

Structural factors rather than reconstruction investment drive depopulation after Japan's 2011 disasters

Corresponding Author: Dr Toshiaki Yoshida

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This manuscript discusses a mixed-methods approach to evaluate post-disaster migration patterns across cities impacted by the GEJET and to examine their correlations with disaster-induced damage and Japanese government investments. While I believe this study is relevant and will appeal to the journal's audience, a few points need strengthening before it is considered for publication.

Main points:

The rationale for the research design is not always clear: the authors employ multiple methods, but it is unclear whether these were part of the design or added later. I believe this is important because if they were part of the original design, it strengthens the argument that the multi-method approach highlighted similar results. Conversely, if the research design were dynamic, it would be important to discuss whether other methods were considered, the results obtained, and why they were not reported.

Specific comments:

General: the authors consistently use 'out-migrate in [town]'. Please double-check this construct from a grammar perspective. The use of opposite prepositions 'out' and 'in' can be hard to interpret. I believe a simpler construct such as 'migrated out of [town]' or 'migrated to [town]' would make your manuscript much simpler to interpret.

Line 173: The authors argue that "[w]hen government officials invest the same amount in two towns with similar levels of damage and population, it is plausible to assume that both towns would have parallel migration trends". Please provide more evidence to justify the plausibility of this statement. A rich body of disaster research indicates that disasters exacerbate pre-existing demographic trends. The authors acknowledge this in the abstract. Hence, I believe that many disaster researchers would not assume 'parallel migration trends' after a disaster if these trends were not already parallel before the disaster.

Line 291: Is the framework by Findlay (2011) something adopted by most researchers in this field? If so, please provide authors unfamiliar with it with a brief description.

Section 3.2: While the authors describe the process for selecting cities for the model, it would be valuable to know how sensitive the model is to the selection criteria. Please indicate whether other configurations were tested and whether the results are consistent.

Section 3.2: I'd appreciate a more in-depth discussion of the results in Fig. 1 to evaluate its contribution to the narrative. It seems that synthetic models yield results that essentially propagate pre-disaster trends forward. Thus, comparing the empirical and model results is equivalent to comparing empirical post-disaster and pre-disaster data. As is, it is unclear whether the methods in Section 3.2 provide *new* insights not already presented.

Line 339: [suggestion] The paragraph starting in this line presents a lot of information/comparison in a short span of sentences and without a clear structure. It is hard to read. I suggest splitting this paragraph into small sections, focusing on one region at a time.

Line 398: please define 'neighbor' in this case. Are neighbours defined by shared boundary or other definition-based spatial statistics?

Line 404: How does the different treatment for the most populated cities in an area affect the model?

Line 429: I am surprised by the use of investment per capita. It is not trivial to determine what this metric assesses, because you may have large areas with large investments as well as small areas with small populations, resulting in a small RIPC. It becomes a compound metric of population and losses, and it is unclear how correlated it is with other metrics in your regression model, which would be a problem.

Table 4: Please assess the independence of your regressors. $\ln(\text{HDR}) * \ln(\text{NNPD})$ is $\ln(\text{HDR} + \text{NNPD})$. A model containing $\ln(\text{HDR})$, $\ln(\text{NNPD})$ and $\ln(\text{HDR} + \text{NNPD})$, especially if HDR and NNPD are not normalized, can suffer from heteroskedasticity.

Reviewer #3

(Remarks to the Author)
Peer Review Results

In this paper, it is stated that the population size of the town before the Great East Japan Earthquake had an impact on population migration after the Earthquake.
The author tries to state that the impact of population size on population migration after the earthquake was greater than that of policy interventions related to disaster recovery, that is, public infrastructure investment and housing reconstruction subsidies for reconstruction.
What this author is trying to argue can be understood from the reality of reconstruction in the disaster-stricken areas.

However, the analytical problems used for this argument must be pointed out.
I will add a critical consideration of the multiple regression analysis in Chapter 4.

1) Time lag between infrastructure investment and effect realization
In particular, the effect of investment in public infrastructure needs to consider the time span from investment consensus building, planning, implementation design, construction, and service, as well as the restoration of urban functions due to the reconstruction of enterprises, commercial facilities, schools, hospitals, etc.
It is thought that the timing of the completion of public infrastructure reconstruction has a significant impact on population migration after the earthquake, not only the amount of investment in public infrastructure.

