Host Communities: siting and effects of facilities

Solid waste facilities and their host communities: demographic comparisons over time

James Baines

Taylor Baines & Associates

Host Communities: siting and effects of facilities

Solid waste facilities and their host communities: demographic comparisons over time

By

James Baines

Taylor Baines & Associates

Working Paper FS20
Public Good Science Fund Contract TBA X0002

November 2001

Acknowledgements

The research team wishes to thank Gerard Fitzgerald for his professional advice concerning social indicators, Dr Patrick Graham for his professional advice on statistical methods, and Colin Goodrich for his critical review at various stages of the research.

Contents

LIST OF TABLES

LIST OF FIGURES

A: Introduction

The research reported here represents the third phase of investigations into the siting and effects of landfill operations in relation to their host communities. Earlier phases of this research programme investigated:

Phase One: Historical patterns of landfill siting from the social perspective of the host communities involved (22 landfill sites) - reported in Taylor Baines & Associates, 1999.

Phase Two: The day-to-day effects of landfill operations experienced by members of the host communities, and community perceptions of the longer-term community impacts of these landfill operations (seven case studies).

Phase One involved statistical comparisons of a specified set of demographic characteristics for host and source communities at the time the original siting decisions were taken. These historical comparisons applied to a sample of 22 landfill sites and their host and source communities. They were based on data sets drawn from the Census of Population and Dwellings carried out in 1981, 1986, 1991 and 1996. Similar historical comparisons were made between 'selected host communities' (i.e. host communities for sites ultimately selected) and 'short-listed alternative host communities' (i.e. alternative prospective host communities which were not ultimately selected). This latter comparison involved a sample of twelve landfill sites.

In both samples, the statistical comparisons were aimed at identifying whether or not there was any systematic bias towards relative disadvantage in host communities. As reported in Working Paper FS1 (ibid.), a comparison of host communities with source communities across the sample of 22 cases revealed no systematic bias towards relatively disadvantaged host communities in New Zealand in the site selection process for landfills over the last twenty years. However, the comparison of selected host communities with alternative candidate host communities did reveal that during the process of site selection, more powerful candidate communities were consistently more effective in avoiding final selection.

This third phase of investigations is intended to provide an update on the statistical comparisons made in Phase One, in order to determine whether or not any of the statistical differences identified previously have endured over time, or indeed, whether or not they have changed.

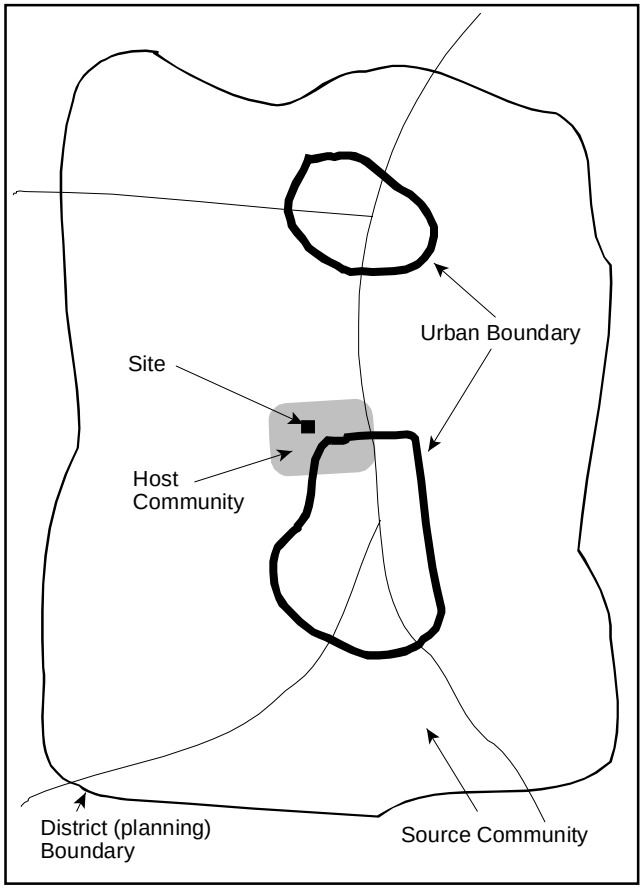
B: Methodology

Whereas the original historical analysis involved field research to investigate aspects of the community setting of selected sites and alternative sites, as well as the statistical comparisons of demographic characteristics for host and source communities, this analysis has been confined simply to the statistical component. The methodology is identical to that described in Section E of Working Paper FS1.

