

12-2010

Cancellation of indigenous Australians from the apprenticeship training contract

John Mangan

Bernard Trendle

Follow this and additional works at: <https://ir.lib.uwo.ca/aprci>



Part of the [Education Economics Commons](#)

Citation of this paper:

Mangan, John and Trendle, Bernard, "Cancellation of indigenous Australians from the apprenticeship training contract" (2010).
Aboriginal Policy Research Consortium International (APRCi). 411.
<https://ir.lib.uwo.ca/aprci/411>

This article was downloaded by: [University of Western Ontario]

On: 18 November 2012, At: 07:13

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Education Economics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/cede20>

Cancellation of indigenous Australians from the apprenticeship training contract

John Mangan ^a & Bernard Trendle ^b

^a Department of Economics, School of Economics, The University of Queensland, Brisbane, QLD 4072, Australia

^b Labour Market Research Unit, Department of Education and Training, Brisbane, Australia

Version of record first published: 15 Apr 2010.

To cite this article: John Mangan & Bernard Trendle (2010): Cancellation of indigenous Australians from the apprenticeship training contract, *Education Economics*, 18:4, 377-394

To link to this article: <http://dx.doi.org/10.1080/09645290903546587>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Cancellation of indigenous Australians from the apprenticeship training contract

John Mangan^{a*} and Bernard Trendle^b

^a*Department of Economics, School of Economics, The University of Queensland, Brisbane QLD 4072, Australia;* ^b*Labour Market Research Unit, Department of Education and Training, Brisbane, Australia*

(Received May 2009; final version received December 2009)

The vocational education and training (VET) sector is a major pathway to post-school education for indigenous students, yet questions are being raised about the capacity of the VET system to provide successful outcomes for the indigenous apprentices and trainees it attracts. Within a system plagued by high cancellation rates in general, indigenous apprentices appear to do particularly badly. This paper combines data from an administrative database on apprenticeship with income data from the 2001 Census of Population and Housing to provide an analysis of attrition rates for apprenticeship training contracts in Queensland, asking: Are cancellation rates for indigenous students significantly higher than those for non-indigenous students, and, if so, what factors are responsible for this?

Keywords: indigenous; vocational education; pathways; cancellation rates; duration analysis

1. Introduction

It is well known that there are a range of issues that affect participation of indigenous Australians (ATSI¹) in post-school education, including access to educational institutions, financial constraints and community expectations. To help overcome some of these barriers, Technical and Further Education (TAFE) and other vocational education and training (VET) providers offer free or heavily subsidised tuition to ATSI, social security holders and other groups thought to be disadvantaged within the Australian education system. This equity and community service aspect is an important component of the VET output, which, in terms of improved access, appears to have paid dividends. For example, in 2001, indigenous persons comprised 2.2% of the population, yet they comprised between 3% and 4% of the VET students.² Furthermore, the Australian Bureau of Statistics (ABS) reports that the number of indigenous students in the VET increased by 60% between 1996 and 2000.³

Despite these apparent successes, there are differing opinions as to how successful the VET-based equity and access programmes have been in terms of achieving desired outcomes and whether or not the current VET system in Australia is well suited to the task of reducing socio-economic inequalities among the disadvantaged groups. Critics point to the high attrition rates among apprentices, which are well in excess of those

*Corresponding author. Email: j.mangan@uq.edu.au

for university students in the same age cohort, and question whether such an apparently unstable environment is suited to the educational needs of the disadvantaged groups. These objections are reinforced by the fact that indigenous students have even higher attrition rates than the non-indigenous students. This paper presents the first econometric study into the differential cancellation rates between indigenous and non-indigenous students. It asks the questions: Are cancellation rates for indigenous students in the VET significantly higher than those for non-indigenous students, and, if so, what factors account for this?

The paper proceeds in the following way: Section 2 provides an overview of the status of indigenous persons in the Queensland labour market and highlights the differences between the observed profiles of indigenous and non-indigenous persons using the data from the 2001 and 2006 Census of Population and Housing. Section 3 provides a preliminary analysis of the differences in the rate of cancellation between the indigenous and the non-indigenous students using the Direct Entry Level Training Administration (DELTA) database, administered by the Department of Education and Training. Section 4 examines the causes of these differences through a multivariate Cox proportional hazard (CPH) model, which provides estimates of the significance and magnitude of the variables impacting on the risk of cancellation. Section 5 contains conclusions and policy analysis.

2. Apprenticeship and indigenous Queenslanders

It is well known that indigenous persons face considerable disadvantages in the labour market both in Australia and in other places such as New Zealand and Canada (Hull 1996; Ball and Pence 2001; Skill New Zealand 2001; Byam 2002). For example, the data from the Australian Census of Population and Housing (2001) indicate that indigenous persons recorded an unemployment rate of 20.0%, far in excess of the average unemployment rate (7.9%) for all persons in Queensland. At the same time, their participation rates at 55.4% were also substantially lower than the average for the entire population (63.5%). In line with the general tightening of the labour market over the inter-censal period, absolute outcomes for indigenous Australians improved, but their relative position did not. In 2006, the unemployment rate for indigenous Queenslanders (13.1%) remained 8.6% points above the average unemployment rate for all persons, whereas the participation rate was 9.2% points lower at 56.2%.

Location plays some role in these outcomes. For example, the proportion of indigenous persons residing in rural or remote regions is well above that for the population as a whole, but other factors contribute significantly to these differences in labour market outcomes. For example, as shown in Table 1, indigenous persons tend to possess a lower level of school and post-school qualifications, with this difference apparent in both 2001 and 2006 census data.

The information presented in Table 1 indicates that, overall, 38.6% of the non-indigenous population had a Year 12 qualification in 2001, rising to 44.0% in 2006. Furthermore, higher proportions of the non-indigenous population had post-school qualifications, with the largest percentage point differences being among those with a bachelor degree or similar qualifications (10.5% in 2006). This result did not vary greatly across the two census years shown in the table, although there was a slight widening of the gap with bachelor and higher-level qualifications (from 8.7% points to 10.5% points) and a narrowing of the gap with certificate-level qualifications (from 7.4% points down to a 4.9% point differential).

Table 1. Education profile of indigenous and non-indigenous persons, 2001 and 2006.

