et al. (2013) use of the ATUS in the *American Economic Review* and, more recently, Lafortune and Low (2023) use of the ATUS in *American Economic Journal: Applied Economics* for examples of analyses using state-level cluster-adjusted standard errors for inference.

Second, we note that HT do not provide a rationale for using the replicate weights to estimate standard errors in fixed-effects regression models, nor a rationale for preferring them. Indeed, the issues that make clustering standard errors at the state level potentially too conservative are even more present when using the replicate weights, particularly in sub-group analyses. The replicate weights provide variance weights designed to replicate the sampling variance introduced by the known stratification in the construction of the sample.[2] The potential issue in using replicate weights arises because the fully specified model in our manuscript includes controls for all of the determinants of selection into the sample. The example use of the replicate weights provided by BLS in the User Guide refers to the estimation of standard errors for unconditional population means in part because their relevance for regression estimation depends upon the controls in the model being estimated by the analyst.[3]

We raise the issue of standard errors because we view our cluster adjustment at the state-level in our fully conditioned model as likely representing conservative estimates of standard errors and generally consistent with standard practice when running regressions using the ATUS; the use of replicate weights alongside our full set of controls for selection into the sample are not obviously resolving an iid violation and are likely far too conservative in some subgroup analyses. In light of this view, we note that all evidence presented by HT in which our results are not found robust rely on estimates using the replicate weights for estimating standard errors, again, often in conjunction with another change to controls or the definition of waiting time

[2]The ATUS uses a three-stage sampling process: state-level adjustment; differential sampling by respondent race, the presence and age of children, and the number of adults in the household; and sampling within household a respondent over age 15.

[3]In correspondence with BLS, they stated that they do not take a position on the use of replicate weights in regression analysis.

occurring simultaneously.

In Table 1, we correct the error involving missing household income and correct the mismatch between the code and the text in our definition of waiting time (i.e., use the textual definition) and present the estimates of the full model (column 6 in our published manuscript) for all outcomes to show the impact of the use of replicate weights in estimating standard errors. Results are robust to using replicate weights, assuming the use of our preferred dependent variable (discussed immediately below). Throughout this response, we will note in table notes when we present standard errors using replicate weights alongside our preferred use of state-level adjusted standard errors.

3.0.2 Defining the Dependent Variable

As mentioned in the coding errors section above, HT point out some inconsistency between the text and the code in the definition of the dependent variable (DV). We correct our code to match the text definition for the purposes of this response, namely by removing time spent waiting on helping household and non-household adults from the calculation of waiting time.

However, HT also re-run all analyses using an expanded definition of waiting time, and, as we show in our results below, this different DV drives nearly all of the differences between the paper and HT's robustness checks and is the primary reason for the sensitivity they observe regarding our claim that the income-gap in waiting time is robust to the inclusion of controls for observables.

Our definition of waiting time was based a conceptual framework. We wanted to capture activities that might be considered essential services; activities one could not easily opt out of. This is because our argument, as articulated in the introduction, is that individuals with less income have less time autonomy. Waiting for unavoidable activities is a useful way to capture this, since everyone must engage in those activities when necessary to some extent, and no one actually wants to spend their time waiting much for the service when they do. During the review process, reviewers were welcome to challenge this definition and ask us to run the analyses using a different dependent variable (note that our code was available to reviewers from the beginning of the review process). However, none did, presumably because they agreed with our conceptual argument.

HT's definition expands to include some activities that are arguably consistent with our conceptual framework that we overlooked in our construction of our outcome variable, such as using veterinary services and using vehicle repair and maintenance services. However, their expansion also includes waiting for activities such as eating and drinking, arts and entertainment, attending sporting events, purchasing real estate, and religious and spiritual activities, all of which we consider beyond our conceptual framework of waiting time. More to the point, we view the inclusion of these activities in the service mix as fundamentally altering the definition and interpretation of the dependent variable and that change makes the use of this expanded dependent variable a poor "robustness check."

For starters, as we note in both the text of the manuscript and the supplemental information, one challenge to studying waiting time is that consumption of services reflects, to some degree, a choice and that choice is shaped by consumer preferences, available alternatives, and ability to pay. Given our focus on the gap between low- and high-income households, low-income households are more likely to be selected out of more discretionary activities like eating out and ticketed entertainment options.

