Voter turnout in Hamilton, Canada: accounting for discrepancies between self-reported and official turnout in three neighbourhood clusters

Dylan Simone1, Peter Kitchen2, Allison Williams3 and Patrick DeLuca2

1Department of Geography and Program in Planning, University of Toronto, Toronto, Ontario, Canada; 2School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario, Canada. Address comments to dylan.simone@utoronto.ca
Received 18 June 2013. Accepted 10 March 2014.


This study investigates the discrepancy between official and self-reported voter turnout. This paper is one of the few in North America to use small-scale data to assess such differences in voter turnout. The 2010 municipal election in Hamilton, Canada, is used as a case study. The research employs data from three sources: official polling station data obtained from the City of Hamilton, the 2006 Canadian Census, and results from the Hamilton Household Quality of Life Survey. The survey sampled 4,002 respondents in three neighbourhood clusters representing areas of high, mixed, and low socio-economic status, and collected data on voter turnout in the 2010 municipal election. Polling station data were aggregated to match the census tract boundaries comprising the three neighbourhood clusters. Results suggest significant differences between official and self-reported voter turnout rates, above and beyond the 15-20% over-reporting that can be expected. For instance, in the Southwest Mountain neighbourhood (having an older, more educated population), we find a 30% difference between the two turnout measures. The literature suggests that populations embodying certain characteristics are both more likely to vote, and more likely to over-report their voting behaviour (i.e., obtaining a University degree, owning one’s home, and being aged 65 or older). We ran statistical tests of proportions to compare five key socio-demographic indicators between our survey sample and the populations residing within the neighbourhood clusters from the 2006 Census. That our sample estimate of self-reported turnout is lifted towards those more likely to vote is consistent with the theory of social desirability bias. That theory posits that respondents over-report socially desirable behaviours to conform to their vision of social norm.

Social desirability bias is the propensity of a survey respondent to over-report actions and behaviours seen as ‘socially desirable’ (e.g., voting, civil service, volunteering), and under-report those seen as social taboos (such as theft, fraud, or other illegal behaviour). See Karp & Brockington, (2005), Holbrook & Krosnick (2010), and Knumpal (2011). Social desirability bias has perpetuated the over-representation of voter turnout in social surveys, and is well acknowledged as a contributing factor towards the ubiquitous discrepancy between self-reported and administrative data regarding electoral turnout. Of interest to many social scientists are the characteristics associated with those most likely to over-report socially desirable behaviours; in the context of voting, persons who are older, have higher levels of education, and who own their dwelling are associated with higher rates of over-reporting (Silver, Anderson, & Abramson, 1986; Bernstein, Chadha, & Montjoy, 2001).

Studies that compare official voter turnout rates to those self-reported in surveys often focus on large scales (such as the nation-state) due to the availability of data (e.g., the American NES). Comparative studies of self-reported and official voter turnout at a micro-scale, such as the neighbourhood, are scarce. To the authors’ knowledge, no published study has attempted such a comparison of turnout rates at this micro-scale in a Canadian context. The purpose of this paper is to investigate the discrepancy that arises between official and self-reported voter turnout at the neighbourhood scale. The theory of social desirability bias is employed to explain these differences.

In this paper, we draw information from three distinct sources to measure and assess social desirability bias. First, we use voter turnout data for individual polling stations from a municipal election in the City of Hamilton, Ontario conducted on October 25, 2010. We then imported these data into a Geographic Information System (GIS). Second, we merged these data with socio-economic information for associated census tracts in Hamilton from the 2006 Census. Third, we used the Hamilton Household Quality of Life Survey (HHQLS), administered by telephone shortly after the October 2010 election, that collected data on health, social capital, quality of life and other issues in three neighbourhood clusters representing areas of high, mixed, and low socio-economic status (SES) in the city. The selection of these neighbourhood clusters is based on nearly a decade of on-going funded research in Hamilton. This has given rise to a body of literature focusing on SES clusters. A number of papers were published: see Kitchen, Williams, & Simone, (2012). Figure 1 displays a map showing the three neighbourhood clusters. Each neighbourhood cluster contains a set of census tracts. One of the telephone survey questions asked respondents if they voted in the October 25, 2010 municipal election.
For each census tract within a neighbourhood cluster, this information was then compared to the voting data made available by the city of Hamilton.