	2001				2006			
	Indigenous		Non-indigenous		Indigenous		Non-indigenous	
	No.	%	No.	%	No.	%	No.	%
School qualification								
Year 12	15,336	22.7	1,022,548	38.6	19,913	25.5	1,250,436	44.0
Year 10	18,907	28.0	776,233	29.3	22,307	28.6	803,528	28.3
Non-school qualification								
Bachelor degree and above	1840	2.7	302,485	11.4	2793	3.6	401,165	14.1
Advanced diploma and diploma	1769	2.6	153,374	5.8	2566	3.3	200,186	7.0
Certificate	6262	9.3	442,055	16.7	10,955	14.1	538,684	19.0
Population total 15+	67,563		2,651,084		77,954		2,841,057	

An examination of labour force participation by occupation, shown in Table 2, indicates that indigenous persons tend to be employed mostly in the lower-skilled and lower-paid occupations. Particularly noteworthy in this table is the large proportion of indigenous persons employed as labourers or related workers in 2001, 28.2% of total

Table 2. Indigenous and non-indigenous employment by first division ASCO occupational category in 2001 and 2006.

	2001				2006			
	Indigenous		Non-indigenous		Indigenous		Non-indigenous	
	No.	%	No.	%	No.	%	No.	%
Managers and administrators	898	3.2	131,701	8.8	1098	3.0	144,160	8.3
Professionals	2928	10.4	247,308	16.5	3768	10.4	298,502	17.1
Associate professionals	2361	8.4	184,629	12.3	3132	8.6	218,470	12.5
Tradespersons and related workers	2949	10.5	196,429	13.1	4367	12.0	233,473	13.4
Advanced clerical and service workers	506	1.8	53,963	3.6	515	1.4	55,064	3.2
Intermediate clerical, sales and service workers	5309	18.8	259,144	17.3	6976	19.2	304,226	17.4
Intermediate production and transport workers	2852	10.1	129,863	8.6	4121	11.3	155,525	8.9
Elementary clerical, sales and service workers	2456	8.7	154,852	10.3	3464	9.5	172,432	9.9
Labourers or related workers	7939	28.2	143,458	9.6	8887	24.5	162,335	9.3
Total	28,198	100	1,501,347	100	36,328	100	1,744,187	100

employment of indigenous persons, compared with only 9.6% for the non-indigenous population. Data from the 2006 census suggest only a marginal improvement, with 24.5% of indigenous persons employed as labourers or related workers compared with a rate of 9.3% for the non-indigenous population. Traditionally, this occupation has been associated with low skill and is estimated to have a median weekly income of only around 80% of the state average, using the 2006 census data (see DETA 2008a). However, the low income of this occupation and the concentration of indigenous people working in it are insufficient to explain the income differential between indigenous and non-indigenous persons, with data from the 2006 census indicating that indigenous persons earned on average only 70% of the average for the total population (see DETA 2008b).

A number of alternative policy options are worthy of consideration when addressing these imbalances. One possibility is by increasing the intake of indigenous persons in trade training. This proposal has the potential to simultaneously increase the education profile, labour force participation rate and income of indigenous persons. Furthermore, given that relatively more indigenous persons reside outside the south-east corner of the state (67.5% of the indigenous population in 2006, compared with 32.0% of the non-indigenous population) where opportunities for professional employment are limited, apprenticeship seems a feasible pathway for improving the labour market outcome of indigenous workers. However, it is also likely that apprenticeship attrition rates for the indigenous population differ substantially from those of the non-indigenous population, a factor that may have important implications for the efficacy of such a scheme. For example, in the American setting, Bilginsoy (2003) notes that race is an important explanator of the observed variation in the risk of cancellation.

3. Differences in the cancellation rates of the indigenous and the non-indigenous population

To examine the differences between the cancellation rates across the indigenous and the non-indigenous cohorts, this study makes use of the DELTA database, administered by the Department of Education and Training. This database traces apprenticeship contracts in Queensland and contains information across three broad categories, which also provided the broad grouping of variables chosen for the empirical analysis: *apprentice characteristics*, including age, sex, education level at school completion, indigenous status, main language spoken at home and disability status; *employer characteristics*, including occupational designation, whether the person is directly employed by the employer or employed through a group training organisation (GTO); and *training characteristics*, i.e., whether the person was trained at a TAFE college, trained by a private registered training provider (RTO) or trained by a non-TAFE state government employer.

The first step of the analysis of the database was the removal of all records of no relevance to the current project, i.e., those not in the Australian Standard Classification of Occupations (ASCO) 4 (tradespersons and related persons occupations). After this step, a number of incomplete records were found and these were also removed from the database prior to the formal analysis, as were all expired training contracts. Finally, a number of sub-categories of some variables that were small in number, e.g., there were less than 30 persons in the occupational category ASCO46, skilled agricultural and horticultural workers, along with the training category and state government

and all records in these two groups were also removed from the database. This cleaning of the database left complete records of 8061 persons. Of these, 74 persons were still active (undergoing training), 4006 persons had completed their training and the remaining 3891 persons had cancelled their training prior to completion. Of these 8061 persons in the database, 7760 (96.3%) were non-indigenous, whereas 301 self-identified as being indigenous. These training contracts have a defined start and end date, while there are also a small number of still active training contracts. Rather than discarding this information using conventional statistical techniques, tools from the field of survival analysis make it possible to use this information.

Table 3 provides details of the hazard of attrition (by dropout) by all the main characteristics captured in the database. The top row of the table presents information for all individuals commencing apprenticeship in Queensland in 2001. We can see that, after the first year, 28.1% have exited (dropped out of) the apprenticeship, with this figure increasing to 41.9% by the end of the second year. After four years, 50.4% of the candidates have left the apprenticeship.

Females are in the minority in the database (1388 females compared with 6673 males) and their chances of dropping out are noticeably higher, with 67.9% of females dropping out during the first four years compared with only 49.1% of males. The effect of age is unusual, with the incidence of dropouts initially decreasing when moving from the youngest age cohort (those aged 15–17) to the second youngest (those aged 18–19); however, for those aged 20–34, the rates of dropouts are much higher, which then taper off among the older age cohorts.

From the database, it is clear that there is also a significant difference between the rates of hazard of attrition by dropout for the indigenous and the non-indigenous cohorts, with 57.7% of the indigenous commencements exiting by dropout compared with 51.9% of the non-indigenous population.