Second, given the discretionary nature of the service itself and number of possible options, a wait may dissuade consumption entirely or competition might offer ways to optimize away from facing a wait. By way of analogy, consider dining out. Since the service itself is highly discretionary, a long wait at a restaurant

may dissuade consumers from using the service at all in the first place or may push them to seek out a nearby alternative restaurant without a wait. This introduces multiple points in which selection renders the choice to consume the service and wait for the service more akin to a leisure time allocation as an expression of preference than a loss of autonomy over ones' time. By comparison, for more necessary services like health care, consumers have less choice in consuming the service when needed and less discretion in whether to wait when facing a wait. This distinction is no small matter and the change is large enough that we view their expanded variable incomparable to our own and not particularly useful as a check for robustness of our results.

Empirically, these recreational activities follow a time use pattern consistent with more selection into participation on ability to pay and more selection out of waiting consistent with continued time autonomy when consuming these services. While high-income households spend about 5.9 more minutes using our textual definition of necessary services per day than low-income households, they spend 12.4 minutes more using these recreational services per day, a pattern consistent with need constraining consumption choice for the services we examine in our paper. The average time spent waiting for these recreational services in a typical day is 0.38 minutes while the average for our textual definition of essential services is 1.20 minutes, again consistent with a pattern you would expect when consumption of the service is more discretionary, driven by preferences, and subject to more available alternative suppliers.

Thus, we continue to prefer our textual definition of the waiting DV and view the use of the expanded definition of waiting time to introduce more recreational activities an extension answering different research questions than those we set out to examine.

Still, because changes in standard error estimation and expanded DV definitions are done simultaneously by HT, we find it useful to disentangle the impact of these elements on our results. Further, we find it useful to examine the extent to which the difference in results is driven by the inclusion of less conceptually related recreational and religious activities in their expanded DV. In Table 2, we present our full model estimates separately by standard error approach with our a definition of waiting time that adds waiting for veterinary services and waiting for vehicle repair and maintenance services to our textual definition described previously. We also estimate the full model with both approaches to the standard errors using HT's expanded DV.

These results make two points. First, adding in the additional services that are defensibly related to our look at essential services does reveal some sensitivity in the likelihood to wait for services at all to controls (see panel B). However, the results for waiting time are robust regardless of standard error estimation (see columns 3 and 4 in panel D).

Second, the sensitivity to controls documented by HT is driven primarily by the inclusion of more recreational activities that we view as less conceptually relevant to the study we had proposed to execute. This is important since, as we will detail later, the basis of HT's conclusion about the sensitivity of our results to controls rests primarily on this change to the definition of the dependent variable, one we feel relies on the inclusion of activities unrelated to our phenomena of interest.

3.1 Race Results

We also disagree with HT's characterization of the robustness of our results regarding race. First, as with the main results, we note that some sensitivity is again driven by the use of replicate weights in the standard errors. We start by estimating the full model with all controls and state fixed effects, corrections to the code of household income and waiting time previously described, separately by race. Results are presented in Table 3. We do this using the corrected text definition of waiting time to match the text description in the

manuscript (Panel A) and our understanding of HT's expanded definition of waiting time (Panel B). Here, we focus on time spent waiting conditional on non-zero waiting time. In Table 4, we present estimates of the within-income race gap HT estimates in their replication report using cluster adjusted standard errors (columns 1-2) and successive difference replicate weights estimation (columns 3-4).

We focus here on non-zero time spent waiting both for time saving and because our manuscript focused primarily on this outcome due to journal imposed length, table, and figure restrictions. Our public working paper provides a more comprehensive discussion and treatment of the race gaps we do and do not observe and what they might suggest conceptually. While we tried to be clear in our final published manuscript that there was not a significant racial gap in the likelihood to experience waiting and instead only arises when considering the length of waiting spells when they occur, our revisions to fit journal requirements and reviewer feedback might not have been as clear as we would have preferred. We note this because the full set of results was estimated and submitted to reviewers and the focus of more attention on waiting time conditional on waiting was driven by reviewer interest and journal-imposed limitations, not by our editorial decisions or preference to focus on one outcome at the cost of understanding the complete picture. To argue that our results regarding racial gaps are not robust because of these limitations imposed upon us in the review and editorial process seems to be making an assertion of non-robustness against a claim we did not intend to make. *Our more specific claim - that the length of waiting spells when waiting occurs is shaped by both race and income - is robust to controls and, again, only somewhat sensitive to the decision regarding standard errors, though our main result, that wealthier Black people wait longer than wealthy white people, remains.*

