Online Supplement
The Hidden Costs of Requiring Accounts

Benjamin Mako Hill
makohill@uw.edu

Aaron Shaw
aaronshaw@northwestern.edu

This file is the online supplement that accompanies the paper "The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence from Peer Production" by Benjamin Mako Hill and Aaron Shaw published in *Communication Research*. The full data and code necessary to reproduce the analysis reported in the paper is available in the Harvard Dataverse at (https://doi.org/10.7910/DVN/CLSFKX).

Descriptive Statistics

As mentioned in the Measures section of the paper, the dataset includes 2,171 observations from a total of 136 wikis. To provide a sense of the size and activity of wikis in our sample, we describe the distributions of these wikis' ages and cumulative number of edits, editors, and pages at the point of the intervention in Table 1. Each of the four outcome variables—*new accounts*,

	Min.	Median	Mean	Max.	Std. Dev.
Age (in weeks)	4	142	148	438	93
Total edits	246	7915	46165	1173872	135632
Total editors	9	114	682	42905	3721
Total pages	148	2380	14356	754883	66815

Table 1: Summary statistics describing the range of size and activity-level of communities included in our analysis. Because our analysis is longitudinal and these measures change over time, statistics are reported for each wiki at the end of the week that the community blocked unregistered contributors ($N = 136$).

	Min.	Median	Mean	Max.	Std. Dev.
new editors	0	2	7	435	25
reverted	0	0	19	7389	175
non-reverted	0	98	581	26596	1917
PWR	0	15556	122828	3918103	314955

Table 2: Summary statistics for the four dependent variables used in our analysis across all observations ("wiki weeks") in the dataset we used to fit the models. Each variable describes the amount of activity within one wiki during one week ($N = 2,171$).

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.201^{***}	-1.423^{***}	-0.428^{***}	-0.612^{***}
	(0.058)	(0.167)	(0.093)	(0.119)
Deviance	2050.599	1528.295	2440.530	2480.181
Num. obs.	2075	2075	2075	2075

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Summary of regression models with originally recorded cut-off dates for eight wikis.

reverted, *non-reverted*, and *persistent word revisions (PWR)*—vary within wiki over time. We present summary statistics for these four measures across the full dataset in Table 2.

Adjusted Cut-offs

When Wikia provided us with a list of wikis that required accounts they included only the date of the most recent configuration change. For 8 wikis, a preliminary visual inspection of the data led us to believe that anonymous editing was first blocked earlier than indicated in the information provided by Wikia. Wikia staff confirmed that information on the date of the original change likely had been overwritten by subsequent maintenance work and the actual intervention could have occurred weeks or months earlier than their list indicated. To identify and adjust for these incorrect dates, we reviewed plots of the number of contributions from unregistered editors over time for every wiki in our dataset. We found that the dates shared by Wikia staff corresponded to a complete cessation or discontinuous decline of editing from users without accounts in the vast majority of cases. Figure 1 shows the number of unregistered contributors over the full history of the 8 communities in which that was not the case. The original cut-offs provided by Wikia are shown in blue. Adjusted cut-offs that correspond to the first abrupt cessation

