Chapter 15
Validity of classification of full files

15.1 INTRODUCTION

The previous comparison of the classification of brief scenarios by research officers and the author (acting as the gold standard) was designed to assess, and improve, the application of the study definitions by the research officers. This was accomplished by providing the research officers with scenarios that had a small, given, set of facts. That proved a useful approach to making a reasonably standard assessment of the understanding of the definitions at several points in the data collection process.

However, such an approach is somewhat artificial, as it does not provide information on the validity of the classification of real coronial files by research officers. The actual classification process required research officers to find the relevant information in files, decide when they had enough information to make a valid classification, interpret the information, and assign an appropriate code.

To assess this component of the data collection process, it was necessary to use proper coronial files rather than short summaries of circumstances. The assessment of this aspect, which involved the dual classification of full coronial files, is considered in this section.

The aim of the work presented here was to:

- determine the level of accuracy of the classification of full coronial files by research officers;
- identify the circumstances that were difficult to classify validly; and
- improve the accuracy of the classification during the study.

15.2 METHODS

The data collection for this comparison took place early in the data collection phase of the study. The intention was for the author to check, for each research officer, the files of about 50 deaths, attempting to ensure that about 15 Cases, 15 Indeterminates and 15 Non-cases, as coded by the research officers, were included. Research officers were instructed to code and process the files as per the standard data collection protocol, described in Chapter 3. This protocol required the files to be selected in random order. Having dealt with the file, and beginning with the first file coded, research officers were instructed to record the code for, and put aside, consecutive files until the required number of files (approximately 15) in each category was obtained. Once enough files were set aside for a particular category, no further files in that category were required (although sometimes a few more were included), but files in other categories continued to be set aside until the quotas for all three categories were filled. The file details and the assigned codes were recorded on paper by the research officers.

The author visited each research officer team early in the data collection stage of the study. During this visit, the files set aside by the research officers were coded blindly by the author. For this aspect of the study, the author was accepted as the gold standard, and his codes were defined as being correct. These codes were compared with the research officer codes at the time of the visit, and discrepancies discussed. This

discussion was used to clarify problems with the interpretation or application of the definitions, and with the identification of relevant information in the files.

Not all research officers met the required quota in all three categories. The main reasons for this were the later exclusion from the analysis of some deaths from multiple death incidents; and a lack of cases in a particular category being coded prior to the visit of the author.

Where more than one death from a multiple death incident had been included by a research officer, only one death with the same relevant circumstances was included in the analysis. This usually meant that only one death from an incident was included in the analysis but, for a few incidents, two deaths were included because they were assigned to different categories (usually one as a worker and one as a bystander).

For the analysis, comparison was made with the exact code, of which there were 19 possible choices (nine case codes, nine indeterminate codes and a non-case code), and with the three general codes of case, indeterminate and non-case. Reported codes for the files (deaths) are those assigned by the author (the gold standard) unless otherwise stated. Sensitivity, specificity and accuracy were the main measures of interest. They are presented here as proportions. The definitions for these terms were as used in Chapter 14:

sensitivity:	the proportion of all deaths in a category that were correctly assigned to that category;
specificity:	the proportion of all deaths not in a category that were correctly assigned not to that category;
accuracy:	the proportion of all deaths assigned to the correct category.

All 21 research officers were included in the comparison (one additional officer who left the study after two weeks was not included). Between 25 and 54 files from each research officer were included, with a mean of 40.3 (sd = 8.2). Since there were more research officers in the larger jurisdictions, more of the files from these jurisdictions were included in the analysis (Table 15.1).