3.2 Robustness of Income Gap in Waiting to Controls

HT finds that our Claim 2 – that the income gap in waiting times cannot be explained by observable characteristics –is not robust in all specifications. HT bases this conclusion, we believe, on three sets of results: i) using an alternative definition of the DV, ii) using an alternative definition of income, and iii) adding some new controls. As we noted previously, we do not feel their expanded definition of waiting time falls within the intended concepts or argument put forward in our paper and we characterized the nature of that sensitivity in our results in Table 2. HT's argument that Claim 2 is not robust then relies on two additional changes to the analysis: using respondent's reported weekly earnings instead of household income and adding additional controls. We discuss each in this section.

3.2.1 Earnings Measure

HT present an alternative measure of earnings, the reported weekly earnings of the respondent, and note differences in the results. They acknowledge some of the limitations of this measure, but we want to more strongly disagree that this presents a comparable enough alternative to warrant consideration as a sensible sensitivity analysis. First, the measure is missing in 45% of the sample and that missingness could reflect non-response or a respondent that is not a significant earner for the household. The extent of the missingness of this measure could raise particular power issues in the context of studying an activity that is not daily in nature (a limitation in the ATUS that we acknowledge in our original paper). But more importantly, the nature of the measure - an individual respondent's earnings - means a respondent in the bottom quartile of earnings could be a 15-year old respondent with a part-time, seasonal job living in a high-income household or a primary earner of a household living below the poverty line. In combination, these

issues raise serious questions about the fundamental comparability of this sensitivity check. That results using this measure disagree with our analysis is perhaps unsurprising given the fundamental differences in what the two measures of income are capturing. However, the difference in these measures is great enough that their usefulness in determining the robustness or sensitivity of our results is unclear to us. Indeed, this is likely why an anonymous reviewer suggested we use probabilistic matching with CPS data - which also provides a continuous measure of household earnings - as a sensitivity check rather than suggesting we use the respondent's earnings.

3.2.2 Controls

In HT's Table 9, they add a number of additional controls: MSA fixed effects, type of activity fixed effects, travel modes used, and age as a categorical variable, rather than a linear one. However, they add all sets of new controls simultaneously, leaving the relative impact of these controls unclear. Moreover, the resulting sample sizes they report raise some questions regarding their implementation of the controls. For the sake of time, we do not examine the inclusion of controls for travel modes used. Our full model controls for time spent traveling to each of the activities associated with waiting and the ATUS includes time spent waiting for transit as travel time to a given activity. We view this control as likely redundant and not likely a contributor to disagreement. Table 5 presents results of the full model using state-level cluster adjusted standard errors for both the text definition of waiting time and the extended definition used by HT. We add the new controls iteratively to examine the impact of each.

First, given the restriction in the ATUS to only the largest 340 MSA's, we wonder whether HT replaced suppressed MSA values as missing or included them in their estimates (given their sample sizes in Table 9 match the sample sizes elsewhere in their replication report). Further, given the limitations in the coverage of this variable and the small cell sizes of many MSAs that are included, we question the use of MSA fixed-effects as a control. Regardless, in Table 5, we show that using the text definition of waiting time, including MSA fixed-effects while removing missing MSAs yields results comparable to our manuscript results.

Second, the only sensitivity we observe arises modestly when activity fixed effects enter the model and more substantive sensitivity arises when all three new controls enter the model. The functional form of age does not meaningfully impact the results. The inclusion of activity fixed effects is of debatable merit given the cell sizes involved in many of activity-specific forms of waiting. We discuss the difficulties and limitations of studying waiting time using daily diary data in both the manuscript and supplemental material, prime among them being the relative infrequency of observing activities such as waiting in a sample constructed with the design of the ATUS. We view them as a bad control in this model, but our results to their inclusion alone remain robust. Indeed, the sensitivity arises when accounting for all three new controls; however, we note that the inclusion of the MSA fixed effects changes the composition and size of the sample and that change may be the primary reason for this modest sensitivity to all three controls. In Table 6, we estimate the model with all controls and only the activity fixed effects and binned measure of age and see that the results are again robust (though the point estimates shrink), consistent with the largest sensitivity in the analysis arising from the sample size and composition change induced by the MSA fixed effects.

