CORRESPONDENCE

Historical changes in baby names in China [version 1; peer review: awaiting peer review]

Yuji Ogihara

Tokyo University of Science, Tokyo, Japan

Abstract
Based on previous research on names and naming practices, I propose three suggestions to Bao et al. (2021), which investigated historical changes in given names of Han Chinese in China between 1920 and 2005. Their study analyzed a one-shot cross-sectional survey conducted in 2005 and reported that unique names increased from 1920 to 2005. The authors concluded that China became more individualistic over time for the period. However, three questions have remained unanswered in Bao et al. (2021). First, were the samples of older birth cohorts truly representative? Second, did unique names increase only after the 1970s? Third, how are the historical changes in average name length interpreted? Answering these three questions would contribute to a further understanding of the historical changes in given names and their underlying psychological/cultural shifts in China.

Keywords
name, uniqueness, historical change, cultural change, individualism, China, need for uniqueness, culture

Corresponding author: Yuji Ogihara (yogihara@rs.tus.ac.jp)

Author roles: Ogihara Y: Conceptualization, Investigation, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2023 Ogihara Y. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ogihara Y. Historical changes in baby names in China [version 1; peer review: awaiting peer review] F1000Research 2023, 12:601 https://doi.org/10.12688/f1000research.131990.1

First published: 05 Jun 2023, 12:601 https://doi.org/10.12688/f1000research.131990.1
I have conducted research on unique names (e.g., Ogihara, 2015, 2021a, 2021b, 2021c, 2023b; Ogihara et al., 2015; Ogihara & Ito, 2022) and related cultural changes (e.g., Ogihara et al., 2016; Ogihara, 2018b, 2023a; for reviews, see Ogihara, 2017, 2018a). Based on this previous research, I suggested three recommendations (Ogihara, 2020) to a prior study on historical changes in baby names in China (Study 2 in Cai et al., 2018). The authors answered some of my comments with an empirical investigation on a new dataset (Bao et al., 2021). I respond to it by focusing on three major points.

Specifically, Bao et al. (2021) examined historical changes in given names of Han Chinese in China between 1920 and 2005 by analyzing a cross-sectional survey. This article provides rich information about historical changes in names in China and their underlying psychological/cultural shifts.

1. Were the samples of older birth cohorts truly representative?
The authors used a random subset of a one-shot cross-sectional survey conducted in 2005 (the 2005 China’s 1% Population Census) and analyzed given names of people born between 1920 and 2005. They emphasized that the sample is representative (e.g., “Using a large representative sample of Chinese names” in Abstract. “We used an unprecedentedly large representative sample of Chinese names, covering a longer period of time from 1920 to 2005” (p. 4) in Discussion, “To obtain a nationally representative sample of Chinese names covering a long period” (p. 2) in Method).

However, the data is from a one-shot cross-sectional survey, not a cross-temporal survey (e.g., birth records). The authors investigated names of Chinese people aged from 0 (newborns) to 85 years. This indicates a possibility that the samples for some populations, especially older birth cohorts, may not be nationally representative (not including all the names given in a year in China). Considering that the average life expectancy in China in 2005 was approximately 73 years (72.99; United Nations, 2022), especially the data for older people would be systematically selected by death, yielding the selection effect. For example, economically wealthy people would live longer (despite diseases and aging, e.g., Wilkinson & Marmot, 2003; Jagger et al., 2008), and physically healthy people would be better suited to survive natural disasters at a higher rate, leading to the possibility that economically not wealthy and physically not healthy older people were underrepresented in the samples. In other words, although a subset of the 2005 China’s 1% Population Census would represent people who lived in 2005, older birth cohorts would not be representative, implying that the results for older years might not reflect the reality. To avoid these systematic biases, previous research examining historical changes in baby names analyzed cross-temporal data. Prior research in China (e.g., Cai et al., 2018), Japan (e.g., Ogihara, 2021a, 2022a; Ogihara et al., 2015; Ogihara & Ito, 2022), the United States (e.g., Ogihara, 2021d; Twenge et al., 2010, 2016), the United Kingdom (e.g., Bush, 2020; Bush et al., 2018), Germany (e.g., Gerhards & Hackenbroch, 2000), and France (e.g., Mignot, 2022) has used a series of yearly cross-temporal data of newborn baby names.