We then compare official and self-reported voter turnout in each census tract. The purpose of this paper is to assess and account for the discrepancies between the two. In order to do so, we first provide a theoretical framework on social desirability bias (pertinent to the context of voting), which is then used to account for the differences seen in our analysis. We present results from tests of proportions that compare key socio-demographic factors within the survey sample to that of the actual population residing in those census tracts, the latter taken from the 2006 Canadian Census. After accounting for the differences in turnout rate, we briefly discuss some limitations of the study, while providing context on how future research utilizing small-scale data can minimize social desirability bias in surveys.

**Social desirability bias**

Statisticians and social scientists have become increasingly interested in the effects that survey respondents’ under-reporting of socially undesirable activities and over-reporting of socially desirable ones have on survey data, in terms of validity and reliability. Two broad dimensions are commonly distinguished within the notion of ‘social desirability’: that of a personality characteristic (such as the ambition to gain social approval), and that of an item characteristic, referring to activities or behaviours which are considered more or less socially desirable when compared to one another (Crowne & Marlowe, 1960; Randall & Fernandes, 1991). Furthermore, the need to manage impressions of oneself through survey responses may well vary according to predominant socio-cultural norms. What is socially desirable in one subgroup may not be in another; e.g., putting collective needs above one’s own or vice-versa (Krumpal, 2011). Why and how individuals behave in these ways are interesting questions, however, our focus is on the empirical question. The empirical literature on the over-reporting of socially desirable behaviours and under-reporting of undesirable ones has been extensive (Holtgraves, Eck & Lasky, 1997; Barnett, 1998; Holtgraves, 2004; Tourangeau & Yan, 2007). In particular, under-reporting commonly occurs for activi-
ties such as smoking, alcohol consumption, and cell phone use while driving, whereas over-reporting tends to occur on topics such as volunteering, voting, and recycling goods. Interestingly, social desirability within surveys does not always involve the sensitivity of the question, but instead, that of the answer. Understandably, survey respondents are often cautious to openly report illicit drug use, or deviant behaviour, for instance, where sanctions may be implemented against them depending on their response.

The characteristics of respondents more likely to over-report socially desirable behaviours, such as voting, have been of significant interest to social scientists for decades (Silver, Anderson, & Abramson, 1986; Bernstein, Chadha, & Montjoy, 2001; Kitchen, Williams, & Simone, 2012). In the context of voting, those more likely to vote happen to also embody the characteristics of persons likely to over-report voting behaviour: being older (aged 65 or older), having a high degree of educational attainment (University), and owning one’s home as opposed to renting, for example. This is primarily thought of as being a product of higher levels of education promoting higher awareness of socially acceptable behaviours, in addition to having a vested interest in the outcome of elections (Karp & Brockington, 2005; Holbrook & Krosnick, 2010). To elucidate, recent work on the HHQLS dataset by way of logistic regression analysis found that those aged 45-64 were 3.5 times more likely to vote in the Hamilton 2010 municipal election, compared to those aged 18-24, while those aged 65+ were over 8.5 times more likely to vote than the reference category of 18-24 years. Similarly, respondents with a University degree were 2.25 times more likely to vote than those who did not complete high school, and home owners were 1.5 times more likely to vote than renters (Kitchen, Williams, & Simone, 2012). Recent work at a national scale in Canada confirms these findings: increasing age, education, and home ownership increased the likelihood of voting (Uppal & LaRochelle-Cote, 2012). The fact that those who are more likely to over-report their voting behaviour are also both those more likely to respond to a survey on voting, in addition to actually casting a ballot, increases the degree of speculation required in analytical studies of voter turnout. This has led to research on methods of reducing over-reporting, such as improving survey methods, reducing sensitivity of questions, and providing alternate responses for one’s voting behaviour (Presser, 1990; Belli, Traugott, Young, & McGonagle, 1999; Duff, Hamner, Park, & White, 2007).