In sum, the claim in the replication report that our results are only robust to controls in some specifications arises out of the introduction of arguably irrelevant categories of activities into the definition of the dependent variable, the use of an alternative earnings measure that is categorically different in the comparisons it makes by nature what it measures, and the introduction of new control variables of dubious merit that only induce sensitivity to our results through a change in the sample to only respondents in large MSAs. The number

of extra checks and sub-analyses that might be conducted in *any* empirical paper are nearly limitless. Our pre-publication reviewers simply asked for different robustness checks than the ones HT conducted. Adding additional robustness checks by a post-publication reviewer is likely to reveal small changes and caveats to almost all quantitative research papers. We take comfort that our results replicate after correcting our coding errors. Our errors in code and any lack of clarity in our writing and editing of the published manuscript are regrettably ours. We hope our supplemental analysis here provides some additional clarity on the nature of where, why, and how the replication report findings diverge from the published manuscript.

4 Conclusion

We conclude by simply thanking HT for their thorough and thoughtful replication. We regret deeply the coding errors in the final submission and will provide a second version of the .do files in the publicly available repository that includes corrections to these coding errors. We find relief that these coding mistakes did not impact the results and the results are only sensitive to a conceptual change in our dependent variable. We would gently suggest deeper consideration to the limited power available for an outcome like waiting time in daily time diaries, particularly when considering additional fixed-effects and subsample analyses.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Aguiar, M., E. Hurst, and L. Karabarbounis (2013). Time use during the great recession. *American Economic Review 103*(5), 1664–1696.

Holt, S. B. and K. Vinopal (2023). Examining inequality in the time cost of waiting. *Nature Human Behaviour 7*(4), 545–555.

Lafortune, J. and C. Low (2023). Collateralized marriage. *American Economic Journal: Applied Economics 15*(4), 252–291.

5 Tables

Table 1: Comparison of full model estimates by standard error approach

	(1) b/se/p	(2) b/se/p	(3) b/se/p	(4) b/se/p	(5) b/se/p	(6) b/se/p
Panel A. Any waiting						
≤$20K	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.02	0.27	0.52	0.51	0.52	0.52
Observations	196,888	196,888	196,888	196,888	196,888	196,888
Panel B. Any waiting ($Pr(T > 0 \mid S > 0)$)						
≤$20K	0.01	0.01	0.01	0.01	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.00	0.02	0.08	0.05	0.08	0.08
Observations	99,472	99,472	99,472	99,472	99,472	99,472
Panel C. Waiting time						
≤$20K	0.52	0.45	0.46	0.46	0.46	0.46
	(0.11)	(0.10)	(0.18)	(0.17)	(0.16)	(0.16)
	0.00	0.00	0.01	0.01	0.01	0.01
Observations	196,888	196,888	196,888	196,888	196,888	196,888
Panel D. Waiting time ($T \mid T > 0$)						
≤$20K	7.02	15.14	11.92	11.92	11.92	11.92
	(2.13)	(3.35)	(4.66)	(4.04)	(4.15)	(4.15)
	0.00	0.00	0.01	0.00	0.00	0.00
Observations	7,872	4,807	4,807	4,807	4,807	4,807
Original DV	✓					
Text DV		✓	✓	✓	✓	✓
Weighted			✓	✓	✓	✓
Robust SE			✓			
Cluster SE	✓	✓		✓		
SDR SE					✓	
Jacknife SE						✓

Note: All models include full set of controls and state fixed-effects, as described in Holt and Vinopal (2023). Original DV = waiting time using the original code. Text DV = corrected waiting time code to match text description. Weighted adds BLS provided sample weights for estimating coefficients. Robust SE = heteroskedastic robust standard errors. Cluster SE = standard errors clustered at the state-level as in Holt and Vinopal (2023). SDR SE = successive differencing replicate standard errors using BLS provided replicate weights. Jacknife SE = jacknife estimated standard errors using BLS provided replicate weights. Reference income category follows Holt and Vinopal (2023) in omitting the highest household income category.