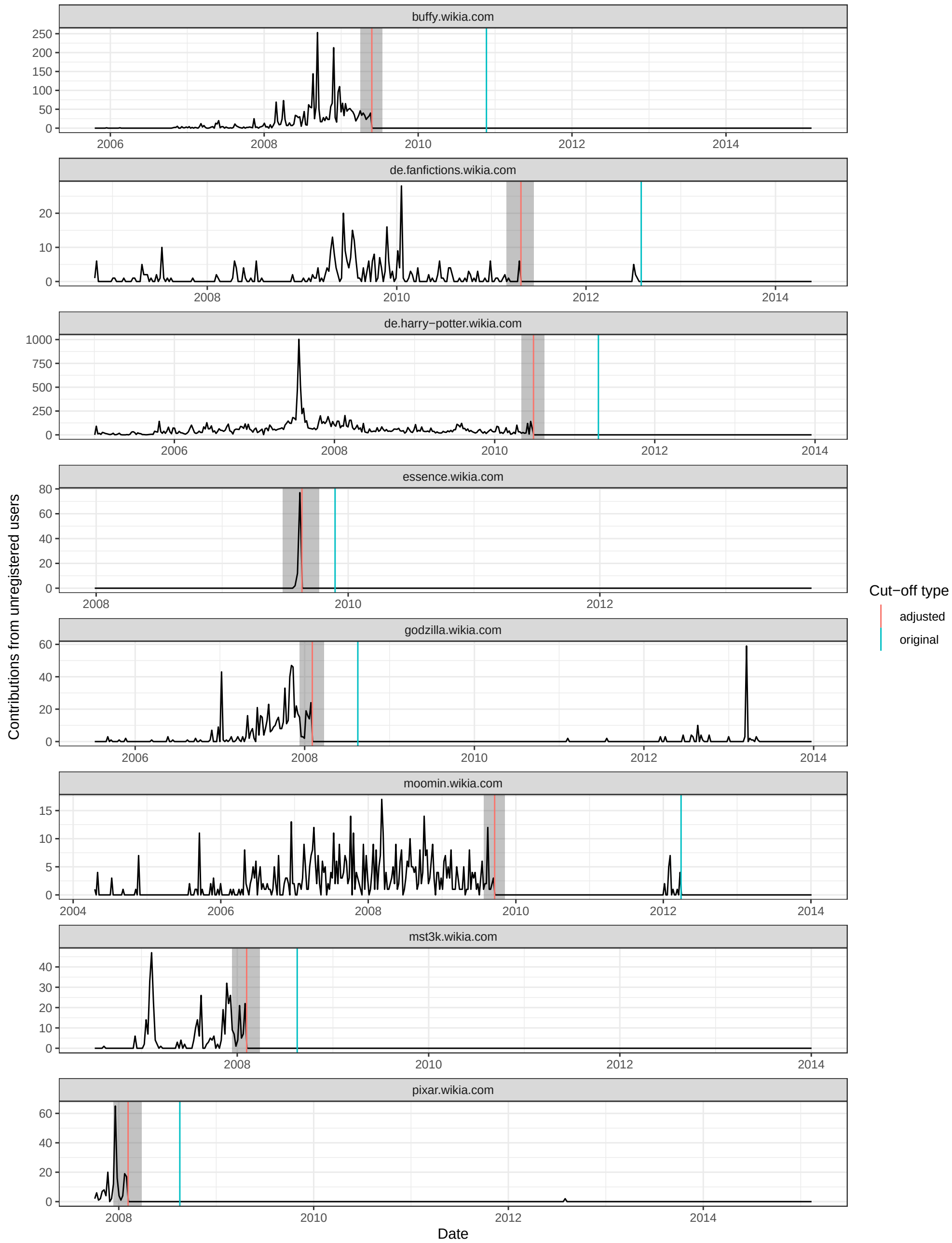

Figure 1: Count of the number of edits from unregistered contributors per week for the 8 communities for which we adjusted the date of the block to an earlier date. The original dates from Wikia are shown in blue and the adjusted dates are shown in red. The shaded region shows the area used in our analysis.

or decrease of unregistered editing are shown in red. In some cases, it seems that unregistered users were unblocked and then blocked again. The main analysis presented in the paper uses the adjusted cut-offs. Table 3 reports the results from the same models using the original, unadjusted cut-off dates. The parameter estimates associated with *acct_req* do not substantively change.

Sample Selection

Wikia provided us with a list of 181 wikis that had blocked unregistered contributions. Our analysis only includes 136 of these wikis.

We excluded 45 wikis from our analysis using the following criteria:

1. Before we collected data, we removed wikis whose database had been deleted by Wikia and was not available ($N = 4$).
2. Before we collected data, we excluded wikis that Wikia's records indicated had never allowed anonymous contributions ($N = 4$).
3. We also excluded wikis where we found evidence that unregistered contributions had never or only very briefly been blocked ($N = 2$). In the cases of these 2 wikis, we saw no evidence that any cut-off had occurred. For one of these wikis, there was a similar number of unregistered contributions before and after the purported cut-off (99% as many contributions from unregistered users) and in the other case there was a large increase (760%). While preparing this manuscript, we were also able to verify that we could contribute to both wikis without an account.
4. Because our analysis seeks to identify the effect of requiring accounts on communities with a history of contributions without accounts, we excluded wikis that were founded less than four weeks before the change was made ($N = 15$).
5. Because wikis with no previous unregistered contributions would not be affected by a requirement for account creation, we excluded wikis for which there were no unregistered edits in the period before the block ($N = 15$).
6. Finally, we dropped wikis that were effectively inactive during the window of analysis—i.e., with no contribution activity at all during at least 70% or more of the 16 week period around the change ($N = 5$).