Study area
The City of Hamilton, at the west end of Lake Ontario, is 75 kilometers southwest of Toronto. In 1815, George Hamilton purchased 104 ha of land in the township and laid out a town site. Large-scale migration from the UK and other regions led to the town developing as a centre for trade and manufacturing. Hamilton was incorporated as a city in 1846. By the mid 20th century, Hamilton emerged as one of Canada’s most important industrial cities with steel production serving as the centerpiece of the region’s economy, hence the longstanding moniker ‘Steeltown’. Following Word War Two, high levels of immigration, particularly from Western Europe, added to the city’s cultural and social diversity. In 2001, as part of the Province of Ontario’s reorganization of municipal governments, the boundaries of Hamilton were enlarged to include adjacent suburban and rural areas including the former towns of Dundas, Ancaster, Flamborough and Stoney Creek as well as the former township of Glenbrook. Hamilton today has a population of 520,000.

Politically, Hamilton is comprised of 15 wards, in which the populace elects both city councilors (one per ward) and the city mayor. As a result of its economic history, Hamilton is known for its highly unionized workforce with the labour movement that grew in prominence in the city in the 1960s and 1970s. Politically, voters in Hamilton have traditionally (although not exclusively) supported parties at the federal and provincial level that are seen to be favourable to workers: notable the New Democratic Party (NDP). As is the case in most civic elections in Canada, it can be argued that in Hamilton, a voter’s preference for a political party at the federal or provincial level is usually not a factor in their choice of a candidate at the municipal level, although in some cases, the party affiliation of those running locally is known.

Over the past several decades, as a result of national and global forces, Hamilton has experienced significant economic restructuring. Thousands of industrial jobs have been lost in recent years and the once dominant steel industry has been weakened considerably by lay-offs and the closure of operations. As the same time, the city has seen a growth in employment in the service and knowledge based industries, particularly health and education. Economic change has resulted in Hamilton’s once robust central core experiencing social decline and a growing socio-economic divide among residents is evident with several neighbourhoods in the central and eastern sections of the city suffering from high levels of poverty and disadvantage (Kitchen, Williams, & Simone, 2012).

As described below, this paper employs polling data from Hamilton’s 2010 municipal election. In the election on 25 October 2010, voters selecting a mayor, councillors in 15 wards as well as public and separate school board trustees. There were a total of 207 polls throughout the city with 142,932 ballots cast (out of 353,317 registered voters) – a turnout rate of 40.5% (City of Hamilton, 2010). Among the 15 Wards, voter turnout ranged from highs of 46% in Ward 13 (Dundas) and 45% in Ward 10 (Lower Stoney Creek) to lows of 35% in Ward 15 (Flamborough) and 31% in Ward 3 (East Hamilton) (City of Hamilton, 2010).

Data and methods
This article uses three sources of data: 1) HHQLS, 2) the 2006 Canadian Census, and 3) polling results from the municipal election in Hamilton, Canada
on October 25, 2010. The HHQLS data was derived from a telephone survey, distributed across three objectively determined neighbourhood clusters based on socio-economic status: the Southwest Mountain (high SES), Central (mixed SES), and Lower City (low SES) areas. Each neighbourhood cluster is comprised of multiple census tracts, thus allowing for the comparison of responses to that of the city’s polling station data. The Institute for Social Research (ISR) at York University administered the survey between November 2010 and March 2011. The ISR drew a random sample of telephone numbers (associated with unique households) as the baseline for their sampling frame, contacting 3,599 households (response rate = 28%), and yielding a sample size of 1,002. The goal was to include approximately 300-350 households in each neighbourhood cluster, with inclusion criteria defined by three main principles: having a population greater than 1,000 in each census tract; having each cluster representing important geographic locations and socio-economic conditions within Hamilton, as determined by the 2006 Canadian Census, and; census tracts in each cluster being contiguous and displaying identifiable boundaries (e.g., major arterial routes and other physical features). In terms of composition, the Southwest Mountain cluster is comprised of three census tracts, the Central cluster encompasses four census tracts, and the Lower City neighbourhood cluster contains five census tracts.

The official voter turnout data by polling station from the city was imported into a GIS (based on the centroid of each polling station), and the polling stations were aggregated into their associated census tract boundary, which can then be aggregated to its respective neighbourhood cluster (see Figure 1). For this exercise, only polling stations whose geographic centroid falls within the census tract boundaries were retained. If a polling station's centroid fell outside the census tract boundaries it was excluded. An alternative approach would have been to allocate shares of the polling station’s voting to each of the census tracts that it overlaps. Either method is imperfect and errors may arise as a result. This data was used in conjunction with the telephone survey data to compare official and self-reported voter turnout rates for the municipal election at the census tract level. The objective of this effort is to take this comparative data, and where differences in turnout rates exist, to account for the discrepancies utilizing both theoretical (social desirability bias) and applied (survey methodology) perspectives.