Definitions

As described on pp. 1-2 of Working Paper FS1, the research employs the concepts of '*source community*' and '*host community*'.

Figure 1: Relating the concepts of 'host community' and 'source community'



The term '*source community*' refers to the entire population resident in the geographic area covered by the TLA which has responsibility for deciding the location of the solid waste facility. The 'source community' incorporates all those who generate solid waste which requires either handling in a transfer station and/or disposal at a landfill (see Figure 1).

The term '*host community*' is used to refer to the community resident in the geographic area most clearly associated with a particular solid waste facility. It is not limited only to those residents who may experience direct physical effects from the operation of such a facility (such as the effects of noise, litter, or odour). It extends to include members of a coherent community - perceived in a social sense - that already exists at the time the facility siting decision was taken, and which will continue to develop in ways that may or may not be affected by the presence of the solid waste facility in its locality.

In a sense, all distinct local communities within a 'source community' are potential 'host communities' at the time a site selection process is initiated.

Samples of facilities

The historical 'host-source' comparison reported in Working Paper FS1 involved a sample of 22 landfill siting decisions. These included fifteen landfills which opened between 1977 and 1997, as well as seven landfill sites for which the final site selection had been confirmed but which were yet to be opened. The historical 'host-alternative hosts' comparison involved a sample of twelve cases.

Since the most recent census data available are drawn from the 1996 Census of Population and Dwellings, there are no more recent demographic statistics to enable an update for any sites whose previous data also came from the 1996 Census. Results from the 2001 Census are not due for release until 2002. Thus the sample of landfill sites remaining for the 'host-source' analysis in this report reduces to thirteen, while the sample for the corresponding 'host-alternative hosts' analysis in this report reduces to five.

Specifying indicators

As reported in Working Paper FS1 (p. 19 and Appendix I), in looking for appropriate indicators for this research, attention was focussed on socio-economic variables that could be considered to correlate with 'less powerful' groups in the community. 'Power' was defined as the ability to understand, take part in, or in any way influence the kind of decisions which lead to the siting of solid waste facilities, which are usually subject to adverse expectations. This analysis takes its cue from research into environmental justice issues (e.g. Bullard, 1994). The seven variables[1] taken to correlate with powerlessness were:

- Educational qualifications — % of population aged 15 years and over with NO tertiary qualification
- Ethnicity — % of the total population that is NON European
- Occupational status — % of the population aged 15 years and over who are NOT legislators, administrators, professionals, technicians
- Employment status — % of the population aged 15 years and over who are NOT employed
- Life stage — % of households with at least one child under 5 years old
- Tenure — % of private dwellings that are NOT owned by the occupant
- Income — median household income[2]

Statistical method

The statistical method is identical to that reported in Working Paper FS1 (Appendix II). Host communities were compared with source communities by examining the ratio between 'observed' and 'expected' values[3] for each demographic variable and for each site. The statistics were pooled[4] for each variable over all sites to assess whether there is any indication of systematic bias in the demographic characteristics of selected sites. In other words, if the pooled statistics describing say the proportion of non-European residents or the proportion of unemployed residents for all the host communities appears different from the corresponding statistics for all source communities, then how strong is the statistical evidence for this difference?

For the reader's convenience, the original methodological note is reproduced in this report as Appendix I.

1 Based on the 'usually resident' population in the host and source communities.

2 This variable could not be subject to the same kind of significance testing as the others, being a median value rather than a bi-modal percentage value.

3 'Expected' values for a host community are derived by taking the corresponding value for the non-host community, assuming there were no difference between host and non-host values. Full details are provided in the methodological note in Appendix II.

4 The pooled statistics are based on weighter averages of the site specific results, with weights reflecting the amount of statistical information provided by each site.

C: Results

Summary of results from the previous historical analyses

The pooled results over all 22 landfill sites revealed that host communities were typically different from source communities when analysed for indicators of 'power to influence decisions'. However, there was no consistent bias towards relative disadvantage in host communities, as shown in Table 1. Three indicators suggest relative disadvantage (lack of education, non-professional occupations, dependents in the household), while three suggest relative advantage (high proportions of Pakeha residents, high proportions in active employment, high proportion of property ownership). This grouping of demographic features is characteristic of rural populations, and is therefore consistent with the fact that landfill sites are most often located in peri-urban or rural areas.

