Appendix 2: CareTrack weighting methods and results

Introduction

CareTrack Australia used a two-stage clustered sample design to recruit participants. Households were selected from the Telstra White pages of two Australian states (New South Wales and South Australia). These households were selected from designated local government areas chosen to be representative of the socioeconomic profile and location of the Australian population. To address the potential sampling and non-sampling biases generated by the nature of the study design, two different options were used as a basis for the weighting process. Five versions of weights were finally created that could be used to generate weighted estimates of compliance that could be compared with the unweighted results, with the aim of demonstrating how sensitive the final results were to a range of assumptions.

The two weighting options are:

Option 1: To assume that the final sample is representative of people with the 22 CareTrack conditions, and

Option 2: To create post-stratification weights from the first interview and adjust these for non-response, with this adjusted weight applied to the final sample data. This was the approach used by the United States (US) study [1].

A third method (condition-based weighting) was also considered. This requires information on the prevalence of each condition by sex and age group from the general population. This method was not undertaken because population-level data was not available and not necessarily appropriate for all 22 conditions. In addition, sample sizes for some of the conditions were too small and there was a lack of independence between the various conditions.

Prior to discussing the weighting methods, we present the prevalence of selected conditions in the sample data compared to prevalence estimates reported by the National Health Survey (NHS) 2007-2008 [2].

Prevalence of conditions by sample and population

The final sample will not be representative of the overall general population because inclusion criteria restricted this sample to respondents who reported having at least one of the 22 conditions. In addition, the presence or absence of a condition was based on whether or not evidence was found in the medical record, rather than being based on self-reported conditions.

Because the sample is not representative of the general population it is expected that the prevalence of conditions in the final sample will differ from the prevalence in the general population. The prevalence of sixteen conditions in the final sample was compared with prevalence data for the Australian population as reported in the 2007–2008 National Health Survey (NHS) (Table 1) [2]. Information on the prevalence of the remaining six conditions included in CareTrack was not available. For those CareTrack conditions with a sample size greater than 100, gender-level prevalence was compared with the relevant NHS-based data. NHS results are also available for 10-year age groups; however age comparison was not conducted, as the condition-level numbers in our sample were deemed too low to draw any justifiable comparisons.

Table 1: CareTrack condition prevalence compared to prevalence in Australian population as reported in the 2007 – 2008 National Health Survey.

Condition	Number persons in CareTrack final sample	Prevalence (unweighted) in CareTrack sample (%)			Prevalence in Australian population (NHS 2007-2008) (%)		
		Overall	Male	Female	Overall	Male	Female
Hypertension	351	30.4	31.6	29.6	9.4	8.8	10
Osteoarthritis	188	16.3	14.3	17.7	7.8	5.9	9.7
Hyperlipidemia	186	16.1	17.6	15	5.7	6.0	5.4
Dyspepsia	180	15.6	12.9	17.6	2.8	2.9	2.8
Low back pain	164	14.2	15.4	13.4	13.8	14.1	13.5
Coronary artery disease	131	11.4	16.4	7.7	3.8	5.1	2.4
Depression	112	9.7	9	10.2	7.4	6.2	8.7
Diabetes	96	8.3	-	-	3.5	-	-
Obesity	67	5.8	-	-	24.6	-	-
Osteoporosis	60	5.2	-	-	3.4	-	-
Asthma	60	5.2	-	-	9.9	-	-
Atrial fibrillation	59	5.1	-	-	0.6	-	-
Chronic heart failure	30	2.6	-	-	1.3	-	-
Chronic obstructive pulmonary disease	28	2.4	-	-	2.4	-	-
Panic disorder	25	2.2	-	-	3.3	-	-
Cerebrovascular accident	19	1.6	-	-	1.2	-	-

Weighting methods

Option 1: Weighting the final data

This version of weighting considers only the 1,154 participants that were included in the final sample. Eligibility criteria for the CareTrack final sample included (i) participant was 18 years old or over and (ii) participant had at least one of the 22 study conditions and/or (iii) the participant had been admitted to hospital during 2009 and 2010.

As a result of the eligibility criteria, the sample will not be representative of the general population and so it is not therefore appropriate to weight to the general population.