The data analysis included two main steps.

The first involved comparing how representative our survey sample was to that of the actual population within the neighbourhood clusters (and census tracts therein). This was achieved by comparing five key socio-demographic indicators between the telephone survey sample and the 2006 Canadian Census. We treat the 2006 census proportion as the population parameter, and test whether the HHQLS sample estimate differs significantly from it. A test of proportions was used to evaluate statistically significant differences between four of the five indicators: 1) percentage of total dwellings owned; 2) percentage of dwellings rented; 3) proportion of the population with a University degree, and; 4) proportion of the population aged 65+. A fifth indicator, median household income, was not included in the test of proportions as it is a raw number, instead of a proportion of the total population. Table 1 summarizes the results of this test for the three neighbourhood clusters. This step was crucial as certain population characteristics (such as being highly educated and affluent) are associated with being both more likely to vote, and to over-report voting behaviour. As such, if these segments of the population are over-represented in our survey sample, this could tilt our estimate of voter self-reported turnout and partially account for the deviation from official turnout.

The second step in the data analysis involved producing a three-way contingency table consisting of the official voter-turnout, the self-reported turnout, and neighbourhood cluster. Table 2 provides a summary of the results from this tabulation, noting both the differences in turnout at the census-tract level, and the aggregated neighbourhood clusters. The specific question asked in the survey to determine if the respondent had voted in

### Table 1. Test of proportions: summary results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Southwest Mountain neighbourhood cluster</th>
<th>Central neighbourhood cluster</th>
<th>Lower City neighbourhood cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High SES</td>
<td>Mixed SES</td>
<td>Low SES</td>
</tr>
<tr>
<td>HHQLS Survey</td>
<td>2006 Census</td>
<td>2006 Census</td>
<td>2006 Census</td>
</tr>
<tr>
<td>Dwellings owned (%)</td>
<td>85 (94)</td>
<td>59 (34)</td>
<td>69 (54)</td>
</tr>
<tr>
<td>Persons aged 65+ (%)</td>
<td>26 (17)</td>
<td>22 (15)</td>
<td>17 (15)</td>
</tr>
<tr>
<td>University degree (%)</td>
<td>20 (16)</td>
<td>24 (29)</td>
<td>14 (7)</td>
</tr>
<tr>
<td>Median household income ($)</td>
<td>78,000</td>
<td>78,558</td>
<td>41,400</td>
</tr>
</tbody>
</table>

Note: 1 Percentage of all occupied private dwellings; 2 percentage of all persons; 3 percentage of all persons aged 15 or older; * significant at 10%; ** significant at 5%; *** significant at 1%

Source: HHQLS, and tabulations from 2006 Canadian Census by author.

Reproduced with permission of the copyright holder. Further reproduction prohibited.
the recent municipal election read: “Did you vote in the last municipal election (held in Hamilton on October 25)?” with the response being dichotomous: yes or no.

Results
From the first step, we see a clear pattern with respect to the representativeness of the sample (see Table 1). In general, the HHQLS sample is significantly older, better educated, and more often home owners than in census data. There are three exceptions to over-response rates here. The significant over-response evaporates for (1) homeownership in Southwest Mountain and for (2) persons aged 65 or older in Lower City. The results for Central neighbourhood cluster indicate (3) a significant under-response by those with a University education.

From the second step, we find a remarkable consistency in the difference between the self-reported survey results and the official voter turnout, across the three neighbourhood clusters (see Table 2). Southwest Mountain was found to have the highest difference between the two measures, at an average of 30% higher self-reported turnout, than was actually measured at the polling stations (for instance, in census tract 2.02, 73% of survey respondents asserted they had voted, while the official turnout rate was 30%; a difference of 43%). Next, the Central neighbourhood cluster displayed a 27% higher self-reported turnout rate than that measured at the polls, while the Lower City area fell slightly below at 25%. The Central neighbourhood cluster displayed the greatest variety in turnout rates – again, attributable to the tremendous mixture of social class and backgrounds of respondents, ranging from a discrepancy of 40% (census tract 42) to 17% (census tract 41). The Lower City neighbourhood cluster, on the other hand, displayed the lowest discrepancy rate, at 15%, in census tract 67.