Table 1: Pooled results for 'host-source' base year[5] comparisons (N=22)

	Education	Ethnicity	Occupation	Employment	Children	Tenure
Pooled Ratio values	1.02	0.96	1.06	0.91	1.33	0.84
Host-source relativity	disadv.	adv.	disadv.	adv.	disadv.	adv.
Chi-square	19.19	9.17	204.54	10.67	241.3	105.48

The differences between host and source community (as shown by the Pooled Ratio values) were typically not large, with the exception of the proportions of households with young children (relatively high in host communities) and the proportions of dwellings which were rented (relatively low in host communities).

Only descriptive analyses for the seventh variable - median household income - were undertaken since formal statistical analysis requires additional distributional information not provided routinely in census products. It should be noted that, for the simple ratios of median household income used in this analysis, ratios less than Unity (<1) indicate 'relative disadvantage' for the host community compared with the corresponding source community while ratios greater than Unity (>1) indicate 'relative advantage'. For the 22 sites, six host communities had lower median household incomes than their corresponding source communities, whereas 16 host communities had higher median household incomes. The average host/source community ratio for median household income over the whole sample is 1.08 (ranging from 0.73 to 1.47), indicating that host communities around landfill sites have typically not been lower-income communities than their corresponding source communities at the time of site selection.

The absence of systematic bias towards relatively disadvantaged host communities in site selection described above was not repeated when host communities were compared with 'alternative host' communities. A pooled analysis for twelve cases was carried out. This revealed consistent bias towards relatively disadvantaged candidate host communities in the short-list situation, as shown in Table 2. The twelve cases included eight where decisions were made under the RMA.

[5] 'Base year' comparisons refer to data (prior to 1996) for site selection decisions.

Table 2: Pooled results for 'host-alternative host' base year comparisons (N=12)

	Education	Ethnicity	Occupation	Employment	Children	Tenure
Pooled Ratio Values	1.06	1.72	1.08	1.59	1.46	1.24
Host-alternative host relativity	disadv.	disadv.	disadv.	disadv.	disadv.	disadv.
Chi-square	31.76	153.69	52.8	22.78	43.91	21.67

Not only is the pattern of relative disadvantage more uniform (when compared with Table 1), the differences between selected host and alternative host communities (as shown by the Pooled Ratio values) were sizeable in four out of the six variables.

The pattern of systematic bias was reinforced further in the comparison of median household incomes between selected and alternative host communities. In eight out of the twelve cases, the median household income level for the selected host community was lower than the alternative candidate host communities, and the average ratio of median household incomes was 0.94 (ranging from 0.60 to 1.45), indicating that selected host communities have typically been lower-income communities than those which were not selected from the shortlist.

Results of 'host-source' comparisons updated to 1996

To make a comparison over time, it was firstly necessary to repeat the base year calculations for the thirteen landfills which had opened prior to the 1996 census. Table 3 summarises the new pooled ratio values and compares them with the corresponding pooled ratio values for the full sample of 22 landfill sites.

Table 3: Pooled results for 'host-source' base year[6] comparisons

	Education	Ethnicity	Occupation	Employment	Children	Tenure
Full sample of 22 landfill sites:						
Pooled Ratio values	1.02	0.96	1.06	0.91	1.33	0.84
Host-source relativity	disadv.	adv.	disadv.	adv.	disadv.	adv.
Chi-square	19.19	9.17	204.51	10.67	241.3	105.48
Sample of 13 pre-1996 sites:						
Pooled Ratio values	1.03	1.08	1.09	0.88	1.52	1.03
Host-source relativity	disadv.	disadv.	disadv.	adv.	disadv.	disadv.
Chi-square	44.44	12.21	316.78	11.3	331.48	44.44

It can be seen from Table 3 that the reduced sample of earlier site selection decisions does indeed indicates a greater incidence of significant disadvantage than for the full sample of 22 decisions. Although only one pooled ratio (for households with young children) shows a large difference, five pooled ratios are larger (i.e. greater disadvantage indicated) and two of them have switched from indicating relative advantage to relative disadvantage. This is consistent with the result reported in

[6] 'Base year' comparisons refer to data (prior to 1996) for site selection decisions.