The population chosen was a compromise between the general population and the population specific to each condition. The results from the 2007-2008 NHS [2] include information on the number of people with a current long-term condition and hence provides a more applicable age distribution to our study than that of the general population, because it is skewed towards higher age groups relative to the population distribution of all people aged 15 years and over (Figure 1). Because information on the proportion of the population with a current long term health condition is only available by age group or by gender (not both), and the lowest age group is 15-24, the following assumptions were made:

1. The proportion of the population with a current long-term condition in the 18-24 year age group is the same as in the 15-24 year age group.
2. The age and gender proportions of the population presented in the NHS report are correct.
3. Within both males and females the age distribution of the proportion with current long-term conditions is the same.

The difference in age profiles between the overall population, the population of people with a current long term condition and the sample are shown in the Figure 2.
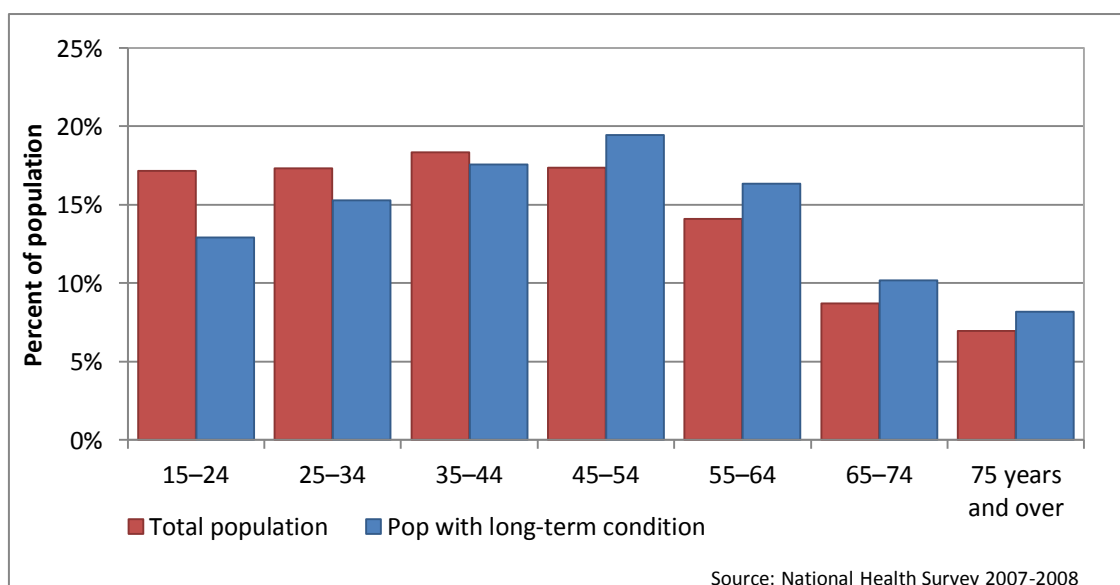

Source: National Health Survey 2007-2008

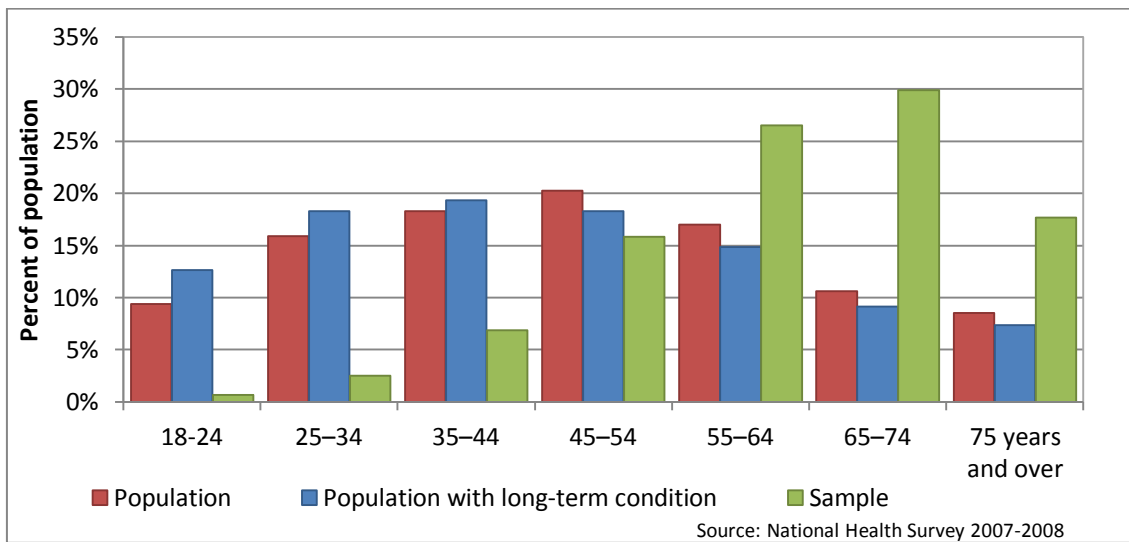**Figure 1: Population distribution by age group (age 15+) with long term condition and overall.**

