INFRASTRUCTURE VICTORIA

Technical paper

Melek Cigdem-Bayram and David Prentice

Responding to high crime rates: what is the mix of prevention, insurance and mitigation individuals choose and its results?

Technical Paper No. 3/18
May 2018

INFRASTRUCTURE VICTORIA

Responding to high crime rates: what is the mix of prevention, insurance and mitigation individuals choose and its results?

Melek Cigdem-Bayram
Infrastructure Victoria/RMIT University
and
David Prentice.[1]
Infrastructure Victoria

Infrastructure Victoria Technical Paper No. 3/18

May 2018

Address:	Infrastructure Victoria
	Level 16, 530 Collins Street
	Melbourne VICTORIA 3000
Telephone:	03 9936 1737
Email:	enquiries@infrastructurevictoria.com.au
Website:	infrastructurevictoria.com.au

[1] This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

Abstract

In this paper we take first steps in providing parameters capturing some wider impacts of crime on individuals for the cost benefit analysis of investments in justice infrastructure. Statistical matching methods are applied to the HILDA dataset in the first broad economic analysis of how individuals respond to living in an acutely high crime environment and the consequences. Compared with individuals living in a postcode with a moderately high crime rate, the matching analysis shows individuals living in postcodes with acutely high crime rates are more likely to be a victim of a violent crime and spend less on insurance. They are also more likely to have a family member incarcerated even if they are no more likely to be incarcerated themselves. There are no significant differences in household incomes or full-time employment rates though those living in an acutely high crime rate postcode are more likely to be unemployed. Finally, although there are no significant differences in measures of mental health, individuals in acutely high crime rate areas spend less on health. This could be because they are less likely to have a long term health condition but could also reflect underinvesting in health care which may have negative consequences for health in the long term.

1. Introduction

There has been much work on the direct effects of crimes on victims. However, there has not been much work on the effect of living in high crime areas per se. Individuals living in a world in which there is crime can be modelled as making decisions under uncertainty. If the individual becomes a victim of crime, they will incur some losses. In theory, compared with a zero-crime world, the individual can maximise their utility by undertaking a mixture of prevention, mitigation and taking out insurance. Prevention could involve investments to reduce the likelihood of becoming a victim e.g. installing an alarm system. Mitigation could feature insurance and changing consumption patterns to reduce losses when a victim e.g. not buying nice things when there is a high probability they will be stolen. There has, though, been few empirical economics analyses of the set of decisions individuals take when faced with greater risks of being a victim of crime and the associated outcomes. What is the mixture of insurance, investment in crime deterrence and consumption modification in response to different levels and types of crimes? And what difference does this make?

Determining the extent to which crime affects individuals, in addition to victims and their associates, is particularly relevant for the research program of Infrastructure Victoria which seeks to improve the measures of the costs and benefits associated with infrastructure. While there is already a broad set of parameters in existence for calculating the individual effects of crime as well as some anticipatory costs (Mayhew, 2003), there is very limited economic evidence on indirect effects on people who are not victims of crime. If these wider effects are significant then they should be considered in a cost-benefit analysis of investments aimed at reducing crime rates. And the estimates of these wider effects of crime can contribute to generating parameters suitable for such a cost-benefit analysis. We take advantage of a set of datasets that enable matching otherwise anonymised individuals to crime levels by postcodes in the Australian state of Victoria. We construct a treatment group of individuals living in the postcodes in the top 2 per cent to 10 per cent of crime rates and compare them with those individuals matched, econometrically, who live in a control group of those within the top twenty per cent to thirty per cent. By matching we deal with observable differences. Comparing individuals living in areas with acutely high crime rates with individuals drawn from other areas with moderately high crime rates is done to deal with two potential threats to identification. The first threat comes from the fear of crime not being material enough to affect individual decisions. Focussing on the acutely high crime rate postcodes deals with this. The second threat comes from unobservable differences across individuals that leads them to self-select across postcodes according to their different sensitivity to crime. The differences in sensitivity could be a result of different tastes for risk, ability to mitigate or other relevant personal characteristics. We consider a broad set of behaviours and outcomes: insurance, mitigation, employment and education, mental and physical health and the likelihood of being incarcerated.

We find that individuals in acutely high crime rate postcodes have a greater experience with crime than individuals in moderately high crime rate postcodes. While the likelihoods of being a victim of property crime are similar, they cannot completely prevent additional exposure to violent crime. They also spend less on insurance. This could either reflect not buying as much insurance or a form of mitigation by reducing consumption so there is less to insure. While they are no more likely to have been incarcerated in the previous twelve months they are more likely to have family members who have been incarcerated in the previous twelve months. Living in an acutely high crime rate area doesn't result in significantly different outcomes in terms of full-time employment rates or household income, unemployment rates are significantly higher. Unlike earlier research we do not find a difference in mental health outcomes but we do find less expenditure on health practitioners and medicine which could either reflect that, on average, individuals are less likely to have a long-term health condition, or that health problems may be more likely in the long run.

These results have implications for the cost benefit analysis of investments in infrastructure that is proposed to reduce crime. This work yields some new impacts that could be combined with estimates of the impact of justice infrastructure on crime to estimate the benefits from investments in such infrastructure in a cost benefit analysis. In addition, our estimates provide a sense of how costs might change as crime rates change. For example, there may be an increase in lost output due to unemployment or incarceration. This is in contrast to the estimates in Mayhew (2003) which are typically calculated as average costs per incident.

There are several limitations to this work. While the In-Confidence version of HILDA has revealed new information associated with the individual effects of crime, we have only filled some gaps rather than provide a comprehensive accounting of the wider impacts of crime on the community. Longitudinal surveys which focus on crime and individual responses to it, would improve our understanding of these issues. One piece of work that would be of considerable interest is to see if the effects differ across age cohorts. For example, to determine if the effects are greater or less for younger people compared with older people. This might be of policy interest in Australia in view of Australia's rapidly aging population and its implications on government expenditure in say health care (Wood et al, 2017).

The paper proceeds as follows. In the next section, we provide an economic framework for the problem and review the (very limited) economics literature that considers a more limited set of outcomes. This is followed by a discussion of the econometric model and the data. Results are presented, discussed and we conclude.

