



MONASH University
Accident Research Centre

**DE-IDENTIFIED LINKAGE OF
VICTORIAN INJURY DATA RECORDS:
A FEASIBILITY STUDY**

by

Angelo D'Elia

Stuart Newstead

May 2010

Report No. 296

MONASH UNIVERSITY ACCIDENT RESEARCH CENTRE
REPORT DOCUMENTATION PAGE

Report No.	Date	ISBN	Pages
296	May 2010	0 7326 2366 9	66

Title: De-identified Linkage of Victorian Injury Data Records: A Feasibility Study

Author(s): Angelo D'Elia and Stuart Newstead

Sponsoring Organisation(s): This project was funded through the Centre's Baseline Research Programme for which grants have been received from:

Department of Justice
Transport Accident Commission

Roads Corporation (VicRoads)

Abstract:

Research utilising the available mass databases relating to real world road crashes compiled by Police, government and insurers in Victoria provides much useful information for injury prevention purposes. Separately, these data sources have been critical for a wide range of important research carried out by the Monash University Accident Research Centre and many other agencies that has led to the development of new countermeasures and evidence-based preventive policies and programmes. The record linkage of these datasets and hospital injury records has the potential to maximise the use of available data by researchers to extend our understanding of the causes, outcomes and costs of road traffic and other injuries. As has occurred in Western Australia, research based on linked data has the potential to enormously enhance the evidence base for road and non-road injury prevention policy making, preventive measures and programme evaluation of interventions. This project explored the feasibility of linking Police-reported motor vehicle crash, Transport Accident Commission (TAC) insurance, in-depth crash data and hospital injury datasets using a de-identified linkage approach. It was found that the de-identified linkage of hospital data to other injury datasets was not feasible and a separate approach to linking this data based on identified information is recommended. De-identified linkage of Police and TAC data was proved to be feasible and identified potential alternative measures of serious injury, other than that reported by Police, that could be used for road safety research and as key outcome measures for Victoria's *arrive alive 2008-2017* road safety strategy.

Key Words:

Record Linkage, Injury Data, Victoria

Disclaimer:

This report is disseminated in the interest of information exchange. The views expressed here are those of the authors, and not necessarily those of Monash University.

Reproduction of this page is authorised.

Monash University Accident Research Centre
Building 70, Clayton Campus
Monash University VIC 3800
Australia

Telephone: +61 3 9905 4371, Fax: +61 3 9905 4363
www.monash.edu.au/muarc

Preface

Project Manager / Team Leader:

Dr. Stuart Newstead

Research Team:

Angelo D'Elia

Contributorship Statement:

Project Concept – SN

Data Linkage and Analysis – AD

Draft Report – AD

Report Review and Finalisation – AD, SN

Contents

EXECUTIVE SUMMARY	xi
1. BACKGROUND AND AIMS	1
1.1 DATASETS.....	1
1.2 OBJECTIVES	2
2. RECORD LINKAGE METHOD	2
2.1 LINKAGE OF THE STAR AND TAC DATASETS	3
2.1.1 <i>Key Variables</i>	3
2.1.2 <i>Review of Matching Variables</i>	4
2.1.3 <i>Matching Protocol</i>	9
2.2 LINKAGE OF THE ANCIS DATASET	9
2.3 LINKAGE OF THE VAED	11
2.3.1 <i>Selection of VAED Cases for Matching</i>	11
2.3.2 <i>Preparation of a Linked STAR-ANCIS-TAC Dataset</i>	12
2.3.3 <i>Determination of a Matching Protocol</i>	12
3. RESULTS.....	14
3.1 STAR AND TAC DATASETS – MATCH RATES	14
3.2 STAR-ANCIS-TAC DATASET – MATCH RATE	14
3.3 CROSS-COMPARISON OF THE STAR, TAC AND ANCIS DATASETS	15
3.4 LINKAGE OF THE VAED	17
4. DISCUSSION AND CONCLUSIONS	17
5. KEY RECOMMENDATIONS	19
6. REFERENCES	19

Appendices

APPENDIX 1	21
FLOW DIAGRAM FOR LINKAGE OF THE STAR AND TAC DATASETS	
APPENDIX 2	25
SYNTAX FOR LINKAGE OF THE STAR AND TAC DATASETS	
APPENDIX 3	33
FLOW DIAGRAM FOR LINKAGE OF THE STAR, TAC AND ANCIS DATASETS	
APPENDIX 4	37
ADDITIONAL AND MODIFIED SYNTAX FOR LINKAGE OF THE STAR, TAC AND ANCIS DATASETS	
APPENDIX 5	43
FLOW DIAGRAM FOR LINKAGE OF THE STAR, TAC, ANCIS AND VAED DATASETS	
APPENDIX 6	47
SYNTAX FOR LINKAGE OF THE STAR, TAC, ANCIS AND VAED DATASETS	

EXECUTIVE SUMMARY

Research utilising the available mass databases relating to real world road crashes compiled by Police, government and insurers in Victoria, all held and accessed by the Monash University Accident Research Centre (MUARC), provides much useful information for injury prevention research purposes. Separately the data sources have been critical for a wide range of important research carried out by MUARC and many other agencies that has led to the development of new countermeasures and evidence-based preventive policies and programs. The linkage of these datasets has the potential to maximise the use of available data by researchers to extend the understanding of the causes, outcomes and costs of road traffic and other injuries.

As has occurred in Western Australia, research based on linked data has the potential to enormously enhance the evidence base for road and non-road injury prevention policy making, preventive measures and evaluation of intervention programs. Further benefits include assessment of data quality issues through cross comparison and assessment of the sampling frames and coverage within each dataset. Effective linking of the datasets would potentially lead to more robust and better qualified results from research that would previously have only analysed the datasets individually.

Objectives

The first major objective of the project was to establish the feasibility of linking records on road crashes in Victoria compiled by Victoria Police and enhanced by VicRoads with corresponding records on injury insurance compensation claims made to the Transport Accident Commission (TAC) and in-depth crash inspection data collected through the Australian National Crash In-depth Study (ANCIS). This would allow data quality issues in the mass crash data sources, such as accident details and the critical issue of injury coding to be assessed. The second major objective of the project was to assess the feasibility of linking hospital admission records with the linked Police-ANCIS-TAC records to further validate and enhance the information available. Success in linking the hospital data would enhance the potential of the dataset for a broader variety of injury research.

Datasets

The feasibility of linking the following datasets was assessed in this project:

- *State Traffic Accident Record (STAR)*: This dataset is an extract from VicRoads' Road Crash Information System (RCIS) and contains records of every road crash involving injury reported to Police in Victoria.
- *Transport Accident Commission (TAC) Claims Data*: This dataset covers all claims made to the TAC for third party injury compensation for injuries arising from transport accidents.
- *Australian National Crash In-depth Study (ANCIS)*: ANCIS collects in-depth data on vehicle crashes including great detail on occupant injury, crash circumstances and vehicle damage.

- *Victorian Admitted Episodes Database (VAED)*: The VAED (injury subset) contains records of injury hospital admissions to Victorian public and private hospitals.

Results and Conclusions

De-identified linking of the STAR, TAC and ANCIS datasets proved to be feasible. Match rates between the STAR and TAC data were determined largely by the accuracy of recording the Police assigned crash number in each data set. In the data provided for the feasibility study, a significant proportion of the TAC claims data was missing a valid Police assigned accident number leading to a reduction in the maximum possible match rate. To improve the overall match rate, the TAC would need to ensure that all claims contain a valid and verified Police accident number. It is understood that this has been or is being rectified by the TAC meaning fully linked TAC claims and Police crash reports should be viable.

An alternative to the linkage of the STAR and TAC datasets by an external agency would be for the TAC to match claims data to Police data in-house. In-house linkage has been performed by matching Police-reported crash data from the Victoria Police Traffic Incident System (TIS) with TAC claims data and STAR. The TAC has reportedly achieved a match rate of 85%, with the non-matched claims explained by Victorian drivers covered by TAC insurance lodging a claim related to a crash that occurred interstate along with claims for injury on public transport which are not recorded in STAR.

Cross-comparison of injury severity measures between the ANCIS and TAC datasets suggested that the use of TAC claims data merged with STAR could provide an alternative and finer measure of serious injury from road safety research compared to the measure of serious injury based on hospitalisation recorded in Police crash reports. A new project has commenced that seeks to collaborate with the TAC in order to develop a linked dataset for ongoing use in road safety research. This will include a review of the injury coding practices of the TAC and a process of consultation with the key agencies to automate and streamline the linkage process into an ongoing system. This work will identify the most appropriate measure or measures of serious injury from the linked dataset for use in road safety performance monitoring and other projects such as time series modelling. In particular, operational biases inherent in serious injury measures derived from Police crash reports (defined as taken to hospital for the period examined for this project, previously defined as being admitted to hospital) should be overcome. Victoria Police have advised that since 2006 this has changed to a combination of taken and admitted to hospital. Indeed, the use of an alternative and finer measure or measures of serious injury should more closely reflect real serious injury trends and provide a more satisfactory key outcome measure for Victoria's *arrive alive 2008 – 2017* road safety strategy.

The second major objective of the project was to assess the feasibility of linking hospital admission records with the linked STAR-ANCIS-TAC dataset using de-identified data. It was found that the de-identified linkage of the VAED to the other Victorian injury data records was not feasible. Problems encountered included the identification of road crash related hospital admissions, lack of information on location of the event leading to hospital admission reducing the number of variables that could be used in a matching key, and identifying the relevant transfer and re-admission cases related to the VAED being a separations database. It was concluded that identifying information would be required to both link relevant cases within the VAED as well as for linkage to other Victorian injury data records. This is

particularly important if comparisons between datasets are to be made such as comparisons between ICD injury codes recorded in both the VAED and TAC claims data.

In order to link hospital data with the other data sources considered, it is recommended that Victoria set up a Data Linkage Unit (DLU) similar to that in Western Australia. Each year in WA, the DLU links road crash reports to hospital morbidity and mortality data using personal identifiers. The identifiers are removed after linkage before the data are made available to external agencies. A Western Australian style data linkage unit using identifying information would also enable the linkage of Victoria's core population health datasets.

Key Recommendations

In summary, the key recommendations resulting from this project were that:

- The TAC perform ongoing in-house linkage of Police-reported crash data from the Victoria Police Traffic Incident System (TIS) with TAC claims data and STAR.
- The dataset linked by the TAC is used to identify alternative measures of serious injury with the potential to be used as key outcome measures for *arrive alive 2008-2017*.
- A Western Australian style data linkage unit is set up to enable the linkage of Victoria's core population health datasets including with mass crash data sources.

1. Background and Aims

Research utilising the available mass databases relating to real world road crashes compiled by Police, government and insurers in Victoria, all held and accessed by the Monash University Accident Research Centre (MUARC), provides much useful information for injury prevention research purposes. Separately the data sources have been critical for a wide range of important research carried out by MUARC and many other agencies that has led to the development of new countermeasures and evidence-based preventive policies and programs. The linkage of these datasets has the potential to maximise the use of available data by researchers to extend the understanding of the causes, outcomes and costs of road traffic and other injuries.

As has occurred in Western Australia, research based on linked data has the potential to enormously enhance the evidence base for road and non-road injury prevention policy making, preventive measures and evaluation of intervention programs. Further benefits include assessment of data quality issues through cross comparison and assessment of the sampling frames and coverage within each dataset. Effective linking of the datasets would potentially lead to more robust and better qualified results from research that would previously have only analysed the datasets individually.

1.1 Datasets

The following datasets containing information relating to road crashes and their injury outcomes are commonly used for road safety research in Victoria:

- ***State Traffic Accident Record (STAR):*** This dataset is an extract from VicRoads' Road Crash Information System (RCIS) and contains records of every road crash involving injury reported to Police in Victoria. Victoria Police collect the data; it is further enhanced by VicRoads who supply de-identified data to MUARC. The data covers crash circumstances, road user and vehicle information in great detail but only has limited information on the injuries sustained by persons involved in the crash. MUARC holds complete records from 1987 onwards.
- ***Transport Accident Commission (TAC) Claims Data:*** This dataset covers all claims made to the TAC for third party injury compensation for injuries arising from transport accidents. The database has limited detail on crash circumstances and the road users involved, but has ICD (International system for the Classification of Diseases) codes for up to the worst 5 injuries assessed from the claim along with less detailed codes on body region of injury by severity. MUARC holds complete de-identified data on TAC claims from 1987 onwards.
- ***Australian National Crash In-depth Study (ANCIS):*** In-depth data on vehicle crashes have been collected at MUARC since 1988 and have proven valuable for a variety of analyses conducted for the Australian Federal Government and the auto industry. The ANCIS study collects great detail on occupant injury, crash circumstances and vehicle damage. Injury is coded using the Abbreviated Injury Scale (AIS). The sampling frame for ANCIS generally covers vehicles aged 10 years old or less that are involved in a crash that resulted in injury to a vehicle occupant that required hospital admission. The ANCIS dataset also includes cases recruited as part of the Holden Crash Investigation project.

- **Victorian Admitted Episodes Database (VAED):** The VAED (injury subset) held by MUARC (and provided by DHS) contains de-identified records of injury hospital admissions to Victorian public hospitals from 1987 and private hospitals from 1994. VAED data are coded using ICD-10 Australian Modification coding system and includes demographic data (age, gender, postcode of residence, country of birth) and cause, activity, location, human intent and diagnoses codes.

1.2 Objectives

The most effective way to link the key Victorian road crash and injury data sources reliably is through the use of unique identifiers. The most reliable of these in road crashes is personal identifiers including name and addresses of those involved and injured in the crashes. Access to personal identifiers in road crash and injury data is very limited due to constraints imposed by privacy legislation and ethical considerations. These problems become more acute when linking data from different sources. Overcoming the problems in using personal identifiers for data linking is both expensive and very time consuming. Consequently, the broad aim of this project was to investigate the feasibility of linking the key Victorian road crash and injury data sources without the use of personal identifier data but via other relevant linking keys available in the data. This is referred to here as a de-identified linkage process.

The first major objective of the project was to link records from ANCIS with corresponding records in the TAC claims and STAR mass crash data sources. This would allow data quality issues in the mass crash data sources, such as accident details and the critical issue of injury coding to be assessed. The second major objective of the project was to link injury hospital admission records from the VAED with the mass crash data records to further validate and enhance the information derived from the analysis of the mass crash data records. This would allow the potential of the linked dataset for a broader variety of injury research to be examined, contingent on the success of linking the hospital data.

2. Record Linkage Method

Linkage was performed on records relating to individuals. Each dataset was provided without identifying information such as name, address and date of birth, necessitating the use of a de-identified linkage approach. A de-identified approach utilises less stringent matching criteria compared with a deterministic one, where a pair of records would only be linked if an exact match occurs on the selected identifiers. The method used for this project required an exact match for the key (non-identifying) variables, but included additional matching stages where selected variables were dropped from the matching “key” in order to improve the overall match rate. This was balanced against achieving as accurate a match as possible, with additional variables used to reject linked records that did not meet specified criteria.

Ethics approval for linkage of the STAR, TAC and ANCIS datasets was obtained from the Monash University Standing Committee on Ethics in Research involving Humans (SCERH). Further ethics approval was then obtained from both the DHS Human Research Ethics Committee (HREC) and the Monash University SCERH for linkage of the VAED.

The general approach for record linkage across datasets included identifying the variables in each dataset that are common, assessing the completeness and accuracy of each common variable, and reviewing their suitability for use in a matching key. The STAR and TAC datasets were linked first, followed by the ANCIS dataset and then the VAED. Linkage was

performed for the period 2000 to 2004 inclusive, except for the VAED where it was performed for 2004 only.

2.1 Linkage of the STAR and TAC Datasets

2.1.1 Key Variables

A review of the STAR and TAC datasets for the year 2000 was performed in order to find the common variables that could be used in the matching process. The following variables were identified:

- SEX – sex of person involved
- AGE – age of person involved
- POSTCODE – postcode of person involved
- NOVEH – number of vehicles involved
- ACCDATE – date of accident
- ACCNO – Police generated accident number
- ACCTIME – time of accident
- DCA – definitions for classifying accidents
- IMPACT – impact code
- USERCODE – type of road user
- LGA – local government area
- REGNO – registration number
- VEHMAKE – vehicle make
- YEARMAN – vehicle year of manufacture
- POWER – vehicle power unit

The TAC dataset contains 19,671 cases and the STAR dataset contains 48,935 cases. It was assumed that for every case listed in the TAC dataset, there existed a corresponding case within STAR apart from claims made for interstate crashes. A summary of the common variables, including the number of missing values, is shown in Table 1.

