Publication bias in the neighbourhood effects literature

Jaap Nieuwenhuis
OTB – Research for the Built Environment, Faculty of Architecture and the Built Environment, Delft University of Technology, P.O. Box 5043, 2600GA Delft, The Netherlands

Abstract
There is a strong belief amongst policy makers regarding the existence of neighbourhood effects, i.e., that above and beyond other influence, the neighbourhood in which they live influences individuals’ life chances. The evidence from scientific reviews leans towards the idea that it is likely that neighbourhoods indeed affect life chances of their residents. Yet, amongst scientists there is no consensus about the neighbourhood effects hypothesis because of differences in methodological and theoretical approaches. I ask why, when there is no consensus, the evidence leans towards supporting the neighbourhood effects hypothesis within policy circles. To test whether the scientific evidence for the neighbourhood effects hypothesis is driven by publication bias, I employ a meta-analysis over all studies that test for a relation between neighbourhood characteristics and educational outcomes. The results show that around the arbitrary boundary for statistical significance (p-value is less than 0.05), there is an unlikely large share of studies that are significant compared to the share of studies that are insignificant. This finding strongly suggests that there is publication bias in the neighbourhood effects literature.

Introduction
Since the publication of Wilson’s (1987) The truly disadvantaged, there has been a continuing interest in neighbourhood effects. The neighbourhood effects literature revolves around the hypothesis that, independent of possible confounders, the neighbourhood exerts an influence on the life chances of individuals living in that neighbourhood. Reviews reveal that most evidence in this literature shows that neighbourhoods do have an influence on individual life chances, and that this effect is mostly in a positive direction (e.g., the richer the neighbourhood, the better residents’ individual outcomes and vice versa; see e.g., Dietz, 2002; Durlauf, 2004; Ellen & Turner, 1997; Galster, 2002; Nieuwenhuis & Hooimeijer, 2015). The idea that neighbourhood characteristics can exert a (positive) influence on neighbourhood residents’ lives not only exists within the scientific community, but also amongst policy makers, where the belief appears even stronger. Amongst policy makers, the
belief that mixing neighbourhoods based on income levels will benefit the neighbourhoods’ residents, has led to many changes in the housing stocks in mainly poor neighbourhoods (Manley et al., 2013; van Ham & Manley, 2012). Neighbourhood mix policies are based on the idea that, when people from different income groups live side by side, the poor will benefit from the rich, e.g., by expanding their social networks or by having more positive role models in the neighbourhood (Musterd & Andersson 2005). Such policies are based on the often contested idea that people from different socio-economic backgrounds will start interacting just by living in close proximity of one another (Atkinson & Kintrea, 2000; Blokland & van Eijk, 2010; Kleinhans, 2004). However, in the end there is no hard proof that people actually behave like that.

There is also a community of researchers who are more sceptical about neighbourhood effects, and argue that we should study why people live in certain neighbourhoods in the first place (see e.g., Slater, 2013). The oft-heard critique is that the neighbourhood effects literature is plagued by selection bias (e.g., Hedman & Galster, 2013; van Ham & Manley, 2012). What is argued is that neighbourhoods are not random selections of households or individuals, but that people select themselves into neighbourhoods based on their preferences and constraints. When this notion is not properly taken into account into statistical modelling, a found neighbourhood effect may be an artefact that is produced by the unmeasured variation that exists in people selecting into neighbourhoods, which is then correlated with the variation in neighbourhoods, but which is also correlated with the outcome that is studied. When not properly controlling on how people select into neighbourhoods, models that test for neighbourhood effects may find spurious effects that are produced by incorrect modelling strategies.

Even though there is no clear consensus amongst scientists about neighbourhood effects, still most evidence leans towards the idea that neighbourhood effects exist. This may be caused by widely spread problem of publication bias (Ioannidis, 2005). Publication bias is a bias in research results that is caused by systematically publishing significant results in favour of insignificant results. The reasoning behind publication bias is twofold. First journals might be more likely to publish significant results (Dwan et al., 2013), perhaps because it is more interesting to be able to publish a finding that validates a new idea, than to publish a finding that is not able to verify a new idea. And second, when researchers are faced with competitive academic cultures where they are evaluated based on their number of publications; in order to get published, they might be more likely to try many different model specifications in order to find significant results (Fanelli, 2010a). Obviously, both mechanisms are related to each other, and both cause an overrepresentation of significant versus insignificant results. Papers with insignificant results may never be published, or models may be tweaked long enough until interesting (read: significant) results are found. It is not difficult to see how these mechanisms lead to an abundance of significant results, and a non-representatively small pool of studies with insignificant results.