Working Paper FS1 at p. 22 that 'there appears to be a trend over time[7] - a greater incidence of significant disadvantage in earlier siting decisions and a lower incidence in more recent decisions'.

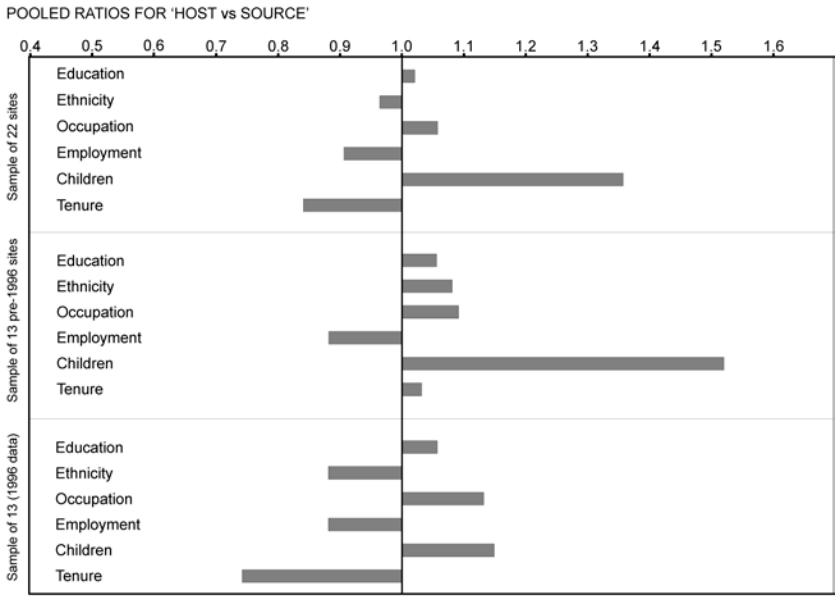
Results from the base year analysis of this sample of 13 siting decisions will now be compared with results from comparing the corresponding demographic data from the 1996 census - see Table 4. All three sets of results (from Tables 3 and 4) have been combined in a single graphical comparison in Figure 2.

Table 4: Pooled results for 'host-source' 1996 comparisons

	Education	Ethnicity	Occupation	Employment	Children	Tenure
Full sample of 13 (pre-1996 data):						
Pooled Ratio values	1.03	1.08	1.09	0.88	1.52	1.03
Host-source relativity	disadv.	disadv.	disadv.	adv.	disadv.	disadv.
Chi-square	44.44	12.21	316.78	11.30	331.48	44.44
Sample of 13 (1996 data):						
Pooled Ratio values	1.06	0.88	1.13	0.88	1.15	0.74
Host-source relativity	disadv.	adv.	disadv.	adv.	disadv.	adv.
Chi-square	142.88	53.12	399.02	14.81	32.66	189.74

The results in Table 4 reveal that with the passage of time two pooled ratios (for educational qualifications and occupational status) have increased slightly (i.e. relatively disadvantaged status has further declined slightly), while three pooled ratios (for ethnicity, households with young children, and rented accommodation) have decreased markedly (i.e. relatively disadvantaged status has improved markedly). The pooled ratio for employment status remained unchanged over time.

Figure 2:



[7] Pooled analysis of sub-samples grouped into the following time periods - 1977-91, 1992-97, yet to be opened (YTBO) - indicates this trend. During the period 1977-91, four out of the six variables show significant disadvantage, whereas for the period 1992-97 three out of six variables, and for the period YTBO two out of six variables show significant disadvantage.

Table 5 presents the corresponding results for host-source comparisons of median household income.

Table 5: Pooled results of 'host-source' comparisons of median household income

	Median household income
Full sample of 22 landfill sites:	
Pooled Ratio values	1.08
Host-source relativity	adv.
Sample of 13 (pre-1996 data):	
Pooled Ratio values	1.04
Host-source relativity	adv.
Sample of 13 (1996 data):	
Pooled Ratio values	1.10
Host-source relativity	adv.

The results in Table 5 show a trend that is consistent with those of the other demographic indicators. The smaller sample of earlier site selection decisions display less comparative advantage than the full sample of 22, but the relative advantage for these 13 cases has increased over time.