The first two inclusion criteria are unavoidable. We cannot include wikis for which we cannot collect data and we cannot estimate the effect of the

	Min.	Median	Mean	Max.	Std. Dev.
Age (in weeks)	-8	8	55	362	79
Total edits	58	400	1704	21044	3867
Total editors	4	10	53	1103	184
Total pages	33	282	592	5822	1030

Table 4: Summary statistics describing the range of size and activity-level of communities excluded from our analysis. Because our analysis is longitudinal and these measures change over time, statistics are reported for each wiki at the end of the week that the community blocked unregistered contributors ($N = 37$).

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.184^{**}	-1.425^{***}	-0.504^{***}	-0.734^{***}
	(0.058)	(0.159)	(0.096)	(0.118)
Deviance	2395.437	1671.311	3013.713	2936.121
Num. obs.	2659	2659	2659	2659

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Summary of regression models with excluded wikis included.

	Negative ($\hat{\tau} \leq -10^{-5}$)	Null effect ($10^{-5} > \hat{\tau} \leq 10^{-5}$)	Positive ($\hat{\tau} > 10^{-5}$)
new editors (M1)	73	3	60
reverted (M2)	102	0	33
non-reverted (M3a)	86	0	50
PWR (M3b)	83	1	52

Table 6: Summary of estimates for the $\hat{\tau}$ associated with blocked fit once per wiki in terms of the degree to which the estimates were positive, negative, or very-near zero.

blocking unregistered users on wikis on which unregistered contributions were never allowed. The remaining inclusion criteria can be relaxed.

Summary statistics for the 37 wikis removed using the final four inclusions criteria are shown in Table 4. If compared to Table 1, it is clear that these wikis are much smaller. Since our inclusion criteria included measures of activity, this is not surprising. Table 5 shows models fit on a dataset that includes the wikis excluded from our main analysis. Our results are somewhat moderated but are substantively unchanged.

Wiki-Level Effects

All of the effects reported in our paper are average effects. To characterize the heterogeneity across the wikis in our sample, we fit models where we interact ***wiki*** with our estimate for *acct_req*. This results in an additional 136 wiki-level fixed effects estimated in each model which each capture the effect of the cut-off on a single wiki. Instead of a single $\hat{\tau}$ capturing the average effect, we have 136 separate point estimates for the parameter. This approach involves fitting 136 separate models on each wiki-level dataset of following form: $Y = \tau acct\_req + \beta + \beta week + \beta week^2 + \varepsilon$.

We summarize the results of these models in Table 6 where we describe the number wiki-level estimates for $\hat{\tau}$ that are negative, positive, and substantively equivalent to a null effect. We define null effects as point estimates for $\hat{\tau}$ between -10^{-5} and 10^{-5}. Because each model includes four parameters and at most 16 data points over our 16-week window, the parameters are generally very poorly estimated with standard errors that are very large relative to the parameter estimates.

Our results suggest a high degree of heterogeneity in effects. For M2, M3a, and M3b, nearly half as many wikis may have experienced effects with the opposite sign to the average estimated effect. In terms of M1, *more* wikis may have seen a decrease in new editors at the cutoff than saw a positive effect. This implies that the positive average effect we report in the full models may be driven by large positive effects in a small subset of communities. The fragility of the results for M1 in other robustness checks reported below, especially those where we drop extreme values, provides additional evidence in support of this conclusion.

Due to estimation issues with model M2, we followed a two-step process that used Bayesian models fit in Stan to estimate the negative binomial model's overdispersion parameter θ before fitting our frequentist models. The estimation process is described in the section of this document on "Spurious Effects."