Table 15.1 Comparison of codes assigned to full coronial files: number of files per jurisdiction

Jurisdiction	Number	Percent
ACT	54	6.4
NSW	193	22.8
NT	66	7.8
QLD	148	17.5
SA	82	9.7
TAS	68	8.0
VIC	153	18.1
WA	82	9.7
Australia	**846**	**100.0**

15.3 RESULTS

OVERALL COMPARISON

Eight hundred and forty six files were included in the analysis, comprising 346 (40.9%) Cases, 184 (21.7%) Indeterminates and 316 (37.4%) Non-cases. Within the group of Cases, the *Worker* (121: 14.3% of all files) and *Road Bystander* (117:13.8% of all files) categories were the largest (Table 15.2).

Table 15.2 Comparison of codes assigned to full coronial files: number of files in each category (using gold standard)

Case status	Code	Case category	Number	Percentage
Case				
	11	Worker	121	14.3
	12	Commuter	37	4.4
	13	Volunteer	1	0.1
	14	Student	4	0.5
	15	Home Duties	16	1.9
	16	Workplace Bystander	23	2.7
	17	Road Bystander	117	13.8
	18	Other Farm	12	1.4
	19	Uncertain	15	1.8
		All cases	346	40.9
Indeterminate				
	21	Worker	8	0.9
	22	Commuter	12	1.4
	24	Student	1	0.1
	25	Home Duties	13	1.5
	26	Workplace Bystander	6	0.7
	27	Road Bystander	45	5.3
	28	Other Farm	6	0.7
	29	Uncertain	93	11.0
		All indeterminates	184	21.7
Non-case				
	0	Non-cases	316	37.4
Total		All files	846	100.0

Comparing the broad categories, 693 (81.9%) files were classified correctly, with 122 (14.4%) differing by one category and 31 (3.7%) differing by two (ie coded as a case by one person and as a non-case by the other). Of the 85 discordant deaths with coding differing by one category, most (68: 55.7%) were Non-cases classified as indeterminates, or indeterminates classified as Non-cases, by the research officers. Twenty (6.3%) of the 316 Non-cases were classified as cases by the research officers, and 11 (3.2%) of the 346 Cases were classified as Non-cases by the research officers. The sensitivity and specificity were, respectively, 0.86 and 0.93 for Cases, 0.74 and 0.89 for Indeterminates, and 0.83 and 0.92 for Non-cases (Table 15.3). Taking into account all 19 possible codes, 603 (71.3%) files were classified correctly.

Table 15.3 Comparison of codes assigned to full coronial files: broad categories

		Research officer code			
		Case	**Indeterminate**	**Non-case**	**Total**
Truth	**Case**	296	39	11	**346**
	Indeterminate	15	136	33	**184**
	Non-case	20	35	261	**316**
	Total	**331**	**210**	**305**	**846**

On an individual research officer basis, accuracy at the broad category level ranged from 0.60 to 0.96, and at the individual category level from 0.48 to 0.91. For the jurisdictions, agreement at the broad level ranged from 0.78 to 0.89, and at the individual category level from 0.60 to 0.82.

WORKING CASES ONLY

Since the primary focus of the overall study was on the work-related deaths of workers, mis-classification of the *Worker* category was the main area of interest and concern in the comparison. For the 121 working deaths, the sensitivity was 0.93 and the specificity 0.98. Of the nine (7.4%) *Worker* deaths that were classified differently, the research officers classified three as commuting cases, three as cases of uncertain category, two as other sorts of cases and one as an indeterminate working case. Of the 11 (1.5%) non-*Worker* deaths classified wrongly as working cases, the correct classifications were commuting for two, cases of uncertain category for three, indeterminate for two and non-cases for four. The overall accuracy for the classification of deaths as working or not was 0.98 (Table 15.4).

Table 15.4 Comparison of codes assigned to full coronial files: Worker cases only

		Research officer code		
		Case	**Non-case**	**Total**
	Case	112	9	**121**
Truth	**Non-case**	11	714	**725**
	Total	123	723	**846**

OTHER CASE CATEGORIES

Of the other specific case categories with more than ten deaths, the sensitivity ranged from 0.58 (for road bystanders) to 0.81 (for home duties deaths), the specificity was 0.98 or higher for all categories, and the accuracy 0.93 or higher for all categories. It should be noted that the number of cases in some categories (eg home duties) was low, so the estimates of sensitivity and specificity for these categories have a large degree of uncertainty (Table 15.5).