Results of 'host-alternative host' comparisons updated to 1996

As discussed earlier, because the most recent census data available are drawn from the 1996 Census of Population and Dwellings, the sample for 'host-alternative' comparison in this report has been reduced from twelve to five sites, opened prior to 1996. Base year calculations for these five sites have been repeated. Table 6 summarises the new pooled ratio values and compares them with the corresponding pooled ratio values for the sample of twelve landfill sites for which alternatives were considered.

Table 6: Pooled results for 'host-alternative host' base year comparisons

	Education	**Ethnicity**	**Occupation**	**Employment**	**Children**	**Tenure**
Sample of 12 sites with alternatives:						
Pooled Ratio values	1.06	1.72	1.08	1.59	1.46	1.24
Host-source relativity	disadv.	disadv.	disadv.	disadv.	disadv.	disadv.
Chi-square	31.76	153.69	52.80	22.78	43.91	21.67
Sample of 5 pre-1996 sites with alternatives:						
Pooled Ratio values	1.04	2.05	1.10	1.32	1.57	1.28
Host-source relativity	disadv.	disadv.	disadv.	disadv.	disadv.	disadv.
Chi-square	8.33	52.92	44.62	4.34	38.6	11.84

Table 6 suggests some differences between the reduced sample of earlier site selection decisions and the full sample of 12 decisions which involved alternative sites. Of the six pooled ratios, four show large differences, while two essentially remain unchanged. Of the four pooled ratios showing large differences, two are markedly higher for the reduced sample than for the full sample, one is slightly higher, and one is markedly lower. Nevertheless, when viewed together, the systematic pattern of relative disadvantage at the time of site selection is still evident.

The base year results of this sample of five siting decisions will now be compared with the corresponding demographic data from the 1996 census (Table 7).

Table 7: Pooled results for 'host-alternative host' 1996 comparisons

	Education	Ethnicity	Occupation	Employment	Children	Tenure
Sample of 5 (pre-1996 data):						
Pooled Ratio values	1.04	2.05	1.10	1.32	1.57	1.28
Host-source relativity	disadv.	disadv.	disadv.	disadv.	disadv.	disadv.
Chi-square	8.33	52.92	44.62	4.34	38.60	11.84
Sample of 5 (1996 data):						
Pooled Ratio values	1.00	1.59	1.07	1.33	1.19	1.14
Host-source relativity		disadv.	disadv.	disadv.	disadv.	disadv.
Chi-square	0.07	54.79	16.97	6.59	5.97	5.28

The results in Table 7 reveal that with the passage of time, three pooled ratios (ethnicity, households with young children and rented accommodation) have decreased markedly (i.e. relatively disadvantaged status has improved markedly) while two pooled ratios (educational status and occupational status) have decreased slightly (i.e. relatively disadvantaged status has improved slightly). The pooled ratio for employment status has essentially remained unchanged over time.

All three sets of results (from Tables 6 and 7) have been combined in a single graphical comparison in Figure 3.

Figure 3:

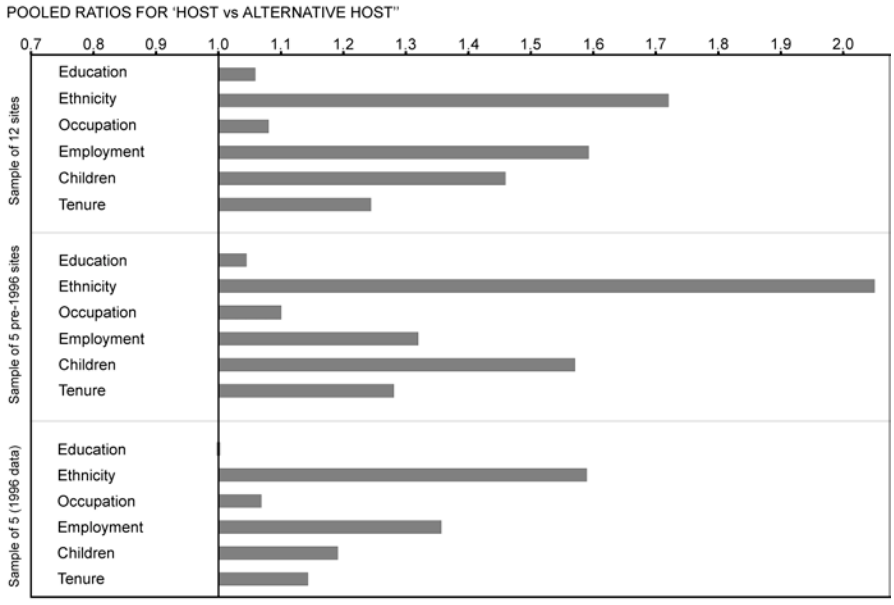


Table 8 presents the corresponding results for host-alternative host comparisons of median household income.