2. Model and previous literature

2.1 Model

In this section we provide a framework within which our analysis can be described and the results interpreted. We present a model of an individual making a set of choices about their employment, education, health and location under uncertainty. Different choices have different returns depending on the state of the world i.e. the way the uncertainty is resolved. An increase in crime increases the likelihood that an individual is going to be in a worse state of the world, whether this is associated with a loss of income, health or life. Denote *x* as a set of broadly defined consumption goods, *e* as the costly effort to reduce the likelihood of a bad state of the world from happening, *i* as the quantity of insurance purchased and *l* as the location. The consumer problem can be represented as choosing *x, e, i* and *l* to maximise expected utility, *EU*:

$$EU = \sum_{j} U\left(x_j, e, i, l | w_j\right)\pi_j(e, l)$$

where *j* denotes the state of the world and π_j is the probability of state of the world *j* occurring. This is modelled as a function of the effort put in to reduce the likelihood of different outcomes occurring and *l*, the location. Utility is modelled as conditional on wealth in each state of the world, w_j so to capture the dependency of consumption choices in each state of the world on wealth.

This framework encompasses the different actions possible to an individual faced with uncertainty as a result of crime. The individual can purchase insurance to restore their wealth in a bad state. They can also choose prevention by taking actions to reduce the likelihood of being a victim (e.g. through greater security). They can also choose to live in locations with lower probability of crime. Depending on their wealth they will choose different combinations of prevention, location and insurance. If we allow for consumption choices over time there is another dimension of mitigation beyond insurance through choices that will involve less loss if a victim of crime e.g. buying a cheap car.

This framework also encompasses the two ways highlighted in Dustmann and Fasani (2016), in which crime can reduce well-being even if the person is not a victim: greater fear; reduced perceived freedom. It is also possible to make assumptions on the probability distribution and utility function to capture the intuition of Dustmann and Fasani (2016) that this all leaves the individual worse off than if there was no crime.

Before proceeding, it is worth elaborating on the location choice. An individual chooses a location to live and work so to maximise their expected utility. In equilibrium, otherwise identical individuals will be sorted across locations with different crime rates according to differences in their aversion to risk and the costs of prevention, insurance and other forms of mitigation. Observed sorting is also influenced to the extent to which individuals are constrained in their choice of location. So otherwise identical individuals may be in a same location not only because of similar risk aversion, prevention, insurance and other mitigation costs but also similar constraints.

2.2 Previous literature

There is no other research in economics that has considered the effect of crime on a wide range of outcomes simultaneously. There are two papers that analyse the effect of crime rates on mental health (Cornaglia et al., 2014; Dustmann and Fasani, 2016) and a third paper relating crime rates to physical activity (Janke et al., 2016) which we now discuss in more detail.

Cornaglia et al. (2014) use a panel of individuals from 2002 to 2006 of the same dataset as we use, the restricted version of the Household, Income and Labour Dynamics in Australia survey (HILDA). As dependent variables, they analyse four measures of mental health as well as an aggregate index. To capture exposure to crime, they consider crime rates (number of incidents per 100,000 people) by Local Government Area (LGA). The effects of violent crime and property crime are considered separately. They conclude that violent crime reduces some types of mental health whereas property crime has no statistically significant effect.

Dustmann and Fasani (2016) select a panel of individuals the British Household Panel Survey from 2002 to 2008, complemented a panel of individuals aged at least 50 from the English Longitudinal Study of Ageing, with four waves between 2002 and 2008. Crime rates (number of incidents per 10,000 people) are by Local Authority (LA) which is a bit smaller area than an LGA.[2] They consider the effects of total, violent and property crime rates. They conclude that it is property crime that is driving a negative effect on some types of mental health.

Janke et al. (2016) analyse nearly one million observations of repeated cross sections of the Active People Survey, in England, for five of the years between 2005 and 2011. They focus on the rates of violent crime with injury by LA.[3] An increase in violent crime reduces walking and overall physical activity which is likely to translate into poorer physical health.

---

[2] The typical resident in Cornaglia et al (2014)'s sample lives in an LGA of about 215,000 whereas the average LA is about 145,000 people.

[3] Violent crime without injury, such as "assault without injury" or "harassment, public".

Finally it is worth noting that there is a more extensive literature that considers the links between crime, fear of crime, happiness and mental health though the causality isn't always treated explicitly.[4]

3. Econometric model and data

3.1 Econometric model

The conceptual framework that underpins the methodological approach employed in this paper is that of the potential outcomes framework. The potential outcomes framework aids program evaluators to make causal inferences about a particular intervention or programme by making inferences on what would have occurred to a programme participant in the absence of the programme. Suppose for instance that individual *i* is participating in some programme or, as in our case, exposed to a high crime area and we want to measure the effect of exposure on some outcome of interest, *Y*. In an ideal paradigm where one can observe the same individual in different states at one time, the impact of living in a high crime area ($D_i = 1$) for individual *i* could be estimated by simply taking the difference between individual i's outcome while living in a high-crime area (Y_{1i}) and the outcome of the same individual in the alternative state were they *not* living in a high-crime area (Y_{0i}). Formally, we can denote average treatment effect on the treated population (our parameter of interest) as:

$$E[\Delta_i|\mathrm{D}_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \qquad (1)$$

In reality however, the obvious limitation is that we cannot observe the same individual in two different states at one time. Thus, for residents of acutely high-crime areas we cannot observe their outcome Y_{0i} which means that we can never directly estimate the impact of crime on residents' outcomes. Rubin's causal model (Rubin, 1973) offers a

[4] See Curry et al. (2008) and Cheng and Smyth (2015) and references therein.

solution to this ‘missing data’ problem and suggests using a suitable counterfactual group to represent residents of acutely high-crime areas (hereby referred to as the treatment group) in $Y_{0,}$ where suitability is determined by how similar the counterfactual group is to the treatment group. By substituting the missing data for the treatment residents, $E[Y_{0i}|D_i = 1]$, with the average outcomes of residents in a counterfactual sample, $E[Y_{0i}|D_i = 0]$, we can modify equation (1) to the following :

$$E[\Delta_i|\mathrm{D}_i = 1] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \qquad (2)$$

As mentioned earlier, an important requirement when identifying a counterfactual sample is that there are no systematic differences between the treatment and counterfactual groups. Data obtained via random assignment of a treatment (i.e. randomised controlled trials) removes any systematic differences by its very nature and is therefore revered as the ‘gold standard’ of data designs. Quasi-experimental data refers to data that was obtained from a non-experimental design (i.e. not randomly assigned) but that attempts to mimic experimental data using econometric techniques to identify an appropriate counterfactual sample. We employ Propensity Score Matching (discussed in Section 3.3) to select a suitable counterfactual group.