Table 1: Summary of common variables

Common Variable Name	Relevant Variable in TAC Data File	Number of Missing Values in TAC File	Relevant Variable/s in STAR Data File	Number of Missing Values in STAR File
SEX	tsex	0	sex	831
AGE	tage	31	age	3,896
POSTCODE	post	142	perpcde	18,452
NOVEH	novehinv	0	no_veh	0
ACCDATE	tdateacc	0	acc_year acc_mnth acc_day	0 0 0
ACCNO	taccno	2,231 [†]	acc_num	0
ACCTIME	tacctime	394	acc_time	55
DCA	dcacode	2,231 [†]	dca	0
IMPACT	impcode	2,892	impact	1,839
USERCODE	claimant	72	seat_pos rdustype veh_type	831 831 831
LGA	lga	2,231 [†]	lga lga_oth	32 -
REGNO	tregno	0	reg_no	2,924
VEHMAKE	vehmake	3,835	veh_make	5,924
YEARMAN	yearman	3,826	veh_year	4,693
POWER	pwrunit	7,136	veh_pwr	12,957

[†] These values were missing from the same cases

2.1.2 Review of Matching Variables

The main objective in the record linkage process was to develop a strategy which produced a high match rate, but which minimised the number of false matchings. For the STAR and TAC data files, the best variables for matching purposes were those assigned or collected by the Police and subsequently used by the TAC in their claims database. Of the common variables listed in Table 1, ACCNO is used in this way and provides the clearest link between the STAR and TAC datasets. In particular, it provides a *unique* way to match a specific crash in each dataset.

However ACCNO was limited as a matching variable due to the large number of missing values present in the TAC dataset. Of the 19,671 cases listed, 2,231 values or 11% of accident numbers were missing. There were no missing accident numbers in the STAR dataset. Discussion with the TAC revealed that if a crash report was not available at the time of the claim, a TAC generated accident number was assigned to the crash. In the past, Police accident numbers were assigned to the claims database retrospectively once they became available. However this practice was reported to have ceased during 1999 leading to a lack of valid Police accident numbers being present in the claims data supplied. Subsequent discussion with the TAC indicates that this problem has or is being rectified.

Following identification of a specific crash, REGNO provided the most unique way of distinguishing between motor vehicles so that the next step in identifying individual claimants for record linkage could occur. (Under the *Road Safety Act*, motor vehicles include motor cars, motor cycles, trucks and buses.) Table 2 shows a breakdown *aside* from registration numbers of the TAC variable tregno and the STAR variable reg_no. It can be seen that the variable tregno does not refer solely to motor vehicles, but also includes values such as 'TRAM' in addition to unknown or not applicable values. On the other hand, the non-registration number breakdown of reg_no shows either missing values of which there were

2,924, or unknown or not applicable values. By coding REGNO as a string variable, non-motor vehicle, missing, unknown or not applicable values for the variables tregno and reg_no could be blanked so that these blank values could be used in the linkage of STAR and TAC cases that had the same accident number, but lacked a motor vehicle registration number that would have prevented a match otherwise.

Table 2: *Non-registration number breakdown of tregno and reg_no*

tregno (TAC)		reg_no (STAR)	
<i>missing</i>	0	<i>missing</i>	2,924
"UNDIS"	1	N.K	1
"	1	N/A	7
BICYCLE	16	N/K	1,232
HITRUN	68	NK	5
HORSE	1	NOT APPLIC	3
HORSED	1		4,172
NOTAPPL	1		
NOVEH	10		
PERMIT	7		
TRAIN	102		
TRAM	138		
U/K	6		
UNKNOWN	1		
UNDIS	325		
UNKNOWN	66		
UNREG	18		
	762		

In order to explain the relatively large number of missing values in the STAR dataset under reg_no, an examination of the next level of detail required to achieve record linkage of an individual claimant was necessary. Within the matching process, this information was provided via the common variable USERCODE. For example, once a particular accident was identified via ACCNO and a specific vehicle was identified via REGNO, the variable USERCODE was used to distinguish between occupants of motor vehicles via their seat position.

For the TAC dataset, USERCODE was based on the variable claimant, whereas for the STAR dataset, USERCODE was based on the variables seat_pos, rdustype and veh_type. Table 3 shows a motor vehicle-related breakdown of these variables and Table 4 shows a breakdown of the non-motor vehicle values.

Table 3: Motor vehicle-related breakdown of variables associated with USERCODE

claimant (TAC)		seat_pos (STAR)	
DRIVER	9,966	CENTRE-FRONT	102
FRONT CENTRE PASSENGER	42	CENTRE-REAR	786
FRONT LEFT PASSENGER	2,867	DRIVER OR RIDER [‡]	31,831
REAR RIGHT PASSENGER	536	LEFT-FRONT	7,723
REAR CENTRE PASSENGER	257	LEFT-RIGHT	2,315
REAR LEFT PASSENGER	723	OTHER-REAR	809
REAR PASSENGER	753	PILLION PASSENGER	131
RIDER	1,705	MOTORCYCLE SIDE CAR PASSENGER	1
PILLION OR SIDE-CAR	104	RIGHT-REAR	1,588
	16,953		45,286

rdustype (STAR)	
DRIVER(CARS/TRUCKS/ETC)	28,110
PASSENGER(CAR/TRUCK/BIKE..)	14,326
MOTOR CYCLIST	1,939
PILLION PASSENGER	144
OTHER PASSENGER	61
	44,580

veh_type (STAR)	
CAR	31,389
STATION WAGON	5,434
TAXI	656
UTILITY	2,067
PANEL VAN	1,060
SEMI-TRAILER	534
TRUCK(EXCLUDING SEMI)	872
BUS/COACH	861
MINIBUS(9-13 SEATS)	103
MOTOR CYCLE	2,064
OTHER VEHICLE	247
	45,287

[‡] This value included bicyclists

Table 4: Non-motor vehicle breakdown of variables associated with USERCODE

claimant (TAC)		seat_pos (STAR)	
<i>missing</i>	72	<i>missing</i>	831
TRAM	107	NOT APPLICABLE	1,852
TRAIN	62	NOT KNOWN	966
PEDESTRIAN	1,688		3,649
PEDAL CYCLIST	662		
HORSE	3		
OTHER	124		
	2,718		

rdustype (STAR)	
<i>missing</i>	832
PEDESTRIAN	1,839
BICYCLIST	1,047
OTH DRIVER(HORSE/TRAM/TRAIN)	128
UNKNOWN	509
	4,355

veh_type (STAR)	
<i>missing</i>	1,839
BICYCLE	1,077
HORSE	9
TRAM	166
TRAIN	14
MOPED	9
MOTOR SCOOTER	17
UNKNOWN	517
	3,648

From Table 3 it can be seen that for persons injured in a motor vehicle, USERCODE became a variable for matching seat position based on the TAC variable claimant and the STAR variable seat_pos. Now considering non-motor vehicle cases (Table 4), reference was again made to the 2,924 missing values under reg_no. A comparison with the variable rdustype revealed that nearly all of the missing values were associated with the values ‘PEDESTRIAN’ and ‘BICYCLIST’. In other words, where rdustype listed a case as a pedestrian or bicyclist, the STAR dataset essentially provided a missing value for the associated registration number field under reg_no.

A closer examination of the TAC variable claimant for the values ‘PEDESTRIAN’ and ‘PEDAL CYCLIST’ (Table 4) revealed that most of these values *were* associated with a registration number under tregno. Since the STAR dataset did not provide related registration number information for these road user types, the TAC variable claimant was utilised to “blank” TAC registration numbers for the relevant cases so that the common variable REGNO could be included as part of a matching key. This was achieved through the use of blank string values for REGNO, allowing record linkage to occur using ACCNO and USERCODE for these cases.

The values ‘PEDESTRIAN’ and ‘PEDAL CYCLIST’ made up around 86% of the non-motor vehicle values under claimant (Table 4). The next largest value was ‘TRAM’, however rather than being associated with a TAC registration number, each case was associated with the value ‘TRAM’ under tregno. This was also the case for the values ‘TRAIN’ and ‘HORSE’, leaving ‘OTHER’ as the only other non-motor vehicle value for claimant.

REGNO and USERCODE

Overall, the review of matching variables suggested that the best way to utilise the common variable REGNO for non-motor vehicle cases was through the use of blank string values. The common variable USERCODE was then used to match the non-motor vehicle type, e.g. ‘TRAM’. For the TAC dataset, the variable claimant formed the basis for blanking non-registration numbers under tregno and was then used as the basis for USERCODE for all cases. For the STAR dataset, although most pedestrian and bicyclist values were already blanked under reg_no, the variables rdustype and veh_type (Table 4) were able to be used to blank remaining non-motor vehicle values under reg_no, with the three variables seat_pos, rdustype and veh_type available to form USERCODE.

The use of the variable claimant (TAC) as the basis for blanking values under tregno not only saw removal of associated registration numbers, but also most of the non-registration values under tregno (Table 2). For example, blanking based on the value ‘PEDAL CYCLIST’ under claimant resulted in 15 out of the 16 values ‘BICYCLE’ under tregno to be removed. A further

blanking based on the missing values of claimant saw the removal of the last value. However after completion of this process, some non-registration values still remained. A direct blanking of the non-registration values of tregno achieved a final “tidying up” of the common variable REGNO derived from the TAC dataset. Similarly for the STAR dataset, a final tidy up of the variable reg_no saw REGNO ready to be used for record linkage.

As mentioned above, for the STAR dataset the variables seat_pos, rdustype and veh_type were available to create USERCODE. For motor vehicle-related values, the following process was followed in the translation of these variables into the one common variable. The variable seat_pos was used to create values consistent with those under claimant (Table 3). However the value ‘DRIVER OR RIDER’ was not included at this point as it included bicyclists. (Although bicyclists are not defined as a motor vehicle under the *Road Safety Act*, they are defined as a vehicle in STAR). Rather, the variable rdustype was used to split the value ‘DRIVER OR RIDER’ by the values ‘DRIVER(CARS/TRUCKS/ETC)’ and ‘MOTOR CYCLIST’ (Table 3) and ‘BICYCLIST’ and ‘OTH DRVER(HORSE/TRAM/TRAIN)’ (Table 4). Table 5 shows a breakdown of the values under rdustype for cases that also had the value ‘DRIVER OR RIDER’ for seat_pos.

Table 5: Values of rdustype for cases that also had the value ‘DRIVER OR RIDER’

seat_pos (STAR)		rdustype (STAR)	
DRIVER OR RIDER	31,831	<i>missing</i>	1
	31,831	DRIVER(CARS/TRUCKS/ETC)	28,091
		PASSENGER (CAR/TRUCK/BIKE..)	164
		MOTOR CYCLIST	1,939
		PILLION PASSENGER	12
		BICYCLIST	1,047
		OTH DRIVER(HORSE/TRAM/TRAIN)	128
		OTHER PASSENGER	1
		UNKNOWN	448
			31,831

A closer examination of this breakdown suggested that the variable rdustype was sufficient for separation of the value ‘DRIVER OR RIDER’ into the categories highlighted in Table 5. In particular, it was found that the variable veh_type was consistent with rdustype. Only the highlighted values were utilised as only these values were sensible as a subset of ‘DRIVER OR RIDER’. In addition, the values ‘MOTOR CYCLIST’, ‘BICYCLIST’ and ‘OTH DRIVER(HORSE/TRAM/TRAIN)’ represented all such values within the STAR dataset. Overall, this suggested that the use of the variable veh_type in the translation of the variables seat_pos, rdustype and veh_type into USERCODE was unnecessary. Following completion of the above process it only remained to use the variable rdustype to recode the value ‘PEDESTRIAN’ (Table 4) so that the variable USERCODE was ready to be used for record linkage.

Use of the Common Variable SEX

As a blank string variable was utilised for REGNO for non-motor vehicle related cases in order to match these cases on the basis of ACCNO and USERCODE, the common variable SEX was added to the matching key in order to improve the accuracy. Adding an extra matching variable should help to achieve this, however a reduction in the match rate could also occur. Considering that non-vehicle related cases were only matched on the basis of ACCNO and USERCODE, the addition of SEX seemed necessary. Assuming that the

variables sex (STAR) and tsex (TAC) were essentially correct, the main limitation of using SEX was the number of missing values. From Table 1 it can be seen that there were no missing values for the variable tsex, but 831 values missing for sex. In addition, for the STAR dataset there were 1,387 cases with unknown values for sex. Combined with the missing values, this was equivalent to around 5% of the total number of cases.

2.1.3 Matching Protocol

Matching of the STAR and TAC datasets was performed using SPSS 15.0 (2006) in three stages using the following common variables as matching keys:

Stage 1

ACCNO, REGNO, USERCODE, SEX

Stage 2

ACCNO, USERCODE, SEX

Stage 3

ACCNO, REGNO, SEX

Stage 2 attempted to address the issue of missing values under REGNO, whilst Stage 3 attempted to deal with missing values under USERCODE. Only cases not matching in a prior stage were used in each subsequent stage.

Each stage of the matching process used the common variables AGE and ACCDATE to reject matched cases that did not meet specific criteria. For ACCDATE this occurred where the difference in the date of the accident was equal to or greater than 31 days, and for AGE, rejection of matched cases occurred where the age difference was equal to or greater than 4 years. For example, at Stage 1 around 2% of initially linked cases were rejected. Appendix 1 shows the flow diagram for linking the STAR and TAC datasets for the year 2000 (the file extensions *.sav and *.sps represent SPSS data files and syntax files respectively). The corresponding SPSS syntax is shown in Appendix 2.

2.2 Linkage of the ANCIS Dataset

The Australian National Crash In-depth Study (ANCIS) dataset provides detailed data for a relatively small number of crashes generally involving at least one injured person. Based on the sampling frame, it was assumed that for every Victorian case listed in ANCIS, there existed a corresponding case within STAR. Inclusion of ANCIS cases in a general linked dataset involved their initial linkage to STAR. The subsequent matching of TAC cases was based on the protocol for the linkage of STAR and TAC datasets as outlined in Section 2.1. However, the addition of TAC cases potentially saw some linked STAR-ANCIS cases drop away in the absence of a matching TAC claim.

Linkage of the STAR and ANCIS datasets was performed in two stages using the following common variables as matching keys:

Stage 1

ACCNOx, USERCODE, SEX, AGE

Stage 2

VIN, USERCODE, SEX, AGE

In the above ACCNOx is the Accident Number (ACCNO) including year and single alphanumeric prefix and VIN is the Vehicle Identification Number.

Overall, the linkage process was designed to generate accurate linkage rather than a high match rate thus allowing for meaningful cross-comparisons to be made between the ANCIS dataset and the STAR and TAC datasets. The inclusion of AGE in the matching key (requiring an exact match), together with the relatively small number of ANCIS cases helped to ensure that this occurred. In addition, the linkage of ANCIS cases utilised the complete set of ANCIS data as it was not possible to split the dataset by year due to the lack of crash date for many cases. This was in contrast to treatment of the STAR and TAC datasets where linkage was performed on one year's worth of crash data at a time.

Table 6: *Summary of common variables*

Common Variable Name	Relevant Variable/s in STAR Data File	Number of Missing Values in STAR File	Relevant Variable in ANCIS Data File	Number of Missing Values in ANCIS File
ACCNOx	acc_num	0	accno	149
VIN	vin	1,955	vin	29
USERCODE	seat_pos rdustype veh_type	831 831 831	position	22
AGE	age	3,896	age	23
SEX	sex	831	sex	23
ACCDATE	acc_year acc_mnth acc_day	0 0 0	dtecrash	114

Table 6 lists the common variables used to link the STAR and ANCIS datasets where the STAR data file contained 48,935 cases and the ANCIS data file contained 430 cases. In principle, the linkage of data would first utilise the variable ACCNOx to identify a particular crash and then VIN to identify a particular vehicle (REGNO is unavailable in the ANCIS dataset). However a linkage stage that includes both variables was deemed to be unnecessary leaving two stages in the matching process as described below.

Stage 1 linked cases with matching accident numbers whilst allowing for linkage where the vehicle identification number was missing. This stage also took into account cases with incorrect or non-applicable values for the vehicle identification number. Stage 2 matched cases where an exact match was found for VIN and allowed for linkage where the accident number had not been recorded. This constituted a large proportion of cases in the ANCIS dataset including cases collected for or supplied by Holden that were also missing the date of accident. Following the use of ACCNOx in Stage 1 and VIN in Stage 2, the common variable USERCODE was used to identify a particular occupant via seat position, followed by the variables SEX and AGE. Consistent with the aim of achieving accurate linkage rather than a high match rate, each stage of the matching process used the common variable ACCDATE to

reject matched cases that did not meet the strictest possible criterion, namely where the difference in the date of the accident (where available) was greater than 1 day.

Appendix 3 shows the flow diagram for linking the STAR, ANCIS and TAC datasets for the year 2000. Appendix 4 provides the additional and modified syntax required for the linkage of ANCIS data to that given in Appendix 2 for matching the STAR and TAC datasets alone. The intermediate data file VIC00ANC.sav contained all of the ANCIS cases able to be linked to the STAR dataset for the year 2000.

2.3 Linkage of the VAED

For the purpose of linking the VAED to the STAR, ANCIS and TAC datasets, the VAED (injury subset) held by MUARC was obtained for the period July 2003 to June 2005 inclusive. A matching protocol was then developed to link 2004 data from the VAED.