Table 8: Pooled results of 'host-alternative host' comparisons of median household income

	Median household income
Full sample of 12 landfill sites with alternatives:	
Pooled Ratio values	0.94
Host-alt.host relativity	disadv.
Sample of 5 sites with alternatives (pre-1996 data):	
Pooled Ratio values	0.88
Host-alt.source relativity	disadv.
Sample of 5 sites with alternatives (1996 data):	
Pooled Ratio values	0.89
Host-alt.source relativity	disadv.

The results in Table 8 show a trend that is consistent with those of the other demographic indicators. The smaller sample of earlier site selection decisions displays more comparative disadvantage than the full sample of 12, while the relative disadvantage for these five cases has not become more accentuated over time.

D: Discussion and Conclusions

Basis for interpretation

The quantitative statistical comparisons presented in this Working Paper are based on a group of discrete demographic indicators which reflect the comparative status of pairs of communities with respect to educational status, ethnicity, occupational status, employment status, the prevalence of families with young children, and the status of dwelling tenure.

Originally, these demographic variables were selected because they were considered to correlate with a community's ability to participate in and influence the planning process. Specifically, the research inferred that communities with a population prevalence of low educational achievement, minority ethnic groups, low occupational status, high unemployment rate, households with young children, low dwelling ownership, and low median household income would be likely to have greater difficulty in understanding, participating in, or in any way influencing the kind of decisions which lead to the siting of solid waste facilities. It was intended that this inference be taken in a comparative rather than an absolute sense, which is particularly relevant to cases where the planning process involves consideration of several alternative or 'competing' prospective sites.

In the context of this longitudinal analysis, the same grouping of indicators can be taken to correlate with a community's ability to participate in and influence on-going management of the solid waste facilities, particularly when Territorial Local Authorities have generally not been pro-active in establishing mechanisms for community involvement.

As was the case with the results presented previously in Working Paper FS1, the emphasis is on the pooled comparisons across samples of sites and their communities. It should also be added that the primary emphasis for interpreting these results should be on the cluster of indicators taken as group rather than individually. A single demographic indicator, considered in isolation, may not be indicating 'powerlessness'. It is the compounding effect of a group of indicators - particularly when a consistent pattern emerges - that is likely to indicate a condition of relative powerlessness.

Trends identified

This analysis has confirmed two trends.

Firstly, by focussing on a smaller sub-sample of earlier site selection decisions for the base year analysis[8], the previous finding - a greater incidence of significant disadvantage in earlier siting decisions and a lower incidence in more recent decisions - has been reinforced.

Secondly, the results of both longitudinal analyses (i.e. 'host-source' comparisons and 'host-alternative host' comparisons) suggest that host communities have <u>not</u> systematically become more disadvantaged as a result of hosting landfill facilities. In fact, the differences evident at the time of site selection appear to have diminished.

[8] Whereas the previous analysis looked explicitly at clusters for 1977-91, 1992-97 and those sites yet to be opened (YTBO), this analysis simply dropped out all sites for which the only demographic data available came from the 1996 census.

For the 'host-source' comparisons, relative disadvantage diminished for four out of seven indicators, remained the same for one indicator, and increased for two indicators, with the magnitudes of the improvements being substantially greater than the magnitudes of the declines. For the 'host-alternative host' comparisons, relative disadvantage diminished for five out of seven indicators, while remaining the same for the other two indicators.

Conclusions

This report has demonstrated that relative improvements in demographic comparisons have occurred within host communities of landfills over time. Whether or not these relative improvements are in any way attributable to the way the facilities and their environs have actually developed over time cannot be deduced from the analysis in this report. It is possible, even likely, that other factors for change have had a bearing over time.

It is noteworthy however, that the off-site effects experienced from landfills are generally not as spatially far-reaching as originally expected, and that the experience of actual effects and their associated impacts is generally less offensive than anticipated (Taylor Baines & Associates).