It would be necessary for the authors to clarify how they overcame these possible biases. The authors already stated that “because the sample sizes for birth years < 1920 were not sufficient, we limited the range of birth years to 1920–2005” (p. 2), but this issue is related to not only sample size but also sample characteristics (selection bias). The sample sizes for the earlier periods between 1920 and 2005 would not be sufficient to claim that the samples are representative, and the samples would be systematically selected and biased. Cross-sectional data should be carefully investigated to discuss cross-temporal changes (e.g., Cai et al., 2018; Ogihara, 2022b; Ogihara & Kusumi, 2020; Twenge, 2011; Twenge & Campbell, 2001).1

2. Did unique names increase only after the 1970s?
The authors concluded that unique names increased in China between 1920 and 2005 and claimed that they replicated their previous study, which insists on an increase in unique names between 1950 and 2009 (Cai et al., 2018).

However, all six indicators the authors analyzed consistently showed that the unique names did not increase from 1920 to 1969. Rather, the indicator of name-character uniqueness, which the authors “preferred” (p. 6) most and stated “the estimation would be more accurate” (p. 6) among all six indicators, shows a gradual decrease in uniqueness from 1920 to 1969 (Figure 2B in Bao et al., 2021). These results were inconsistent with their previous finding that insists on a continuous increase in unique names from 1950 to 2009 (Cai et al., 2018). The authors did not mention this point clearly.2

---

1 Furthermore, because some people changed their given names, their names in 2005 are different from the names given at births. This possibility should also be considered when cross-sectional data are used to examine cross-temporal changes in names.

2 Cai et al. (2018) also showed this possibility (unique names increased only after 1970). The average name character frequency per birth cohort of 1970-1979 was higher than that of 1960-1969 (Figure 2 in Cai et al., 2018). However, this might be solely due to small sample sizes (Ogihara, 2020).
The study would be improved if the authors made efforts to explain why unique names did not increase between 1920 and 1969 and why the study did not replicate the previous finding. One possible reason is the above-mentioned plausible biases in the samples. As I explained above, the samples in the older birth cohorts would likely include a higher proportion of more economically wealthy people. Previous research has demonstrated that people of high economic status tend to express more uniqueness (e.g., Ma *et al.*, 2017; Snibbe & Markus, 2005; Stephens *et al.*, 2007; Wang *et al.*, 2020). Thus, the values of the uniqueness indicators in the older birth cohorts would be higher than the actual values and should be lower in reality. If this is true, an increase in unique names would be observed from 1920 to 1969 as well as from 1970 to 2005, showing that unique names would continue to increase from 1920 to 2005.

3. How are the historical changes in average name length interpreted?

The historical changes in average name length (described in Figure 2F in Bao *et al.*, 2021) were newly added to a previous study (Cai *et al.*, 2018). They showed a different pattern of changes from those of character-based indices and seem to be divided into three periods: 1) 1920-1960: almost stable (maintained), 2) 1961-1990: sharp decrease, and 3) 1991-2005: sharp increase (Table 1).

However, the authors did not explain these changes and possible interpretations sufficiently. These drastic changes might be related to various changes in official rules regarding names, political policies, and so on (e.g., Ogihara, 2020). These changes in social, economic, and political aspects should also be considered when cultural changes are discussed.

The analysis shows that given names of Han Chinese in China typically consisted of two Chinese characters at least between 1920 and 2005 (Figure 1 in Bao *et al.*, 2021). From 1920 to 1960, the proportions of one-character and three-character names did not change extensively, leading to the stability of the average name length. From 1961 to 1990, the proportion of one-character names remarkably increased (from approximately 10% to over 30%), but the proportion of three-character names did not vary, which decreased the average name length. It would be beneficial to investigate why only the proportion of one-character names remarkably increased during this period. From 1991 to 2005, the proportion of three-character names increased and the proportion of one-character names decreased, causing the increase in the average name length of this period. It would also be important to examine why the proportion of three-character names increased but the proportion of one-character names decreased.

**Conclusion**

I propose three suggestions that would further increase the validity and impact of the article (Bao *et al.*, 2021). First, it would be better to answer whether the samples of older birth cohorts were truly representative. Second, it would be preferrable to answer whether unique names increased only after the 1970s. Third, it should be clarified how the historical changes in average name length are interpreted. These suggestions would hopefully contribute to a further understanding of the historical changes in baby names and their underlying psychological/cultural shifts in China.

**Ethics statement**

Not applicable.

**Data availability**

No data is associated with this article.

---

Table 1. Historical changes in average name length of given names and proportions of one-character and three-character given names in China (Bao *et al.*, 2021).