Table 15.5 Comparison of codes assigned to full coronial files: sensitivity, specificity and accuracy for cases

Case category	Number	Sensitivity	Specificity	Accuracy
Worker	121	0.93	0.99	0.98
Commuter	37	0.73	0.99	0.98
Home Duties	16	0.81	0.99	0.99
Workplace Bystander	23	0.74	0.99	0.98
Road Bystander	117	0.58	0.98	0.93
Other Farm	12	0.67	0.98	0.97
Uncertain	15	0.20	0.98	0.97

SOURCES OF DISAGREEMENT

The main sources of disagreement for wrongly coded files were the research officer missing relevant information in the file (especially for commuting cases), or differently interpreting the same information. These differences in interpretation included when

work was deemed to have started or stopped (working and commuting deaths); whether the circumstances really met the case definitions (mainly working cases); and which cases definition was most closely met by the circumstances (especially with cases coded to the *Uncertain* category). Details of discordant cases coded as workers by either the author or a research officer are shown in Table 15.6, along with comments on the misclassifications.

The assigned codes were occasionally later changed as a result of checking of information by the study team during the data coding stage. The change usually occurred because the reviewer identified extra information in the file, or interpreted the information differently. However, some changes were made because the study team, having reviewed a large number of files with varying circumstances, decided that all deaths with specific circumstances on the boundary of a category's definition either should be, or should not be, included as cases. Deaths that were difficult to classify were discussed by the five member study team, who usually reached a consensus on the appropriate code. However, the final code was sometimes decided by a majority vote.

This meant that sometimes the author's original classification was later changed when the death was entered into (or excluded from) the study database. Of the nine working cases that were differently coded by the research officers, all remained as working cases upon review. Of the 11 deaths coded as working cases by the research officers but to some other category by the author, after review five were coded as working. Four of these five had been coded as uncertain which category or indeterminate working by the author, with the working category being strongly considered for all four. The remaining death later coded as working was coded by the author as an Indeterminate *Other-Farm* death, because the information in the file was inconclusive. No further information

became available, but the full team decided, on review, that the circumstances warranted the death being included as a working case. The six other deaths coded as working by the research officer but not the author retained the codes assigned by the author.

There was no formal adjustment of estimates to take account of the proportion of files form each States and Territory. However, those jurisdictions with the larger number of files also had more research officers, so more of the sample files were included from the jurisdictions with more total files. Therefore, the unadjusted estimates presented here are considered to be a reasonable reflection of the true sensitivity and specificity for coding of all files.

15.4 CONCLUSIONS

As for the scenarios described in Chapter 14, the comparison of classification of full files showed a high level of agreement between the true classification (as determined by the author) and the classification assigned by the research officers. This was especially so for working deaths, which were the primary focus of the study. Agreement for other case-classifications was generally good, apart from the *Road-Bystander* and *Uncertain* categories.

Although the comparison provides good evidence that deaths were being appropriately classified, the main function of the process was to improve the accuracy of the classification during the study. This was achieved by identifying, at an early stage, any problems with the classification process, including misunderstanding and different interpretation of the definitions, conceptually difficult circumstances for which

clarification of the definitions was required, and practical problems with the identification and interpretation of information in files.

The main problem areas were found to be identifying relevant information on the purpose of journeys for motor vehicle incident deaths; determining when work activity should be deemed to have started or stopped; determining the strength of evidence required before a definitive code could be assigned; and the allocation of deaths to a case category when the circumstances had elements of more than one category.