A number of factors fed into this conclusion. Many host communities were experiencing the transition from old and often unsanitary dumps to newer and more sanitary landfills, brought about both by rising community expectations and the requirements of the Resource Management Act. It is also likely that neighbours become accustomed to effects over time. It was a common experience in the case studies carried out that other activities in the locality of a landfill produce similar kinds of effects, but with less negative connotations (e.g. logging trucks, quarrying activity, commercial orchard spraying, chicken and pig farming, and so on). However, stereotype connotations can and do change over time. The real transformations in waste management practices will in time translate into different popular experience, and thence into different perceptions and attitudes. Furthermore, another finding of the case study research was the positive experience of buffer zones around many landfills - areas of land in very close proximity to landfill boundaries, in which a variety of recreational uses have evolved because occasional occupation leads to far lower incidences of unpleasant off-site effects than would be so for permanent residence. Such buffer zones have often come to be viewed as positive local recreational amenities.

With these factors in mind, this analysis does suggest that the siting of landfill facilities has not resulted in some kind of long-term social and community blight, when assessed using the indicators employed in this piece of social research.

References

Taylor Baines & Associates (1999): Solid Waste facilities and their host communities. Working Paper FS1, April 1999. 36p.

Taylor Baines & Associates (2000): Contemporary host community experience of solid waste facilities in New Zealand. Working Paper FS10, June 2000.

Bullard, Robert D (1994): "Dumping in Dixie: Race, Class and Environmental Quality" Boulder Colorado: Westview Press.

Appendix I: Methodological note

Notes on statistical analysis for Siting of Landfills project

Patrick Graham

pgraham@chmeds.ac.nz

25th June 1998

1 Preliminary comments

1.1 estimation or hypothesis testing

Given that census data is being used for the comparisons, it is likely that sample sizes will be large and that classical significance tests will yield "significance" even for differences which are fairly trivial. It may be better to focus on estimation of differences, via confidence intervals, rather than hypothesis testing.

1.2 Types of comparisons

Comparisons can either be made between host areas and source areas or between host and non-host areas and both relative and absolute differences

between host and comparison areas can be considered.

1.3 Notation

p_h –proportion with the characteristic (e.g.) tertiary qualified in the host area.

p_{nh} – proportion with the characteristic in non-host areas.

N_h – number of individuals resident in the host area.

N_{nh} – number of individual resident in the non-host area.

N_s – number of individuals resident in the source area.

$p_s = (N_h p_h + N_{nh} p_{nh})/N_s$ – proportion with the characteristic in the source area.

2 Statistical methods for comparing host areas with source areas.

2.1 inference for observed:expected ratio

The ratio of the proportion with the characteristic in the host area to the proportion expected if the overall source area prevalence of the characteristic held also in the host area is $R = p_h/p_s$. Suppose $R = 1.3$ and the characteristic of interest is the proportion renting their home, then we are saying that the proportion of people renting is 30% greater in the host area than in the source area as a whole. Note this is *not* a difference of 30 percentage

points, i.e. we are talking about the difference between, say, 10% and 13% not 10% and 40% In order to develop methods for statistical inference for the observed: expected ratio it is easiest to work on a log scale (natural logarithm) and convert back to the original scale at the end. This is illustrated below.

The variance of the natural logarithm of the observed:expected ratio, R, is

$$V(\ln(R)) = \frac{V(p_h)}{p_h^2} - 2\frac{Cov(p_h, p_s)}{p_h p_s} + \frac{V(p_s)}{p_s^2}$$

where

$$V(p_h) = \frac{p_h(1-p_h)}{N_h}$$

$$Cov(p_h, p_s) = \frac{p_h(1-p_h)}{N_s}$$

$$V(p_s) = \frac{n_h p_h(1-p_h) + N_{nh} p_{nh}(1-p_{nh})}{N_s^2}.$$

A 95% confidence interval for the observed to expected ratio is obtained as

$$\exp(ln(R) \pm 1.96\sqrt{V(ln(R))})$$

A significance test can be based on the statistic

$$\chi^2 = \frac{(ln(R))^2}{V(ln(R))}$$

which can be compared to the chi-square distribution with 1 degree of freedom. Values of χ^2 greater than 3.84 are said to be statistically significant at the 5% level. (i.e. the dreaded $p < 0.05$).