2) Inconsistency of input independent variables
Table 4 shows that the RIPC coefficient values are very small at -0.03, 0.03, and 0.01.
In other words, the effect on NETMGR1222 is very small and of course not significant.
The relationship between NETMGR1222 and RIPC is considered to be small.

In addition to Fig. 4, it is desirable to show correlation plots plotting the relationship between NETMGR1222 and RIPC, NETMGR1222 and HDR, and NETMGR1222 and NNPD.
Without using multiple regression analysis, it may be possible to clarify that there is no relationship between NETMGR1222 and RIPC or HDR.

In other words, in Table 4, it is thought that there is a problem in the selection of independent variables that were input before determining statistical significance in the application of multiple regression models.

3) Problems with HDR calculation methods
In this paper, the house destroyed rate (HDR) is calculated for each local government. The number of affected houses is divided by the number of households in the municipality and calculated as a percentile value.
In municipalities with a small number of households, this proportion increases when many of them are affected.
On the other hand, in municipalities with a large number of households, this proportion is small even if they suffer partial catastrophic damage.
When the number of affected houses is the same, it is an indicator that it is small in a municipality with a large number of households and large in a small municipality.
It is questionable whether it can be said to be an indicator of the magnitude of the initial damage.

From the above, it is pointed out that this paper contains several problems in acceptance.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have satisfactorily addressed all my comments.

Reviewer #3

(Remarks to the Author)

In response to my three comments, the author gave me a polite response.

I also checked the revisions to the paper.

In response to the analysis issues I pointed out, the author added scatter plots, added tables, redone statistical calculations, added analysis, and corrected sentences.

The results of statistical analysis of the author's hypothesis and its consideration can be evaluated as a complete achievement.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We would like to thank you for your time to provide a review on our manuscript. The reviews helped us to refine the paper.

Reviewer 1: The rationale for the research design is not always clear: the authors employ multiple methods, but it is unclear whether these were part of the design or added later. I believe this is important because if they were part of the original design, it strengthens the argument that the multi-method approach highlighted similar results. Conversely, if the research design were dynamic, it would be important to discuss whether other methods were considered, the results obtained, and why they were not reported.	Updated: For the most part, we conducted our research as planned. We carried out fieldwork, case studies, OLS analysis, and GIS visualization. We also considered synthetic control methods (and panel data analysis) from the beginning, but as supplementary analyses. We assumed that reconstruction investments do not affect migration immediately or on a year-by-year basis. Rather, they influence migration cumulatively, as the overall competition of projects progresses. We expected that OLS and case studies would support much of our argument, but we also collected data for synthetic control and panel analyses as robustness checks.
Reviewer 1: The authors consistently use ‘out-migrate in [town]’. Please double-check this construct from a grammar perspective. The use of opposite prepositions ‘out’ and ‘in’ can be hard to interpret. I believe a simpler construct such as ‘migrated out of [town]’ or ‘migrated to [town]’ would make your manuscript much simpler to interpret.	Thank you for this suggestion. We have updated the text accordingly. We agree that the suggested change improves the readability of the manuscript. We retained the expression of out-migration in two instances: (1) in the Introduction, to keep the literature review short and concise; and (2) in Section 3.3 (Migration Destinations), where we need to explain excess out-migration—for example, from Minamisanriku Town to Tome City (Ryūshutsu chōka).
Reviewer 1: Line 173: The authors argue that “[w]hen government officials invest the same amount in two towns with similar levels of damage and population, it is plausible to assume that both towns would have parallel migration trends”. Please provide more evidence to justify the plausibility of this statement. A rich body of disaster research indicates that disasters exacerbate pre-existing	Deleted and reorganized. The sentence was unnecessary and added confusion.