In an earlier paper (Kitchen, Williams, and Simone, 2012), we provide a regression model for having voted in the election (dependent variable). Both the SW Mountain neighbourhood (Odds Ratio = 1.44) and Central neighbourhood (Odds Ratio = 1.46) were found to be more likely to have stated that they voted in the election, compared to the respondents in the Lower City neighbourhood.

Discussion
The purpose of this paper was to compare official voter turnout of the 2010 municipal election in Hamilton, Ontario, Canada to self-reported voter turnout in a household survey conducted in the months following the election. A consistent 25-30% difference was seen between the two turnout rates, across each of the three neighbourhood clusters (high, mixed, and low socio-economic status). This is partially explained through a limitation of this study, in that our survey sample was tilted towards respondents who are more likely to vote: the educated, elderly, and those owning their homes. However, this is only a single piece to the larger puzzle of why there exists such a large discrepancy. Social desirability bias, in which survey respondents over-report socially desirable behaviour and actions (such as voting), while under-reporting socially undesirable ones (such as smoking and alcohol consumption), provides a frame-work for understanding this discrepancy. Further complicating matters is the observed pattern that those more likely to over-report voting behaviour share the same characteristics as those who actually vote; in addition, these characteristics are further shared by people who are more likely to respond to a social survey in the first place (Karp & Brockington, 2005; Holbrook & Krosnick, 2010; Krumpal, 2011).

A second limitation of this study relates to the different scalar bases being utilized. Polling data are at the smallest level of geography (polling station boundaries) utilized in this study. Census variables are at the census tract level of geography and the HHQLS is at a neighbourhood cluster level. The polling station boundaries do not fit perfectly within the census tract boundaries, so as mentioned previously there was some data error that impact the results. If, for example, a polling station centroid was not within a census tract, any voters that correspond to the polling station area within the census tract were not included. Likewise, if a polling station straddled census tract boundaries with its centroid located within a census tract, then all the voters from the polling station were included, regardless if they live within the census tract or not.

Indeed, the importance of aggregation effects cannot be understated in research utilizing geographically sensitive data. Given differences between individual response rates and a collective conscious (through high levels of social capital in a neighbourhood), participation in formal versus informal political events (social desir-
ability bias – to be seen acting in socially desirable ways based on local cultures and norms), and the pertinence of place-based issues in mobilizing citizens to participate in the electoral process (in places like Hamilton, the rich history of Steel work, health inequalities between neighbourhoods, and the role of post-secondary education institutions in the economy have been of significant interest), understanding how data aggregation occurs, and its implications for analysis are of utmost importance.

Future research on voting behaviour (and other sensitive topics), in which a respondent may feel obliged to misrepresent their actual response in order to conform to their notion of social norm, should be sure to pay particular attention to survey design, the wording of questions, and data collection. For instance, a validated measure to decrease over-reporting in surveys of voting behaviour is to provide an option as to why the respondent did not vote (Bell, Traugott, Young, & McGonagle, 1999; Duff, Hamner, Park, & White, 2007). This allows the respondent the opportunity to conform to social norms in another manner – such as being allowed to answer that they did not vote due to activities such as childcare, or employment commitments consuming their free time. Providing such outlets as opposed to the dichotomous ‘yes/no’ answer to whether one voted is an effective measure in aiding the reduction of the misreporting of voter turnout (Tourangeau & Smith, 1996; Krumpal, 2011).

Additionally, as the presence of interviewers or bystanders may influence the psychosocial effect of choosing to misreport behaviour, alternative, self-administered methods such as Internet surveys can be explored, in which the respondent may feel less pressure to conform to social norms in their answers (Des Jarlais et al, 1999; Metzger et al, 2000; Okamoto et al, 2002; Holbrook, Green & Krosnick, 2003). Future research should continue to investigate patterns between self-reported voter participation rates and official turnout data, both across space and between spatial scales. With the release of the 2011 Canadian Census and National Household Survey, new questions around voting behaviour can be investigated utilizing pertinent data. In this research, the power of place and importance of geography could very well be the key to uncovering new insight into patterns of conformity to social norms in survey research, as well as improved methods to aid in decreasing the burden of misreported answers within the social sciences.

References