Figure 2: Comparison of age distribution for the final sample, the total population and the population with current long term condition.

The final sample obtained in CareTrack was skewed to older age groups, particularly because of the conditions required for inclusion at the final survey. There were only nine observations in the 18-24 year age group in the final sample, so these were aggregated with the 25-34 year age group for weighting purposes. This weighting method is referred to as FinalOnly1. When combined with the 25-34 year age group, the respondents in this combined age group became highly weighted. In order to assess the sensitivity to these large weights a second, alternative set of weights (FinalOnly2) collapsed the 18-34 and 35-44 year age groups. Collapsing age-sex cells to reduce the range in the weights is a standard method in survey analysis [3]. This reduced the range of weights and also reduced the estimated design effect (the increase in sample size required to obtain the same standard error as a simple random sample would provide) from 3.6 for FinalOnly1 to 2.4 for FinalOnly2. Although the age distribution of the FinalOnly2 may not agree exactly with the age distribution of the standard population (unless the collapsed age groups are used), (Figure 3), the overall comparison of the weighted distribution in comparison to the population with current long-term conditions is more similar than before the weighting was undertaken (compare Figure 2 and Figure 3). Findings for both FinalOnly1 and FinalOnly2 are presented in the Results section.
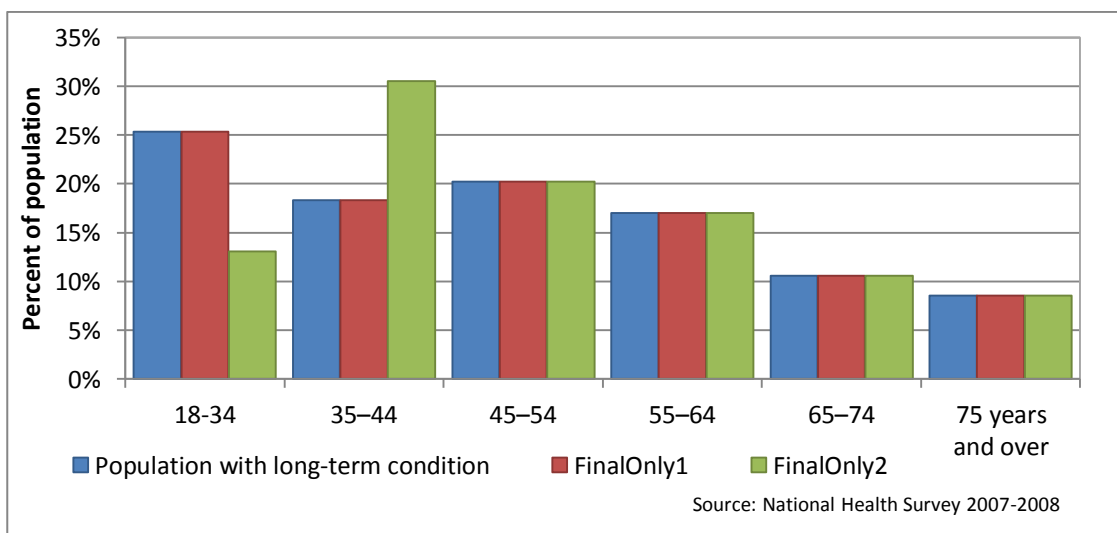

Figure 3: Comparison of age distribution of Australians with a long-term condition with the age distributions of the CareTrack sample using two weighting options based only on the final survey data.