A first consideration of the question of how different the risks of cancellation are for indigenous and non-indigenous persons can be provided by a simple, non-parametric survival technique, the Kaplan–Meier (K–M) estimate of the survival curve. The K–M survival techniques are a univariate, non-parametric form of survival analysis. These methods are simple to implement and produce informative graphs, but allow for the presence of covariates by stratification only, i.e., cancellation by gender, age, etc. Stratification is feasible only when there are a small number of strata and the sample size is large enough to ensure that there are sufficient training contracts to obtain reliable estimates of survival within each stratum.

Overall, the rate of cancellation for indigenous persons from apprenticeship training contracts in 2001 was 56.3% compared with 49.2% for the non-indigenous persons. This difference is highlighted by the different endpoints of the two series shown in Figure 1. When separate K–M estimates are fitted within strata defined by covariate groups, the indigenous status in this exercise, the log-rank test can be used to ascertain whether or not the estimated survival is the same in each stratum. This test compares the overall survival function and produces a p value based on the Chi-square distribution to test the null hypothesis of equal survival functions across the groups in the strata. A small p value leads to a rejection of the null hypothesis and to the conclusion that at least one group has survival times different from those of the others.⁴ The relatively low p value presented in Table 4 provides evidence of the difference in the risk of cancellation between the indigenous and the non-indigenous apprentices (at somewhere between the 5% and the 10% level of significance).

Table 3. Kaplan–Meier estimates of the cumulative proportion of dropouts from apprenticeship.

		<i>n</i>	Year 1	Year 2	Year 3	Year 4
All apprentices		8061	0.281	0.419	0.483	0.504
Gender	Male	6673	0.254	0.386	0.457	0.491
	Female	1388	0.384	0.572	0.654	0.679
Age	15–17	1949	0.270	0.442	0.527	0.560
	18–19	3207	0.247	0.379	0.443	0.471
	20–24	1916	0.330	0.461	0.539	0.571
	25–34	678	0.314	0.455	0.519	0.566
	35–44	227	0.208	0.341	0.397	0.466
	45+	84	0.169	0.194	0.297	0.351
ATSI	Indigenous	301	0.300	0.441	0.538	0.577
	Non-indigenous	7760	0.275	0.416	0.487	0.519
Disability	Able	7477	0.274	0.415	0.486	0.519
	Disabled	584	0.303	0.437	0.528	0.558
Language	English	7930	0.275	0.417	0.488	0.520
	NESB	131	0.302	0.440	0.548	0.635
Highest year	<Yr10	390	0.374	0.525	0.625	0.651
	Yr10	2350	0.320	0.474	0.553	0.586
	Yr11	1260	0.304	0.463	0.537	0.564
	Yr12	4061	0.232	0.360	0.426	0.460
Occupation	ASCO41	953	0.139	0.250	0.306	0.342
	ASCO42	1314	0.209	0.340	0.418	0.448
	ASCO43	893	0.143	0.259	0.333	0.408
	ASCO44	1563	0.251	0.372	0.447	0.468
	ASCO45	1815	0.458	0.647	0.715	0.743
	ASCO46	16	0.188	0.188	0.188	0.188
	ASCO49	1507	0.309	0.463	0.547	0.572
GTO	Other	6107	0.264	0.413	0.487	0.510
	GTO	1954	0.312	0.431	0.499	0.557
PSECTOR	Other	7783	0.283	0.427	0.500	0.533
	PSECTOR	278	0.079	0.155	0.206	0.229
RTO	Other	6063	0.268	0.409	0.477	0.504
	RTO	1998	0.301	0.441	0.531	0.583
CLOSED	Other	7556	0.248	0.386	0.462	0.496
	CLOSED	505	0.707	0.908	0.935	0.935
SGOV	Other	8048	0.276	0.417	0.489	0.522
	SGOV	13	0.385	0.385	0.473	0.473
Weekly income	<\$419.9	92	0.120	0.201	0.296	0.296
Q2	\$419.9–\$682.9	4462	0.348	0.511	0.585	0.612
Q3	\$682.9–\$945.9	3026	0.202	0.322	0.395	0.433
Q4	\$945.9–\$1208.9	473	0.102	0.201	0.250	0.295
Q5	\$1208.9–\$1471.9	8	0.475	1.000	1.000	1.000
Region	SEQ	3229	0.237	0.374	0.481	0.488
	Regional QLD	4832	0.302	0.446	0.516	0.544

Table 4. Log-rank test of differences between indigenous and non-indigenous apprentices.

	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$
Non-indigenous persons	7760	3812	3833	0.115	3.29
Indigenous persons	301	169	148	2.987	3.29

Note: $\chi^2 = 3.3$ on 1 df, $p = 0.069$.

3.1. Modelling differences in the risk of cancellation between the indigenous and the non-indigenous population

The differences in the risk of cancellation can be attributed to a wide range of factors, such as variation in the age composition, education level, potential income of the occupation and place of residence. For this reason, a clearer indication of the influence of indigenous status on the risk of cancellation needs to account for these differences across the two cohorts of this analysis.

From a theoretical perspective, a common way to represent voluntary student attrition is with a utility model in which an individual student i evaluates his or her utility in each period t from two alternatives, $j = 1, 2$, continuing on or dropping out from the course.

$$U_{ijt} = U(W_{ijt}, X_{ijt}) \quad (1)$$

where W_{ijt} represents student i 's expected wage from alternative j at time t and X_{ijt} represents the non-pecuniary attributes of alternative j at time t , such as study and working conditions, the geographic location and interaction with employers and other apprentices. X_{ijt} may also include characteristics of the student that influence how he or she evaluates utility from course attributes such as age, marital status and education