3.2 Data

The empirical analysis relies on the following three data sources: (i) the Household, Income and Labour Dynamics Survey (HILDA); (ii) Crime data; and (iii) the Australian Bureau of Statistics. We discuss each of these data sources in turn below.

3.2.1 HILDA

The first of these is the In-Confidence Household, Income and Labour Dynamics Survey (HILDA). HILDA is a nationally representative household longitudinal survey of individuals who are at least fifteen years old. Survey participants are followed annually and asked an array of questions relating to the households and its occupants. Survey questions cover a wide range of subject areas including (but not pertaining to) income, education, employment, wealth and health. Although HILDA includes a wide range of questions relevant to our research questions, not all questions are asked in all survey waves. Because the crime data is only available from 2005 on, we only use HILDA data from the fifth wave onwards.

As of 2017, the survey features fifteen waves between 2001 and 2015. In 2001 there were 7,682 households and 19,914 individuals. The sample was topped up in 2011 to improve the sample's representation of immigrants arriving in Australia after 2001; this added a further 2,153 households and 5,477 individuals. By wave 15, there are a total of 9,631 households and 23,292 individuals. We use the survey responses provided by individuals to estimate the effect of area-level crime on a range of individual decisions and outcomes but also use it to obtain neighbourhood-level demographic data for inclusion in the probit model (see section 3.3). The In-Confidence version of the HILDA dataset provides geographical information based on the residential home address of household respondents. Geographical data is provided at various spatial units with the smallest of these being the Collection District (CD). As the crime data is only available at the postcode level, we use Postal Code information to join the HILDA individual data with the postcode crime and population data and also use it (as well as Collection-District information) for aggregating demographic characteristics at a larger geographical unit.

Although HILDA is a national survey, there is sufficient sample size for the sample to be restricted to individuals solely living in Victoria.[5] We confine our sample to Victoria as it is the only state for which we have crime data.

3.2.2 Crime data

The second primary dataset is the number of offences, by postcode, between 2005 and 2016, as recorded on the Victoria Police Law Enforcement Assistance Program (LEAP). This includes offences collected by the police, victim reports and family incidents. It does not include traffic infringements, offences for which other agencies, such as the Federal Police, are responsible and offences by Victorians outside of Victoria. This data is published by the Victorian Crime Statistics Agency (CSA).[6]

The crime data is reported in six categories:

1. Crime against the person
2. Property and deception offences (Arson, Property damage, Burglary/breaking and entering, Theft, Deception, Bribery)
3. Drug offences (Drug dealing and trafficking, Cultivate or manufacture drugs, Drug use and possession)
4. Public order and security offences (weapons and explosives, disorderly and offensive conduct, public nuisance offences, public security offences)
5. Justice Procedures offences (justice procedures, breaches of orders)
6. Other offences (regulatory driving offences, transport regulation offences, other government regulatory offences, miscellaneous offences)

[5] For instance, in the most recent wave of HILDA there are 2,396 (5,780) households (individuals) residing in Victoria.
[6] For more details see https://www.crimestatistics.vic.gov.au/about-the-data (last accessed May 22, 2017).

The CSA does not provide disaggregated postcode level crime data, in terms of offences, due to privacy reasons. We aggregate the above six categories to arrive at a total crime figure per 1,000 persons and use the total crime rates to define acutely high crime rate areas.[7]

At this point it is worth discussing the tension between using actual crime rates and the theoretical model in section 2.1 in which decisions are driven by expectations about crime. It has been observed that perceived crime rates are systematically greater than actual crime rates (Standing Committee on Legal and Constitutional Affairs, 2004). However, there aren't findings whether the gap between real and perceived narrows or widens as the rate of crime increases. If it narrows, then, in the approach outlined below, this should bias against finding any effects and vice versa. If the gap is constant then our results will be valid. This should be kept in mind when interpreting the results that follow.

3.2.3 Census of population and housing

The third primary dataset level is the Census of Population and Housing. The population data is used to convert the crime statistics to per 1000 heads of population. This removes the effect of population size per se. The ABS has compiled population data to match postcodes. The matching geographic units are referred to as ABS Postal Areas.[8] We begin with the data for the 2006 and 2011 censuses. Population estimates for the ABS Postal Area populations in the other years are generated by interpolating and extrapolating from the two Census years.

[7] Converting crime figures into rates surfaced extreme values which we dealt with by removing postcodes with total crime rates per 1,000 that lie in the top 1 percentile.

[8] For more details see http://www.abs.gov.au/websitedbs/censushome.nsf/home/factsheetspoa?opendocument&navpos=450 (last accessed May 22, 2017)

3.3 Definitions of treatment and control groups

We refer to individuals living in a postcode with an acutely high crime rate as the 'treatment group'. To determine which postcodes have acutely high crime rates for each year we rank the postcodes by crimes per 1000 people. An area is defined as having an acutely high crime rate if it is a postcode that is in the top ten per cent of postcodes (excluding the top one percent). We then select the treatment group as the pool of individuals surveyed within HILDA who live in these postcodes.

The pool of individuals for the control group is also chosen from HILDA. To make these individuals as comparable as possible to those in the treatment group we select the control group in two stages. First, we draw the control group from individuals in households that live in postcodes that fall within the third decile – as calculated for each year. These postcodes are referred to as having moderately high crime rates – still above average but not as high as those in the top decile. Second, we use propensity score matching, checking for balancing, to select individuals living in the third decile postcodes that are as similar as possible to the individuals living in postcodes with acutely high crime rates. By restricting the pool of potential controls to individuals living in postcodes with moderately high crime rates we reduce the potential for there to be systematic unobservable differences across individuals in the treatment and control groups.

As crime rates can be quite variable there is a small number of postcodes which are sometimes in the first decile and sometimes in the third decile. To prevent observations being both treated and controls we remove all observations for which ever state is relatively unimportant. For example, if a postcode makes up 5 per cent of the treatment group but 0.05 per cent of the control group, we drop the observations associated with that postcode in the control group. This results in the loss of a few treatment observations and about 10 per cent of control observations.