Internal Validity Concerns

Several threats to the internal validity of our findings stem from challenges common to regression discontinuity designs (Hahn et al., 2001) and the related method of interrupted time series analysis (McDowall, 1980). These threats include the potential for non-compliance and crossover, spurious effects, the absence of control cases, the presence of extreme values or influential points, the bandwidth (or window) of data used for estimating effects, and the granularity of the measure of the forcing variable (Imbens & Lemieux, 2008; Jacob et al., 2012; Lee & Lemieux, 2010; Murnane & Willett, 2011). Below, we briefly explain and address each of these potential threats.

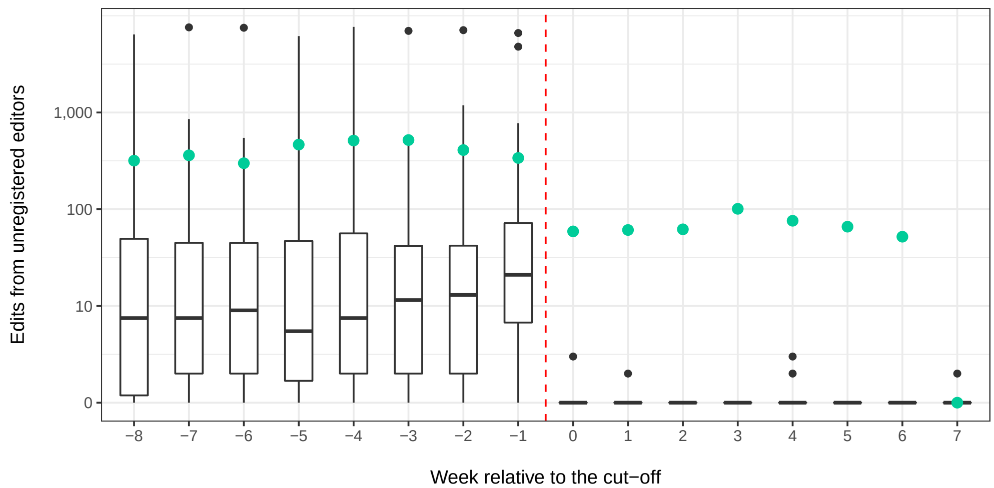

Figure 2: Contributions from unregistered editors per week during the period before and after the cut-off for all the wikis in our analysis. The vertical dashed line indicates the week of cut-off. The points in green are observations from one wiki ("Halo Wiki") that experienced many contributions from unregistered editors after the cut-off.

Compliance in Receiving the Treatment

Figure 2 shows the number of contributions from unregistered users over the 16 week window centered around the time of each community's configuration change. The figure clearly illustrates that the change produced a sudden, discontinuous decline in the number of contributions from unregistered editors in the vast majority of the wikis in our sample. Although there is a range of unregistered contributions before the cut-off, the number of such contributions drops to zero and the entire boxplot collapses onto the X axis afterward. The fact that there not all values on the right side of the figure are equal to `0` suggests that there was least some degree of "non-compliance."

Non-compliance may reflect bugs in the software or resourceful individuals' ability to work around the block. In either case, the intervention in question was intended to eliminate the possibility of unregistered contributions and, in every case, successfully provoked a decrease the number of such contributions.

Our analysis takes non-compliance into account by including the wikis that showed any evidence of undergoing the feature change. As a result, we provide *intent-to-treat* (ITT) estimates (Jacob et al., 2012; Murnane & Willett, 2011). Other estimation procedures, including "fuzzy" RDDs, could identify other estimands, such as the local average treatment effect among compliers. We do not consider such local estimands the primary quantities of theoretical interest here. Because some degree of non-compliance may be a realistic part of any attempt to block contributions from users without accounts in similar settings, we believe our procedure produces valid estimates of the impact of the intervention on the communities that received it. Future analysis might seek to identify alternative estimands and compare them against those we identify in this paper.