2.3.1 Selection of VAED Cases for Matching

Prior to linking the VAED to a linked STAR-ANCIS-TAC dataset, the relevant VAED cases were selected from the data provided. The initial data file contained 377,829 cases for the period July 2003 to June 2005 inclusive, including re-admission data. As a de-identified linkage process was required, it was critical that the number of cases was reduced in order to obtain the highest level of accuracy possible. This was achieved using the following process:

- The period January 2004 to January 2005 inclusive was selected using the VAED variable *admdate* (admission date) leaving 201,961 cases;
- Land transport accidents (ICD-10 codes V01 – V89) were flagged using *f_t_extc* (first external cause of morbidity and mortality). Unspecified transport accidents (V99) were also included. A total of 15,094 cases were flagged noting that the variable *f_t_extc* contained no missing values;
- Cases with patient account classification of Compensable – Transport Accident Commission (TAC) were also flagged using *accstat1* (account class 1). Of the 201,961 cases there were 114,116 missing values for this variable and 4,010 cases flagged as having the TAC account classification;
- Of the 4,010 cases flagged as having the TAC account classification, 612 were identified as being related to non-land transport accidents. First external cause (*f_t_extc*) was investigated for these cases and it was found that:
 - 35% had an external cause falling under Y85 – Sequelae of transport accidents;
 - 34% of cases fell under Y83 – Surgical operation and other surgical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure; and
 - 10% fell under X59 – Exposure to unspecified factor; the remaining cases had various other external causes.

In general, non-land transport accidents with a TAC account classification appeared to be re-admissions. As the VAED is a separations database, it was not possible to link re-admission cases to earlier, original hospital admissions in the absence of identifying information, e.g. a

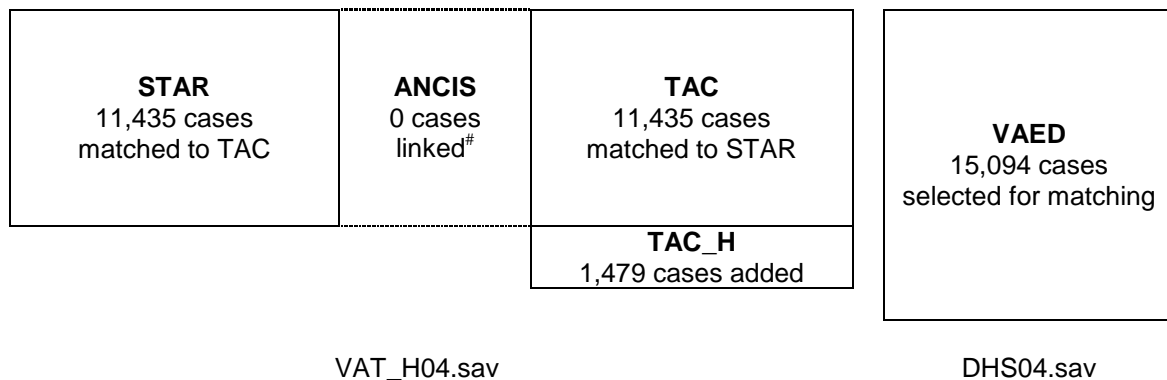
person's name and address. Therefore, these cases could not be included in the VAED data subset for matching purposes.

⇒ Land transport accidents (and unspecified transport accidents) were selected for matching purposes, i.e. 15,094 cases.

2.3.2 Preparation of a Linked STAR-ANCIS-TAC Dataset

Before linking the selected VAED hospital admission records, it was necessary to supplement the matched STAR-ANCIS-TAC dataset with unmatched (leftover) TAC cases. However, it was only necessary to add the relevant unmatched TAC claims, namely those with hospital admission. In the linkage of STAR and TAC data, TAC claims involving both hospital and non-hospital admission were matched, however as described above, for the linkage of hospital data it was critical to keep the number of cases to a minimum in order to obtain the highest level of accuracy possible, with unmatched TAC claims not involving hospital admission no longer required. Non-hospital cases were identified through the TAC variable lenstay, where lenstay = 0 indicated non-admission.

A conceptual diagram of the supplemented STAR-ANCIS-TAC dataset and the selected VAED cases prior to linkage is shown in Figure 1. For 2004, a total of 11,435 STAR and TAC cases were matched leaving 5,367 TAC claims unmatched (TACLeft3.sav). Of these, 1,479 were claims with hospital admission (TAC_H). The original STAR-ANCIS-TAC data file supplemented with unmatched hospital admission cases led to the creation of a new file with 12,914 cases (VAT_H04.sav). As a consequence, an additional 1,479 missing values were created for each STAR variable. The next section describes the development of a matching protocol for the linkage of VAED data to the mass crash data records.



Of the 430 ANCIS cases provided (as at June 2006), only 4 cases were able to be clearly identified as being from 2004 with 0 linking to the STAR dataset.

Figure 1: Conceptual diagram of STAR-ANCIS-TAC and VAED datasets prior to linkage

2.3.3 Determination of a Matching Protocol

Possible variables in the VAED (injury subset) held by MUARC that could be utilised in a matching key included:

- admdate – admission date
- age – age in years (non-integer)
- modage – age in years truncated to an integer

- campname – name of campus
- pstcode – postcode of patient’s usual residence
- sex – sex of patient

All of these variables were essentially complete, containing no missing values.

As described, the addition of leftover TAC hospital cases to the original linked STAR-ANCIS-TAC data file saw the same number of missing values created for each STAR variable. Clearly in identifying common variables for matching between the supplemented STAR-ANCIS-TAC and the VAED datasets, the use of TAC rather than STAR variables was preferable considering that the relevant TAC variables contained few if any missing values. However the use of several matching stages meant that a matching key could still include a common variable derived from STAR in initial steps where the aim was to produce a higher degree of accuracy. In particular, the TAC claims data held by MUARC did not contain the name of the admitted hospital, therefore the STAR variable HospitalName was utilised to create the common variable HSP which was included in two initial stages. A summary of the common variables used in the matching process is shown in Table 7.

Table 7: Summary of common variables

Common Variable Name	Relevant Variable in VAED Data File	Number of Missing Values in VAED File	Most Appropriate Variable in VAT_H04.sav	Number of Missing Values in VAT_H04.sav
ADMDATE	admdate	0	tdateacc (TAC)	0
MODAGE	modage	0	tage (TAC)	0
HSP	campname	0	HospitalName (STAR)	5,130
PSTCODE	pstcode	0	post (TAC)	34
PERSEX	sex	0	tsex (TAC)	0

Linkage of the supplemented STAR-ANCIS-TAC and VAED datasets was performed in six stages using the following common variables as matching keys:

Stage 1

ADMDATE HSP PSTCODE MODAGE PERSEX

Stage 2

ADMDATE PSTCODE MODAGE PERSEX

Stage 3

ADMDATE1 HSP PSTCODE MODAGE PERSEX

Stage 4

ADMDATE1 PSTCODE MODAGE PERSEX

Stage 5

ADMDATE PSTCODE PERSEX

Stage 6

ADMDATE1 PSTCODE PERSEX

Reflecting the use of de-identified data and in order to reduce the likelihood of false matchings, it was necessary to use all available common variables in the linkage process in addition to reducing the number of VAED cases available for matching as described above. As per the earlier protocols each additional stage attempted to link leftover cases, thus each stage of the matching process was designed to become slightly less restrictive.

Stages 1 and 2 required an exact match for hospital admission date. In cases where admission was listed as the day after an accident, Stages 3 and 4 used identical matching keys, however shifted the date for matching appropriately (ADMDATE1). Stages 1 and 3 both used the variable HSP for accuracy, with this variable being dropped in Stages 2 and 4 respectively due to missing information. This included the situation where hospital names in STAR did not have an equivalent in the VAED. In addition, cases where the admitted hospital was private were recoded 'Private' for matching purposes consistent with the VAED which identified private hospital admission in this way. Stages 5 and 6 did not include HSP and loosened the restrictions on age by not including MODAGE in the matching key, however did reject matched cases where an age difference of greater than 1 year occurred.

Appendix 5 shows the flow diagram for linking the supplemented STAR-ANCIS-TAC and VAED datasets for the year 2004. The corresponding SPSS syntax is shown in Appendix 6.

3. Results

3.1 STAR and TAC Datasets – Match Rates

Table 8 shows the match rates achieved for the years 2000 to 2004 inclusive for the STAR and TAC datasets. The match rate was defined as the number of matched cases divided by the number of cases in the TAC dataset, expressed as a percentage. It was reasonable to expect that for every TAC claimant, a corresponding case was contained in the STAR data file apart from claims made for interstate crashes. It followed that for the de-identified linkage process utilised, match rates lower than 100% were due to such crashes along with missing information and inaccuracies in the data. In particular, the Police generated accident number, which was a key matching variable, caused some problem due to the number of missing values present in the TAC dataset (up to 15% of cases), leading to a reduction in the maximum possible match rate. To improve the overall rate, the TAC would need to ensure that all claims contain a valid and verified Police accident number. It is understood that this problem has been or is being rectified by the TAC.

Table 8: Match rates for 2000 to 2004

Year	Match Rate (%)
2000	75
2001	75
2002	76
2003	72
2004	68

3.2 STAR-ANCIS-TAC Dataset – Match Rate

As described in Section 2.2, the linkage of ANCIS cases utilised the complete set of ANCIS data as it was not possible to split the dataset by crash year due to lack of date of accident for

a high proportion of cases including all of the cases collected for or supplied by Holden. In addition, the linkage process was designed to generate accurate linkage rather than a high match rate thus allowing for meaningful cross-comparisons to be made between the ANCIS dataset and the STAR and TAC datasets. Although the date of crash was missing from a large number of cases it was estimated that 368 cases were *collected* during the period 2000 to 2004 inclusive allowing for an approximate match rate to be calculated. The ANCIS dataset contained a total of 430 cases, of which 241 cases were initially linked to STAR giving an initial match rate of 65%. Following the additional linkage of the TAC dataset, this was reduced to 156 cases due to the absence of a matching TAC claim giving an overall match rate of 42%.

3.3 Cross-Comparison of the STAR, TAC and ANCIS Datasets

The linkage of records from ANCIS with corresponding records in the STAR and TAC datasets allowed data quality issues in the mass crash data sources, such as accident details and the critical issue of injury coding to be assessed. This was achieved through the cross-comparison of STAR and TAC variables with corresponding ANCIS variables in the linked STAR-ANCIS-TAC dataset for the year 2000. In a preliminary step, the quality of linkage between the STAR and TAC datasets (based on 14,692 matched cases) was assessed through a comparison of Police-collected accident details, namely the number of vehicles involved in an accident (NOVEH) and the Definitions for Classifying Accidents (DCA) code. As the corresponding variables in each dataset came from the same source and were not utilised in the matching process, it may be expected that they have the same values where a correct match has occurred. Equal values occurred for 96% of matched cases for NOVEH and 77% of matched cases for DCA. Why only 77% of the DCA codes matched is unclear and could only be answered through in-depth probing of the processes involved in assembling the TAC data which were beyond the scope and access of the project. This point would be worth further investigation by the TAC but may relate to the TAC taking their DCA coding straight from the Victoria Police Traffic Incident System (TIS) rather than from VicRoads who it is understood review and in some cases re-code DCA in assembling STAR. A further comparison of DCA coding between the STAR and ANCIS datasets for 2001 saw equal values for 71% of matched cases so the inconsistency is not confined to one year alone.

NOVEH and DCA were chosen for comparison between STAR, TAC and ANCIS as they were present in each dataset. For the cross-comparison, 45 ANCIS cases were successfully linked to corresponding STAR and TAC cases in the combined STAR-ANCIS-TAC dataset. Equal values occurred between STAR and ANCIS for 80% of cases for NOVEH and 42% of cases for DCA. The low correspondence rate of the DCA coding is again a concern. In the ANCIS dataset the values of these variables were determined by the in-depth crash investigators post-accident rather than coming from Police reports as per the STAR and TAC datasets. The observed lack of concordance between the two coders of DCA suggests that there is significant room for error or difference in opinion in coding crash DCA and that the variable should be treated with some caution. However, the 2001 concordance rate for DCA in the STAR and ANCIS databases was much higher at 82% suggesting a change in DCA coding practice by the ANCIS coders given it is understood that there was no comprehensive change in DCA coding practices by the Police or VicRoads during this period. If this is the case, the DCA recorded in STAR might be more accurate with ANCIS needing to review DCA coding practices to ensure consistency. The exact reasons for the lack of concordance in DCA coding and changes in concordance over time needs further closer investigation to establish the relative accuracy of this variable in each data source. Investigation may include

review of the coding in both data sources by a third independent coder expert in DCA coding with reference to the original Police report crash descriptions and diagrams.

With regards to injury coding, it was useful to compare variables in the TAC dataset that could be used as measures of serious injury with corresponding variables that were available in the ANCIS dataset. A report by Hoareau et al. (2006) analysing injury severity trends in TAC claims between 2000 and 2004, referred to hospitalisation as recorded in Police crash reports as a relatively coarse definition of serious injury. A number of other potentially finer serious injury measures derived from the TAC data included: length of hospital stay, Maximum Abbreviated Injury Scale (MAIS), Injury Severity Score (ISS) and ICD-based Injury Severity Score (ICISS).

The measures MAIS and ISS were both available in the ANCIS dataset, hence a cross-comparison of these alternative measures of serious injury was performed. Figures 2 and 3 show cross-comparisons for MAIS and ISS respectively derived from each data source. A good correlation ($R^2 = 0.72$) existed between the TAC and ANCIS datasets for the injury measure ISS with a weaker correlation ($R^2 = 0.50$) present for MAIS. A broader assessment of TAC injury severity measures involving many more cases, i.e. between matched TAC and VAED data was contingent on the successful linkage of the VAED to the STAR, ANCIS and TAC datasets (Section 2.3).

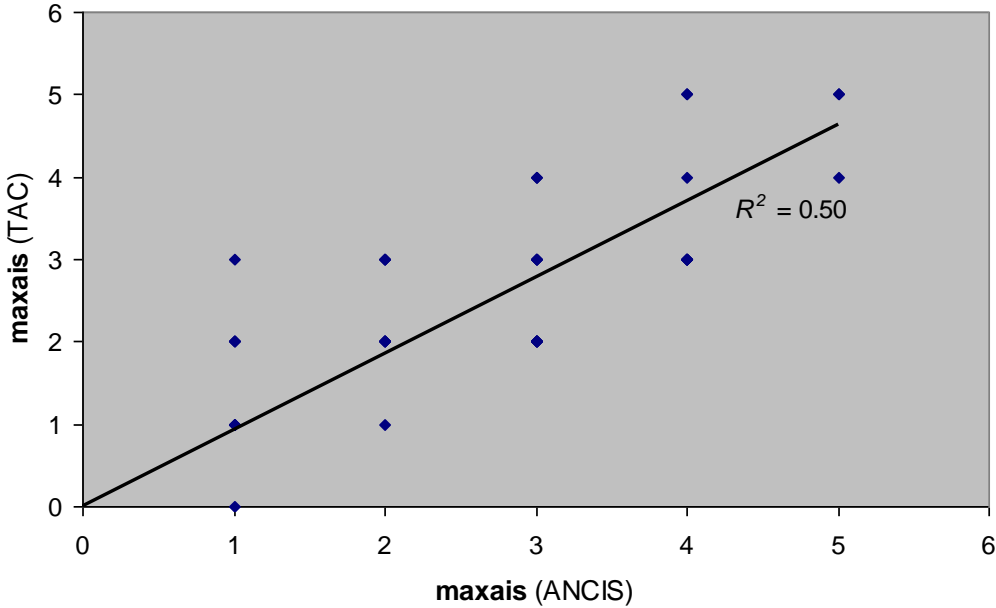


Figure 2: Cross-comparison of MAIS between the ANCIS and TAC datasets

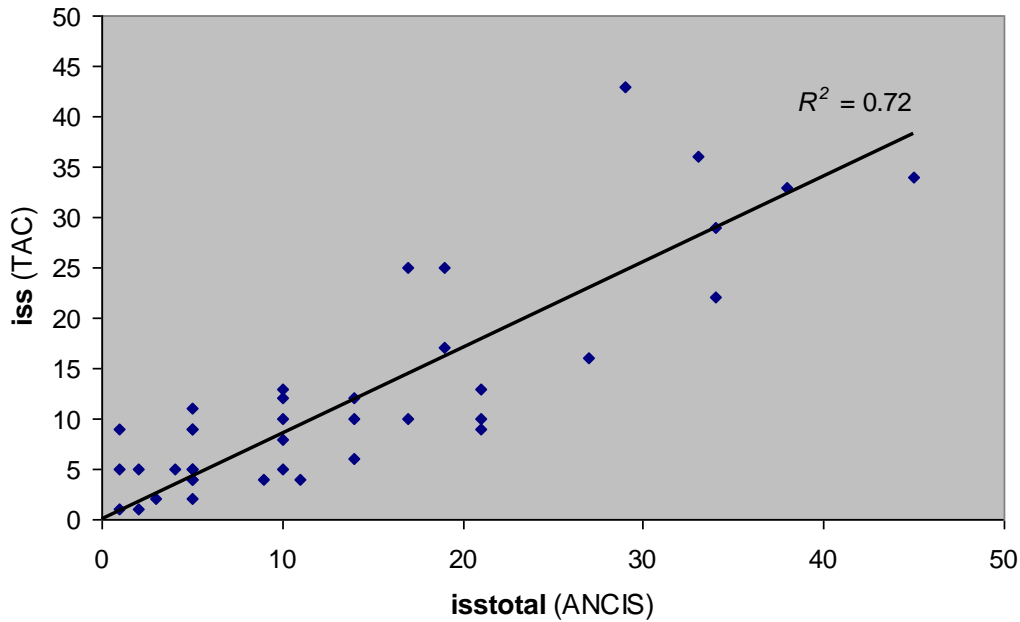


Figure 3: Cross-comparison of ISS between the ANCIS and TAC datasets

3.4 Linkage of the VAED

A poor match rate of 36% was achieved for the linkage of the VAED to the supplemented STAR-ANCIS-TAC dataset. This indicated that the de-identified linkage of the VAED to the STAR-ANCIS-TAC dataset was not feasible without identifying information. Based on this result it seems the effective matching of VAED data to other Victorian injury data records requires a matching process based on person details. The low match rate reflected some of the principal problems encountered in linking the VAED including the identification of road crash related hospital admissions, lack of information on location of the event leading to hospital admission reducing the number of variables that could be used in a matching key, and identifying the relevant transfer and re-admission cases related to the VAED being a separations database.