This process results in 68 postcodes being in the treatment group and 123 postcodes in the control group. The set of postcodes, when ranked from highest to lowest, that make up about 70 per cent of observations in each group, are listed in Table A.4. in the appendix. They are also mapped in Figure One. Note that we have a good mixture of postcodes from metropolitan Melbourne and regional Victoria. And a number of the treatment and control areas are even adjacent or close by e.g. Seaford and Frankston; Port Melbourne and Albert Park; Redan and Wendouree.

Comparing individuals living in areas with acutely high crime rates with individuals drawn from areas with moderately high crime rates is done to deal with two potential threats to identification. The first threat comes from the fear of crime not being material enough to affect individual decisions. Focussing on above average crime rate postcodes is likely to deal with this as people are more likely to be aware of crime differences. The second threat comes from unobservable differences across individuals that lead them to self-select across high and low crime postcodes. The differences could be different tastes for risk, ability to prevent crime, mitigate its effects or other relevant personal characteristics. By selecting from individuals who are living in postcodes with above average crime rates we reduce the likelihood that the individuals in the two groups have differences in unobservable characteristics that are problematic. All of the individuals in our sample are living in areas with above average crime rates. To explain this within the framework of section 2.1 we are trying to minimise the impact of sorting of individuals, based on unobservable characteristics across areas with different crime rates might have on the decisions and outcomes so to isolate the effect of crime rates.

We construct the final sample for estimation by drawing individuals from each of the two pools using propensity score matching (PSM).[9][10] Propensity score matching is a technique developed by Rosenbaum and Rubin (1983) which aims to mimic a randomised controlled trial for programmes where treatment is assigned non-randomly. It achieves this with a balancing score known as a propensity score which allows us to match treatment participants with a counterfactual group on a wide range of observable characteristics using just a single index. The index is estimated via a probit equation which sets membership in the treatment group as the dependent variable and all observable characteristics that may determine treatment assignment as independent variables. These are defined and listed in Table A.2. in the appendix. Most of the explanatory variables are at the area rather than the individual level. The estimated score then provides us with an estimate of probability of treatment assignment.

The Propensity Score Matching in this paper is conducted in three stages:

1. Estimate the probability of being observed in the treatment group by using a probit, where the control variables include a range of area-level characteristics (see Table A.2. for a complete list of variables included in the probit model).

2. Use a rule to construct matching observations for the treatment observations, subject to the treatment and controls having a common support, thereby ensuring that there is sufficient overlap in the combination of characteristics in the treatment and control samples. We employ the Nearest Neighbour (NN) Matching algorithm to match persons in the treatment sample with those in the control group with the closest propensity score. This

[9] Basic application of the propensity score matching methodology has recently been criticised as potentially worsening balance (King and Nielsen, 2016). However, as reported in Table 2 we have checked the quality of the matching and find that the balance has improved.
[10] There are many options available to someone applying Propensity Score Matching. Our approach is largely consistent with that recommended in Caliendo and Kopeinig (2008).

is done with replacement so to allow each treatment observation to be used more than one so as to be matched to the closest control observation. Because HILDA is a longitudinal survey we need to need to confine matching within waves to avoid potential issues with unobservable characteristics varying over time and practically to prevent matching individuals with themselves over time. It is also at this stage we correct the sample such that postcodes are either in the treated or control groups but not both over time.

3. Check the extent to which the resulting treatment and control groups are similar in the mean values of their observable characteristics and the standardised bias reduction in the matched sample (this is referred to as balancing). If they are not sufficiently similar or reduction in standardised bias is sub-optimal, then change the specification in stage 1 and start again until the treatment and control groups are balanced

The main way we test the robustness of our finding is to use an alternative matching specification. We use a Kernel Matching Algorithm which utilises all members of the pool for the treated and control samples but places greater weight on control persons whose propensity scores are closer to the treatment group and less weight on those that are further apart.

3.4 Outcome variables

As HILDA is not designed specifically to provide data for analysing the effects of crime, the survey questions do not automatically address each of the choices and outcomes within the framework presented in section 2.1. Nevertheless, we argue that HILDA does inform on a broad set of outcomes and also choices about insurance and the success of prevention. Table 1 lists the outcome variables analysed in this paper, with the descriptive statistics along with some complementary information on location. The definitions of all decision and outcome variables are included in Table A.1.

The first decision we analyse is whether households in acutely high crime rate areas spend more on insurance. There are three things to consider about this measure. First, note the limitation of this variable is that it also includes insurance against other outcomes e.g. natural disasters. The second consideration is that lower expenditure could reflect two underlying phenomena. On the one hand it could be that households are choosing to insure less, leaving themselves exposed to risk. On the other hand, it could be the case that households are mitigating potential losses from crime by buying insurable goods that will be less affected by crime e.g. cheaper goods. We cannot distinguish between these two cases. But, in either case, individuals are potentially worse off than they would have been because of crime. The third point to note is that if areas with acutely high crime rates also have lower house prices (due to crime or other factors) then this will also reduce premiums for home insurance. However, this will be ameliorated by the combination of using propensity score matching so we are comparing similar individuals, in terms of observable characteristics, across the two postcodes. Comparing individuals in acutely high crime rate postcodes with those in moderately high crime rate postcodes also reduces the size of any such effects. Finally, note that any estimate of the effect on insurance will be a lower bound effect. Those living in higher crime areas will presumably be paying higher premiums for their insurance. Hence, if we find that people in high crime areas spend less on insurance, this suggests that they purchase a more than proportionally lower quantity of insurance.

The second set of variables relate to the prevention decision. HILDA does not provide a panel of observations on mitigation efforts but it does provide data on its obverse – victimisation of crime.

The third set of variables is those related to employment and education outcomes. We include three indicator variables on employment related outcomes, a measure of household income and an indicator variable signifying completing some higher education. Next, in part to compare our work to the earlier literature, we report a set of outcomes around health. We include three measures of physical and mental health outcomes and three measures of health related expenditure. A final set of outcomes that doesn't fit as neatly into the framework in section 2.1 but is of interest is incarceration outcomes. These provide a broader characterisation of how individuals are exposed to crime and its consequences.