I studied the notion of publication bias in the neighbourhood effects literature, and to be more specific in the entire body of literature that studied neighbourhood effects on individual educational outcomes. I used a meta-analytical approach to study 88 studies to answer the question whether publication bias is apparent in the neighbourhood effects literature.
Method

Data

I used the data from an existing meta-analysis by Nieuwenhuis and Hooimeijer (2015). The data are based on a search of Scopus on October 2011. The search was done over the title, abstract and keywords of the studies in the Scopus database, and encompassed two themes: ‘neighbourhoods’ and ‘education’. The ‘neighbourhood’ theme included: neighb*rhood or “community characteristic*” or “residen* characteristic*” or “environment* characteristic*” or “context* characteristic*”. And the theme ‘education’ included: education* or school or grade* or drop*out or “drop out” or academic*. The asterisk symbol is used to allow for every variant of a search term. Filters were used to limit the search results to social scientific studies from peer-reviewed journals. The search result was 5,516 papers.

The relevant studies were identified in a two-step process. First, after examining the titles and abstracts, 244 potential candidate studies remained. In the second step, I conducted a full-text review of the 244 remaining studies. The studies were included if they met the following criteria: 1) ‘educational achievement’ on an individual level was the dependent variable; 2) the independent variables contained at least one neighbourhood characteristic; 3) (non-experimental) multivariate analysis was used; 4) a ‘neighbourhood’ was defined as “the neighbourhood in which the respondent lives/lived” rather than as the area around the school that the respondent attends; 5) the sample used does not consist of pre-schoolers; 6) the sample was from a developed country; 7) the study used recent data, defined as data from 1960 to the present; and 8) the article provided information to obtain the coefficient and standard error. In the final meta-analysis, 88 studies were included that met all of the inclusion criteria. Because many studies contain analyses over subgroups (e.g., men/women or blacks/whites), the final N is larger.1

In the final set of studies, ‘education outcome’ could be defined as: high school graduation; high school dropout; grades/test scores; school performance (including teacher assessments); grade retention; years of education; highest achieved education; college attendance; and college graduation. The recorded neighbourhood characteristics were defined within four broad categories: neighbourhood poverty (e.g., share of low income families); the educational climate (e.g., the share of high school dropouts); proportion of migrant/ethnic groups; and social disorganisation (e.g., crime rates or low social cohesion).

For this study on publication bias, I recorded three characteristics from the included studies: the unstandardised coefficients, standard errors, and p-values of the independent neighbourhood variables. Where no exact p-values were provided, I calculated these.

Analyses

To study publication bias in the neighbourhood effects literature, I used a series of caliper tests (Gerber & Malhotra, 2008a). The idea behind the caliper test is that the 5% significance standard (or z = 1.96) used in scientific research is an arbitrary boundary between significance and insignificance. Under the assumption that the published articles are a

1 For more details about the data (e.g., a list of all included studies), I would like to direct the reader to Nieuwenhuis and Hooimeijer (2015; Appendix 1).
random draw from a sampling distribution; when looking at equal-sized intervals just above and below the significance threshold, the chance for a result to be just above or below is equal. So when the number of studies just above the threshold is much larger than just below the threshold, this is an indication for publication bias.

**Results**

To examine the current consensus in research of neighbourhood effect on education outcomes, I start by looking at how many positive, negative, and insignificant neighbourhood effects are found in the literature and how big the average standardised point estimates are for these categories (Table 1).

Table 1: Average standardised point estimates of the results of neighbourhood disadvantage on individual educational outcomes.

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th>Perc.</th>
<th>Weighted point est.(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative effect</td>
<td>130</td>
<td>42.76%</td>
<td>-3.9539</td>
</tr>
<tr>
<td>No (significant) effect</td>
<td>139</td>
<td>45.72%</td>
<td>-0.2514</td>
</tr>
<tr>
<td>Positive effect</td>
<td>35</td>
<td>11.51%</td>
<td>0.7792</td>
</tr>
<tr>
<td>Total results</td>
<td>304</td>
<td>100%</td>
<td>-1.7160</td>
</tr>
</tbody>
</table>

\(^a\)Point estimates are standardised using the inverse of the standard error (1/s.e.)