The Potential for Crossover

Another threat to RDD designs is "crossover." Because our forcing variable is time, this would be manifested in our case as situations in which subjects have manipulated the timing of the intervention. Two related aspects of the discontinuity we analyze here raise the potential for crossover. First, the implementation of the design change was presumably requested or consented to by at least one administrator within each wiki. Second, that knowledge of the timing or character of the design change may have diffused more widely among editors and would-be editors. In either case, these could lead individuals to alter their behavior in anticipation of the treatment intervention, resulting in biased estimates.

In the body of the paper, we argue that the threat of fore-knowledge of the design change was minimized because the new feature would primarily affect the least experienced and committed members of the community who had not registered accounts or logged into the site. Our conversations with several Wikia wiki administrators and staff led us to conclude that these users likely lack the sort of insider knowledge necessary to learn about the change ahead of time. We substantiated this by conducting a full-text search on a 10% sample of the wikis in the study looking for evidence of discussion about the account requirement prior to each wiki's respective cut-off date. We found no evidence of any such discussion.

Because administrators were much more likely to participate in the decision to implement the feature change, we anticipate that any bias induced by fore-knowledge of the timing and nature of the intervention would bias the estimates through an impact on administrator behavior. To address the potential threat posed by administrator fore-knowledge of the discontinuity, we construct alternate versions of our dataset by dropping all contributions from administrators (before and after the cut-offs) and re-estimating our empirical

	reverted (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	-1.479^{***}	-0.412^{***}	-0.586^{***}
	(0.163)	(0.092)	(0.118)
Deviance	1612.624	2543.870	2599.352
Num. obs.	2171	2171	2171

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: Summary of regression models with contributions from administrators removed from the dataset.

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.011	−0.179	0.097	0.010
	(0.057)	(0.112)	(0.072)	(0.095)
Deviance	1785.323	1250.119	2286.931	2369.242
Num. obs.	2030	2030	2030	2030

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: Summary of regression models estimated around placebo cut-off dates 90 days before the actual cut-off dates for each wiki.

models. We present results of these new models in Table 7.

Dropping administrators from the analysis does not affect the number of new accounts (all administrators must have registered accounts) so the results of Model 1 are unchanged and are not reproduced. For the other three models, the results are substantively similar. In all three cases the effects are slightly larger in magnitude. This suggests that administrator contributions do not drive the direction or magnitude of the impact of the intervention.

Spurious Effects: Placebo Tests

The effects we estimate at the cut-offs may also reflect spurious correlations. This is particular concern in an RDD when one's forcing variable is time. To assess the sensitivity of our results to this threat, we construct "placebo" tests specifying identical models for two sets of fictional cut-off dates, one 90 days prior to the actual cut-off and the second 90 days after the actual cut-off. We report the results of these placebo tests in Table 8 (90 days before) and Table 9 (90 days after). Because some wikis began requiring accounts relatively early or late in their lifetimes, shifting the analytic window in this way results in a slightly smaller number of observations.

None of the models reported in Tables 8 and 9 indicate effects at the placebo cut-off dates. This suggests that the effects we estimate around the true cut-offs reflect variation introduced by the interventions.

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.006	−0.154	0.104	0.026
	(0.057)	(0.113)	(0.072)	(0.094)
Deviance	1831.922	1267.008	2346.899	2427.850
Num. obs.	2143	2143	2143	2143

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 9: Summary of regression models estimated around placebo cut-off dates 90 days after the actual cut-off dates for each wiki.

Estimating model M2 using R's `glm.nb()` function in the MASS package introduced several estimation challenges. This is not surprising for a number of reasons including the large number of zeros in the data and the fact that, as placebo tests, these models are an intentionally bad fit for our data.

To estimate M2, we first fit Bayesian versions of the models using the Stan statistical programming language and the `rstanarm` interface to Stan for R using the default/recommended settings and priors.[1] Because our analysis is not Bayesian, we then then fit identical frequentist maximum likelihood models using R's `glm()` function using means of the Stan posterior distributions as start values. In order to estimate these placebo models for M2, we also fixed the overdispersion parameter θ used in the negative binomial model as equal to the mean of the posterior for θ in the Bayesian model fit with Stan. In the case of the second placebo for M2 (90 days after), `glm()` issued a warning about possible non-convergence.