<table>
<thead>
<tr>
<th></th>
<th>Average name length</th>
<th>Proportion of one-character names</th>
<th>Proportion of three-character names</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920-1960</td>
<td>Stable</td>
<td>Stable</td>
<td>Stable</td>
</tr>
<tr>
<td>1961-1990</td>
<td>Decrease</td>
<td>Increase</td>
<td>Stable</td>
</tr>
<tr>
<td>1991-2005</td>
<td>Increase</td>
<td>Decrease</td>
<td>Increase</td>
</tr>
</tbody>
</table>

Note. This table is based on Bao *et al.* (2021)'s Figure 2.

---

However, the possible bias in the samples, which I discuss before, should be considered here, too.
References


Dear Dr. Ogihara,

Thank you for proposing these suggestions to our article. I agree that answering these three questions could help clarify the implications of our findings and facilitate the understanding of cultural changes in China. As the first author of this article (Bao et al., 2021), I am responsible for providing a reply to address your concerns. Overall, it is important to note that no research is perfect without limitations. While there are studies using more comprehensive data on baby names to test cultural changes in the United States (Twenge et al., 2010, 2016) and France (Mignot, 2022), the complete annual baby names data for the other countries (e.g., the United Kingdom, Japan, and China) are difficult if not impossible to access. Thus, the 2005 China’s 1% Population Census data used in our study (Bao et al., 2021) is still the most qualified dataset to our knowledge. Below is my detailed point-by-point rebuttal.

1. Were the samples of older birth cohorts truly representative?

Let me first reiterate the data source as described in Bao et al. (2021):

To obtain a nationally representative sample of Chinese names covering a long period, we accessed data from the 2005 China’s 1% Population Census (National Bureau of Statistics [NBS] of China, 2005). The 2005 China Census was conducted using a three-stage stratified cluster sampling method, with respondents randomly selected from each of 340 prefectural-level cities or regions in China. Our sample was a random subset (N = 2,585,481) drawn by the NBS, which had been widely used in previous economic and population research.

Notably, the 2005 China Census was conducted by the National Bureau of Statistics of China, and the random subset we used was also drawn by the NBS. Given the three-stage stratified cluster sampling method for the census, it is reasonable to describe the 2005 China Census as a “nationally representative sample” in terms of its high representativeness of the Chinese population alive in 2005. Indeed, we highlighted its representativeness of the Chinese population but did not overstate its representativeness of any specific birth cohort.

More importantly, four points deserve special consideration, rendering the doubt about the representativeness of older birth cohorts unwarranted or less necessary.

First, although the 2005 China Census is a “one-shot cross-sectional survey” rather than a “cross-temporal survey”, this would not necessarily bias the research on names. The main reason for research using cross-temporal surveys (e.g., cross-temporal meta-analysis) to study psychological changes is to disentangle cohort effects from age effects in participants’ self-report responses. However, names are not a self-report measure and thus unlikely to be confounded with age effects. While a small proportion of people may change their names over the lifetime, most people...
retain their birth names throughout their lives. Therefore, it is not much problematic to use cross-sectional name data.

Second, it should be clarified that the Study 2 in Cai et al. (2018) was also based on cross-sectional (but not cross-temporal) name data, because “the name database in the Chinese Public Security Department” (Cai et al., 2018, p. 3) only included people alive in that year but did not include the deceased population. Chinese who have died are usually removed from the household registration records in the Chinese Public Security Department. Therefore, the large sample of 2 million names with at least 2,000 names per year (Bao et al., 2021) was certainly more representative, reliable, and valid than the small sample of 600 names with only 10 random names per year (Cai et al., 2018).

Third, Ogihara mentioned that “the average life expectancy in China in 2005 was approximately 73 years”, which might introduce selection bias due to the health and wealth of older people who managed to live longer. The rationale for this argument is probably that name uniqueness can dramatically affect, or at least be strongly associated with, one’s (physical) longevity and (economical) prosperity, which actually lacks empirical support. Moreover, it is critical to think about what “reality” is when speaking of older generations. One important factor that Ogihara did not mention, but which has caused the unexpected deaths of Chinese born before 1945, is war. In the Chinese People’s War of Resistance Against Japanese Aggression (1931~1945), tens of millions of Chinese people were killed. How do we account for the possible impact of the war on the representativeness of the older Chinese? Is including or excluding the Chinese who died in the war more likely to reflect the reality?