3.5 Descriptive statistics

Table 1 contains the descriptive statistics on the actions, outcomes and locations of the individuals in HILDA in the treatment group, compared with those Victorians in HILDA from the rest of Victoria. We highlight the potential for a reverse causality problem of high crime rates or other negative factors lowering high prices attracting lower income individuals with other disadvantages to live there, rather than higher crime causing greater disadvantage. This supports our not using other Victorians as our control group but applying propensity score matching to individuals from postcodes with moderately high crime rates. This makes it more likely our control group are like the treatment group in unobservable and unobservable characteristics except for their exposure to crime. In other words, we are comparing 'like with like' so to isolate the effect of crime on individual and household outcomes of interest. For example, rather than the treated households spending less on healthcare because the treated households have lower income than the control households, it is households with similar incomes but higher crime rates that choose to spend less on healthcare.

We first note that households in acutely high crime rate areas on average spend less on general insurance than other Victorians. We also confirm they are also more likely to be victims of crime which is useful for the validity of our analysis.

Looking next at education and employment and outcomes, the share of Victorians with bachelors degrees does not vary much across the two groups. While participation rates do not vary much either the average Victorian is more likely to be employed full time and less likely to be unemployed than a Victorian living in an acutely high crime rate area. Victorians in acutely high crime areas earn, on average, considerably less than other Victorians.

Looking at health outcomes, the most notable differences are in expenditure. Victorians in acutely high crime rate areas spend less on medicines, health practitioners and health insurance. This could be due to their being less likely to have a long-term health condition. Also note that there appears to be very little difference in mental health outcomes unlike in the findings of Cornaglia et al (2014). Finally Victorians in acutely high crime rate areas are more likely than the average Victorian to be incarcerated themselves or have a family member incarceration.

Though this pattern of descriptive statistics is not unexpected, it cannot be taken on face value as evidence of the effects of crime that are not ameliorated by prevention, insurance or other forms of mitigation. It is also consistent with reverse causality. These statistics are consistent with high crime rates or other negative factors causing lower house prices resulting in more people with existing socioeconomic disadvantages (or disadvantages resulting from other aspects of the area) as well as a greater propensity to be involved in crime living there (Weatherburn, 2001).[11] This is consistent with the geographical distribution of crime which is concentrated in regional and outer regional areas of Victoria. For instance, 51% of respondents who live in acutely high crime rate areas are from regional areas whereas only 27% of respondents who live in the rest of the State are from regional areas.

4. Results

4.1 Matching results

We employ post-estimation tests to measure the effectiveness of the propensity score matching in selecting a suitable counterfactual sample. The results from the propensity score matching are exhibited in Table 2 below where we compare the mean values of the treated with the matched and unmatched control samples on key area-level and individual characteristics included in the probit. [12] It can be seen that the propensity score matching succeeds in identifying control observations whose area-wide characteristics are comparable to that of persons who live in high-crime areas. For instance, household income among persons in the treatment and matched control group is around 10 percent below that calculated for all individuals who live in the moderately high crime rate areas. The proportion of public housing tenants by collection district in the matched areas is almost double that of unmatched areas. Moreover, homeownership is close to 20 percent lower for persons who live in matched areas as compared to those in unmatched areas. There are two key 'take home' points which are revealed in the matching results presented in Table 2: firstly, they draw our attention once again to the relative socioeconomic disadvantage prevailing in acutely high crime rate areas; and secondly, they signal that the

[11] Although Cigdem-Bayram and Prentice (2018) fail to find any statistically significant relationship between crime rates and house prices, except for crimes against the person in regional Victoria.
[12] In the probit model we divide continuous area-level characteristics into quintiles to improve the quality of the matches between the treatment and control groups.

Table 1 *Summary statistics on Treatment Group and the Rest of Victoria*

Outcomes	Treatment Group			Rest Of Victoria		
	Mean	Standard deviation	Count	Mean	Standard deviation	Count
Insurance						
General insurance expenditure	1164	1583	2669	1525	1785	32011
Mitigation						
Victim Of Violence	2.2%	14.8%	2372	1.3%	11.3%	30000
Victim Of Property Crime	5.0%	21.7%	2373	3.6%	18.6%	30054
Employment and Education						
In labour force	68.3%	46.6%	2841	69.0%	46.3%	34568
Employed full-time	40.9%	49.2%	2838	42.9%	49.5%	34525
Household disposable income	76692	69049	2841	88601	73615	34568
Unemployed	5.2%	22.2%	2841	3.5%	18.3%	34568
Bachelors degree	25.2%	43.4%	2841	26.3%	44.0%	34557
Health Related						
Long-term health condition	19.1%	39.3%	2830	21.8%	41.3%	34546
Mental health	73.1	17.6	2385	74.4	16.8	30269
Psychological distress	16.6	7.2	1110	15.6	6.2	14305
Expenditure on medicines	356	533	2669	463	977	32011
Expenditure on health practitioners	692	1327	1570	1019	2294	18043
Health insurance expenditure	741	1188	1695	1159	1580	19802
Incarceration						
Own incarceration	0.4%	6.1%	2373	0.2%	4.3%	30046
Family member incarceration	1.9%	13.6%	2373	1.2%	10.9%	30043
Location						
Major City	49.0%	50.0%	2839	73.3%	44.2%	34564
Inner Regional Areas	32.5%	46.8%	2839	23.7%	42.5%	34564
Outer Regional Areas	18.6%	38.9%	2839	3.0%	17.0%	34564
Remote Areas	0	0	2839	0.01%	0.9%	34493

Table 2 *Average values for area and individual level characteristics of treatment and matched and unmatched control groups*

Variable	Treatment group	Matched control	Unmatched control
Area-level variables			
Median Annual Household Disposable Income By Postcode, $	74,703	73,710	80,368***
Mean age By Collection District	44.67	45.03	46.52***
% Males By Postcode	0.45	0.46	0.47***
% Australian Born By Collection District	0.79	0.79	0.78
% Public Housing Tenants By Collection District	0.04	0.05	0.03***
% Owner occupiers By Collection District	0.59	0.60	0.70***
% In Labour Force By Postcode	0.68	0.68	0.68*
Individual-level variables			
Male	0.44	0.44	0.48***
Married	0.48	0.49	0.55***
Father Has University Qualification	1.45	1.47	1.48**