To ensure that the reported results are valid, we inspected the parameter estimates for the β associated with *acct_req* and found that the mean of the estimated posterior distribution from the Stan models was extremely similar to the point estimates from `glm()`. To double-check that the reported estimates were correct, we also estimated models using Stata and found point very similar estimates.

Shorter and Longer Analytic Windows

Another threat pertains to the number of observations before and after the interventions. While we draw our inference at the point of the intervention, the possibility remains that some portion of our estimates could be due to underlying trends in the data rather than true discontinuities.

In RDD studies, this is often discussed in terms of the "bandwidth" of

[1] https://mc-stan.org/ https://cran.r-project.org/web/packages/rstanarm/index.html

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.099	-1.597^{***}	-0.366^{***}	-0.643^{***}
	(0.071)	(0.189)	(0.097)	(0.124)
Deviance	835.858	1098.371	1213.625	1236.564
Num. obs.	1088	1088	1088	1088

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 10: Summary of regression models estimated on a shorter (4 weeks pre and post cut-off) analytic window.

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.227^{***}	-1.568^{***}	-0.358^{***}	-0.544^{***}
	(0.052)	(0.155)	(0.082)	(0.118)
Deviance	3250.060	2159.168	3802.431	3917.286
Num. obs.	3238	3238	3238	3238

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 11: Summary of regression models estimated on a longer (12 weeks pre and post cut-off) analytic window.

the analysis and researchers have presented multiple techniques for evaluating the sensitivity of the design to different bandwidths (Jacob et al., 2012). We adopt the most suitable such approach in the context of our data and perform a sensitivity analysis by re-fitting our models for analytic windows of 8 and 24 weeks. Table 10 (8-week window) and Table 11 (24-week window) show the results of these alternate bandwidth specifications.

Table 10 indicates that reducing the analytic window to 4 weeks before and after the intervention attenuates the effect on new editors per week (M1) and renders this effect close to zero and not statistically significant. The effects on reverted edits (M2), non-reverted edits (M3a), and PWR (M3b) remain very similar to those estimated across the original analytic window. Table 11 expands the analytic window to 12 weeks before and after the intervention and shows effects estimates almost identical to those in the original models in the paper.

Due to estimation issues with model M2, we followed a two-step process that used Bayesian models fit in Stan to estimate the negative binomial model's overdispersion parameter θ before fitting our frequentist models. The estimation process is described in the section of this document on "Spurious Effects."

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.199^{***}	-1.496^{***}	-0.398^{***}	-0.558^{***}
	(0.054)	(0.158)	(0.085)	(0.114)
Deviance	3561.096	2289.062	4493.600	4543.903
Num. obs.	3796	3796	3796	3796

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 12: Summary of regression models estimated using data collapsed into 4-day bins instead of 7-day bins.

	new editors (M1)	*reverted* (M2)	*non-reverted* (M3a)	*PWR* (M3b)
acct_req	0.203^{***}	-1.614^{***}	-0.365^{***}	-0.530^{***}
	(0.058)	(0.159)	(0.091)	(0.118)
Deviance	1784.945	1809.577	2217.716	2250.951
Num. obs.	1897	1897	1897	1897

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 13: Summary of regression models estimated using data collapsed into 9-day bins instead of 7-day bins.

Smaller and Larger Time Windows for Binning Data

The size of the week-long time windows into which we bin the observations pose a related threat. Bins that are too large might over-smooth the model fit in the region immediately around the discontinuity. Likewise, bins that are too small could produce more volatility in our measures that would undermine the assumption of local linearity around the discontinuity (Jacob et al., 2012; Lee & Lemieux, 2010). RDD studies confronting similar issues have used several corresponding sensitivity analysis techniques, including the one we adopt below: re-estimation with alternate bin-widths.

To evaluate the sensitivity of our estimates to the bin-width, we re-estimate our models with data binned into 4-and 9-day periods. The results for 4-day periods appear in Table 12 and the results for 9-day periods in Table 13. Table 12 shows that smaller, 4-day bins attenuate the effect of the intervention on new editors (M1), accentuate the effect on reverted edits (M2), and have little impact on the estimates of the effect on quality contributions. Longer, 9-day bins show a somewhat attenuated effect for our reverted edits model (M2) but otherwise very similar pattern of results to those reported in the paper.