Transfer and re-admission cases are included in the VAED as new additional cases. An analysis of duplicates based on the variables admdate, age, postcode and sex found that 92% of the 306 duplicate cases with an earlier separation date were labelled “Separation and transfer to acute hospital/extended care”. In addition, re-admissions with a TAC claim may occur more than a year after initial admission. As linkage was performed on a year-by-year basis, it is possible that some re-admission cases should be matched to an earlier year of TAC claims and mass crash data along with the initial admission record.

4. Discussion and Conclusions

Cross-comparison of the injury severity measure ISS between the ANCIS and TAC datasets ($R^2 = 0.72$) suggested that the use of TAC claims data merged with STAR could provide an alternative and finer measure of serious injury for road safety research to hospitalisation as recorded in Police crash reports. Use of the common variable ACCNO was caused some problem as a matching variable due to the number of missing values present in the TAC dataset, leading to a reduction in the maximum possible match rate. To improve the overall rate, the TAC would need to ensure that all claims contain a valid and verified Police accident

number. It is understood that this has been or is being rectified by the TAC meaning fully linked TAC claims and Police crash reports should be viable.

An alternative to the linkage of the STAR and TAC datasets would be for the TAC to match claims data to Police data in-house. In-house linkage has been performed by matching Police-reported crash data from the Victoria Police Traffic Incident System (TIS) with TAC claims data and STAR. The TAC has reportedly achieved a match rate of 85%, with the non-matched claims explained by interstate and public transport cases.

These results suggest that linkage of Police reported crash data with the TAC claims dataset is feasible and results in a combined database more capable of measuring injury outcome in detail. This is particularly significant considering that the de-identified linkage of hospital admissions data was not found to be feasible without identifying information. Although the TAC claims dataset was put together primarily to manage claims, a new project has commenced that seeks to collaborate with the TAC in order to develop a linked dataset for use in road safety research. This will include a review of the injury coding practices of the TAC and a process of consultation with the key agencies to automate and streamline the linkage process into an ongoing system.

Future work will identify the most appropriate measure or measures of serious injury from the linked dataset for use in road safety performance monitoring and other projects such as time series modelling. In particular, operational biases inherent in serious injury measures derived from Police crash reports (defined as taken to hospital for the period examined for this project, previously defined as being admitted to hospital) should be overcome. Victoria Police have advised that since 2006 this has changed to a combination of taken and admitted to hospital.¹ Indeed, the use of an alternative and finer measure or measures of serious injury should more closely reflect real serious injury trends and provide a more satisfactory key outcome measure for Victoria's *arrive alive 2008 – 2017* road safety strategy.

The second major objective of the project was to assess the feasibility of linking hospital admission records with the linked STAR-ANCIS-TAC dataset using de-identified data. It was found that the de-identified linkage of the VAED to the other Victorian injury data records was not feasible. Problems encountered included the identification of road crash related hospital admissions, lack of information on location of the event leading to hospital admission reducing the number of variables that could be used in a matching keys, and identifying the relevant transfer and re-admission cases related to the VAED being a separations database. It was concluded that identifying information would be required to both link relevant cases within the VAED as well as for linkage to other Victorian injury data records. This is particularly important if comparisons between datasets are to be made such as comparisons between ICD injury codes recorded in both the VAED and TAC claims data.

In order to link hospital data with the other data sources considered, it is recommended that Victoria set up a Data Linkage Unit (DLU) similar to that in Western Australia. Each year in WA, the DLU links road crash reports to hospital morbidity and mortality data using personal identifiers. Main Roads WA provides the identified road crash data to the DLU and the identifiers are removed after linkage. This dataset is then used to produce the annual report on road crashes in Western Australia (WA Police Service et al., 2004). The linkage of hospital and Police crash datasets using identifying information has also occurred in NSW. The Injury

¹ Victoria Police has also advised that some TAC data items relating to admitted to hospital are based on an "administrative admission" where a person spent four hours at hospital, not necessarily being admitted to a ward.

Risk Management Research Centre matched records using information that was common to the datasets, such as names, residential addresses and dates of birth. When hospital inpatient separations data was restricted to records relating to motor vehicle traffic crashes on public roads a match rate of 69% was achieved (Boufous et al., 2008). In addition to linking road crash and hospital data, a Western Australian style data linkage unit using identifying information would also enable the linkage of Victoria's core population health datasets. This currently occurs in WA with access possible for bona fide research and evaluation projects that have been approved by an institutional ethics committee.

5. Key Recommendations

Key recommendations resulting from this project are that:

- The TAC perform ongoing in-house linkage of Police-reported crash data from the Victoria Police Traffic Incident System (TIS) with TAC claims data and STAR.
- The dataset linked by the TAC is used to identify alternative measures of serious injury with the potential to be used as key outcome measures for *arrive alive 2008-2017*.
- A Western Australian style data linkage unit is set up to enable the linkage of Victoria's core population health datasets including with mass crash data sources.

6. References

Boufous, S., Finch, C., Hayen, A. and Williamson (2008). *Data linkage of hospital and Police crash datasets in NSW*. Injury Risk Management Research Centre, University of New South Wales.

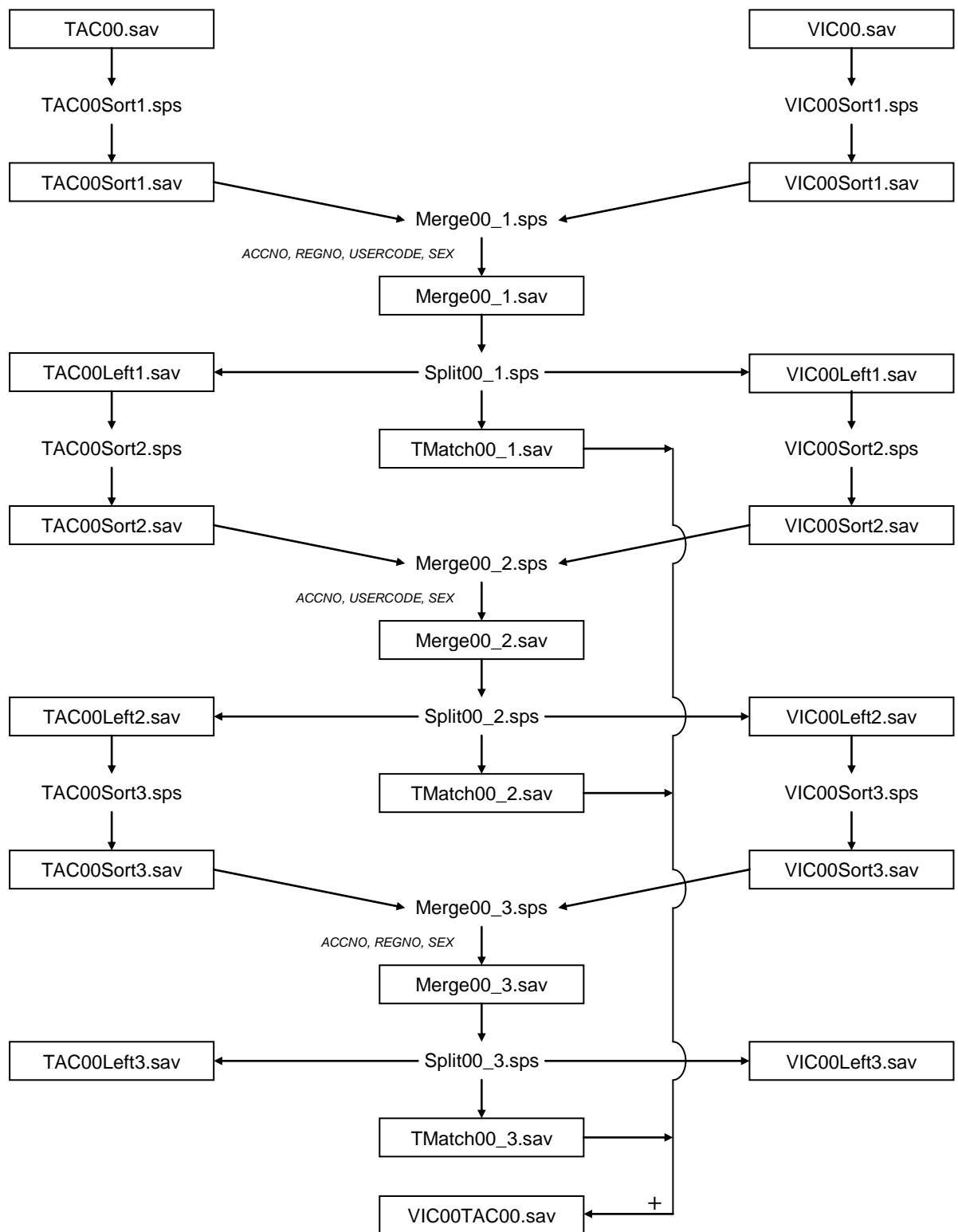
Hoareau, E., Newstead, S. & D'Elia, A. 2006. *An Analysis of Injury Severity Trends in Transport Accident Commission Claims: 2000-2004*. Monash University Accident Research Centre, Report to the Transport Accident Commission.

SPSS for Windows, Rel. 15.0.0. 2006. Chicago: SPSS Inc.

Western Australian Police Service, Main Roads Western Australia, Insurance Commission of Western Australia, Department of Health, Department for Planning and Infrastructure and Office of Road Safety (2004). *Road crash data sharing between government agencies*, Report for the Road Safety Council.

Appendix 1

Flow Diagram for Linkage of the STAR and TAC Datasets



+ VIC00TAC00.sps

Appendix 2

Syntax for Linkage of the STAR and TAC Datasets

VIC00Sort1.sps

GET FILE 'VIC00.sav'.

NUMERIC ACCNO(F6.0).
COMPUTE ACCNO=NUMBER(SUBSTR(ACC_NUM,6,6),F6.0).
EXECUTE.

NUMERIC ACCDATE(F12.0) RNEWDATE(F12.0).
COMPUTE ACCDATE=
YRMODA(ACC_YEAR,NUMBER(ACC_MNTH,F2.0),NUMBER(ACC_DAY,F2.0))
- YRMODA(1960,1,1).
COMPUTE RNEWDATE = (ACCDATE + YRMODY(1960,1,1)) * 86400.
FORMAT RNEWDATE(DATE9).
EXECUTE.

NUMERIC USERCODE(F2.0).

IF (SEAT_POS='CF') USERCODE=22.
IF (SEAT_POS='LF') USERCODE=23.
IF (SEAT_POS='RR') USERCODE=24.
IF (SEAT_POS='CR') USERCODE=25.
IF (SEAT_POS='LR') USERCODE=26.
IF (SEAT_POS='OR') USERCODE=27.
IF (SEAT_POS='PL' OR SEAT_POS='PS') USERCODE=42.

IF (SEAT_POS='D ' AND RDUSTYPE=2) USERCODE=21.
IF (SEAT_POS='D ' AND RDUSTYPE=4) USERCODE=41.
IF (SEAT_POS='D ' AND RDUSTYPE=6) USERCODE=91.
IF (SEAT_POS='D ' AND RDUSTYPE=7) USERCODE=60.

IF (RDUSTYPE=1) USERCODE=90.
EXECUTE.

VALUE LABEL USERCODE 20 'VEH POSN UNKNOWN'
21 'DRIVER'
22 'FRONT CENTRE PASS'
23 'FRONT LEFT PASS'
24 'REAR RIGHT PASS'
25 'REAR CENTRE PASS'
26 'REAR LEFT PASS'
27 'REAR PASS'
40 'MCYCLE POSN UNKNOWN'
41 'RIDER'
42 'PILLION OR SIDE-CAR'
60 'TRAM/TRAIN/HORSE'
90 'PEDESTRIAN'
91 'PEDAL CYCLIST'
99 'OTHER'.

EXECUTE.

STRING REGNO(A6).
COMPUTE REGNO=REG_NO.
EXECUTE.

IF ((REGNO='N/K ') OR (REGNO='N/KNOW') OR (REGNO='TRAM ') OR
(REGNO='UN-REG') OR (REGNO='TRAM 2') OR (REGNO='UNREGI') OR
(REGNO='UN REG') OR (REGNO='UNREG ') OR (REGNO='N/A ') OR
(REGNO='N.K ') OR (REGNO='UNKNOW') OR (REGNO='HORSE ') OR
(REGNO='NK ') OR (REGNO='NOT AP') OR (REGNO='TRAIN ') OR

```
(REGNO='TRAM #')) REGNO='  '.  
EXECUTE.
```

```
STRING RREGNO (A6).  
STRING RSEX (A1).  
NUMERIC RAGE(F3.0).  
COMPUTE RACCNO=ACCNO.  
COMPUTE RREGNO=REGNO.  
COMPUTE RACCDATE=ACCDATE.  
COMPUTE RSEX=SEX.  
COMPUTE RAGE=AGE.  
COMPUTE RUSERCOD=USERCODE.  
EXECUTE.
```

```
SORT CASES BY ACCNO REGNO USERCODE SEX.  
EXECUTE.
```

```
SAVE OUTFILE 'VIC00Sort1.sav'.  
EXECUTE.
```

```
FREQ VAR = SEX USERCODE RDUSTYPE ACC_YEAR SEAT_POS VEH_TYPE.
```

```
LIST VAR = REG_NO ACC_NUM SEX AGE ACCDATE ACC_DAY ACC_MNTH ACC_YEAR  
          USERCODE SEAT_POS VEH_TYPE RDUSTYPE  
          /CASES=FROM 1 TO 50.
```

TAC00Sort1.sps

```
GET FILE 'TAC00.sav'.
```

```
NUMERIC ACCNO (F6.0).  
STRING ACCNOS (A6) YEARACC (A4).  
EXECUTE.
```

```
COMPUTE ACCNOS = SUBSTR(TACCNO,6,6).  
COMPUTE YEARACC = SUBSTR(TACCNO,2,4).  
EXECUTE.
```

```
COMPUTE ACCNO=NUMBER(SUBSTR(TACCNO,6,6),F6.0).  
FREQ VAR = YEARACC CLAIMANT.
```

```
STRING SEX (A1).  
COMPUTE SEX=TSEX.  
COMPUTE AGE=TAGE.  
COMPUTE ACCDATE=TDATEACC.  
COMPUTE USERCODE=CLAIMANT.  
RECODE USERCODE (50,60,92=60).  
EXECUTE.
```

```
VALUE LABEL USERCODE 20 'VEH POSN UNKNOWN'  
                21 'DRIVER'  
                22 'FRONT CENTRE PASS'  
                23 'FRONT LEFT PASS'  
                24 'REAR RIGHT PASS'  
                25 'REAR CENTRE PASS'  
                26 'REAR LEFT PASS'  
                27 'REAR PASS'  
                40 'MCYCLE POSN UNKNOWN'  
                41 'RIDER'  
                42 'PILLION OR SIDE-CAR'
```

```
60 'TRAM/TRAIN/HORSE'  
90 'PEDESTRIAN'  
91 'PEDAL CYCLIST'  
99 'OTHER'.
```

```
STRING REGNO (A6).  
STRING P (A1).  
COMPUTE P=SUBSTR(TREGNO,1,1).  
IF (P = ' ') REGNO=LTRIM(SUBSTR(TREGNO,2,6)).  
IF (P <> ' ') REGNO=LTRIM(SUBSTR(TREGNO,1,6)).  
EXECUTE.
```

```
IF ((REGNO='UKNOW') OR (REGNO='TRAM ') OR (REGNO='UNREG ') OR  
(REGNO='NKNOWN') OR (REGNO='UNKNOW') OR (REGNO='HITRUN') OR  
(REGNO='BICYCL') OR (REGNO='UNDIS ') OR (REGNO='TRAIN ') OR  
(REGNO='U/K ') OR (REGNO='UNDIS') OR (REGNO=' ') OR  
(REGNO='HORSE ') OR (REGNO='HORSED') OR (REGNO='NOTAPP') OR  
(REGNO='NOVEH ') OR (REGNO='PERMIT')) REGNO=' '.
```

```
IF ((CLAIMANT=50) OR (CLAIMANT=60) OR (CLAIMANT=90) OR  
(CLAIMANT=91) OR (CLAIMANT=92) OR (CLAIMANT=99)) REGNO=' ' .  
EXECUTE.
```

```
STRING TCREGNO (A6).
```

```
COMPUTE TCREGNO =REGNO.  
COMPUTE TUSERCOD=USERCODE.  
COMPUTE TNEWDATE=(TDATEACC+YRMODY(1960,1,1)) * 86400.  
FORMAT TNEWDATE (DATE9).  
EXECUTE.
```

```
SORT CASES BY ACCNO REGNO USERCODE SEX.  
EXECUTE.
```

```
SAVE OUTFILE 'TAC00Sort1.sav'  
/DROP= P ACCNOS YEARACC /COMPRESSED.  
EXECUTE.
```

```
FREQ VAR = SEX AGE USERCODE.
```

```
LIST VAR=ACCNO SEX AGE ACCDATE TDATEACC TNEWDATE  
USERCODE CLAIMANT REGNO TCREGNO  
/CASES FROM 1 TO 50.
```

Merge00_1.sps

```
GET FILE 'TAC00Sort1.sav' /RENAME=(LGA=TACLGA).
```

```
NUMERIC YRMAN (F4).  
COMPUTE YRMAN=YEARMAN.  
EXECUTE.
```

```
MATCH FILES FILE='VIC00Sort1.sav' /IN=INVIC  
/FILE=* /IN=INTAC  
/BY ACCNO REGNO USERCODE SEX  
/MAP.  
EXECUTE.
```

```
SAVE OUTFILE = 'Merge00_1.sav'.
```

DISPLAY SORTED VARIABLES.