Table 2: Comparison of full model estimates between expanded DV and expanded DV without recreation and religious activities

	(1)	(2)	(3)	(4)
	Expanded DV		New DV	
	b/se/p	b/se/p	b/se/p	b/se/p
Panel A. Any waiting				
≤$20K	-0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
	0.80	0.77	0.82	0.81
Observations	196,888	196,888	196,888	196,888
Panel B. Any waiting ($Pr(T > 0 \mid S > 0)$)				
≤$20K	-0.00	-0.00	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.01)
	0.79	0.76	0.22	0.25
Observations	192,489	192,489	99,921	99,921
Panel C. Waiting time				
≤$20K	0.40	0.40	0.52	0.52
	(0.22)	(0.23)	(0.19)	(0.17)
	0.08	0.08	0.01	0.00
Observations	196,888	196,888	196,888	196,888
Panel D. Waiting time ($T \mid T > 0$)				
≤$20K	9.43	9.43	15.00	15.00
	(3.28)	(3.84)	(3.93)	(3.96)
	0.01	0.02	0.00	0.00
Observations	8,350	8,350	5,529	5,529
Cluster SE	✓		✓	
SDR SE		✓		✓

Note: All models include full set of controls and state fixed-effects, as described in Holt and Vinopal (2023), and are weighted using BLS provided sampling weights. New DV = waiting time defined in text with waiting for vehicle maintenance and veterinary services included. Expanded DV follows the definition used by HT in their replication report and includes everything in the New DV but adds time spent waiting for dining out, real estate purchases, entertainment activities (both sports and arts events), and religious services. Cluster SE = standard errors clustered at the state-level as in Holt and Vinopal (2023). SDR SE = successive differencing replicate standard errors using BLS provided replicate weights. Jacknife SE = jacknife estimated standard errors using BLS provided replicate weights. Reference income category follows Holt and Vinopal (2023) in omitting the highest household income category.

Table 3: Income within race

	(1) White b/se/p	(2) Black b/se/p	(3) Hispanic b/se/p	(4) Asian b/se/p	(5) White b/se/p	(6) Black b/se/p	(7) Hispanic b/se/p	(8) Asian b/se/p
Panel A. Text DV, Waiting time ($T\|T>0$)								
≤$20K	17.08	-14.49	24.17	10.70	17.08	-14.49	24.17	10.70
	(4.29)	(10.38)	(12.30)	(25.00)	(4.73)	(18.76)	(12.16)	(29.14)
	0.00	0.17	0.06	0.67	0.00	0.44	0.05	0.71
Observations	3,799	766	649	129	3,799	766	649	129
Panel B. Expanded DV, Waiting time ($T\|T>0$)								
≤$20K	11.99	-3.08	21.46	10.66	11.99	-3.08	21.46	10.66
	(4.01)	(9.56)	(6.82)	(16.15)	(4.51)	(11.41)	(8.05)	(13.45)
	0.00	0.75	0.00	0.51	0.01	0.79	0.01	0.43
Observations	6,683	1,205	1,173	256	6,683	1,205	1,173	256
Cluster SE	✓	✓	✓	✓				
SDR SE					✓	✓	✓	✓

Note: All models include full set of controls and state fixed-effects, as described in Holt and Vinopal (2023). Text DV = corrected waiting time code to match text description. Expanded DV = expanded definition of waiting time adopted by HT. Cluster SE = standard errors clustered at the state-level as in Holt and Vinopal (2023). SDR SE = successive differencing replicate standard errors using BLS provided replicate weights.

Table 4: Race within Income

	(1) ≤$20K b/se/p	(2) ≥$150K b/se/p	(3) ≤$20K b/se/p	(4) ≥$150K b/se/p
Panel A. Text DV, Waiting time $(T\|T>0)$				
Black	-8.67	26.98	-8.67	26.98
	(4.73)	(10.87)	(5.48)	(10.90)
	0.07	0.02	0.12	0.01
Native American	60.62	13.29	60.62	13.29
	(70.60)	(25.18)	(31.43)	(18.16)
	0.39	0.60	0.06	0.47
Asian	-16.74	2.83	-16.74	2.83
	(10.65)	(12.59)	(11.64)	(10.47)
	0.12	0.82	0.15	0.79
Pacific Islander/Hawaiian	261.70	28.43	261.70	28.43
	(30.48)	(21.89)	(31.90)	(15.20)
	0.00	0.20	0.00	0.06
Multiple races	-5.90	-21.05	-5.90	-21.05
	(13.24)	(13.32)	(9.29)	(15.72)
	0.66	0.12	0.53	0.18
Latino(a)	7.58	4.42	7.58	4.42
	(6.13)	(7.11)	(6.22)	(6.86)
	0.22	0.54	0.22	0.52
Observations	1,031	330	1,031	330
Panel B. Expanded DV, Waiting time $(T\|T>0)$				
Black	-2.70	17.28	-2.70	17.28
	(4.30)	(7.37)	(5.69)	(9.03)
	0.53	0.02	0.64	0.06
Native American	43.50	11.36	43.50	11.36
	(54.00)	(16.11)	(22.43)	(14.89)
	0.42	0.48	0.05	0.45
Asian	-13.88	0.74	-13.88	0.74
	(9.59)	(10.55)	(8.72)	(7.72)
	0.15	0.94	0.11	0.92
Pacific Islander/Hawaiian	257.66	13.28	257.66	13.28
	(21.16)	(16.49)	(18.00)	(17.40)
	0.00	0.42	0.00	0.45
Multiple races	-4.94	-30.56	-4.94	-30.56
	(10.16)	(14.33)	(8.00)	(21.98)
	0.63	0.04	0.54	0.17
Latino(a)	5.56	-1.37	5.56	-1.37
	(5.77)	(5.80)	(5.36)	(5.77)
	0.34	0.81	0.30	0.81
Observations	1,479	648	1,479	648
Cluster SE	✓	✓		
SDR SE			✓	✓