Due to estimation issues with model M2, we followed a two-step process that used Bayesian models fit in Stan to estimate the negative binomial model's overdispersion parameter θ before fitting our frequentist models. The estimation process is described in the section of this document on "Spurious Effects."

Absence of Case Controls

The within-subjects design of this study also introduces concerns due to the absence of true control cases. Because we draw inference from the pre- and post-intervention data within the same wikis, we assume the absence or irrelevance of other factors and events that may have coincided temporally with the administration of the interventions across the wikis in our population. Classic RDD, interrupted time series (ITS), and differences-in-differences (DiD) designs each offer solutions to this concern.

In classic regression discontinuity designs, the discontinuity occurs at an arbitrary (or "as-if random") point along a forcing variable. On one side of the discontinuity, cases remain untreated, whereas separate cases on the other side receive treatment. In the context of both ITS and DiD studies, the analysis of "case controls" that do not experience an interruption serves an analogous purpose, providing a counterfactual baseline of comparison if the intervention in question occurred as a result of a random, or as-if random, assignment procedure. Could data from other, untreated wikis provide a population of control cases against which we could compare the effects of the intervention on the treated cases? We do not lack for data on other wikis. So why not pursue these strategies?

We do not pursue these strategies because we believe that the nature of the intervention and the possibly endogenous circumstances of its administration undermine the "selection on observables" assumption necessary for matching. As we discussed earlier, the decision to implement the requirement that would-be editors of these wikis log in to accounts did not occur randomly. Instead, the intervention resulted from consent on the part of, or a request from, at least one of each community's administrators. The interventions sometimes followed on the heels of specific events (e.g., an influx of vandalism or spam) that likely reflect endogenous or unobservable factors that differentiate these wikis from others. This challenges the assumptions required for many forms of causal identification but makes matching on unobservables particularly unappealing because we believe that superficially similar wikis are likely to not be good controls cases that are otherwise equal in expectation.

Our panel RDD approach makes other assumptions and suffers from its own threats to validity which we have described in depth in the "Threats to Validity" section of the manuscript. We are comfortable with an approach that forgoes case controls for several reasons. First, the sheer size and diversity of the population supports the idea that the variations we observe reflect the average effects of implementing the intervention rather than some single time-dependent or community-specific factor. Second, as described in robustness checked reported in this supplement, we find no evidence that administrator activity drives or alters the findings (see our section on "The Potential for Crossover" above). Third, we also find almost no evidence of discussion of the intervention before it was implemented and the vast majority of the communities' would-be contributors most impacted by the intervention appear not to have known about the timing or character of the design change in advance. Each of these factors strengthen the credibility of inference about the effect of the intervention on contributions from unregistered would-be contributors in the weeks immediately around the cut-off.

A downside of our approach is that our estimates may not generalize to arbitrary peer production communities at arbitrary points in time. However, because peer production communities do not implement barriers to entry at arbitrary points in time, we feel that such generalization is unnecessary.

Influential Cases

A final threat concerns the potential for disproportionate influence on our estimates from extreme values of our outcomes. This threat is particularly relevant given the highly skewed distributions we observe across different wikis for our dependent variables. Since this threat stems from large wikis at the upper end of these distributions, we address this by re-estimating each of our models and successively dropping all observations from the wikis with the largest 1%, 5%, and 10% average values for each dependent variable within the analytic window.

We present a model fit using the full dataset alongside the same specification for each of the restricted samples in order to facilitate comparison. We report these results in Tables 14–17, where each table presents the estimates for a single DV using a progressively more restricted sample. Overall, we find that the estimates on new editor accounts (M1) and (to a much lesser extent) reverted edits (M2) are sensitive to removing extremely active wikis.