Table 15.6 Discordant working "cases" from full coronial files: description, valid code, research officer code and comment

Case	Description	Correct code	RO's Code	Comment
QLD 1	Adult farmer who died when his motorcycle crashed while he was mustering horses	11	19	It appears that the man was working on the property of a friend, but it was unclear if he was working in a paid capacity. However, the work clearly arose directly out of his occupation as a farmer. So, the incident is most appropriately seen as being work related.
SA 1	Labourer, at work, hit by car as he crossed the road to retrieve a bag from the truck of a workmate. He was travelling to another job, and the driver of the other car was almost certainly travelling to work.	11	12	There was uncertainty as to when work should be deemed to have started. In this instance, the work team had met at a central location, before getting in their work vehicle and travelling to a job. They are correctly seen as travelling for work, rather than commuting, when the crash occurred.
SA 2	One of seven men working for a drilling company who died of multiple injuries when the van they were travelling in ran onto the wrong side of the road and into the path of a truck. The men were living in a hotel in town and had to drive out to the drill site, "effectively lengthening the shifts to 13.5 hours", according to the coroner.	11	12	There was uncertainty as to when work should be deemed to have started. Since the men were required to live in town and travel long distances each day to get to and from the work site, their work day really began as soon as they began travelling to the drill site. They should be considered to have been travelling for work, rather than commuting, when the crash occurred.
SA 3	Fruiterer who was travelling to the markets, probably for work, when he was run over by a delivery truck.	11	19	There was uncertainty as to when work should be deemed to have started. In this instance, the man's workplace was his fruit shop, and he was travelling elsewhere for work purposes. He should be considered to have been travelling for work, rather than commuting, when the crash occurred.
SA 4	"Farmer" operating a bull-dozer at a family-owned gypsum mine, when the dozer rolled and crushed him.	11	19	This was a clear work-related death of a worker. The research officer didn't realise that the man owned the mine, and couldn't remember why he coded the case 19 rather than 11.
TAS 1	Hobby farmer killed in a two-car collision when the car he was driving, returning from buying feed for livestock and groceries, collided with a vehicle being driven by someone who was not working.	11	12	There was uncertainty as to when work should be deemed to have started or stopped. In this instance, the man's workplace was his farm. He was travelling elsewhere for work purposes (although the purchase of groceries along with the livestock feed created some uncertainty regarding the working status of the journey). He should be considered to have been travelling for work, rather than commuting, when the crash occurred.

Table 15.6 continued

Case	Description	Correct code	RO's Code	Comment
TAS 2	Elderly farmer who died of a ruptured thoracic aorta, probably as a result of being kicked by a sheep a few weeks earlier.	11	21	There was uncertainty as to the strength of evidence connecting the kick by the sheep (which clearly occurred at work) and the subsequent ruptured aorta. However, the file evidence was strong enough to make the connection, so the death should be coded as a definite case. The research officer recognised the possibility of the connection and suggested copying the file for later consideration by the study team, which was an acceptable approach.
VIC 1	Teenage student who was pulling out thistles on his father's farm, long with a friend, when the dune buggy he was in rolled over.	11	18	The research officer identified the same facts of the incident, but had not realised that the activity constituted work.
NSW 1	Grazier who was asphyxiated when the car he was fixing on his farm rolled off the jacks and crushed him.	11	15	There was uncertainty about whether this sort of activity was best considered to have been working or home duties. Since the car was a farm vehicle, the man should be considered to have been working.
SA 5	Pest controller driving to work when he veered into the path of a semi-trailer carrying sheep.	12	11	There was uncertainty as to when work should be deemed to have started. In this instance, the man usually travelled to his base of employment prior to setting out to perform pest control at various locations. Since, at the time of the collision, he was travelling to this base of employment to start work, he should be considered to have been commuting rather than travelling for work.
WA 1	Accountant who was killed in a single vehicle collision when the car he was driving home from work crashed.	12	11	There was uncertainty as to when work should be deemed to have stopped. In this instance, the man had a set place of work and had left it for the day. Since, at the time of the collision, he was travelling from his base of employment to his home, he should be considered to have been commuting rather than travelling for work.