Finally, and reciprocally, a critical scholar may also ask, for instance, whether the name samples used in Ogihara et al. (2015) were truly representative. Specifically, Ogihara et al. (2015) retrieved Japanese names from two companies: one was from the “Benesse Corporation”, which collected “baby names from customers who bought their products”; the other was from the “Meiji Yasuda Life Insurance Company”, which collected “baby names from people holding insurance policies with the company”. In both cases, the problem of unrepresentativeness seemed much more serious because the two samples were likely drawn from wealthier people who had money to buy these products and insurance. It would also be important to address this systematic selection bias in the samples of Japanese names.

Nevertheless, it is safer to claim that the 2005 China Census is at least representative of Chinese born between 1949 and 2005, and that the results before 1949 are exploratory or preliminary but still informative for understanding Chinese cultural changes over a longer time span. We hope that future research can access more comprehensive data on baby names in China.

2. Did unique names increase only after the 1970s?

Simply put, yes, but the shifting pattern of name uniqueness varies across regions within China, which warrants future research and explanations.

In both articles (Bao et al., 2021; Cai et al., 2018), despite the huge difference in sample sizes, we consistently found that name uniqueness in China began increasing after the 1970s. This is
possibly a result of the China's Reform and Opening-Up Policy since 1978. Therefore, the statement that “the study did not replicate the previous finding” is unwarranted and incorrect. We did replicate the previous finding by using two similar indices of character uniqueness (in the third point below, I will explain why the name length indices were different from character uniqueness and improper to be used to indicate name uniqueness).

Nonetheless, this is only an overall result at the national level. My unpublished findings, which are also based on the 2005 China Census, indicate that the patterns of historical changes in name-character uniqueness actually varied across different regions within mainland China. While there was a continuous sharp increase in name uniqueness after the 1970s (consistently since 1978) in all regions of mainland China, I found that from 1920 to 1978, name uniqueness (1) increased steadily in Northeast China, (2) remained almost unchanged in North and Southwest China, and (3) decreased in South, Central, East, and Northwest China.

I will refrain from reporting and explaining more here, as these results remain unpublished. But a more comprehensive understanding of Chinese cultural changes requires distinguishing between different regions within China, since China is a country spanning about 9.6 million square kilometers and there are substantial within-culture variations in China.

3. How are the historical changes in average name length interpreted?

The changes in the four indices of name length were not sufficiently interpreted in our article (Bao et al., 2021). Here I provide further probable explanations for the increased proportion of one-character given name and the decreased average name length (note that these two indices should be understood as reverse because the average name length was largely determined by the relative proportions of one-character and two-character names).

The main reason for this pattern may be the joint effect of the China’s One-Child Policy (1979~2015) and the declined use of “generation names” in China (probably since the 1960s), both of which are factors unrelated to the need for uniqueness, making the indices based on name length less representative of cultural uniqueness and individualism. Take my own family as an illustrative example: my grandparents have three children born in the 1960s and 1970s, and all of them have one-character given names (i.e., Yue, Tai, and Dai). This actually deviates from the traditional naming practice in China of using “generation characters” in given names. We have also mentioned this in our article (Bao et al., 2021, p. 2): “Traditionally, the majority of Chinese given names consist of two characters: one represents the generation in a family and the other denotes the unique identity of a person (Zhu and Millward, 1987; see also Figure 1).” Meanwhile, the China’s One-Child Policy also matters. It is likely that families with only one child no longer need to use generation names for multiple children, which may explain why the proportion of one-character names reached its highest level in the 1980s. In addition, the reason why the proportion of three-character names increased but the proportion of one-character names decreased from 1990 to 2005 may be that people began to be aware of the high probability of name duplication if their names only had one character, and thus they pursued two-character or even three-character baby names. Indeed, this is a plausible explanation based on my observations and informal interviews of my family members and peers.
All of these plausible interpretations, however, require more rigorous empirical examination. We hope that future research can empirically test the impact of political policies on naming practices in China. Nonetheless, because name length indices are more susceptible to such social or political impacts, they are less appropriate to indicate cultural emphasis on uniqueness or individualism. Therefore, we suggest that name-character uniqueness is the most appropriate index for research on Chinese cultural change.

I hope that the above responses have largely addressed your concerns and can enhance the understanding of historical changes in baby names and related cultural phenomena in China.

Best regards,
Han-Wu-Shuang Bao

**Competing Interests:** No competing interests.