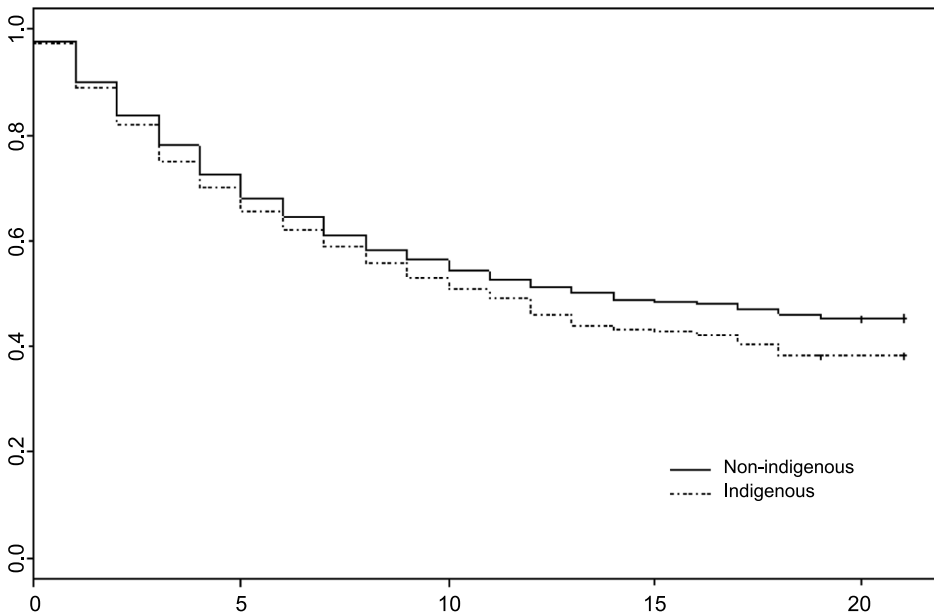


Figure 1. Kaplan-Meier survival curves for measuring the risk of cancellation for indigenous and non-indigenous persons.

level. Duration or survival models estimate the conditional probability that an apprentice drops out or remains. This problem cannot be dealt with using a univariate technique, which allows stratification by only one variable at a time. For this reason, a multivariate framework is adopted, allowing the testing of multiple (and competing) hypotheses. Towards this task, a model explaining the risk of cancellation is formulated, with the model attempting to explain the probability of completion. In this way, the influence of individual factors (represented in X_{ijt}) included in the model can be isolated, allowing a clearer understanding of the role of indigenous status on the risk of cancellation.

The personal characteristics of the apprentice may play a role in the probability of cancellation or completion, and the DELTA database enables the inclusion of several different variables that may systematically impact on the probability of a successful outcome from a training contract. Included in this database is the person's *gender*, *age* at commencement, *disability*, home *language* status and, importantly for the current work, their indigenous status, represented by the variable *atsi*. Education levels may also play a role in determining the probability of a completion. There is some variation in the level of education of persons commencing training contracts in 2001, with the highest year of schooling ranging from Year 8 to Year 12. It may be that persons with a higher level of schooling have a higher completion rate, and this is formally tested by including the highest level of schooling (*highyr*).

The characteristics of the job may also play an important role in the cancellation/completion decision, and the DELTA database includes variables capturing the occupation profile of persons undertaking formal apprenticeship training contracts. These are included in the model as 0, 1 dummy variables, i.e., a person who is employed as a mechanical and engineering tradesperson (*asco41*) would be recorded as a 1 under this variable, whereas persons employed in all other occupations would take 0 for that particular variable. In the formal modelling, one of the occupational dummy variables is omitted; in this exercise, it was *asco45* (ood tradespersons). This omission is necessary in order to circumvent the problem of perfect collinearity and means that the coefficients of the remaining occupational dummy variables measure the risk of completion relative to employment in *asco45*. The occupations represented in the model, apart from those already mentioned, comprise *asco42*, automotive tradespersons; *asco43*, electrical and electronic tradespersons; *asco44*, construction tradespersons; and *asco49*, other tradespersons and related workers.

The characteristics of the employer are also recorded on the training contract. For example, the person can be employed directly (*emprivate*), employed through a GTO (*empgto*) or employed by a state, local or commonwealth government organisation (*emppublic*). Like the occupation category variables, these variables are recorded as 0, 1 dummy variables, taking the value of 1 if the person is employed by a particular type of employer and 0 otherwise. The choice of training provider is also recorded and included in our model. For example, persons can undertake formal studies at TAFE (*TAFE*), a private training organisation (*trainprivate*), a private RTO that closed prior to the training contract being completed (*trainclosed*) or a non-TAFE public sector organisation (*trainpublic*). In estimation, it is also necessary to exclude one of these variables for the same reason as for the omission of one of the occupational dummy variables. For the variables capturing the characteristics of the employer, the variable *emprivate* has been omitted in the formal modelling. This means that the estimated coefficients of the remaining variables capturing the characteristics of the employer measure the risk of cancellation relative to the risk of those employed directly by their

employer. Similarly, the trainer type *TAFE* is omitted in estimation so that the estimated coefficients of the other trainer types provide an estimate of the impact of these trainer types on the risk of cancellation relative to the risk faced by those attending TAFE institutes. The DELTA database records the postcode of employers and there may be a systematic variation in the location of employment for indigenous and non-indigenous persons. Finally, the income associated with a particular occupation is likely to influence the probability of a successful training contract completion. Cancellation from a training contract in occupations with higher income implies a greater forgone benefit. Thus, we would expect that, all else being equal, higher incomes are associated with lower rates of cancellation. In this study, the log of income has been used (*lnincome*). Income has been derived using the data from the 2001 Census of Population and Housing for each trade occupation appearing in the DELTA database.

4. Empirical estimates

In order to undertake the analysis, a CPH model was formulated and estimated. Gregoriou (2002) and Box-Steffensmeier and Zorn (2001) note that the CPH model has become a popular tool in the analysis of survival data in social sciences. The CPH model evaluates the effect of covariates to determine the magnitude and significance of these effects on the risk of cancellation from the apprenticeship training contract. Unlike the univariate K–M estimator, the CPH model can be used in a multivariate setting to test the hypothesis that survival does not depend on the level of a covariate, while adjusting for the other covariates in the model.

Unlike parametric methods of duration analysis, which compare regression coefficients under the constraints of a given mathematical model for the underlying survival times, the CPH model does not require that the precise nature of the survival function be known, nor that it be constrained by the assumption of a particular mathematical form. A limitation of the model, however, is that it assumes that the hazard functions for the levels of a given factor or treatment are proportional. Fortunately, this assumption can be examined graphically or by the implementation of various statistical tests, and in this study, residual-based tests have been used.