2.2 Statistical inference for the difference between observed and expected prevalence

Let $D = p_h - p_s$. Then $D = 0.05$ indicates that the prevalence of the characteristic is five percentage points greater in the host area than in the source area as a whole.

The variance for D is given by:

$$V(D) = V(p_h) - 2Cov(p_h, p_s) + V(p_s)$$

where $V(p_h)$, $Cov(p_h, p_s)$ and $V(p_s)$ are as defined above. A 95% confidence interval for D is given by:

$$D \pm 1.96\sqrt{V(D)}$$

and a significance test can be based on:

$$\chi^2 = \frac{D^2}{V(D)}$$

which can again be compared to tables of the chi-square distribution with one degree of freedom.

3 Statistical methods for comparing host and non-host areas.

When comparing host and non-host areas the ratio and difference measures are $R = p_h/p_{nh}$ and $D = p_h - p_{nh}$. The variance for these measures is given by:

$$V(\ln(R)) = \frac{V(p_h)}{p_h^2} + \frac{V(p_{nh})}{p_{nh}^2}$$

$$V(D) = V(p_h) + V(p_{nh})$$

where

$$V(p_{nh}) = \frac{p_{nh}(1 - p_{nh})}{N_{nh}}$$

and $V(p_h)$ is as defined above.

Confidence intervals and significance tests can be computed using the same formulae given in section 2, but with these modified versions of R, D, $V(\ln(R))$ and $V(D)$.

The methods described here for comparing host and non-host areas are also appropriate for comparing host and "almost-host" areas.

4 Pooling results over sites

The above calculations will be carried out for each site and each characteristic. To summarise these results it would be reasonable to compute an overall summary effect for each characteristic, in effect pooling the information from the various sites. There are several possibilities here but the simplest approach is described below.

4.1 Ratio measures.

Let R_i denote the ratio measure (host v source or host v non-host) for some characteristic for the i^{th} site. Let $W_i = 1/V(ln(R_i)$, the inverse of the variance for the natural logarithm of the ratio for the i^{th} site and let

$$\overline{L} = \frac{\sum_i W_i ln(R_i)}{\sum_i W_i}$$

then $V(\overline{L}) = (\sum_i W_i)^{-1}$ and an overall summary measure of the ratio of host to source area prevalence is given by: $\overline{R} = \exp(\overline{L})$ and a 95% confidence interval is given by:

$$\exp(\overline{L} \pm 1.96\sqrt{V(\overline{L})}.$$

A one degree of freedom chi-square test can be based on the statistic

$$\chi^2 = \frac{\overline{L}^2}{V(\overline{L})}.$$

4.2 Difference measures

Let D_i denote the difference measure (host v source or host v non-host) for some characteristic for the i^{th} site and now let $W_i = 1/V((D_i)$, denote the inverse of the variance for difference measure for the i^{th} site. An overall summary measure of the difference between host and comparison area prevalence is:

$$\overline{D} = \frac{\sum_i W_i D_i}{\sum_i W_i}$$

and $V(\overline{D}) = (\sum_i W_i)^{-1}$. A 95% confidence interval is given by

$$\overline{D} \pm 1.96\sqrt{V(\overline{D})}.$$

A one degree of freedom chi-square test can be based on the statistic

$$\chi^2 = \frac{\overline{D}^2}{V(\overline{D})}.$$

4.3 Heterogeneity

The ratio and/ or difference measures may vary considerably over site, in which case it may not be sensible to compute a pooled summary effect. Al-

though there are statistical tests available for testing heterogeneity of effects across sites these will probably not be useful here, because, with the large sample sizes likely to be available in this study, even trivial heterogeneity will be declared statistically significant. Some judgement will have to be exercised in order to decide whether pooling is sensible. If the ratio or difference measures do seem to vary markedly over sites then it would be interesting to try and relate this to characteristics of the sites (e.g rural v urban may be one site characteristic). This could be achieved by grouping the sites into sub-groups defined according to certain characteristics, checking whether the effect measures seemed relatively homogeneous within subgroups, computing pooled summary estimates for each subgroup and comparing these pooled subgroup estimates.

Other more formal statistical methods may also be relevant here but this would depend to some extent on the software you have available and the number and type of site characteristics considered relevant. This could be discussed at a later date after you have had a preliminary look at the data.