CROSSTAB INTAC BY INVIC.
FREQ VAR USERCODE SEX INTAC INVIC.

LIST VAR = INTAC INVIC TACCNO RACCNO ACCNO TCREGNO RREGNO REGNO
TUSERCOD RUSERCOD SEX AGE TSEX TAGE RSEX RAGE
/CASE=FROM 1 TO 100.

Split00_1.sps

GET FILE 'Merge00_1.sav'.

COMPUTE AGEDIFF = ABS(RAGE - TAGE).
COMPUTE DATEDIFF = ABS(TDATEACC-RACCDATE).
EXECUTE.

FREQ VAR = AGEDIFF DATEDIFF.

RECODE AGEDIFF (4 THRU 999=1)(ELSE=0).
RECODE DATEDIFF (31 THRU 2999=1)(ELSE=0).

FREQ VAR=AGEDIFF DATEDIFF.

TEMP.
SELECT IF (AGEDIFF=1 OR DATEDIFF=1 OR (INTAC=0 AND INVIC=1)).
SAVE OUTFILE='VIC00Left1.sav'
/KEEP = PERS_ID TO RUSERCOD
/COMPRESSED.
EXECUTE.

TEMP.
SELECT IF (AGEDIFF=1 OR DATEDIFF=1 OR (INTAC=1 AND INVIC=0)).
SAVE OUTFILE = 'TAC00Left1.sav'
/KEEP = TSEX TO YRMAN SEX AGE ACCDATE ACCNO USERCODE REGNO YEARMAN
/COMPRESSED.
EXECUTE.

SELECT IF (AGEDIFF=0 AND DATEDIFF=0 AND INVIC=INTAC).
SAVE OUTFILE='TMatch00_1.sav'
/DROP=AGEDIFF DATEDIFF INVIC INTAC
/COMPRESS.
EXECUTE.

TAC00Sort2.sps

GET FILE 'TAC00Left1.sav'.

COMPUTE ACCNO = NUMBER(SUBSTR(TACCNO,6,6),F6.0).
COMPUTE USERCODE = TUSERCOD.
COMPUTE SEX = TSEX.
COMPUTE REGNO = TCREGNO.
COMPUTE AGE = TAGE.
COMPUTE ACCDATE = TDATEACC.
EXECUTE.

SORT CASES BY ACCNO USERCODE SEX REGNO AGE ACCDATE.

SAVE OUTFILE = 'TAC00Sort2.sav'.
EXECUTE.

VIC00Sort2.sps

```
GET FILE 'VIC00Left1.sav'.
```

```
COMPUTE ACCNO = RACCNO.  
COMPUTE USERCODE = RUSERCOD.  
COMPUTE SEX = RSEX.  
COMPUTE REGNO = RREGNO.  
COMPUTE AGE = RAGE.  
COMPUTE ACCDATE = RACCDATE.  
EXECUTE.
```

```
SORT CASES BY ACCNO USERCODE SEX REGNO AGE ACCDATE.
```

```
SAVE OUTFILE='VIC00Sort2.sav'.  
EXECUTE.
```

Merge00_2.sps

```
MATCH FILES FILE='VIC00Sort2.sav' /IN=INVIC  
/FILE='TAC00Sort2.sav' /IN=INTAC  
/BY ACCNO USERCODE SEX  
/MAP.  
EXECUTE.
```

```
SAVE OUTFILE='Merge00_2.sav'.  
EXECUTE.
```

```
CROSSTAB INVIC BY INTAC.
```

```
FREQ VAR USERCODE SEX INVIC INTAC.
```

```
LIST VAR=ACCNO TACCNO RACCNO REGNO TCREGNO RREGNO USERCODE TUSERCOD  
RUSERCOD  
SEX TSEX RSEX AGE TAGE RAGE  
/CASES FROM 1 TO 50.
```

Split00_2.sps

```
GET FILE 'Merge00_2.sav'.
```

```
COMPUTE AGEDIFF = ABS(RAGE - TAGE).  
EXECUTE.  
COMPUTE DATEDIFF = ABS(TDATEACC-RACCDATE).  
EXECUTE.
```

```
DO IF ((RREGNO NE ' ') AND (TCREGNO NE ' ') AND (RREGNO NE TCREGNO)).  
COMPUTE REGDIFF=1.  
ELSE.  
COMPUTE REGDIFF=0.  
END IF.  
EXECUTE.
```

```
RECODE AGEDIFF (4 THRU 999=1)(ELSE=0).  
RECODE DATEDIFF (31 THRU 2999=1)(ELSE=0).  
FREQ AGEDIFF DATEDIFF.
```

```
TEMP.
```

```
SELECT IF (AGEDIFF=0 AND DATEDIFF=0 AND REGDIFF=0 AND INTAC=INVIC).
SAVE OUTFILE 'TMatch00_2.sav'
  /DROP=AGEDIFF DATEDIFF REGDIFF INVIC INTAC
  /COMPRESSED.
EXECUTE.
```

```
TEMP.
SELECT IF (AGEDIFF=1 OR DATEDIFF=1 OR REGDIFF=1 OR (INTAC=0 AND INVIC=1)).
SAVE OUTFILE 'VIC00Left2.sav'
  /KEEP = PERS_ID TO RUSERCOD
  /COMPRESSED.
EXECUTE.
```

```
TEMP.
SELECT IF (AGEDIFF=1 OR DATEDIFF=1 OR REGDIFF=1 OR (INTAC=1 AND INVIC=0)).
SAVE OUTFILE 'TAC00Left2.sav'
  /KEEP = TSEX TO YRMAN SEX AGE ACCDATE ACCNO USERCODE REGNO YEARMAN
  /COMPRESSED.
EXECUTE.
```

TAC00Sort3.sps

```
GET FILE 'TAC00Left2.sav'.
```

```
COMPUTE ACCNO = NUMBER(SUBSTR(TACCNO,6,6),F6.0).
COMPUTE USERCODE = TUSERCOD.
COMPUTE SEX = TSEX.
COMPUTE REGNO = TCREGNO.
COMPUTE AGE = TAGE.
COMPUTE ACCDATE = TDATEACC.
EXECUTE.
```

```
SORT CASES BY ACCNO REGNO SEX AGE USERCODE ACCDATE.
```

```
SAVE OUTFILE = 'TAC00Sort3.sav'
  /COMPRESSED.
```

```
EXECUTE.
```

VIC00Sort3.sps

```
GET FILE 'VIC00Left2.sav'.
```

```
COMPUTE ACCNO = RACCNO.
COMPUTE USERCODE = RUSERCOD.
COMPUTE SEX = RSEX.
COMPUTE REGNO = RREGNO.
COMPUTE AGE = RAGE.
COMPUTE ACCDATE = RACCDATE.
EXECUTE.
```

```
freq rsex rage.
```

```
SORT CASES BY ACCNO REGNO SEX AGE USERCODE ACCDATE.
```

```
SAVE OUTFILE= 'VIC00Sort3.sav'.
EXECUTE.
```

Merge00_3.sps

```
MATCH FILES FILE= 'VIC00Sort3.sav' /IN=INVIC
      /FILE='TAC00Sort3.sav' /IN=INTAC
      /BY ACCNO REGNO SEX
      /MAP.
EXECUTE.
```

```
SAVE OUTFILE= 'Merge00_3.sav'.
EXECUTE.
CROSSTAB INVIC BY INTAC.
```

Split00_3.sps

```
GET FILE= 'Merge00_3.sav'.
```

```
COMPUTE AGEDIFF = ABS(RAGE - TAGE).
EXECUTE.
COMPUTE DATEDIFF = ABS(TDATEACC-RACCDATE).
EXECUTE.
```

```
FREQ VAR = AGEDIFF DATEDIFF.
```

```
RECODE AGEDIFF (4 THRU 999=1)(ELSE=0).
RECODE DATEDIFF (31 THRU 2999=1)(ELSE=0).
FREQ VAR = AGEDIFF DATEDIFF.
```

```
TEMPORARY.
SELECT IF (AGEDIFF=0 AND DATEDIFF=0 AND INTAC=INVIC).
XSAVE OUTFILE='TMatch00_3.sav'
      /DROP=AGEDIFF DATEDIFF INVIC INTAC
      /COMPRESS.
EXECUTE.
```

```
TEMPORARY.
SELECT IF (AGEDIFF=1 OR DATEDIFF=1 OR (INTAC=0 AND INVIC=1)).
XSAVE OUTFILE= 'VIC00Left3.sav'
      /KEEP = PERS_ID TO RUSERCOD
      /COMPRESSED.
EXECUTE.
```

```
TEMPORARY.
SELECT IF (AGEDIFF=1 OR DATEDIFF=1 OR (INTAC=1 AND INVIC=0)).
XSAVE OUTFILE = 'TAC00Left3.sav'
      /KEEP = TSEX TO YRMAN SEX AGE ACCDATE ACCNO USERCODE REGNO YEARMAN
      /COMPRESSED.
EXECUTE.
```

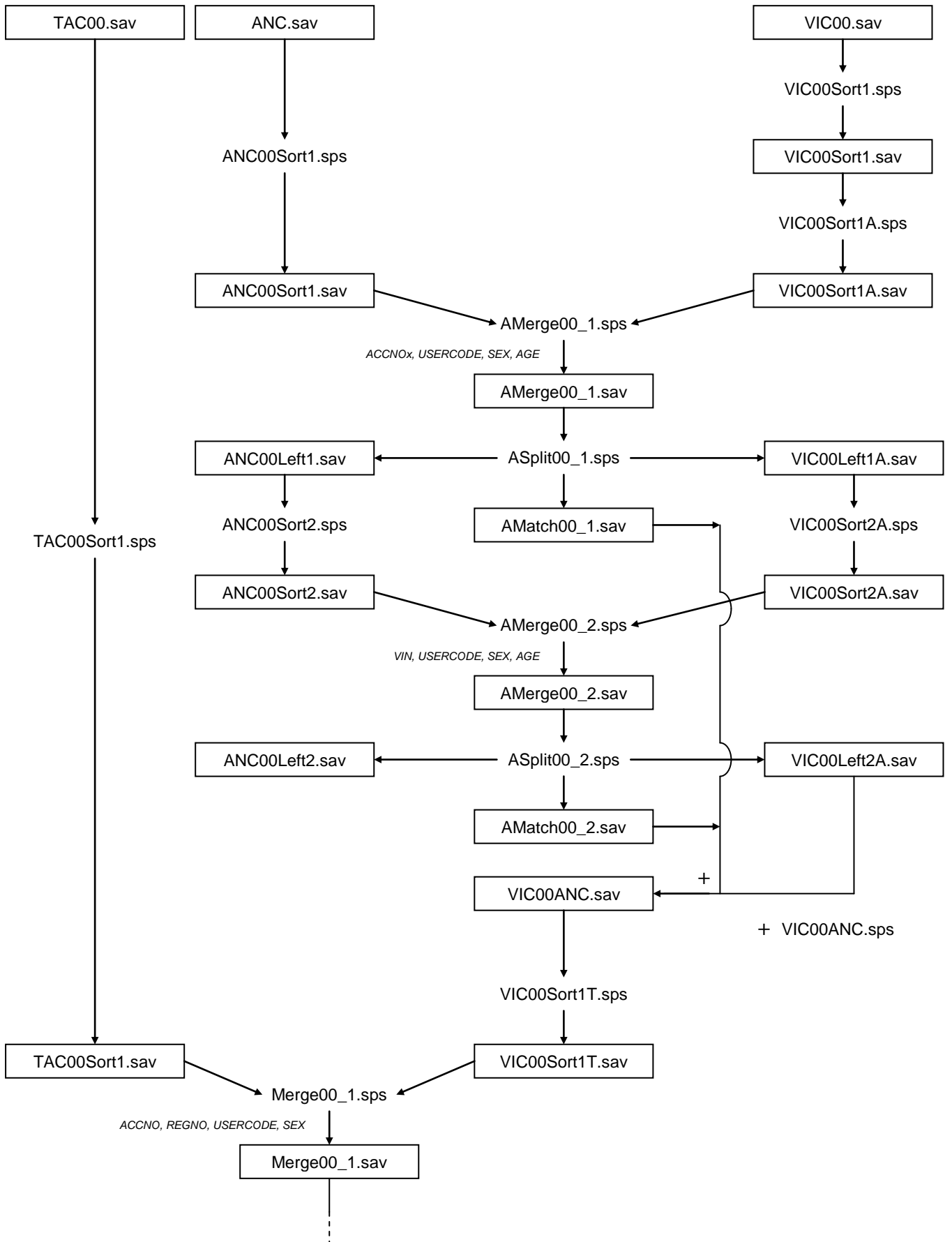
VIC00TAC00.sps

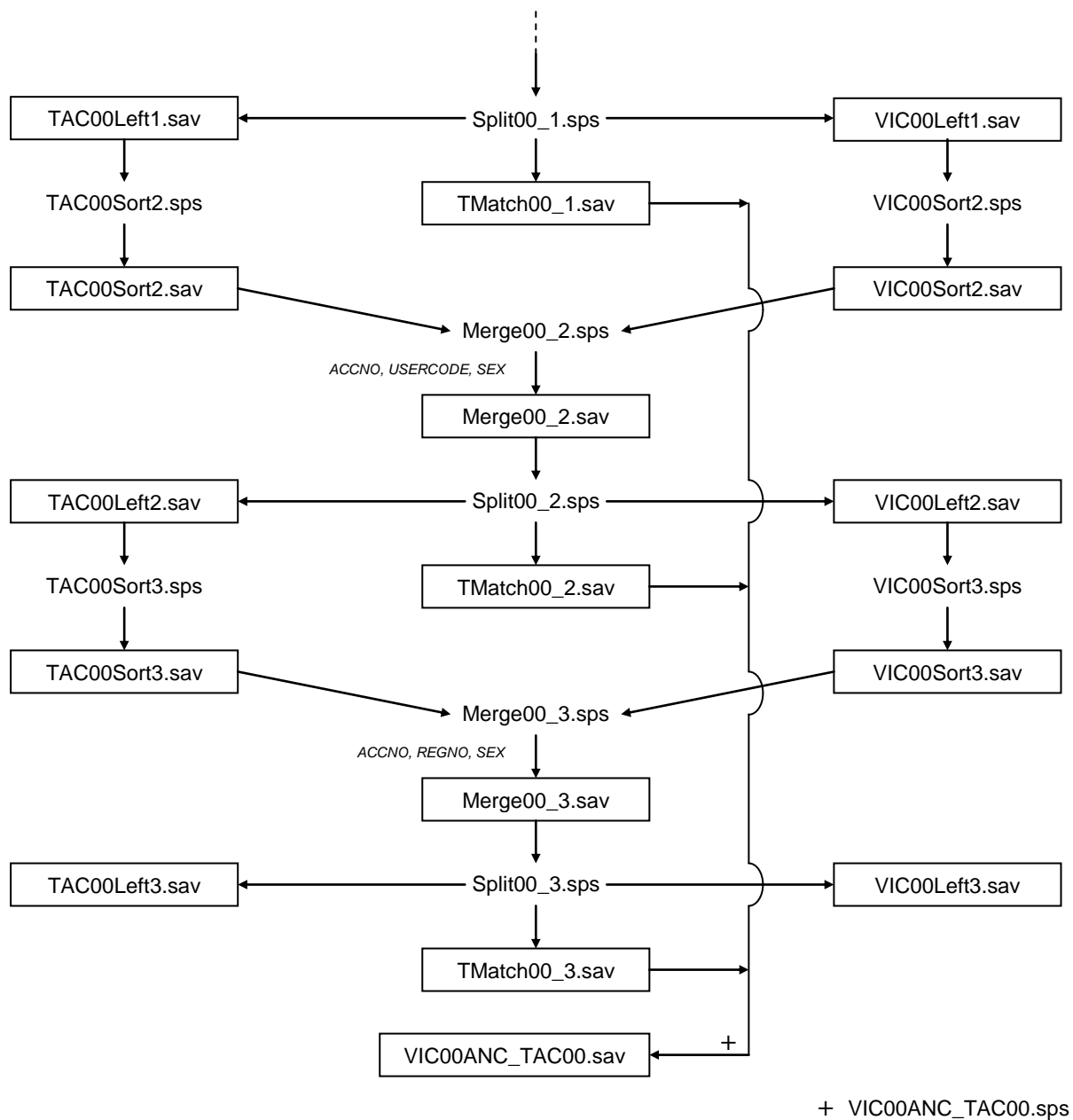
```
ADD FILES FILE='TMATCH00_1.sav'
      /FILE='TMATCH00_2.sav'
      /FILE='TMATCH00_3.sav'.
EXECUTE.
```

```
SAVE OUTFILE='VIC00TAC00.sav'.
```

Appendix 3

Flow Diagram for Linkage of the STAR, TAC and ANCIS Datasets





Appendix 4

Additional and Modified Syntax for Linkage of the STAR, TAC and ANCIS Datasets

VIC00Sort1A.sps

GET FILE 'VIC00Sort1.sav'.