The CPH model is semi-parametric because all training contracts have a common, arbitrary baseline hazard function. The most frequently used CPH model specifies a hazard function for the i th training contract as:

$$\begin{aligned}\lambda_i(t) &= \lambda_0(t) \exp(z_i' \beta) \\ &= \lambda_0(t) \exp(z_{i1} \beta_1 + z_{i2} \beta_2 + \dots + z_{ik} \beta_k)\end{aligned}\quad (2)$$

where z_i denotes a vector of covariates of dimension k for the i th training contract, $\lambda_0(t)$ is the baseline hazard (the hazard in the absence of covariates, corresponding to $z = 0$) and β is a vector of regression parameters. In the CPH model, the ratio of the hazard functions of the i th and the j th training contract, denoted by $\lambda_i(t)$ and $\lambda_j(t)$, is called the hazard ratio and is given by $\lambda_i(t)/\lambda_j(t) = \exp[(z_i - z_j)' \beta]$. It should be noted that the hazard ratio does not depend on t and is therefore constant over time. The CPH model provides an estimate of the vector β , the components of which are the hazard ratios corresponding to each covariate. When all but the r th covariate are equal, the right-hand side of $\lambda_i(t)/\lambda_j(t)$ becomes $\exp[(z_i - z_j)' \beta_r]$ so that β_r corresponds to the hazard ratio for the r th covariate.

Table 5 provides the results from the empirical implementation of this modelling technique. In this case, the dependent variable is dichotomous, taking the value 0 for a completion or active training contract and 1 for a cancellation. Variables that have coefficients with an estimated positive (+ve) sign are found to increase the risk of cancellation, whereas those with a negative (-ve) sign are associated with a decrease in the risk of cancellation.

These coefficients are difficult to interpret on their own and so the exponentials of the coefficients are presented in the second column. These exponentials are interpreted as the multiplicative effects on the hazard of cancellation. Thus, for example, holding all else constant in our model, an additional year of age decreases the hazard of cancellation from 49.5% to 49.1%, i.e. $(0.992 \times 49.5 = 49.1)$. In contrast, changing the status from female to male is associated with a decrease in the risk of cancellation, with the risk of cancellation faced by males estimated at 0.913 of the risk faced by females.

The remaining columns of the table present the standard errors, z scores and probability levels, allowing testing of the hypothesis concerning the significance of individual variables. Here, we can see that all variables are significant at normal statistical levels in determining the risk of cancellation of the apprenticeship training contract with the exception of *gender*, suggesting males and females face the same risk of dropout; *disability*, indicating no significant additional risks of dropout for the disabled; *language*, implying no difference in the risk of cancellation between English and non-English speaking persons; *asco43*, suggesting insignificant difference in the risk of cancellation between *asco43* and *asco45*; and *empgto*, suggesting no statistically significant difference in the risk of cancellation between people employed directly and people employed by a GTO.

The final three columns of this table provide the Grambsch and Therneau non-proportionality test.⁵ These tests indicate a particular problem with the variable *empgto*, however, because of the insignificance of this variable, no action has been taken to deal with this potential problem. The bottom panel of the model provides the model diagnostics. These diagnostics allow more general tests, e.g., whether the model explains any of the variations in attrition observed in the data. The results presented in this panel indicate that the model explains a significant amount of variation in the risk of cancellation. For example, we see that the likelihood ratio test, Wald test and score test are all significant, indicated by the low p values.

The effect of education, in this model, is captured by three dummy variables, the first being Yr10, a 0, 1 dummy variable taking the value of 1 when the individual has a Year 10 level of education and 0 otherwise. The next variable is Yr11, a 0, 1 dummy variable taking the value of 1 when the individual has a Year 11 level of education and 0 otherwise. The last variable is Yr12, a 0, 1 dummy variable taking the value of 1 when the individual has a Year 12 level of education and 0 otherwise. The dummy variable for Years 8 and 9 is omitted to avoid the problem of perfect multicollinearity and thus the results measure the effect of the level of education relative to the completion of Year 8 or 9. For the dummy variables capturing the effect of the different levels of education, the results indicate that those entering apprenticeship with their highest level of schooling at Year 10 face a lower risk of dropout than those with less than this level of schooling (i.e. only 86.0% of the risk of dropout). One additional year of schooling, i.e., moving from Year 10 to Year 11, sees the risk of dropout declining to 81.1% of the risk faced by those who did not attain a Year 10 level of schooling, whereas completion of Year 12 sees the risk of cancellation falling to 64.6% of the risk faced by those who did not complete Year 10.

Table 5. CPH model of the probability of training contract completion.

	CPH model results				Grambsch–Therneau test for non-proportionality			
	Coefficient	Exp(Coefficient)	SE(Coefficient)	z	p	rho	χ^2	p
gender	-0.091	0.913	0.062	-1.477	0.140	-0.023	2.357	0.125
age	-0.008	0.992	0.003	-2.635	0.008	-0.029	2.966	0.085
atsi	0.229	1.258	0.080	2.855	0.004	0.017	1.103	0.294
disability	0.065	1.067	0.061	1.066	0.290	-0.017	1.203	0.273
language	0.117	1.124	0.124	0.944	0.350	0.021	1.795	0.180
yr10	-0.151	0.860	0.073	-2.083	0.037	-0.005	0.091	0.763
yr11	-0.210	0.811	0.078	-2.686	0.007	0.002	0.021	0.885
yr12	-0.444	0.642	0.072	-6.137	0.000	-0.007	0.188	0.664
asco41	-0.437	0.646	0.104	-4.216	0.000	0.011	0.435	0.510
asco42	-0.543	0.581	0.058	-9.426	0.000	0.030	3.533	0.060
asco43	-0.018	0.982	0.121	-0.148	0.880	0.034	4.566	0.033
asco44	-0.148	0.863	0.084	-1.761	0.078	-0.022	1.890	0.169
asco49	-0.599	0.550	0.051	-11.736	0.000	0.021	1.736	0.188
empgto	0.039	1.040	0.038	1.029	0.300	-0.042	7.250	0.007
psector	-0.816	0.442	0.132	-6.187	0.000	0.017	1.077	0.299
rto	0.165	1.180	0.038	4.357	0.000	0.020	1.717	0.190
closed	1.366	3.920	0.054	25.287	0.000	0.032	4.340	0.037
income	-3.454	0.032	0.487	-7.092	0.000	0.028	3.409	0.065
GLOBAL	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	84.665	0.000

Note: $R^2 = 0.161$ (max = 1), number of observations = 8061; likelihood ratio test = 1416 on 18 df, $p = 0.000$; Wald's test = 1589 on 18 df, $p = 0.000$; score (log-rank) test = 1802 on 18 df, $p = 0.000$.