demographic trends. The authors acknowledge this in the abstract. Hence, I believe that many disaster researchers would not assume ‘parallel migration trends’ after a disaster if these trends were not already parallel before the disaster.	
Reviewer 1: Line 291: Is the framework by Findlay (2011) something adopted by most researchers in this field? If so, please provide authors unfamiliar with it with a brief description.	Updated in the text. Among the six principles that govern migration destinations, three are relevant to domestic and environmental migration; these have been incorporated into the text. The remaining three pertain to broader, international migration and are not included, given the scope of this paper.
Reviewer 1: Section 3.2: While the authors describe the process for selecting cities for the model, it would be valuable to know how sensitive the model is to the selection criteria. Please indicate whether other configurations were tested and whether the results are consistent.	Updated: We explored an alternative specification of the model by dropping one of the criteria, such as population, and adding another criterion—the ratio of commuters to the working population. The results of the synthetic control did not change.
Reviewer 1:Section 3.2: I’d appreciate a more in-depth discussion of the results in Fig. 1 to evaluate its contribution to the narrative. It seems that synthetic models yield results that essentially propagate pre-disaster trends forward. Thus, comparing the empirical and model results is equivalent to comparing empirical post-disaster and pre-disaster data. As is, it is unclear whether the methods in Section 3.2 provide *new* insights not already presented.	Updated: Thank you for the comment. In general, and particularly for our article, synthetic control methods provide a more “accurate” narrative rather than a “new” one by adding the power of causal inference. Specifically, the method helps pinpoint when the effects of the intervention begin to manifest. In our case, it is clear that the Tohoku disaster marked the onset of depopulation. The value of using this method in our article, therefore, lies in allowing us to more precisely trace differences in negative net migration trends between Minamisanriku and Otsuchi, beyond what can be observed from empirical or descriptive trends alone. The method also enables us to quantify the extent of depopulation—for example, in percentage

	terms—and to show how depopulation persisted in Minamisanriku. If this section does not add sufficient new insights and instead complicates the narrative (or conflicts with word limits), we are willing to move it to the appendix.
Reviewer 1: Line 339: [suggestion] The paragraph starting in this line presents a lot of information/comparison in a short span of sentences and without a clear structure. It is hard to read. I suggest splitting this paragraph into small sections, focusing on one region at a time.	We appreciate this suggestion. We reorganized it to improve readability.
Reviewer 1: Line 398: please define 'neighbor' in this case. Are neighbours defined by shared boundary or other definition-based spatial statistics?	Updated.
Reviewer 1. Line 404: How does the different treatment for the most populated cities in an area affect the model?	Thank you for this comment, which helped clarify our thinking. We have updated the identification protocols and the text accordingly. Briefly, our hypothesis is that the directional force of migration is driven by regional population centers. If a town in the dataset is not a population center, we select the most populated city within a 30-minute range (by driving or public transportation) from that town to calculate the population difference. If a town in the dataset is itself the largest population center in the area, we select the second most populated city within a 30-minute range to calculate the population difference. We believe that, rather than simple geographic distance, travel time—and the ease of commuting back and forth to one's hometown—better governs migration decisions.

	We have also renamed the variable PAN (Population Advantage of the Neighbor). We believe that this refined approach captures the concepts of “short distance” and “intervening opportunity” summarized by Findlay (2011).
Reviewer 1: Line 429: I am surprised by the use of investment per capita. It is not trivial to determine what this metric assesses, because you may have large areas with large investments as well as small areas with small populations, resulting in a small RIPC. It becomes a compound metric of population and losses, and it is unclear how correlated it is with other metrics in your regression model, which would be a problem. Reviewer 3: 2) Inconsistency of input independent variables. Table 4 shows that the RIPC coefficient values are very small at -0.03, 0.03, and 0.01. In other words, the effect on NETMGR1222 is very small and of course not significant. The relationship between NETMGR1222 and RIPC is considered to be small.	The independence of RIPC from other regressors is confirmed in the revision responding to the next comment. Additionally, we conducted two robustness checks. First, instead of using RIPC, we used the total amount of reconstruction investment (RIT) allocated to each municipality. The total investment shows a small and insignificant effect on net migration. Second, we included the logged population in 2011 to further control for population size. In both cases, the results remain largely the same. We have added the results of the first robustness check to the appendix.
Reviewer 1: Table 4: Please assess the independence of your regressors. A model containing $\ln(\text{HDR})$, $\ln(\text{NNPD})$ and $\ln(\text{HDR} + \text{NNPD})$, especially if HDR and NNPD are not normalized, can suffer from heteroskedasticity. $\ln(\text{HDR}) * \ln(\text{NNPD})$ is $\ln(\text{HDR} + \text{NNPD})$.	Updated with an assessment of the independence of regressors using correlation plots and mean-centering. Mean-centering reduces the inflation of VIF caused by the inclusion of interaction terms. Even after log-transforming the variables to approximate normal distributions, models analyzing population dynamics (partly due to the log transformation) typically exhibit heteroskedasticity. To address this issue, we revised our models to use robust standard errors instead of conventional standard errors. We also clarify that the interaction term in