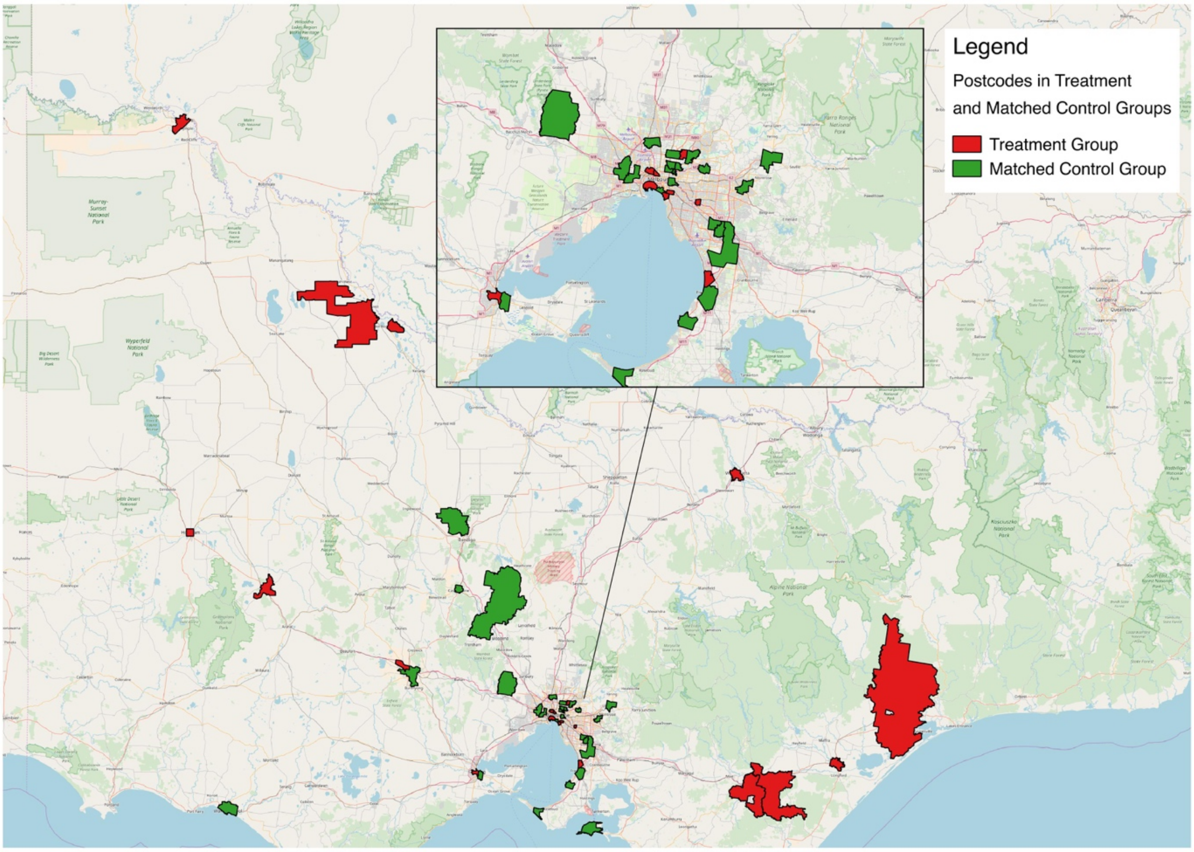
Notes: Difference in average treatment and matched control variables is significantly different at: *** 1%; **5%;***10%.

propensity score matching works considerably well in identifying a control group that is similarly disadvantaged in terms of its neighbourhood characteristics.

We map out the location of treatment and matched control postcodes in Figure 1, with areas shaded in red representing high crime areas and those in green represented the matched control areas. Contrary to expectations, the distribution of high crime areas appears to be uniformly spread across urban and regional Victoria and not confined strictly to populated urban areas.

What also becomes apparent in Figure 1 is the close proximity between treatment areas and the matched control areas. This suggests that propensity score matching method selects the counterfactual group among persons who live in postcodes that neighbour high crime areas. This outcome is not surprising and signals that there are strong socioeconomic parallels between neighbouring postcodes.

Figure 1 *Location of a selection of treatment (green) and control (red) groups*



4.2 Results

The results on the key outcome variables for the treatment and matched control groups are presented in Table 3. The first main outcome is that individuals in acutely high crime rate areas spend significantly less on insurance than similar individuals in moderately high crime rate areas while being twice as likely to be a victim of violence. There is, though, no significant difference in the likelihood of being a victim of property crime.[13] The result for

[13] There are two reasons why it is not automatically the case that people in acutely high crime areas are automatically more likely to be a victim of each type of crime. First, matching is by total crime rate rather than crime rates by category. Second, we are matching similar individuals across the two sets of postcodes. So even though the average person is more likely to be a victim of crime in a high crime areas, this doesn't

insurance could mean that households are insuring less or what they have for insurance is less valuable. Given we show below that there are no significant differences in average income levels between the two groups, if they are spending on less valuable items, this a suboptimal outcome induced by crime just as insuring less is. In other words, if crime rates were lower, these individuals would invest more in their homes, cars and consumer durables.

On average, living in an acutely high crime area does not significantly alter labour market outcomes. There are no significant differences in participation and full-time employment rates or in annual household income. The different results for unemployment but not other labour market outcomes suggest different part-time or casual employment rates. This is unexpected and could reward further exploration. On the other hand, these outcomes could be more fragile as individuals in acutely high crime rate areas are significantly more likely to be unemployed – about as twice as high. There is no significant difference in higher education outcomes.

Unlike the previous literature we find no evidence that living in an acutely high crime rate area, relative to a moderately high crime area, has a negative effect on mental health. This could mean either there aren't significant impacts of crime on the mental health of non-victims. Alternatively it could also mean that the negative impacts of crime come from increases from low to moderate crime rates. Finally, it may signal that persons who have been exposed to high- crime areas may develop an immunity to the concomitant risks over time. This is an area that needs further investigation.

---

automatically follow for a particular type of individual in an area. The results for property crime are consistent with this.