In Table 1 there is a similar frequency of negative neighbourhood effects and results that find no effect. In order to examine whether the number of studies that find a negative effect is caused by publication bias, I look at the frequencies of studies just above and below the arbitrary significance threshold of 5% (or z = 1.96). There is no statistical reason to expect a substantially larger amount of studies just above the threshold than just below. So when I do find substantially more studies above than below the threshold, it would mean that the 5% significance level is affecting what is being published (Gerber & Malhotra, 2008a). When looking at the histogram plotting the frequency distribution of z-values (Figure 1), the frequency of z-values just under the critical value of 1.96 (equivalent to 5% significance) is about twice as small as the frequency just over 1.96. To test the likelihood of this happening, a caliper test (Gerber & Malhotra, 2008a) with bandwidths of 3, 5, 10, 15, and 20% over and under caliper is employed (Table 2).
The results of the caliper tests (Table 2) show that, independent of the chosen bandwidth, the distributions of result just under and just over caliper are unlikely to happen. The caliper tests show a high likelihood for publication bias in the neighbourhood effects literature due to the significance threshold.

**Discussion and conclusion**

The results indicate clear evidence for publication bias due to arbitrary significance levels in the neighbourhood effects literature. The reasons behind the discrepancy between the number of studies just above and just below the significance threshold could lie in two sources: the journals and the researchers. Journals may have a tendency to be more favourable towards studies with significant results (Dwan et al., 2013), perhaps because significant results are
more interesting than studies that fail to support their hypothesis. Furthermore, due to the competitive environment in academics (“publish or perish”), researcher are faced with a pressure to publish, and may therefore have an incentive to produce significant results (Fanelli, 2010a). This can be achieved for example by additional data sampling when a result is almost significant. Or by trying additional model specifications until a significant result is found. Studies that in the end fail to achieve statistically significant findings, may end up in the file-drawer and never be published.

The fact that the effects from the caliper test are not very large may first be caused by the small sample size. However secondly, it may also be related to the lack of consensus within the neighbourhood effects literature. It has been argued that in scientific fields that are younger or deal with highly complex phenomena, such as the field of neighbourhood effects, theoretical and methodological consensus is more difficultly reached. When there is no consensus on how to collect, analyse, and interpret data, it was shown that the publication of significant results is more likely, compared to older scientific fields that have a stronger theoretical and methodological consensus (Fanelli, 2010b).

In this study I focussed on a sample of studies that had related neighbourhoods to educational outcomes. However, it is very likely that the findings also hold up for studies of neighbourhood effects on any other outcome variable. The mechanisms behind the results are in no way related the studied dependent and independent variables. As was shown by Gerber and Malhotra, the results hold for very diverse studies in social science and political science (Gerber & Malhotra, 2008a, 2008b). Therefore, the findings could be interesting for all researchers of neighbourhood effects.

For researchers of neighbourhood effects and policy makers concerned with housing, the results should be considered as a critical note to the existing hypothesis that neighbourhoods influence the chances in life of individual residents. Due to publication bias, the evidence for this hypothesis may be based on thin grounds. This result shows that the neighbourhood effects hypothesis should be studied with even more caution than is already done. Moreover, the field would benefit from larger sample sizes and more sophisticated modelling strategies to come to more reliable outcomes. Finally, p-values were never intended to have a significance/insignificance threshold, but were intended to show the strength of the evidence (Sterne & Smith, 2001). If the arbitrary significance threshold of 5% would be abandoned and results would be presented with exact p-values, there would be no incentive for journals or researcher to favour results just over the threshold, over results just under the threshold. Under that scenario, publication bias due to significance thresholds would not be an issue, and “insignificant” results could be judged on their scientific value instead of their p-value.

Acknowledgements
I would like to thank Jarno Hoekman for his significant contribution to the idea behind this paper and his helpful comments on various draft versions. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 615159 (ERC Consolidator Grant DEPRIVEDHOODS, Socio-spatial inequality, deprived neighbourhoods, and neighbourhood effects).
References