STRING ACCNOx(A11).
COMPUTE ACCNOx = SUBSTR(ACC_NUM,1,11).
EXECUTE.

SORT CASES BY ACCNO USERCODE SEX.

SAVE OUTFILE 'VIC00Sort1A.sav'.

ANC00Sort1.sps

GET FILE 'ANC.sav'.

RENAME VARIABLES ACCNO = AACCNO.
STRING ACCNOx(A11).
EXECUTE.

IF (AACCNO <> 999.00) ACCNOx = STRING(AACCNO,F11.2).
EXECUTE.

RENAME VARIABLES SEX = ASEX.
STRING SEX(A1).
IF (ASEX = 1) SEX = 'M'.
IF (ASEX = 2) SEX = 'F'.
EXECUTE.

RENAME VARIABLES VIN = AVIN.

NUMERIC AAGE(F1.0).
COMPUTE AAGE = AGE.
EXECUTE.

NUMERIC USERCODE(F2.0).
IF (POSITION=1) USERCODE=21.
IF (POSITION=2) USERCODE=23.
IF (POSITION=3) USERCODE=22.
IF (POSITION=4) USERCODE=26.
IF (POSITION=5) USERCODE=25.
IF (POSITION=6) USERCODE=24.
IF (POSITION=7) USERCODE=27.
IF (POSITION=8) USERCODE=27.
IF (POSITION=9) USERCODE=27.
EXECUTE.

NUMERIC AUSRCODE(F2.0).
COMPUTE AUSRCODE=USERCODE.
EXECUTE.

SORT CASES BY ACCNOx USERCODE SEX AGE.

SAVE OUTFILE 'ANC00Sort1.sav'.

AMerge00_1.sps

GET FILE 'ANC00Sort1.sav'.

MATCH FILES FILE = 'VIC00Sort1A.sav' /IN=INVIC
/FILE = * /IN=INANC
/BY ACCNOx USERCODE SEX AGE
/MAP.

SAVE OUTFILE = 'AMerge00_1.sav'.

ASplit00_1.sps

GET FILE='AMerge00_1.sav'.

NUMERIC DATEDIFF(F1.0).
COMPUTE DATEDIFF=ABS(RACCDATE-AACCDATE).
EXECUTE.

RECODE DATEDIFF(2 THRU 9999=1)(ELSE=0).
EXECUTE.

TEMP.
SELECT IF (DATEDIFF=1 OR (INANC=1 AND INVIC=0)).
SAVE OUTFILE='ANC00Left1.sav'
/KEEP caseno TO AUSRCODE.

TEMP.
SELECT IF (DATEDIFF=1 OR (INANC=0 AND INVIC=1)).
SAVE OUTFILE='VIC00Left1A.sav'
/KEEP pers_id TO RUSERCOD.

SELECT IF (DATEDIFF=0 AND INANC=1 AND INVIC=1).
SAVE OUTFILE='AMatch00_1.sav'
/DROP DATEDIFF INANC INVIC.

VIC00Sort2A.sps

GET FILE 'VIC00Left1A.sav'.

COMPUTE USERCODE = RUSERCOD.
COMPUTE AGE = RAGE.
COMPUTE SEX = RSEX.
EXECUTE.

SORT CASES BY VIN USERCODE SEX AGE.

SAVE OUTFILE 'VIC00Sort2A.sav'.

ANC00Sort2.sps

GET FILE 'ANC00Left1.sav'.

STRING VIN(A20).
IF (AVIN <> '999') VIN = AVIN.
EXECUTE.

IF VIN = '' VIN = 'MISSING'.

EXECUTE.

NUMERIC USERCODE(F2.0).
COMPUTE USERCODE=AUSRCODE.
EXECUTE.

STRING SEX(A1).
IF (ASEX = 1) SEX = 'M'.
IF (ASEX = 2) SEX = 'F'.
EXECUTE.

NUMERIC AGE(F3.0).
COMPUTE AGE = AAGE.
EXECUTE.

SORT CASES BY VIN USERCODE SEX AGE.

SAVE OUTFILE 'ANC00Sort2.sav'.

AMerge00_2.sps

GET FILE 'ANC00Sort2.sav'.

MATCH FILES FILE = 'VIC00Sort2A.sav' /IN=INVIC
/FILE = * /IN=INANC
/BY VIN USERCODE SEX AGE
/MAP.

SAVE OUTFILE = 'AMerge00_2.sav'.

ASplit00_2.sps

GET FILE='AMerge00_2.sav'.

NUMERIC DATEDIFF(F1.0).
COMPUTE DATEDIFF=ABS(RACCDATE-AACCDATE).
EXECUTE.

RECODE DATEDIFF(2 THRU 9999=1)(ELSE=0).
EXECUTE.

TEMP.
SELECT IF (DATEDIFF=1 OR (INANC=1 AND INVIC=0)).
SAVE OUTFILE='ANC00Left2.sav'
/KEEP caseno TO AUSRCODE.

TEMP.
SELECT IF (DATEDIFF=1 OR (INANC=0 AND INVIC=1)).
SAVE OUTFILE='VIC00Left2A.sav'
/KEEP pers_id TO RUSERCOD.

SELECT IF (DATEDIFF=0 AND INANC=1 AND INVIC=1).
SAVE OUTFILE='AMatch00_2.sav'
/DROP DATEDIFF INANC INVIC.

VIC00ANC.sps

```
ADD FILES FILE='AMATCH00_1.sav'  
          /FILE='AMATCH00_2.sav'  
          /FILE='VIC00LEFT2A.sav'.  
EXECUTE.
```

```
SAVE OUTFILE='VIC00ANC.sav'.
```

VIC00Sort1T.sps

```
GET FILE 'VIC00ANC.sav'.
```

```
SORT CASES BY ACCNO REGNO USERCODE SEX AGE ACCDATE.
```

```
SAVE OUTFILE 'VIC00Sort1T.sav'.
```

Merge00_1.sps

```
GET FILE 'TAC00Sort1.sav' /RENAME=(LGA=TACLGA).
```

```
NUMERIC YRMAN (F4).  
COMPUTE YRMAN=YEARMAN.  
EXECUTE.
```

```
MATCH FILES FILE='VIC00Sort1T.sav' /IN=INVIC  
            /FILE=* /IN=INTAC  
            /BY ACCNO REGNO USERCODE SEX  
            /MAP.  
EXECUTE.
```

```
SAVE OUTFILE = 'Merge00_1.sav'.
```

```
DISPLAY SORTED VARIABLES.
```

```
CROSSTAB INTAC BY INVIC.  
FREQ VAR USERCODE SEX INTAC INVIC.
```

```
LIST VAR = INTAC INVIC TACCNO RACCNO ACCNO TCREGNO RREGNO REGNO  
          TUSERCOD RUSERCOD SEX AGE TSEX TAGE RSEX RAGE  
          /CASE=FROM 1 TO 100.
```

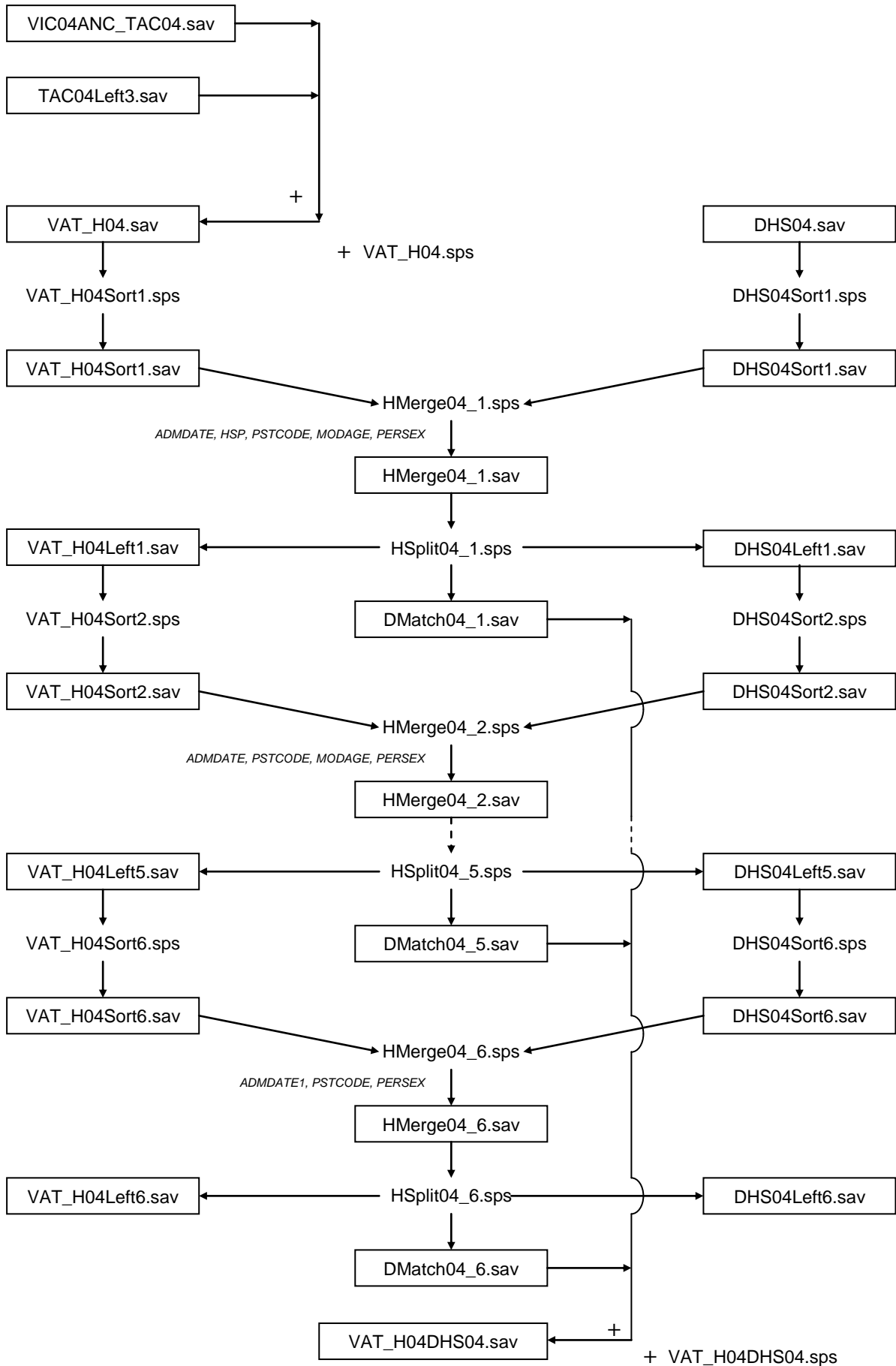
VIC00ANC_TAC00.sps

```
ADD FILES FILE='TMATCH00_1.sav'  
          /FILE='TMATCH00_2.sav'  
          /FILE='TMATCH00_3.sav'.  
EXECUTE.
```

```
SAVE OUTFILE='VIC00ANC_TAC00.sav'.
```


Appendix 5

Flow Diagram for Linkage of the STAR, TAC, ANCIS and VAED Datasets



Appendix 6

Syntax for Linkage of the STAR, TAC, ANCIS and VAED Datasets

VAT_H04.sps

GET FILE 'TAC04LEFT3.sav'.

SELECT IF lenstay > 0.
EXECUTE.

ADD FILES FILE = 'VIC04ANC_TAC04.sav'
/FILE = *.
EXECUTE.

SAVE OUTFILE = 'VAT_H04.sav'.

VAT_H04Sort1.sps

GET FILE 'VAT_H04.sav'.

NUMERIC ADMDATE(F6.0) ADMDATE1(F6.0).

COMPUTE ADMDATE = tdateacc.
COMPUTE ADMDATE1 = tdateacc + 1.
EXECUTE.

NUMERIC MODAGE(F2.0).

COMPUTE MODAGE = tage.
EXECUTE.

NUMERIC PSTCODE(F4.0).

COMPUTE PSTCODE = post.
EXECUTE.

STRING PERSEX(A1).
EXECUTE.

COMPUTE PERSEX = tsex.
EXECUTE.

STRING HSP(A25).
EXECUTE.

COMPUTE HSP = 'VAT_H Missing'.
EXECUTE.

IF HospitalName = '2840' HSP = 'VAT_H Missing'.
IF HospitalName = '2CC' HSP = 'VAT_H Missing'.
IF HospitalName = '3130' HSP = 'VAT_H Missing'.
IF HospitalName = '3152' HSP = 'VAT_H Missing'.
IF HospitalName = '3152' HSP = 'VAT_H Missing'.
IF HospitalName = 'A;FRED' HSP = 'ALFRED'.
IF HospitalName = 'AL' HSP = 'VAT_H Missing'.
IF HospitalName = 'ALBERT HOSPITAL' HSP = 'Private'.
IF HospitalName = 'ALEXANDRA' HSP = 'ALEXANDRA'.
IF HospitalName = 'ALEXANDRA' HSP = 'ALEXANDRA'.
IF HospitalName = 'ALEXANDRA DIST' HSP = 'ALEXANDRA'.
IF HospitalName = 'ALEXANDRA DISTR' HSP = 'ALEXANDRA'.
IF HospitalName = 'ALFRED-TRAUMA' HSP = 'ALFRED'.
IF HospitalName = 'ALFRED' HSP = 'ALFRED'.

IF HospitalName = 'ALFRED HOS' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED HOSP' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED HOSPITAL' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED HOSPITAL' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED MED' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED MELB' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED MELBOURN' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED TRAUMA' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED TRAUMA U' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED.HOSPITAL' HSP = 'ALFRED'.
 IF HospitalName = 'ALFRED/ST VINCE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'ALFREDN' HSP = 'ALFRED'.
 IF HospitalName = 'ANGLESEA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'ANGLIS' HSP = 'ANGLISS'.
 IF HospitalName = 'ANGLIS FERNTREE' HSP = 'ANGLISS'.
 IF HospitalName = 'ANGLIS FTG' HSP = 'ANGLISS'.
 IF HospitalName = 'ANGLIS HOSPITAL' HSP = 'ANGLISS'.
 IF HospitalName = 'ANGLISS' HSP = 'ANGLISS'.
 IF HospitalName = 'ANGLISS HOSPITA' HSP = 'ANGLISS'.
 IF HospitalName = 'APOLLO BAY' HSP = 'APOLLO BAY'.
 IF HospitalName = 'ARARAT' HSP = 'ARARAT'.
 IF HospitalName = 'ARARAT BASE' HSP = 'ARARAT'.
 IF HospitalName = 'ARARAT HOSP' HSP = 'ARARAT'.
 IF HospitalName = 'ARARAT HOSPITAL' HSP = 'ARARAT'.
 IF HospitalName = 'ARMC' HSP = 'VAT_H Missing'.
 IF HospitalName = 'AU' HSP = 'AUSTIN'.
 IF HospitalName = 'AUSTEN' HSP = 'AUSTIN'.
 IF HospitalName = 'AUSTIN' HSP = 'AUSTIN'.
 IF HospitalName = 'AUSTIN HEALTH' HSP = 'AUSTIN'.
 IF HospitalName = 'AUSTIN HOSP' HSP = 'AUSTIN'.
 IF HospitalName = 'AUSTIN HOSPITAL' HSP = 'AUSTIN'.
 IF HospitalName = 'AUSTINE' HSP = 'AUSTIN'.
 IF HospitalName = 'AX' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BACCHUS MARSH' HSP = 'BACCHUS MARSH'.
 IF HospitalName = 'BAIRNSDALE' HSP = 'BAIRNSDALE'.
 IF HospitalName = 'BAIRNSDALE HOS' HSP = 'BAIRNSDALE'.
 IF HospitalName = 'BAIRNSDALE HOSP' HSP = 'BAIRNSDALE'.
 IF HospitalName = 'BALLARAT' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT .BASE' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT BASE' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT BASE H' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT BASE M' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT H.S' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT HEALTH' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT HOSP' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLARAT.BASE' HSP = 'BALLARAT HEALTH'.
 IF HospitalName = 'BALLART ST JOHN' HSP = 'Private'.
 IF HospitalName = 'BARHOM' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BASE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BASE HOSP' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BASE HOSPITAL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BASS COAST REGI' HSP = 'BASS'.
 IF HospitalName = 'BAYS HOSPITAL' HSP = 'Private'.
 IF HospitalName = 'BBH' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BENALLA' HSP = 'BENALLA'.
 IF HospitalName = 'BENALLA BASE HO' HSP = 'BENALLA'.
 IF HospitalName = 'BENDGIO HOSP' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO A & E' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO A+E' HSP = 'BENDIGO'.