Importantly for the current analysis, the variable *atsi*, indicating indigenous status, is found to be significant. The exponential of the coefficient (1.309) indicates that, holding all else constant in our model, the risk of cancellation increases with a change in the status from non-indigenous to indigenous. In terms of magnitude, indigenous persons are estimated to have a 25.8% higher risk of cancellation than the non-indigenous persons.

Although the baseline hazard is left unspecified in the CPH model, Fox (2002) notes that having estimated the model, it is possible to extract an estimator of the baseline hazard. It is also possible to display the impact of changing the value of one of the variables, with all others held at their average values, on the risk of survival and this is what is provided in Figure 2. This figure shows the estimated baseline hazard for two groups, indigenous and non-indigenous persons. This is provided with the point-wise 95% confidence envelope around the survival function.

In this figure, we can see that the risk of cancellation is clearly different for the two cohorts, reinforcing the results derived earlier. What is shown clearly in this figure is that, holding all else at its average value, the probability of a successful completion of the apprenticeship training contract is higher for those from a non-indigenous background than it is for indigenous persons undertaking a formal training contract. This difference is highlighted by the fact that the 95% confidence intervals of the estimated hazard functions for both cohorts hardly overlap beyond the first quarters. This in itself is indicative of a statistically significant difference in the risk of cancellation for the two groups.

A factor that further complicates the analysis presented in Table 5 is the potential that there is some variation across the population that is not observed. Unobserved heterogeneity is recognised as one of the most important sources of variation in economic and social behaviour (in this case, the risk of dropout). In many implementations,

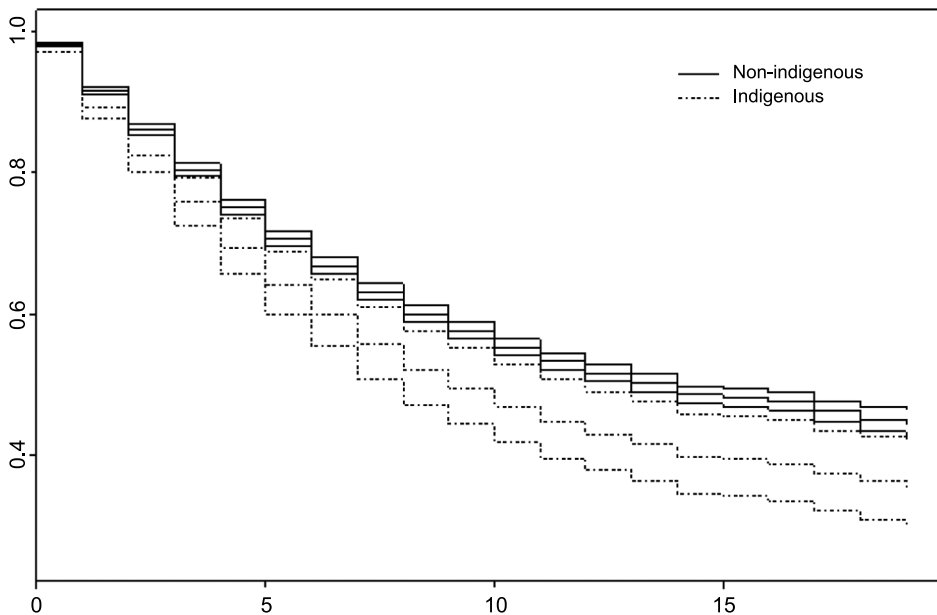


Figure 2. The impact of distribution of survival times by changing the status from non-indigenous to indigenous.

including the model presented in Table 5, this issue is ignored. However, if this unobserved variation is present, it may bias the coefficient estimates in a regression model. A more robust way of dealing with the problem, when conducting survival analysis, is to estimate survival models that attempt to account for the effect of this unobserved heterogeneity. One of the most frequently adopted measures to deal with this problem is to introduce a positive-valued random variable, v , into the hazard specification. In the context of the proportional hazard approach, the augmented hazard function, which incorporates a multiplicative mixture term, is given by:

$$\ln\{-\ln[1-h_{ij}(t_j|x_i)]\} = x_i'\beta + v\gamma(t) + u_i \quad (3)$$

where $u_i = \log(v)$. As Pozzoli (2009) notes, it is not possible to estimate the values of (v) themselves, as by construction, they are unobserved. However, if we suppose that the distribution of v has a shape whose functional form is summarised in terms of only a few key parameters, then it is possible to estimate those parameters with the available data. So after having specified a distribution for the random variable v , we derive the 'frailty' survivor corresponding to this mixture distribution and write the log likelihood function so that it refers to the original parameters and the mixing distribution parameters rather than to the random variable v . The unobserved heterogeneity term is assumed to be independent of the observed covariate x_i and the duration variable T .

Following Lancaster (1990) and Arranz and Muro (2004), we make assumptions about the distribution of the unobserved heterogeneity. In this study, we use the gamma distribution, which is frequently used in these types of studies. Like Arranz and Muro (2004), the choice of gamma distribution is based on a recent work that indicates that this option may be correct because a wide set of distributions used for the control of the unobserved heterogeneity converge to the gamma distribution.

Table 6 presents a proportional hazard model in which we attempt to account for the unobserved heterogeneity by including a frailty term. The simplest way to deal with the problem of unobserved heterogeneity is to assume that all persons in a category have a shared risk (frailty). In the example here, with candidates for apprenticeship clustered in regions, it is assumed that all persons within a postcode have a shared frailty. Individual frailty was not incorporated in this model because the hypothesis of normal unobserved heterogeneity between graduates was not supported by the data.

The results indicate that the variance of the random frailty effect is greater than zero and significant (as shown by the variance (0.058) and the L -likelihood value). The results for the model incorporating the shared frailty are similar to the stratified proportional hazard model, though there are a few differences. The variable *atsi* remains significant in both models, with the risk of an indigenous person dropping out being 26.2% higher than that of a non-indigenous person in the stratified CPH model and 24.9% higher in the model of shared frailty to account for the unobserved heterogeneity.

Other notable differences come about via the effect of education. The impact of education is slightly less pronounced in the model shown in Table 6 than in the model shown in Table 5, in which no attempt to account for the unobserved heterogeneity is made. Persons with a Year 10 level of education face a risk of dropout that is 85.0% of the risk of dropout for someone who finished school in Year 8 or 9, compared with 86.0% in the CPH model. Persons who left school in Year 11 face a risk of dropout that is 80.5% of the risk faced by individuals who finished school in Year 8 or 9. This

Table 6. Proportional hazard model with shared frailty.