Option 2: United States (McGlynn) paper weighting method

The initial sample used in the US study [1] was a sample of adults that had responded to another health-related study: the Community Tracking Study, and were re-contacted for the US study [1]. Logistic regression was used to adjust for the different probabilities of response. The covariates included in the analysis were reported as age, sex, race, education level, income, self-reported use of hospital and GP services, insurance status and health status, and participation in the study. They used the coefficients from fitting this logistic regression to "adjust the scores for non-response, and weighted the data for the participants to be representative of the population from which they were drawn" [1]. In other words, the initial sample data were weighted and then these weights were adjusted using the results of the logistic regression.

In the context of our study, to replicate this method, the following steps were required:

Step 1: To weight the CareTrack first interview data to the population from which the sample was drawn.

Step 2: Use logistic regression to adjust for non-response bias.

Step 3: Combine the product of both weights in the analysis of the final CareTrack sample.

Step 1. To weight the first interview data

The sample for the first interview can be considered to be a random sample of the general population. Sampling weights were created to take into account that the probability of selection differed between the strata and clusters. Some aggregation of age groups was necessary in order to restrict the weights to a reasonable range

The sampling weights were then adjusted to the age-sex population for Australia, with age groups in four broad categories (18-39, 45-54, 55-74, 75+).

The resultant weights ranged from 388 to 9,598 with an estimated design effect (DEFF) of 1.77.

Step 2. Use logistic regression to adjust for non response

Logistic regression was undertaken to determine the probability of responding at the final sample. Responses to a series of questions asked of respondents at first interview were used as independent variables. These included age, sex, household size, whether English was the main language spoken at home, and a series of questions that measured understanding of health information. The people who responded were more likely to be older, to speak English at home, to have good health status, to be in the central quintiles for the index of relative socioeconomic status, to have a college or university level education, be over 65, have confidence in filling out health-related paperwork and be retired or in a employment status group that included students, people whose work was 'home duties' and unemployed people (Table 2). The probability of response ranged from 0.01 to 0.5, with a median of 0.146.

Step 3: Combine weights in the analysis of the final CareTrack sample.

The product of the weight from the first survey and the propensity weighting provides weights for the 1,154 respondents in the final survey. These weights ranged from 203 to 25,566, with a median of 1,130. In the combined results table below these are referred to as McGlynn1. These weights were trimmed (McGlynn2) and then readjusted to the Australian population (McGlynn3). The most appropriate version of the US weights is McGlynn2, as it takes the combined weights trimmed to remove the extremes. The McGlynn3 weights had an extreme range, (from 777 to 232,880) partly due to the small sample sizes at the final sample in the remote areas and in the youngest male age group, and also because the weights were adjusted to the Australian general population. Whether it is appropriate to adjust these weights to the Australian population is a moot point, because the final sample was not representative of the general population, however given that the aim of the weighting process was to assess how sensitive the results were to the various weight assumptions, all three of the US study-based weights were used to create weighted estimates of overall and condition level.

Table 2: Results of logistic analysis of the probability of response at final survey

Effect	Comparison	Odds Ratio	95% Confidence Limits	
Broad age group	18 to 39 years	1.00	-	-
	40 to 54 years	2.66	1.97	3.59
	55 to 74 years	4.03	2.99	5.44
	75 + years	4.34	3.02	6.25
English main language spoken at home	Yes vs No	2.19	1.09	4.42
Household size	Other responses	0.94	0.29	3.07
	1	0.96	0.59	1.54
	2	1.22	0.76	1.95
	3	1.05	0.62	1.77
	4 – 8	1.00	-	-
Health status	Excellent or very good vs remainder	0.70	0.61	0.80
IRSD quintile 1 = disadvantaged 5 = least disadvantaged	1	0.97	0.32	2.89
	2	1.89	1.37	2.60
	3	1.80	1.27	2.54
	4	1.73	1.27	2.34
	5	1.00	-	-
Been to University or College	Yes vs No	1.32	1.14	1.52
Region sampled	Eastern Adelaide	1.00	-	-
	Eyre and Western	1.47	0.63	3.39
	Sthn Adelaide	0.74	0.54	1.02
	Limestone Coast	1.08	0.79	1.49
	Metro Hunter	0.67	0.50	0.91
	Regional Hunter	0.50	0.36	0.70
	Western NSW	1.01	0.29	3.52
Age over 65	Yes vs No	1.44	1.18	1.75
Confidence in filling out medical-forms by yourself	Yes vs No	1.35	1.12	1.63
Occupation code	Employed	1.00	-	-
	Retired	1.40	1.14	1.72
	Other (unemployed, student, household duties etc)	1.28	1.02	1.60

Results for different weighting approaches and sensitivity analyses

Table 3 reports the percentage compliance for each condition in the CareTrack final sample using weighted and unweighted results, thus providing an estimate of the sensitivity of the final results to the various weighting approaches.