IF HospitalName = 'BENDIGO BASE' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO BASE HO' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO EMERGEN' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO HEALTH' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO HOSP' HSP = 'BENDIGO'.
 IF HospitalName = 'BENDIGO HOSPITA' HSP = 'BENDIGO'.
 IF HospitalName = 'BETTER HEALTH' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BGO HEALTH CARE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BHS' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BIRCHIP' HSP = 'BIRCHIP'.
 IF HospitalName = 'BLACKBURN HOSP' HSP = 'Private'.
 IF HospitalName = 'BOORT' HSP = 'BOORT'.
 IF HospitalName = 'BORDERTOWN' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BOX' HSP = 'BOX HILL'.
 IF HospitalName = 'BOX HILL' HSP = 'BOX HILL'.
 IF HospitalName = 'BOX HILL HOSP' HSP = 'BOX HILL'.
 IF HospitalName = 'BOX HILL HOSPIT' HSP = 'BOX HILL'.
 IF HospitalName = 'BOX HILL SOUTH' HSP = 'BOX HILL'.
 IF HospitalName = 'BOX HOSPITAL' HSP = 'BOX HILL'.
 IF HospitalName = 'BOXHILL' HSP = 'BOX HILL'.
 IF HospitalName = 'BRHS' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BRIGHT' HSP = 'BRIGHT'.
 IF HospitalName = 'BRIGHT HOSPITAL' HSP = 'BRIGHT'.
 IF HospitalName = 'BROOKE ST MEDIC' HSP = 'VAT_H Missing'.
 IF HospitalName = 'BURWOOD HEALTH' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CABRINI' HSP = 'Private'.
 IF HospitalName = 'CAMBERDOWN' HSP = 'SOUTH WEST'.
 IF HospitalName = 'CAMPERDOWN' HSP = 'SOUTH WEST'.
 IF HospitalName = 'CAMPERDOWN BASE' HSP = 'SOUTH WEST'.
 IF HospitalName = 'CANBERRA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CASTERTON' HSP = 'CASTERTON'.
 IF HospitalName = 'CASTERTON MEMOR' HSP = 'CASTERTON'.
 IF HospitalName = 'CASTLEMAINE' HSP = 'CASTLEMAINE'.
 IF HospitalName = 'CENTRAL GIPPSLA' HSP = 'SALE'.
 IF HospitalName = 'CENTRAL GIPSLAN' HSP = 'SALE'.
 IF HospitalName = 'CENTRAL HEALTH' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CENTRAL WELLING' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CHARLTON' HSP = 'CHARLTON'.
 IF HospitalName = 'CHARLTON DISTRI' HSP = 'CHARLTON'.
 IF HospitalName = 'CHILDRENS' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'CHILDRENS HOSP' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'CHILDRENS HOSPI' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'CLUNES MED' HSP = 'VAT_H Missing'.
 IF HospitalName = 'COBRAM' HSP = 'COBRAM'.
 IF HospitalName = 'COBRAM HOSP' HSP = 'COBRAM'.
 IF HospitalName = 'COHUNA' HSP = 'COHUNA'.
 IF HospitalName = 'COHUNA & DISTRI' HSP = 'COHUNA'.
 IF HospitalName = 'COHUNA DISTRICT' HSP = 'COHUNA'.
 IF HospitalName = 'COLAC' HSP = 'COLAC'.
 IF HospitalName = 'COLAC AREA' HSP = 'COLAC'.
 IF HospitalName = 'COLAC AREA HEAL' HSP = 'COLAC'.
 IF HospitalName = 'COLAC HOSP' HSP = 'COLAC'.
 IF HospitalName = 'COLAC HOSPITAL' HSP = 'COLAC'.
 IF HospitalName = 'COLERAINE DISTR' HSP = 'COLERAINE'.
 IF HospitalName = 'CORONA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CORONERS' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CORONOR' HSP = 'VAT_H Missing'.
 IF HospitalName = 'COROWA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CORRYONG' HSP = 'CORRYONG'.
 IF HospitalName = 'COURT' HSP = 'VAT_H Missing'.

IF HospitalName = 'COWES' HSP = 'VAT_H Missing'.
 IF HospitalName = 'COWOWA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'CWHS' HSP = 'VAT_H Missing'.
 IF HospitalName = 'DADENONG' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG & DIS' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG (PUBL' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG DIS' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG DIST' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG DISTR' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG HOSP' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG HOSPI' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG PRIVA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'DANDENONG PUBLI' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONG VALLE' HSP = 'Private'.
 IF HospitalName = 'DANDENONG.SOUTH' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDENONTG' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDEONG' HSP = 'DANDENONG'.
 IF HospitalName = 'DANDEY VALLEY' HSP = 'Private'.
 IF HospitalName = 'DANDY VALLEY' HSP = 'Private'.
 IF HospitalName = 'DAYLESFORD' HSP = 'DAYLESFORD'.
 IF HospitalName = 'DELEGATE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'DG' HSP = 'VAT_H Missing'.
 IF HospitalName = 'DIMBOOL DIST' HSP = 'DIMBOOLA'.
 IF HospitalName = 'DONALD' HSP = 'DONALD'.
 IF HospitalName = 'DONALD THEN BAL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'DONALD THEN BEN' HSP = 'DONALD'.
 IF HospitalName = 'DONALD THEN MEL' HSP = 'DONALD'.
 IF HospitalName = 'EAST GRAMPIANS' HSP = 'ARARAT'.
 IF HospitalName = 'ECHUCA' HSP = 'ECHUCA'.
 IF HospitalName = 'ECHUCA DIST' HSP = 'ECHUCA'.
 IF HospitalName = 'ECHUCA DISTRICT' HSP = 'ECHUCA'.
 IF HospitalName = 'ECHUCA HOSPITAL' HSP = 'ECHUCA'.
 IF HospitalName = 'ECHUCA REGIONAL' HSP = 'ECHUCA'.
 IF HospitalName = 'EDEHOPE' HSP = 'EDENHOPE'.
 IF HospitalName = 'EDENHOPE - DIST' HSP = 'EDENHOPE'.
 IF HospitalName = 'EDENHOPE' HSP = 'EDENHOPE'.
 IF HospitalName = 'EGB HOSP' HSP = 'VAT_H Missing'.
 IF HospitalName = 'EGHS' HSP = 'ARARAT'.
 IF HospitalName = 'EGHS MOTEL' HSP = 'ARARAT'.
 IF HospitalName = 'EPPING' HSP = 'EPPING'.
 IF HospitalName = 'EPPING HOSPITAL' HSP = 'EPPING'.
 IF HospitalName = 'EPWORTH' HSP = 'Private'.
 IF HospitalName = 'EPWORTH HOSPITA' HSP = 'Private'.
 IF HospitalName = 'FAMILY DOCTOR' HSP = 'VAT_H Missing'.
 IF HospitalName = 'FDA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'FERNTREE GULLY' HSP = 'ANGLISS'.
 IF HospitalName = 'FOOTSCRAY' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'FOOTSCRAY GEN H' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'FOOTSCRAY GENER' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'FOOTSCRAY HOSP' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'FOOTSCRAY HOSPI' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'FOSTER' HSP = 'FOSTER'.
 IF HospitalName = 'FRANKSTON"Y' HSP = 'FRANKSTON'.
 IF HospitalName = 'FRANKSTON' HSP = 'FRANKSTON'.
 IF HospitalName = 'FRANKSTON HOSP' HSP = 'FRANKSTON'.
 IF HospitalName = 'FRANKSTON HOSPI' HSP = 'FRANKSTON'.
 IF HospitalName = 'FRANKSTON PENIN' HSP = 'FRANKSTON'.
 IF HospitalName = 'FRANKSTON PUBLI' HSP = 'FRANKSTON'.
 IF HospitalName = 'FREEMASON' HSP = 'Private'.

IF HospitalName = 'FREEMASONS' HSP = 'Private'.
 IF HospitalName = 'FREEMASONS MELB' HSP = 'Private'.
 IF HospitalName = 'G V BASE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G V HEALTH' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G VALLEY BASE H' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.B HEALTH' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.V BASE HOSPIT' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.V HEALTH' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.V. BASE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.V. HEALTH' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.V. HOSPITAL' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G.V.B. HOSPITAL' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'G/V BASE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GBH' HSP = 'SALE'.
 IF HospitalName = 'GBH SALE' HSP = 'SALE'.
 IF HospitalName = 'GEEL9ONG' HSP = 'GEELONG'.
 IF HospitalName = 'GEELNG HOSP' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG A&E' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG HOS' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG HOSP' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG HOSPITA' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG PUBLIC' HSP = 'GEELONG'.
 IF HospitalName = 'GEELONG WEST' HSP = 'GEELONG'.
 IF HospitalName = 'GEEONG' HSP = 'GEELONG'.
 IF HospitalName = 'GIPPS BASE' HSP = 'SALE'.
 IF HospitalName = 'GIPPSLAND' HSP = 'SALE'.
 IF HospitalName = 'GIPPSLAND BASE' HSP = 'SALE'.
 IF HospitalName = 'GISBORNE MEDICA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'GL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'GOULBOURN VALLE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GOULBURN VALLE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GOULBURN VALLEY' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GOULBURN VEALLE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GOULBYRN VALLEY' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GOULDBURN VALLE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GRAMPIANS' HSP = 'ARARAT'.
 IF HospitalName = 'GRAMPIANS HEALT' HSP = 'ARARAT'.
 IF HospitalName = 'GTV HEALTH SHEP' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV BASE' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV BASE HOSP' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV BASE HOSPITA' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV BAUL HOSP' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV HEALTH' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV HOSP' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV HOSPITAL' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GV MEDICAL CENT' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'GVBH' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'H V HEALTH' HSP = 'VAT_H Missing'.
 IF HospitalName = 'HAMILTON' HSP = 'HAMILTON'.
 IF HospitalName = 'HAMILTON BASE' HSP = 'HAMILTON'.
 IF HospitalName = 'HEALESVILLE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'HEATHCOTE' HSP = 'HEATHCOTE'.
 IF HospitalName = 'HEYFIELD' HSP = 'VAT_H Missing'.
 IF HospitalName = 'HOSPITAL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'ILFRED' HSP = 'ALFRED'.
 IF HospitalName = 'INGLEWOOD' HSP = 'INGLEWOOD'.
 IF HospitalName = 'JOHN FALKINER' HSP = 'Private'.
 IF HospitalName = 'JOHN FAULKNER' HSP = 'Private'.
 IF HospitalName = 'JOHN FAWKNER' HSP = 'Private'.

IF HospitalName = 'KANIVA' HSP = 'WEST WIMMERA'.
 IF HospitalName = 'KERANG' HSP = 'KERANG'.
 IF HospitalName = 'KERANG DISTRICT' HSP = 'KERANG'.
 IF HospitalName = 'KILMORE' HSP = 'KILMORE'.
 IF HospitalName = 'KILMORE DIST HO' HSP = 'KILMORE'.
 IF HospitalName = 'KILMORE DISTRIC' HSP = 'KILMORE'.
 IF HospitalName = 'KILMORE HOS' HSP = 'KILMORE'.
 IF HospitalName = 'KILMORE HOSP' HSP = 'KILMORE'.
 IF HospitalName = 'KILMORE HOSPITA' HSP = 'KILMORE'.
 IF HospitalName = 'KL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'KNOX' HSP = 'Private'.
 IF HospitalName = 'KNOX HOSPITAL' HSP = 'Private'.
 IF HospitalName = 'KNOX Private' HSP = 'Private'.
 IF HospitalName = 'KORUMBIURRA' HSP = 'KORUMBURRA'.
 IF HospitalName = 'KORUMBURRA' HSP = 'KORUMBURRA'.
 IF HospitalName = 'KYABRAM' HSP = 'KYABRAM'.
 IF HospitalName = 'KYABRAM HOSPITA' HSP = 'KYABRAM'.
 IF HospitalName = 'KYNETON' HSP = 'KYNETON'.
 IF HospitalName = 'KYNETON HOSP' HSP = 'KYNETON'.
 IF HospitalName = 'KYNETON HOSPITA' HSP = 'KYNETON'.
 IF HospitalName = 'L R H' HSP = 'TRARALGON'.
 IF HospitalName = 'LA TROBE REGION' HSP = 'TRARALGON'.
 IF HospitalName = 'LAKES COMMUNITY' HSP = 'VAT_H Missing'.
 IF HospitalName = 'LAKES ENTRANCE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'LARA MEDICAL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'LATORBE REGIONA' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROB EVALLEY' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROBE' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROBE HOSPITA' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROBE REG' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROBE REG 1' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROBE REGIONA' HSP = 'TRARALGON'.
 IF HospitalName = 'LATROBE VALLEY' HSP = 'TRARALGON'.
 IF HospitalName = 'LEONGATHA' HSP = 'LEONGATHA'.
 IF HospitalName = 'LEONGATHA BASE' HSP = 'LEONGATHA'.
 IF HospitalName = 'LEONGATHA MEMOR' HSP = 'LEONGATHA'.
 IF HospitalName = 'LEONTGATHA' HSP = 'LEONGATHA'.
 IF HospitalName = 'LORNE' HSP = 'LORNE'.
 IF HospitalName = 'LORNE HOSP' HSP = 'LORNE'.
 IF HospitalName = 'LRH' HSP = 'TRARALGON'.
 IF HospitalName = 'LRH TRARALGON' HSP = 'TRARALGON'.
 IF HospitalName = 'LRM TRARALGON' HSP = 'TRARALGON'.
 IF HospitalName = 'LRMH' HSP = 'TRARALGON'.
 IF HospitalName = 'MALLEE TRACK' HSP = 'MALLEE TRACK'.
 IF HospitalName = 'MALLEY TRACK HE' HSP = 'MALLEE TRACK'.
 IF HospitalName = 'MANSFIELD' HSP = 'MANSFIELD'.
 IF HospitalName = 'MANSFIELD DIS' HSP = 'MANSFIELD'.
 IF HospitalName = 'MANSFIELD DISTR' HSP = 'MANSFIELD'.
 IF HospitalName = 'MANSFIELD HOSPI' HSP = 'MANSFIELD'.
 IF HospitalName = 'MANSIFELD' HSP = 'MANSFIELD'.
 IF HospitalName = 'MANSIFIELD' HSP = 'MANSFIELD'.
 IF HospitalName = 'MARIBYRNONG' HSP = 'VAT_H Missing'.
 IF HospitalName = 'MARONDAH' HSP = 'MAROONDAH'.
 IF HospitalName = 'MAROONAH' HSP = 'MAROONDAH'.
 IF HospitalName = 'MAROONDAH' HSP = 'MAROONDAH'.
 IF HospitalName = 'MAROONDAH HOSP' HSP = 'MAROONDAH'.
 IF HospitalName = 'MAROONDAH HOSPI' HSP = 'MAROONDAH'.
 IF HospitalName = 'MAROONDAY' HSP = 'MAROONDAH'.
 IF HospitalName = 'MARYBOROUGH' HSP = 'MARYBOROUGH'.
 IF HospitalName = 'MARYBOROUGH/DIS' HSP = 'MARYBOROUGH'.