	Coefficient	Exp(coefficient)	SE(coefficient)	χ^2	df	<i>p</i>
gender	-0.088	0.915	0.062	2.000	1	0.160
age	-0.007	0.993	0.003	4.810	1	0.028
atsi	0.221	1.247	0.082	7.130	1	0.008
disability	0.066	1.068	0.061	1.130	1	0.290
language	0.027	1.028	0.127	0.050	1	0.830
yr10	-0.163	0.850	0.073	4.890	1	0.027
yr11	-0.216	0.805	0.079	7.470	1	0.006
yr12	-0.465	0.628	0.073	40.250	1	0.000
asco41	-0.423	0.655	0.105	15.870	1	0.000
asco42	-0.563	0.569	0.059	89.480	1	0.000
asco43	-0.028	0.972	0.122	0.050	1	0.820
asco44	-0.153	0.858	0.086	3.080	1	0.079
asco49	-0.622	0.537	0.053	134.730	1	0.000
empgto	0.066	1.068	0.044	1.980	1	0.160
psector	-0.829	0.437	0.137	34.480	1	0.000
rto	0.200	1.221	0.039	25.670	1	0.000
closed	1.401	4.058	0.055	623.560	1	0.000
lincome	-3.463	0.031	0.497	47.450	1	0.000
frailty(postcode, distribution = 'gamma')				169.610	87.4	0.000

Note: Iterations = 10 outer, 31 Newton-Raphson; variance of random effect = 0.0585, *I*-likelihood = -33587.2; likelihood ratio test = 1668 on 105 df, *p* = 0.000, *n* = 8061.

compares with a risk of 81.1% in the CPH model. Persons who left school in Year 12 face a risk of dropout that is 62.8% of the risk faced by candidates for apprenticeship who finished school in Year 8 or 9, compared with 64.2% in the CPH model.

The large difference in the risk of cancellation between indigenous and non-indigenous persons shown in all versions of the model and observed in Figure 2 leads to the consideration of the possibility that for every variable in the model, the coefficients are different for the indigenous and the non-indigenous persons. Borooah (2001) notes that there are two ways of formally checking this in a modelling exercise like the one undertaken in this study. The first is to estimate a single equation, while allowing for each coefficient to be different for each group. The second is to estimate two separate equations for indigenous and non-indigenous persons. The former approach is adopted in this exercise, with the results presented in Table 5. The advantage of this strategy is that it allows easy testing of the hypothesis of whether the same variable had the same or different impacts over the two population groups. To implement this methodology, interaction terms are included for all variables. The resulting coefficient, if significant, shows the additional impact of these variables for persons of an indigenous origin.

The results presented in Table 7 indicate that there are four interaction terms that are statistically significant, being for *age*, *language* and the industry dummy variables of *asco41* and *asco43*. The estimated interaction term for *age* indicates that, unlike the situation for the general population, an increase in age for indigenous persons in formal training contracts is associated with an increased risk of cancellation. The variable *language* was insignificant in the model for the entire sample, presented in

Table 7. CPH model of risk of cancellation with indigenous person interaction terms.

	Coefficient	Exp(coefficient)	Se(coefficient)	χ^2	df	p	Coefficient	Exp(coefficient)	SE(coefficient)	χ^2	df	p
gender	-0.087	0.917	0.063	1.9	1	0.170	-0.011	0.989	0.277	0.000	1	0.970
age	-0.008	0.992	0.003	6.22	1	0.013	0.025	1.026	0.015	2.730	1	0.099
disability	0.036	1.037	0.066	0.3	1	0.580	0.050	1.051	0.212	0.060	1	0.810
language	-0.051	0.950	0.137	0.14	1	0.710	0.811	2.250	0.403	4.040	1	0.044
yr10	-0.143	0.867	0.077	3.43	1	0.064	-0.203	0.816	0.289	0.490	1	0.480
yr11	-0.204	0.816	0.083	6.07	1	0.014	-0.113	0.893	0.328	0.120	1	0.730
yr12	-0.446	0.640	0.077	33.66	1	0.000	-0.183	0.833	0.286	0.410	1	0.520
asco41	-0.447	0.640	0.107	17.48	1	0.000	0.801	2.228	0.378	4.500	1	0.034
asco42	-0.581	0.559	0.060	92.5	1	0.000	0.425	1.529	0.314	1.830	1	0.180
asco43	-0.053	0.949	0.125	0.18	1	0.670	0.938	2.555	0.358	6.850	1	0.009
asco44	-0.189	0.828	0.088	4.61	1	0.032	0.627	1.872	0.261	5.770	1	0.016
asco49	-0.640	0.527	0.054	138.41	1	0.000	0.260	1.297	0.313	0.690	1	0.410
empgto	0.057	1.058	0.048	1.38	1	0.240	-0.127	0.881	0.200	0.400	1	0.530
psector	-0.870	0.419	0.168	26.87	1	0.000	-0.270	0.763	0.342	0.620	1	0.430
rto	0.198	1.219	0.040	24.11	1	0.000	-0.001	0.999	0.206	0.000	1	1.000
closed	1.428	4.170	0.057	629.45	1	0.000	-0.645	0.525	0.378	2.910	1	0.088
income	-3.510	0.030	0.505	48.39	1	0.000	-0.149	0.861	0.197	0.570	1	0.450
frailty(postcode)										177.490	93.1	0.000

Note: Iterations = 6 outer; 23 Newton-Raphson; variance of random effect = 0.0661; *I*-likelihood = -33571.6; likelihood ratio test = 1712 on 126 df, *p* = 0, *n* = 8061.

Table 4 (p -value = 0.41), and remains so for the non-indigenous population (p -value = 0.81 in Table 5). However, for indigenous people, *language* is significant, with non-English speaking indigenous persons facing over twice the risk of cancellation for the general population.⁶ Similarly, for the industries *asco41* (mechanical and fabrication engineering tradespersons), *asco43* (electrical and electronic tradespersons) and *asco44* (construction tradespersons), we see that indigenous persons in these occupations face a much greater risk of dropout than the non-indigenous persons. Once again, this model is estimated with a shared frailty term (postcode) in an attempt to deal with the issue of unobserved heterogeneity.