Note: All models include full set of controls and state fixed-effects, as described in Holt and Vinopal (2023). Text DV = corrected waiting time code to match text description. Expanded DV = expanded definition of waiting time adopted by HT. Cluster SE = standard errors clustered at the state-level as in Holt and Vinopal (2023). SDR SE = successive differencing replicate standard errors using BLS provided replicate weights.

Table 5: Including New Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Text DV				Extended DV			
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
Panel A. Any waiting								
≤$20K	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.56	0.70	0.21	0.63	0.64	0.37	0.85	0.70
Observations	135,532	196,888	196,888	135,532	135,532	196,888	196,888	135,532
Panel B. Any waiting ($Pr(T > 0 \mid S > 0)$)								
≤$20K	0.01	0.00	0.01	0.00	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.05	0.29	0.02	0.26	0.63	0.39	0.86	0.71
Observations	69,401	99,472	99,472	69,401	132,788	192,489	192,489	132,788
Panel C. Waiting time								
≤$20K	0.50	0.24	0.47	0.16	0.45	0.41	0.42	0.34
	(0.17)	(0.10)	(0.17)	(0.10)	(0.22)	(0.12)	(0.22)	(0.14)
	0.00	0.03	0.01	0.11	0.05	0.00	0.06	0.02
Observations	135,532	196,888	196,888	135,532	135,532	196,888	196,888	135,532
Panel D. Waiting time ($T \mid T > 0$)								
≤$20K	11.87	8.01	10.77	6.23	11.27	8.47	9.29	8.63
	(4.83)	(3.59)	(4.10)	(3.50)	(4.41)	(2.45)	(3.15)	(3.39)
	0.02	0.03	0.01	0.08	0.01	0.00	0.00	0.01
Observations	3,239	4,807	4,807	3,239	5,707	8,350	8,350	5,707
Add MSA FE	✓			✓	✓			✓
Add activity FE		✓		✓		✓		✓
Add age bins			✓	✓			✓	✓

Note: Standard errors clustered at the state-level in parentheses. Text DV = waiting time using text definition (columns 1-4). Extended DV = waiting time using expanded definition in HT (columns 5-8). All models include all controls from full specification in Holt and Vinopal (2023). MSA FE = MSA fixed effects. Activity FE = fixed effects for activity associated with waiting. Age bins = indicators for 5-year binned categories of age.

Table 6: Including Age Bins and Activity Fixed Effects Together

	(1) Text DV b/se/p	(2) Extended DV b/se/p
Panel A. Any waiting		
≤$20K	0.00	0.00
	(0.00)	(0.00)
	0.73	0.67
Observations	196,888	196,888
Panel B. Any waiting ($Pr(T > 0\|S > 0)$)		
≤$20K	0.00	0.00
	(0.00)	(0.00)
	0.37	0.69
Observations	99,472	192,489
Panel C. Waiting time		
≤$20K	0.23	0.37
	(0.10)	(0.12)
	0.03	0.00
Observations	196,888	196,888
Panel D. Waiting time ($T\|T > 0$)		
≤$20K	7.63	8.42
	(3.63)	(2.40)
	0.04	0.00
Observations	4,807	8,350

Note: Standard errors clustered at the state-level in parentheses. Text DV = waiting time using text definition. Extended DV = waiting time using expanded definition in HT. All models include all controls from full specification in Holt and Vinopal (2023) and add activity fixed effects for the activities associated with the waiting and replace linear age with indicators for 5-year age bin categories.