Table 14 shows results for M1 and suggests the estimated magnitude of our effect for is positive across all restricted models but decreases in magnitude substantially as data from the largest wikis are removed. The effects are not statistically significant in the models with 5% and 10% of the largest wikis removed. Although the decreasing effect suggest provides some evidence that the effect of new editors may be located in the largest wikis, we know of no

	Full sample	Bottom 99%	Bottom 95%	Bottom 90%
acct_req	0.201***	0.183**	0.123	0.089
	(0.057)	(0.063)	(0.072)	(0.084)
Deviance	2157.498	2082.805	1982.336	1858.206
Num. obs.	2171	2139	2059	1947

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 14: Summary of regression models estimating the effect of the cut-off on the number of new editor accounts (M1) dropping the wikis with extreme values of new editors (M1).

	Full sample	Bottom 99%	Bottom 95%	Bottom 90%
acct_req	−1.479***	−1.476***	−1.769***	−1.748***
	(0.163)	(0.164)	(0.169)	(0.188)
Deviance	1612.625	1575.224	1012.949	1336.300
Num. obs.	2171	2139	2059	1947

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 15: Summary of regression models estimating the effect of the cut-off on the number of reverted edits (M2) dropping the wikis with extreme values of reverted edits.

	Full sample	Bottom 99%	Bottom 95%	Bottom 90%
acct_req	−0.412***	−0.411***	−0.438***	−0.458***
	(0.092)	(0.093)	(0.097)	(0.103)
Deviance	2543.870	2505.104	2408.806	2274.242
Num. obs.	2171	2139	2059	1947

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 16: Summary of regression models estimating the effect of the cut-off on the number of non-reverted edits (M3a) dropping the wikis with extreme values of non-reverted edits.

	Full sample	Bottom 99%	Bottom 95%	Bottom 90%
acct_req	−0.586***	−0.609***	−0.639***	−0.665***
	(0.118)	(0.118)	(0.123)	(0.132)
Deviance	2599.352	2561.360	2465.367	2329.600
Num. obs.	2171	2139	2059	1947

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 17: Summary of regression models estimating the effect of the cut-off on the number of persistent word revisions or PWR (M3b) dropping the wikis with extreme values of PWR.

	reverted>0 (M2)
acct_req	-2.777^{***}
	(0.451)
Deviance	913.402
Num. obs.	2171

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 18: Logistic regression model estimating the probability that a wiki will experience at least one reverted edit in a given wiki.

theoretical reason to anticipate this. We believe that the fragility of the estimate is an artifact of fact that the effect on new editors is relatively small and noisy.

Estimates for M2, M3a and M3a are robust to the removal of extreme values and even increase in magnitude in the restricted samples. Table 15 showing reverted edits, Table 16 showing non-reverted edits, and Table 17 showing PWR all show effect sizes that are stable or even larger when the largest values of the dependent variables are removed and standard errors associated with those estimates that are only modestly larger.

Logistic Specification for Reverts

Previous research by TeBlunthuis et al. (2018) has shown that reverts in Wikia wikis are relatively rare. As we explain in our paper, a large number of the wiki week periods in our data contain no reverts at all and 10 wikis in our sample experience no reverts at all during our analytic window. The fact that much of our variation in reverted edits is between 0 and 1 raises a concern about the appropriateness of the negative binomial specification of our models of *reverted*. To address this threat, we fit models using an alternative specification of Model M2 where we estimated a logistic regression model where the dependent variable is a dichotomous variable set to 1 if *reverted* > 0 and 0 otherwise.

The results of this alternative specification are shown in Table 18. The parameter estimate for *acc_req* ($\beta = -2.78$) suggests that after the block, the odds of a wiki having at least one reverted it are 6% the odds of experiencing at least one reverted edit before. This estimates are consistent with our pattern of effects reported above.

References

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209. https://doi.org/10.1111/1468-0262.00183

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Jacob, R. T., Zhu, P., Somers, M.-A., & Bloom, H. (2012). A practical guide to regression discontinuity. *MDRC Working Papers on Research Methodology*. http://www.mdrc.org/practical-guide-regression-discontinuity

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

McDowall, D. (1980). *Interrupted time series analysis*. Beverly Hills, CA, Sage Publications.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY, Oxford University Press.

TeBlunthuis, N., Shaw, A., & Hill, B. M. (2018). Revisiting "The rise and decline" in a population of peer production projects, In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, New York, NY, ACM. https://doi.org/10.1145/3173574.3173929