5. Summary and conclusions

The paper asked two questions: Are cancellation rates for indigenous students in the VET above those for non-indigenous students in a statistically significant manner, and, if so, what factors account for this? The data from the DELTA database indicate that the cancellation rates for indigenous persons, at 56.3%, are 7.1% points above the rates recorded for non-indigenous persons (49.2%). The log-rank tests on the respective survival curves confirm that the rate of cancellation for the indigenous persons is significantly different from that for non-indigenous persons.

To cast further light on the nature of this difference, a multivariate CPH model of cancellation was estimated in which indigenous status was included as one of the explanatory variables. The impact of indigenous status on completion was estimated to be much higher than what is observed in the raw data, with a 30.9% increase in the risk of cancellation when all other variables in the database are held constant at their average values. Furthermore, this penalty associated with indigenous status was found to be statistically significant, with the t value of the estimated coefficient indicating significance below the 1% level.

An additional analysis was undertaken to see whether the variables incorporated in our model impacted on indigenous persons in a different way compared with non-indigenous persons. The results presented in Table 7 indicate that there are four interaction terms that indicate significant difference: *age*, *language* and the industry dummy variables of *asco41* and *asco43*.

The estimated interaction term for *age* indicates that for indigenous persons, an increase in age is associated with an increased risk of cancellation. This is opposite to the impact of age on non-indigenous persons. Furthermore, the variable *language* was insignificant in the model estimated over the complete sample, presented in Table 4 (p -value = 0.41), and remains so for the non-indigenous population (p -value = 0.81 in Table 3). However, for indigenous people, *language* is significant, with non-English speaking indigenous persons facing over twice the risk of cancellation for the general population. This may indicate language difficulties being confronted by indigenous populations but may also act as a proxy for other variables, such as the cultural background of the students. This finding seems to warrant further analysis, which is, however, outside the scope of the current study and beyond the capability of the DELTA database used here. Finally, for some of the industries, i.e., *asco41* (mechanical and fabrication engineering tradespersons) and *asco43* (electrical and electronic tradespersons), we see that indigenous persons face higher risks of cancellation than non-indigenous persons.

The observed higher risk of cancellation for indigenous students and the factors (other than being indigenous) suggest that the current environment in Australia has

serious limitations if the VET is being put forward as a major platform for improving educational outcomes for indigenous students. Moreover, the results indicate that specialist language skills modules should be introduced into the VET for indigenous students and increased attention should be paid to attract and retain mature, aged indigenous students. However, with this work focussing on disentangling the role of cultural background and language, additional research may need to be undertaken, with emphasis on capturing of variables not available in the current DELTA administrative database.

Disclaimer

The views expressed in this document are those of the authors and should not be considered as necessarily representing the views of the Department of Education and Training of the Queensland Government.

Notes

1. Referred to as Aboriginal and Torres Strait Islanders (ATSI).
2. This compares with 1% of university students.
3. See, Australian Social Trends 2002, ABS Cat 4102.0, Canberra.
4. In this analysis, the K–M estimator is used to estimate $S(t)$ for all training contracts, with the estimator stratified by the highest year of school education at the commencement of the training contract. The log-rank tests are then applied to compare all $S(1)$. This estimator of $S(t)$ is non-parametric because a distribution for T , that is $f(t)$, need not be specified. In calculating the K–M estimator, the data are partitioned into time intervals (here, quarters from commencement). If r_j denotes the number of training contracts alive immediately preceding the j th failure time (which contains d deaths or exits), then $I - d/r$ represents the conditional probability of survival within the interval ending at the j th failure, given that the training contract was alive at the beginning of the interval. Multiplying all intervals yields the K–M estimator, which is the unconditional probability of survival up to time t , i.e., $S(t) = \prod_{j=1}^m \left(1 - \frac{d_j}{r_j}\right)$ for the time interval $(t, \text{to}, +)$.
5. For a description of these tests, see Box-Steffensmeier and Zorn (2001).
6. The authors note that the robustness of this result must be considered in the light of the small numbers being considered, i.e., there were only 12 non-English speaking indigenous persons in the sample.

References

- Arranz, J., and J. Muro. 2004. Recurrent unemployment, welfare benefits and heterogeneity. *International Review of Applied Economics* 18, no. 4: 423–41.
- Ball, J., and A. Pence. 2001. *First Nations Partnership Programs generative curriculum model: Program evaluation report*. Canada: University of Victoria.
- Bilginsoy, C. 2003. The hazards of training: Attrition and retention in construction industry apprenticeship programs. *Industrial and Labour Relations Review* 57, no. 1: 54–66.
- Borooah, V. 2001. *Logit and probit: Ordered and multinomial models*. Sage University Paper No. 138. London: Sage.
- Box-Steffensmeier, J.M., and C.J.W. Zorn. 2001. Duration models and proportional hazards in political science. *American Journal of Political Science* 45, no. 4: 972–88.
- Byam, N. 2002. *Evaluation of the First Nations and Inuit Youth Employment Strategy*. Canada: Department of Indian Affairs and Northern Development.
- DETA. 2008a. *Queensland's labour market progress: A 2006 Census of Population and Housing profile. Issue no. 5: Income and hours worked*. <http://www.training.qld.gov.au/resources/employers/pdf/issue5-income-hours.pdf>

- DETA. 2008b. *Queensland's labour market progress: A 2006 Census of Population and Housing profile. Issue no. 9: Indigenous People in Queensland*. <http://www.training.qld.gov.au/resources/employers/pdf/issue9-indigenous-people.pdf>
- Fox, J. 2002. *An R and S-plus companion to applied regression*. London: Sage.
- Gregoriou, G. 2002. Hedge fund survival times. *Journal of Asset Management* 3, no. 3: 237–52.
- Hull, J. 1996. *Aboriginal post-secondary and labour market outcomes: The importance of education and training*. Ottawa: Department of Indian Affairs and Northern Development.
- Lancaster, T. 1979. Econometric methods of the duration of unemployment. *Econometrica* 47, no. 4: 939–56.
- Pozzoli, D. 2009. The transition to work for Italian university graduates. *Labour: Review of Labour Economics and Industrial Relations* 23, no. 1: 131–69.
- Skill New Zealand. 2001. *Sharing for success: Good practice and issues for Maori Education-Pukenga*. Wellington: Aotearoa.