Table 3 *Mean outcomes for treatment and matched control groups*

Outcomes	Average treatment effect on treated population				
	Nearest Neighbour				Kernel Matching
	Treatment	Matched Control	Difference	t-stat	Difference
Insurance					
General insurance expenditure	1288	1388	-100	-1.73*	-73
Prevention					
Victim of violence	0.018	0.009	0.009	2.19**	0.008**
Victim of property crime	0.046	0.042	0.004	0.51	0.001
Employment and Education					
In labour force	0.695	0.691	0.003	0.19	0.008
Employed full-time	0.713	0.716	-0.003	-0.12	-0.005
Household disposable income	81682	82849	-1167	-0.51	-2310
Unemployed	0.035	0.017	0.018	3.68***	0.013***
Bachelors degree	0.316	0.322	-0.006	-0.40	-0.007
Health					
Long-term health condition	0.209	0.243	-0.033	-2.31**	-0.024*
Mental Health	74.41	74.57	-0.16	-0.26	0.051
Psychological distress	15.80	15.39	0.40	1.28	0.39
Expenditure on medicines	384	425	-41	-1.69*	-44**
Expenditure on health practitioners	801	1004	-202	-2.73***	-195***
Health insurance expenditure	910	985	-75	-1.38	-81
Incarceration					
Own incarceration	0.004	0.001	0.003	1.55	0.002
Family member Incarceration	0.014	0.005	0.009	2.71***	0.005

Notes: Difference in average treatment and matched control variables is significantly different at: *** 1%; **5%;***10%.

The result that individuals living in acutely high crime rate areas are less likely to have a long-term health condition seems more likely to result from people with long-term health conditions systematically choosing to live elsewhere rather than a direct effect of higher crime rates. However, households in acutely high crime rate areas spend significantly less on health practitioners (about 10 per cent) and medicine (about 20 per cent less) though not on private health insurance). This could reflect lower rates of long-term health conditions. If not, it could reflect an increased likelihood of more serious health conditions in the future.

The last set of findings relate to incarceration rates. While there are no statistically significant differences in individual incarceration rates between similar individuals living in acutely high and moderately high crime rate areas, individuals living in the acutely high crime rate areas are about fifteen per cent more likely to have family members who are incarcerated. This difference is statistically significant and consistent with greater exposure to crime in the highest crime rate areas. The latter result raises the question as to whether incarceration rates of HILDA respondents are underestimated in acute crime areas i.e. because they are currently incarcerated and unable to complete the survey. An alternative explanation is that households in areas with acutely high crime rates are larger and therefore more likely to have family members who have been recently incarcerated. The difference in the size of the family required for this does seem considerable though.

Finally, the last column of Table 3 reports the treatment effects calculated using a Kernel Matching Algorithm rather than nearest neighbour. In general, the results are similar. The values of the treatment effects and their statistical significance are similar except for General insurance expenditure and Family member incarceration. While the sign and size of these effects are similar they are no longer statistically significantly different from zero.

4. Conclusion

Cost benefit analysis of infrastructure investments to deliver services to reduce crime can currently draw on a range of parameters connected to the direct effects of crime on victims. However, there is very limited evidence on the effects of living in areas with high crime rates on individuals and households that are not victims of crime. In this paper we draw on data to match individuals from HILDA and postcode crime statistics to analyse the causal effects of crime on individual and household decisions about insurance and outcomes with respect to prevention, employment and education, health and incarceration. To deal with potential endogeneity from low incomes and other forms of disadvantage and crime rates being simultaneously determined we compare outcomes for a selection, using propensity scoring, of individuals living in acutely high crime rate postcodes with similar individuals in moderately high crime rate postcodes.

We find that individuals living in acutely high crime rate postcodes compared with individuals in moderately high crime rate postcodes spend significantly less on insurance and are more likely to be a victim of violence (though not property crime). Either they are not insuring as much or else they are spending less on housing, motor vehicles and consumer durables than similar households in moderately high crime rate areas. On average the labour market outcomes are not different but they are potentially more fragile as unemployment rates are significantly higher in acutely high crime rate areas. Unlike earlier work by Cornaglia et al (2014) and Dustmann and Fasani (2016) we find no significant difference in mental health outcomes. We also find there are no significant differences in what households in the two sets of postcodes spend on private health insurance but that households in acutely high crime rate postcodes spend less on health practitioners and medicines. This could either be due to the fact that individuals in acutely high crime areas are less likely to have a long-term health condition or that they are under-investing in health care in the short run, which may have negative long run consequences. Finally, we find no significant differences in own-incarceration rates but that individuals in acutely high crime rate postcodes are more likely to have family members that are incarcerated. Most of these results are robust to different matching methods.

These results suggest that the cost benefit analysis of investment in infrastructure designed to reduce crime rates should consider the broader impacts of higher crime rates such as lower expenditure on insurance and health, the likelihood of being unemployed and the incarceration of family members. We have not found any evidence that the broader impacts on employment, participation, education and mental health on non-victims need to be considered in cost benefit analysis. The mixed results in the literature on the effects on mental health and the possibility of more cohort specific effects of these nature suggest that more research is required before a definitive conclusion on their role in cost benefit analysis can be stated. This work provides a demonstration of the type of approach that would need to be adapted to determine these results. But, in addition, more data, for example, on specific cohorts, would be needed as well.

These results can assist in developing the parameters needed for such an analysis. For example, the change in unemployment rates can multiplied by the productivity and numbers of individuals to yield a measure of loss output. Alternatively, this work provides a guide for additional research to yield more specific parameter values if these are more appropriate for cost-benefit analysis.

References

Standing Committee on Legal and Constitutional Affairs, Australian House of Representatives (2004), *Crime in the Community*, July 2004, Canberra.

Caliendo, Marco and Sabine Kopeinig (2008), "Some practical guidance for the implementation of propensity score matching", Journal of Economic Surveys, 22(1), 31-72.

Cheng, Zhiming and Russell Smyth (2015), "Crime victimization, neighborhood safety and happiness in China", Economic Modelling, 51, 424-435.

Cigdem-Bayram, Melek and David Prentice (2018), "How do crime rates affect property prices?", Infrastructure Victoria working paper.

Cornaglia, Francesca, Naomi E. Feldman and Andrew Leigh (2014), "Crime and Mental Well-Being", Journal of Human Resources, 49(1), 110-140.

Curry, Aaron, Carl Latkin and Melissa Davey-Rothwell (2008), "Pathways to Depression: The Impact of Neighborhood Violent Crime on Inner-City Residents in Baltimore, Maryland, USA", Social Science Medicine, 67(1), 23-30.

Dustmann, Christian and Francesco Fasani (2016), "The effect of local area crime on mental health", Economic Journal, 126 (June), 978-1017.

Janke, Katharina, Carol Propper and Michael A. Shields (2016), "Assaults, murders and walkers: The impact of violent crime on physical activity", Journal of Health Economics, 47, 34-49.

King, Gary and Richard Nielsen (2016), "Why Propensity Scores Should Not Be Used for Matching", mimeo, Harvard University.