Table 3: Percentage compliance with Care Track Indicators by condition and overall, by different versions of weighting.

Condition	Published compliance rate	Option 1. Weighting of final data only		Option 2. Weighting using the United States approach		
	Unweighted	FinalOnly1	FinalOnly2	McGlynn1	McGlynn2	McGlynn3
Alcohol dependence	12.8	5.8	6.4	5.4	5.6	4.8
Antibiotic use	19.0	25.0	18.5	11.9	12.8	10.8
Obesity	24.0	22.3	21.6	22.6	22.7	22.4
Hyperlipidemia	34.8	35.5	34.8	37.1	36.9	36.7
Asthma	37.6	36.1	41.1	41.3	41.0	45.0
Surgical site infection	37.9	36.3	39.8	36.8	37.2	35.8
Preventive care	42.2	42.2	43.2	42.2	42.2	44.9
Osteoarthritis	43.0	45.6	46.4	41.9	41.9	42.1
Community acquired pneumonia	52.4	43.2	47.2	44.4	45.6	41.3
Cerebrovascular accident	53.1	53.5	53.5	46.7	46.7	51.4
Atrial fibrillation	55.0	57.0	57.0	52.5	52.5	52.9
Depression	55.2	65.3	62.4	64.6	63.2	73.8
Osteoporosis	55.3	55.4	55.0	52.8	52.8	54.2
Venous thromboembolism	58.2	56.9	57.7	55.8	57.3	56.1
Diabetes	62.9	59.2	58.8	59.1	58.8	59.4
Chronic obstructive pulmonary disease	70.8	70.6	70.6	68.1	68.1	66.7
Hypertension	71.6	65.2	65.6	68.8	68.1	64.9
Low back pain	71.6	69.5	70.6	69.7	70.4	70.0
Panic disorder	71.6	78.6	73.8	69.1	69.2	71.0
Chronic heart failure	76.2	78.4	78.4	77.3	77.3	77.0
Dyspepsia	78.1	79.9	82.2	79.4	79.2	83.8
Coronary artery disease	89.9	90.2	90.2	89.5	89.4	88.7
Overall	57.2	56.4	54.0	56.3	56.3	57.2