	the model is specified as the product of the log-transformed variables, $\ln(\text{HDR}) \times \ln(\text{NNPD})$, rather than $\ln(\text{HDR} + \text{NNPD})$.
Reviewer 3: 1) Time lag between infrastructure investment and effect realization In particular, the effect of investment in public infrastructure needs to consider the time span from investment consensus building, planning, implementation design, construction, and service, as well as the restoration of urban functions due to the reconstruction of enterprises, commercial facilities, schools, hospitals, etc. It is thought that the timing of the completion of public infrastructure reconstruction has a significant impact on population migration after the earthquake, not only the amount of investment in public infrastructure.	Thank you for this important comment. We additionally ran a series of panel data analyses with one-, two-, three-, and four-year lagged reconstruction investment per capita variables. As the reviewer indicated, the model shows a weak but statistically significant and positive impact on net migration. This result suggests that reconstruction investment can bring people (back) in. Existing studies on the “reconstruction paradox” argue that such investments exacerbate depopulation. Our model offers a corrective perspective: while infrastructure investment has a positive effect on net migration, the effect size is smaller than the negative impact generated by disaster damage. As a result, prior studies may have observed what appears to be a paradoxical effect of reconstruction. Our approach relies on long-term averages to capture these cumulative and delayed effects. Therefore, we briefly mention the panel data results in the Results section.
Reviewer 3: In addition to Fig. 4, it is desirable to show correlation plots plotting the relationship between NETMGR1222 and RIPC, NETMGR1222 and HDR, and NETMGR1222 and NNPD. Without using multiple regression analysis, it may be possible to clarify that there is no relationship between NETMGR1222 and RIPC or HDR. In other words, in Table 4, it is thought that there is a problem in the selection of independent variables that were input	We have added the following scatterplots as online appendix materials: 1. NNETMGR0010 and NNETMGR1222, with points scaled by HDR 2. NNETMGR0010 and NNETMGR1222, with points scaled by PAN (revised NNPD) 3. HDR and RIPC, with points scaled by NNETMGR1222 Figure 4, along with Figures 1 and 2 in the appendix (NNETMGR_0010 and

before determining statistical significance in the application of multiple regression models.	NNETMGR_1222, with points scaled by RIPC, HDR, and PAN), collectively shows that PAN, HDR, and (although rejected in the OLS models) RIPC alone do not explain the larger net migration observed 11 years after the disaster. These figures provide additional support for our analysis. Only towns with large point sizes in both HDR and PAN exhibit faster depopulation. Appendix Figure 3 shows that HDR and RIPC are not correlated to the extent that would raise concerns about multicollinearity or violations of the independence of independent variables. Additional checks for the independence of regressors are also included in the appendix.
3) Problems with HDR calculation methods: In this paper, the house destroyed rate (HDR) is calculated for each local government. The number of affected houses is divided by the number of households in the municipality and calculated as a percentile value. In municipalities with a small number of households, this proportion increases when many of them are affected. On the other hand, in municipalities with a large number of households, this proportion is small even if they suffer partial catastrophic damage. When the number of affected houses is the same, it is an indicator that it is small in a municipality with a large number of households and large in a small municipality. It is questionable whether it can be said to be an indicator of the magnitude of the initial damage.	We agree that the house destruction rate (HDR) is intended to capture relative damage to each town rather than absolute damage or magnitude. Using a variable such as the total number of damaged houses raises a different concern: larger towns may report more damaged houses simply because they have more households, even if the relative impact is limited. To address this, we conducted an additional robustness check using the total number of damaged houses instead of HDR. The results are not statistically significant, indicating that absolute damage measures are less informative in this context. We have also renamed the variable to House Destruction Rate (from Destroyed Rate) for consistency.