Table 15.6 continued

Case	Description	Correct code	RO's Code	Comment
NSW 2	Driver of an ice-cream vending van who was killed when the van he was driving crossed over the medium strip and collided with two other vehicles. There was no mention of why or where he was travelling and he had his son in the car.	19	11	There was uncertainty as to when work should be deemed to have started or stopped. In this instance, there was some ambiguous information in the file, making final classification difficult. Since he was almost certainly either working or commuting, he was clearly a case, so 19 seemed the appropriate code. The research officer considered there to be enough evidence that the man was working. During the coding of the data, when a definitive decision about case status had to be made, it was decided to code this as a working case.
NSW 3	"Retired" (according to the police information in the file) man in his 50's who died when his clothing became trapped in an auger he was using to dig post-holes on his mother's property.	19	11	There was uncertainty as to whether this activity should be considered to have been work or home duties. During the coding of the data, when a definitive decision about case status had to be made, it was decided to code this as a working case.
QLD 2	Deer farmer who died of internal injuries sustained when the car he was driving crossed to the wrong side of a dirt road and collided head-on with a truck. He was driving to town with a passenger, but it was not clear why the subject was travelling. A workers' compensation claim was submitted regarding the man's death.	19	11	There was uncertainty as to why this man was travelling. The workers' compensation claim strongly suggests that it was either for work or commuting, but there was no information in the file to make a definitive code. During the coding of the data, when a definitive decision about case status had to be made, it was decided to code this as a working case.
VIC 2	Banker who died of a laryngeal obstruction, due to a bolus of food, whilst on a passenger flight between two cities. He was travelling with his business partner and friend but there was no evidence as to why he was travelling.	21	11	There was uncertainty as to why this man was travelling. The research officer considered there to be enough evidence that the man was travelling for work. If this man was on a business trip, this travel would have been considered to have been work and the death would have been considered work-related. Since the reason for travel is not known, there is insufficient evidence to accept this as a case, or to exclude it as a non-case. So he should be considered an indeterminate working case.

Table 15.6 continued

Case	Description	Correct code	RO's Code	Comment
TAS 3	Self-employed bushman, returning home from his girlfriend's home, when his truck ran off the road on the property where he lived. It was not clear whether this was a working property, and there were fence posts at the back of the truck.	28	11	The research officer thought the presence of fence posts in the man's truck, which were probably there as a result of his employment, meant that he should be considered to have been working. However, there was good evidence in the file that this man was travelling for reasons other than work. The only way he could be included as a case was if the property was a farm, thereby putting the man in the "other-farm" category. Since there was uncertainty about the status of the property, he should be considered an indeterminate other-farm case.
NSW 4	Adult who died of an accidental overdose of dextropropoxyphene, to which she was addicted. There was some suggestion of over-prescribing.	0	11	This woman was clearly not working. The research officer had made a simple mistake in the coding, and the concern about over-prescribing was not strong enough for the death to be considered a bystander to work. Therefore, she should be coded as a non-case.
NSW 5	Spray-painter driving in a sprint car championship, probably for recreational reasons, was killed during a collision while racing.	0	11	There was a little uncertainty regarding why this man was driving. The research officer thought that he might have been driving professionally. However, the file had considerable evidence suggesting that the driving was just a hobby, so he should be considered a non-case.
QLD 3	Dozer driver who died in a speedway incident when the car he was driving rolled. He was probably driving for recreational reasons.	0	11	There was a little uncertainty regarding why this man was driving. The research officer thought that he might have been driving professionally. However, the file had considerable evidence suggesting that the driving was just a hobby, so he should be considered a non-case.
VIC 3	Retired glider pilot killed in a glider crash. Flying for recreation.	0	11	This person was clearly not working, so he should have been coded as a non-case.