Mayhew, P. (2003), *Counting the costs of crime in Australia: Technical Report*, Technical and Background Series, no. 4, Australian Institute of Criminology, Canberra.

Weatherburn, Don (2001), "What causes crime?", Crime and Justice Bulletin, NSW Bureau of Crime Statistics and Research, no. 54.

Wood, G., Cigdem-Bayram, M., and Ong, Rachel. (2017), "Australian demographic trends and implications for housing assistance programs", AHURI Final Report 286, Australian Housing and Urban Research Institute Limited, Melbourne.

Appendix

Table A.1. *Outcome variables and their type and definition*

Variable name	Definition	Variable type
Insurance		
General insurance expenditure#	Annual household expenditure on home/contents/motor vehicle insurance in dollars	Continuous
Prevention		
Victim of violence	Denotes whether individual was victim of violence crime in previous 12 months	Dichotomous
Victim of property crime	Denotes whether individual was victim of property crime in previous 12 months	Dichotomous
Employment and education		
In labour force	Denotes whether individual was in labour force	Dichotomous
Employed full time	Denotes whether individual was in labour force	Dichotomous
Household disposable income	Denotes individual's annual household disposable income	Continuous
Unemployed	Denotes whether individual was unemployed	Dichotomous
Bachelors degree	Denotes whether individual achieved a Bachelor's degree or higher	Dichotomous
Health		
Long-term health condition	Denotes whether individual has a long term health condition	Dichotomous
Mental health	Denotes individual's mental health using the SF-36 measure (transformed)	Ordinal
Psychological distress	Denotes individual's psychological distress using the Kessler Psychological Distress scale (K1)	Ordinal
Expenditure on medicines#	Household annual expenditure on medicines, prescriptions, pharmaceuticals and alternative medicines in dollars	Continuous
Expenditure on health practitioners#	Household annual expenditure on medicines, prescriptions, pharmaceuticals and alternative medicines in dollars	Continuous
Health insurance expenditure#	Annual household expenditure on private health insurance in dollars	Continuous
Incarceration		
Own incarceration	Denotes whether individual was incarcerated in previous 12 months	Dichotomous
Family member incarceration	Denotes whether family member of individual was incarcerated in previous 12 months	Dichotomous

Note: all observations are for each wave unless specified otherwise.

***** The Kessler scale is only available in waves 7, 9, 11 and 13. When analysing psychological distress (and other outcomes for which information is available in every other wave) we pool data from the waves where data is available;

#Expenditure on medicines, health practitioners and health and contents insurance is only available in wave 6 to wave 15.

Table A.2. *List of variables used in probit model*

Variable name	Definition	Variable type
Dependant Variable		
High crime rate areas	Denotes respondents who live in postcodes that falls in the top 10 percentile of the total crime distribution in wave x	Dichotomous
Area-level explanatory variables		
Median household disposable income (quintiles)	Denotes respondents who live in postcodes with low/high median household disposable, ordered by quintiles	Ordinal
Median household disposable income (quartiles)	Denotes respondents who live in postcodes with low/high median household disposable, ordered by quartiles	Ordinal
Mean age (quintiles)	Denotes respondents who live in postcodes with young/aged persons, ordered by quintiles	Ordinal
% Males (quintiles)	Denotes respondents who live in postcodes with low/high proportion of males, ordered by quintiles	Ordinal
% Males (quintiles)	Denotes respondents who live in postcodes with low/high proportion of males, ordered by quintiles	Ordinal
% Males (quintiles)	Denotes respondents who live in Collection Districts with low/high proportion of males, ordered by quintiles	Ordinal
% Australian born (quartiles)	Denotes respondents who live in postcodes with low/high proportion of Australian-born persons, ordered by quintiles	Ordinal
% Public housing tenants (quintiles)	Denotes respondents who live in Collection Districts with low/high proportion of public housing tenants, ordered by quintiles	Ordinal
% Public housing tenants (quintiles)	Denotes respondents who live in postcodes with low/high proportion of public housing tenants, ordered by quintiles	Ordinal
% homeownership (quintiles)	Denotes respondents who live in postcodes with low/high proportion of homeowners, ordered by quintiles	Ordinal
% In labour force (quintiles)	Denotes respondents who live in postcodes with low/high proportion of	Ordinal

Variable name	Definition	Variable type
	persons in the labour force, ordered by quintiles	
% Unemployed (quartiles)	Denotes respondents who live in postcodes with low/high proportion of unemployed, ordered by quartiles	Ordinal
Individual-level variables		
Age range*	Denotes the age range that the respondent falls within	Ordinal
Male	Denotes whether respondent is male	Dichotomous
Married	Denotes whether respondent is married	Dichotomous
Father has university qualification	Denotes whether respondent's father has a Bachelor's degree or higher	Dichotomous
Australian-born	Denotes whether respondent is Australian born	Dichotomous
Age	Age of respondent	Continuous

* We divide individuals into six age ranges: (i) 15 to 24; (ii) 25-34; (iii) 35-44; (iv) 45-54; (v) 55-64; and (vi) 65 and over.

Table A.3. *Postcodes in selected treatment and control groups*

Treatment Postcode	Representative Name	Control Postcode	Representative Name
3031	Flemington	3012	Maidstone
3051	North Melbourne	3020	Sunshine
3081	Heidelberg West	3046	Glenroy
3148	Chadstone	3068	Clifton Hill
3181	Prahan	3070	Northcote
3182	St. Kilda	3072	Preston
3198	Seaford	3078	Fairfield
3207	Port Melbourne	3084	Heidelberg
3220	Geelong	3108	Doncaster
3355	Wendouree	3121	Richmond
3380	Stawell	3140	Lilydale
3400	Horsham	3153	Bayswater
3500	Mildura	3171	Springvale
3585	Swan Hill	3174	Noble Park
3677	Wangaratta	3175	Dandenong
3840	Morwell	3199	Frankston
3844	Traralgon	3206	Albert Park
3850	Sale	3219	East Geelong
3875	Bairnsdale	3280	Warrnambool
		3337	Melton
		3350	Redan
		3444	Kyneton
		3450	Castlemaine
		3556	Eaglehawk
		3922	Cowes
		3931	Mornington
		3941	Rye

Note: Reported postcodes are those for which individuals and households make up at least 70 per cent of the sample.
Representative name reported because some postcodes cover multiple suburbs.