These results are shown graphically in Figure 4, together with the 95% confidence intervals for the unweighted results. It can be seen that the only condition for which there is any marked difference in compliance between weighted and unweighted results is for *depression* with the McGlynn3 version. The confidence interval around the McGlynn3 estimate for depression ranges from 56.0 to 87.2, which overlaps with the CI around the unweighted estimate, and therefore, there is no significant difference between the two estimates. Even if this result was significant, in comparing so many estimates it would be expected that some would be significant, just by chance. The overall conclusion is that despite using two distinctly different weighting approaches, and various versions of the weights created by these approaches, the estimates are very similar, and given the complexity of explaining either weighting method and the lack of difference the weighting method has on the results it was decided to present the unweighted results in the paper.
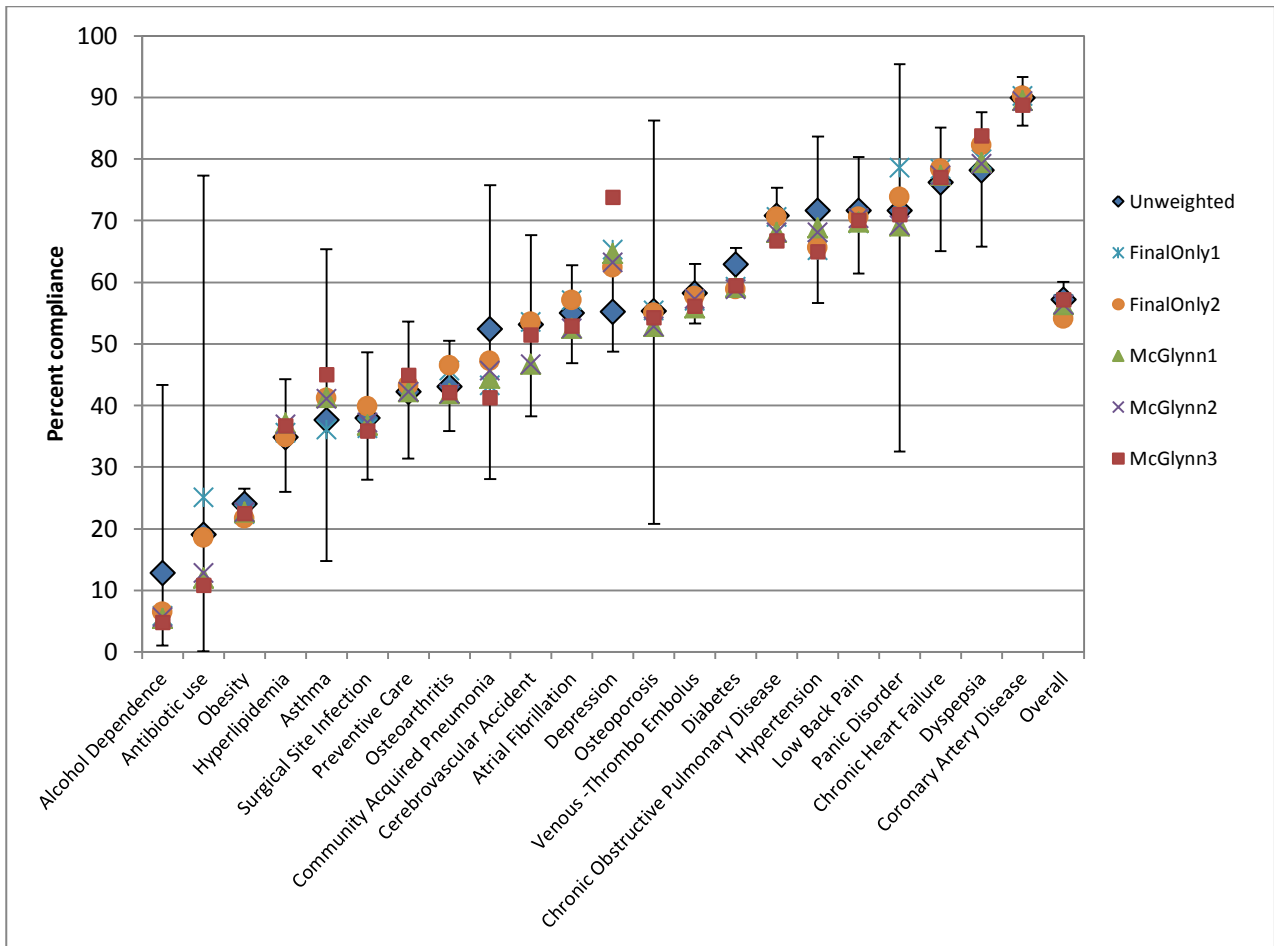
Figure 4. Comparison of weighted results with the unweighted results presented in the manuscript. 95% confidence interval around the unweighted results is displayed.

Conclusion

Obtaining a "representative" sample of the Australian population that is logistically accessible for medical record review within the scope of the allocated budget was a difficult process. These issues, as well as other inclusion and exclusion criteria (having seen a healthcare provider in timeframe, healthcare provider consenting to access to medical records, aged > 18) introduce numerous potential sampling and non-sampling biases. To address this, CareTrack examined the effect on the results for compliance of weights created using two distinctly different approaches and compared these results with the unweighted data. This determined that results obtained from all the weighting methods (Figure 4 and Table 3) were not significantly different to the unweighted results for all conditions and overall. Thus we conclude that the CareTrack results are robust.

References

1. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003 Jun 26;348(26):2635-2645.
2. Australian Bureau of Statistics. National Health Survey: Summary of Results, 2007-2008 (Reissue). Accessed online April 2012 @ http://wwwabsgovau/AUSSTATS/abs@nsf/Lookup/43640Main+Features12007-2008%20(Reissue)?OpenDocument. 2009.
3. Lohr, Sharon (1999) Sampling: design and analysis. Duxbury Press, Pacific Grove California.