IF HospitalName = 'MASADA' HSP = 'Private'.
 IF HospitalName = 'MEDICAL CLINIC' HSP = 'VAT_H Missing'.
 IF HospitalName = 'MELB-AIRAMB' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'MELB' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'MELB HOSP' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'MELB HOSPITAL' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'MELB ROYAL' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'MELB WOMENS' HSP = 'ROYAL WOMENS'.
 IF HospitalName = 'MELBOURNE' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'MERCY' HSP = 'MERCY'.
 IF HospitalName = 'MERCY HOSPITAL' HSP = 'MERCY'.
 IF HospitalName = 'MERCY HOSPITAL' HSP = 'MERCY'.
 IF HospitalName = 'MIDLURA BASE' HSP = 'MILDURA'.
 IF HospitalName = 'MILDURA' HSP = 'MILDURA'.
 IF HospitalName = 'MILDURA BASE' HSP = 'MILDURA'.
 IF HospitalName = 'MILDURA BASE AL' HSP = 'MILDURA'.
 IF HospitalName = 'MILDURA BASE HO' HSP = 'MILDURA'.
 IF HospitalName = 'MILDURA BASEY' HSP = 'MILDURA'.
 IF HospitalName = 'MILDURA HOSP' HSP = 'MILDURA'.
 IF HospitalName = 'MITCHAM Private' HSP = 'Private'.
 IF HospitalName = 'MO' HSP = 'MONASH'.
 IF HospitalName = 'MONASH' HSP = 'MONASH'.
 IF HospitalName = 'MONASH CLAYTON' HSP = 'MONASH'.
 IF HospitalName = 'MONASH HOSP' HSP = 'MONASH'.
 IF HospitalName = 'MONASH HOSP' HSP = 'MONASH'.
 IF HospitalName = 'MONASH HOSPITAL' HSP = 'MONASH'.
 IF HospitalName = 'MONASH M.C' HSP = 'MONASH'.
 IF HospitalName = 'MONASH M.L' HSP = 'MONASH'.
 IF HospitalName = 'MONASH M/C' HSP = 'MONASH'.
 IF HospitalName = 'MONASH MC' HSP = 'MONASH'.
 IF HospitalName = 'MONASH MED' HSP = 'MONASH'.
 IF HospitalName = 'MONASH MED CENT' HSP = 'MONASH'.
 IF HospitalName = 'MONASH MED CTR' HSP = 'MONASH'.
 IF HospitalName = 'MONASH MEDICAL' HSP = 'MONASH'.
 IF HospitalName = 'MONASH MEDICAL.' HSP = 'MONASH'.
 IF HospitalName = 'MONSAH' HSP = 'MONASH'.
 IF HospitalName = 'MONSH' HSP = 'MONASH'.
 IF HospitalName = 'MOOROONDAH' HSP = 'MAROONDAH'.
 IF HospitalName = 'MORNINGTON PENI' HSP = 'VAT_H Missing'.
 IF HospitalName = 'MOUNT ALEXANDER' HSP = 'CASTLEMAINE'.
 IF HospitalName = 'MOUNT BEAUTY' HSP = 'MOUNT BEAUTY'.
 IF HospitalName = 'MOUNT GAMBIER' HSP = 'VAT_H Missing'.
 IF HospitalName = 'MT ALEXANDER' HSP = 'CASTLEMAINE'.
 IF HospitalName = 'MT ALEXANDER HO' HSP = 'CASTLEMAINE'.
 IF HospitalName = 'MT ALVERNIA' HSP = 'VAT_H Missing'.
 IF HospitalName = 'MT BEAUTY' HSP = 'MOUNT BEAUTY'.
 IF HospitalName = 'MT BEAUTY HOSPI' HSP = 'MOUNT BEAUTY'.
 IF HospitalName = 'MT GAMBIER' HSP = 'VAT_H Missing'.
 IF HospitalName = 'MYRTLEFORD' HSP = 'MYRTLEFORD'.
 IF HospitalName = 'N/K' HSP = 'VAT_H Missing'.
 IF HospitalName = 'NARACOORTE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'NHILL' HSP = 'WEST WIMMERA'.
 IF HospitalName = 'NHILL & DISTRIC' HSP = 'WEST WIMMERA'.
 IF HospitalName = 'NHILL HOSP' HSP = 'WEST WIMMERA'.
 IF HospitalName = 'NHILL HOSPITAL' HSP = 'WEST WIMMERA'.
 IF HospitalName = 'NNK' HSP = 'VAT_H Missing'.
 IF HospitalName = 'NORCHEN' HSP = 'EPPING'.
 IF HospitalName = 'NORTH EAST HEAL' HSP = 'WANGARATTA'.
 IF HospitalName = 'NORTH EAST MED' HSP = 'WANGARATTA'.
 IF HospitalName = 'NORTH EAST MEDI' HSP = 'WANGARATTA'.

IF HospitalName = 'NORTH EASTERN H' HSP = 'WANGARATTA'.
 IF HospitalName = 'NORTHEAST HEALT' HSP = 'WANGARATTA'.
 IF HospitalName = 'NORTHERN' HSP = 'EPPING'.
 IF HospitalName = 'NORTHERN HOS' HSP = 'EPPING'.
 IF HospitalName = 'NORTHERN HOSP' HSP = 'EPPING'.
 IF HospitalName = 'NORTHERN HOSPIT' HSP = 'EPPING'.
 IF HospitalName = 'NORTHERN HWY' HSP = 'VAT_H Missing'.
 IF HospitalName = 'NORTHERN REGION' HSP = 'VAT_H Missing'.
 IF HospitalName = 'NORTHEWRN HOSPI' HSP = 'RURAL NORTHWEST'.
 IF HospitalName = 'NOT KNOWN' HSP = 'VAT_H Missing'.
 IF HospitalName = 'NTH HOSP' HSP = 'EPPING'.
 IF HospitalName = 'NTH HOSPITAL' HSP = 'EPPING'.
 IF HospitalName = 'OMEEO DISTRICT' HSP = 'OMEEO'.
 IF HospitalName = 'OMEEO HOSPITAL' HSP = 'OMEEO'.
 IF HospitalName = 'ORBOST' HSP = 'ORBOST'.
 IF HospitalName = 'ORBOST HOSPITAL' HSP = 'ORBOST'.
 IF HospitalName = 'ORBOST REGIONAL' HSP = 'ORBOST'.
 IF HospitalName = 'OTWAY HEALTH' HSP = 'APOLLO BAY'.
 IF HospitalName = 'OUYEN' HSP = 'VAT_H Missing'.
 IF HospitalName = 'OWN G P' HSP = 'VAT_H Missing'.
 IF HospitalName = 'OYAL MELBOURNE' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'PORTLAND' HSP = 'PORTLAND'.
 IF HospitalName = 'PORTLAND & DIST' HSP = 'PORTLAND'.
 IF HospitalName = 'PORTLAND A & E' HSP = 'PORTLAND'.
 IF HospitalName = 'PORTLAND DISTRI' HSP = 'PORTLAND'.
 IF HospitalName = 'PRINCE ALFRED' HSP = 'VAT_H Missing'.
 IF HospitalName = 'Private DOCTOR' HSP = 'VAT_H Missing'.
 IF HospitalName = 'R.M.H' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'RC' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'RCH' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'REFUSED AMBULAN' HSP = 'VAT_H Missing'.
 IF HospitalName = 'RM' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'RMH' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROBINVALE' HSP = 'ROBINVALE'.
 IF HospitalName = 'ROBINVALE BASE' HSP = 'ROBINVALE'.
 IF HospitalName = 'ROCHESTER' HSP = 'ROCHESTER'.
 IF HospitalName = 'ROSEBUD' HSP = 'ROSEBUD'.
 IF HospitalName = 'ROSEBUD.HOSP' HSP = 'ROSEBUD'.
 IF HospitalName = 'ROYAL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'ROYAL ADELAIDE' HSP = 'VAT_H Missing'.
 IF HospitalName = 'ROYAL ALFRED' HSP = 'ALFRED'.
 IF HospitalName = 'ROYAL CHILDREBS' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'ROYAL CHILDREN' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'ROYAL CHILDRENS' HSP = 'ROYAL CHILDRENS'.
 IF HospitalName = 'ROYAL MEL' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELB' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELB HOSP' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELBG' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELBO9URN' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELBOUNE' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELBOURN' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELBOURNE' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL MELBUORNE' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'ROYAL WOMANS' HSP = 'ROYAL WOMENS'.
 IF HospitalName = 'ROYAL WOMENS' HSP = 'ROYAL WOMENS'.
 IF HospitalName = 'ROYAL WOMENS HO' HSP = 'ROYAL WOMENS'.
 IF HospitalName = 'ROYAM MELBOURNE' HSP = 'ROYAL MELBOURNE'.
 IF HospitalName = 'S W HEALTHCARE' HSP = 'SOUTH WEST'.
 IF HospitalName = 'S/WEST HEALTHCA' HSP = 'SOUTH WEST'.
 IF HospitalName = 'SAKE' HSP = 'SALE'.

IF HospitalName = 'SALE' HSP = 'SALE'.
 IF HospitalName = 'SALE GBH' HSP = 'SALE'.
 IF HospitalName = 'SALE HOSP' HSP = 'SALE'.
 IF HospitalName = 'SALE HOSPITAL' HSP = 'SALE'.
 IF HospitalName = 'SANDRINGHAM' HSP = 'SANDRINGHAM'.
 IF HospitalName = 'SANDRINGHAM' HSP = 'SANDRINGHAM'.
 IF HospitalName = 'SANDRINGHAM HOS' HSP = 'SANDRINGHAM'.
 IF HospitalName = 'SEPARIT HOSP' HSP = 'VAT_H Missing'.
 IF HospitalName = 'SEYMOUR' HSP = 'SEYMOUR'.
 IF HospitalName = 'SEYMOUR BASE' HSP = 'SEYMOUR'.
 IF HospitalName = 'SHEPPARTON' HSP = 'GOULBURN VALLEY'.
 IF HospitalName = 'SIA MED CENT' HSP = 'VAT_H Missing'.
 IF HospitalName = 'SOUTH EASTERN' HSP = 'Private'.
 IF HospitalName = 'SOUTH GIPPSLAND' HSP = 'FOSTER'.
 IF HospitalName = 'SOUTH WEST' HSP = 'SOUTH WEST'.
 IF HospitalName = 'SOUTH WEST HEAL' HSP = 'SOUTH WEST'.
 IF HospitalName = 'SOUTHERN HEALTH' HSP = 'VAT_H Missing'.
 IF HospitalName = 'SOUTHWEST' HSP = 'SOUTH WEST'.
 IF HospitalName = 'SOUTHWEST HEALT' HSP = 'SOUTH WEST'.
 IF HospitalName = 'ST ARNAUD' HSP = 'ST ARNAUD'.
 IF HospitalName = 'ST JOHN"S' HSP = 'Private'.
 IF HospitalName = 'ST JOHN"S OF GO' HSP = 'Private'.
 IF HospitalName = 'ST JOHN BALLARA' HSP = 'Private'.
 IF HospitalName = 'ST JOHN HOSP' HSP = 'Private'.
 IF HospitalName = 'ST JOHN OF BALL' HSP = 'Private'.
 IF HospitalName = 'ST JOHN OF GOD' HSP = 'Private'.
 IF HospitalName = 'ST JOHN OF GOLD' HSP = 'Private'.
 IF HospitalName = 'ST JOHNS' HSP = 'Private'.
 IF HospitalName = 'ST JOHNS BALLAR' HSP = 'Private'.
 IF HospitalName = 'ST JOHNS HOSPIT' HSP = 'Private'.
 IF HospitalName = 'ST JOHNS OF GOD' HSP = 'Private'.
 IF HospitalName = 'ST VINCENT"S' HSP = 'ST VINCENTS'.
 IF HospitalName = 'ST VINCENT' HSP = 'ST VINCENTS'.
 IF HospitalName = 'ST VINCENTS' HSP = 'ST VINCENTS'.
 IF HospitalName = 'ST VINCENTS HOS' HSP = 'ST VINCENTS'.
 IF HospitalName = 'ST VINCENTS MEL' HSP = 'ST VINCENTS'.
 IF HospitalName = 'ST VINCENTS PUB' HSP = 'ST VINCENTS'.
 IF HospitalName = 'STAWELL' HSP = 'STAWELL'.
 IF HospitalName = 'STAWELL BASE' HSP = 'STAWELL'.
 IF HospitalName = 'STAWELL HOSP' HSP = 'STAWELL'.
 IF HospitalName = 'SUNSHIJNE' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSHINE' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSHINE GENERA' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSHINE HOSP' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSHINE HOSPIT' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSHINE WEST G' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSHINE WESTER' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SUNSIHNE' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'SWAN HILL' HSP = 'SWAN HILL'.
 IF HospitalName = 'SWAN HILL DISTR' HSP = 'SWAN HILL'.
 IF HospitalName = 'SWHC' HSP = 'SWAN HILL'.
 IF HospitalName = 'TERANG' HSP = 'TERANG'.
 IF HospitalName = 'THE ALFRED' HSP = 'ALFRED'.
 IF HospitalName = 'THE ANGLISS' HSP = 'ANGLISS'.
 IF HospitalName = 'THE MEDICAL CEN' HSP = 'VAT_H Missing'.
 IF HospitalName = 'THE NORTHERN' HSP = 'EPPING'.
 IF HospitalName = 'THE VALLEY' HSP = 'Private'.
 IF HospitalName = 'THE VALLEY HOSP' HSP = 'Private'.
 IF HospitalName = 'THE VALLEY PRIV' HSP = 'Private'.
 IF HospitalName = 'TIMBOON' HSP = 'TIMBOON'.

IF HospitalName = 'TIMBOON HOSPITA' HSP = 'TIMBOON'.
 IF HospitalName = 'TRARALGON' HSP = 'TRARALGON'.
 IF HospitalName = 'TRARALGON HOSPI' HSP = 'TRARALGON'.
 IF HospitalName = 'TRARALGON REGIO' HSP = 'TRARALGON'.
 IF HospitalName = 'UMH&CS' HSP = 'CORRYONG'.
 IF HospitalName = 'UNKNOWN' HSP = 'VAT_H Missing'.
 IF HospitalName = 'UPPER MURRAY HE' HSP = 'CORRYONG'.
 IF HospitalName = 'VALLEY Private' HSP = 'Private'.
 IF HospitalName = 'VAUCLUSE HOSP' HSP = 'Private'.
 IF HospitalName = 'VINCENTS' HSP = 'ST VINCENTS'.
 IF HospitalName = 'W & G' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'W & G SUNSHINE' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'W D B H' HSP = 'VAT_H Missing'.
 IF HospitalName = 'W.G. HOSP' HSP = 'VAT_H Missing'.
 IF HospitalName = 'W.G.H. FOOTSCRA' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'W.M.H' HSP = 'VAT_H Missing'.
 IF HospitalName = 'W/BOOL BASE' HSP = 'SOUTH WEST'.
 IF HospitalName = 'WAGARATTA' HSP = 'WANGARATTA'.
 IF HospitalName = 'WANGARATA BASE' HSP = 'WANGARATTA'.
 IF HospitalName = 'WANGARATTA' HSP = 'WANGARATTA'.
 IF HospitalName = 'WANGARATTA BASE' HSP = 'WANGARATTA'.
 IF HospitalName = 'WANGARATTA DIST' HSP = 'WANGARATTA'.
 IF HospitalName = 'WANGARATTA HOSP' HSP = 'WANGARATTA'.
 IF HospitalName = 'WANGARATTA.BASE' HSP = 'WANGARATTA'.
 IF HospitalName = 'WARLEY-COWES' HSP = 'VAT_H Missing'.
 IF HospitalName = 'WARLEY' HSP = 'VAT_H Missing'.
 IF HospitalName = 'WARLEY HOSP' HSP = 'VAT_H Missing'.
 IF HospitalName = 'WARNAMBOOL ' HSP = 'SOUTH WEST'.
 IF HospitalName = 'WARNAMBOOL BASE' HSP = 'SOUTH WEST'.
 IF HospitalName = 'WARRACKNABEAL' HSP = 'RURAL NORTHWEST'.
 IF HospitalName = 'WARRAGUL' HSP = 'WARRAGUL'.
 IF HospitalName = 'WARRAGUL BASE' HSP = 'WARRAGUL'.
 IF HospitalName = 'WARRAGUL HOSPIT' HSP = 'WARRAGUL'.
 IF HospitalName = 'WARRAGUL.HOSPIT' HSP = 'WARRAGUL'.
 IF HospitalName = 'WARRANAMBOOL BA' HSP = 'SOUTH WEST'.
 IF HospitalName = 'WARRIGAL' HSP = 'WARRAGUL'.
 IF HospitalName = 'SOUTH WEST HEALTHCARE' HSP = 'SOUTH WEST'.
 IF HospitalName = 'SOUTH WEST HEALTHCARE BAS' HSP = 'SOUTH WEST'.
 IF HospitalName = 'WELLINGTON HEAL' HSP = 'VAT_H Missing'.
 IF HospitalName = 'WEMMERA BASE' HSP = 'HORSHAM'.
 IF HospitalName = 'WERIBEE MERCY' HSP = 'MERCY'.
 IF HospitalName = 'WERRIBEE' HSP = 'MERCY'.
 IF HospitalName = 'WERRIBEE HOSPIT' HSP = 'MERCY'.
 IF HospitalName = 'WERRIBEE MERCY' HSP = 'MERCY'.
 IF HospitalName = 'WEST FOOTSCRAY' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WEST GENERAL' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WEST GIIPSLAND' HSP = 'WARRAGUL'.
 IF HospitalName = 'WEST GIPPSLAND' HSP = 'WARRAGUL'.
 IF HospitalName = 'WEST WIMMERA BA' HSP = 'WEST WIMMERA'.
 IF HospitalName = 'WESTERN' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN FOOTSCR' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN GEN' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN GENE' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN GENERAL' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN GENREAL' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN HEALTH' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN HOSP' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN HOSPITA' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WESTERN SUNSHIN' HSP = 'WESTERN HEALTH'.
 IF HospitalName = 'WG' HSP = 'WESTERN HEALTH'.

```

IF HospitalName = 'WG HOSP' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WGH - SUNSHINE' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WGH' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WGH FOOTSCRAY' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WGH HOSP' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WGH, FOOTSCRAY' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WILLIAM ANGLISS' HSP = 'ANGLISS'.
IF HospitalName = 'WILLIAM ANGLIAS' HSP = 'ANGLISS'.
IF HospitalName = 'WILLIAM ANGLIS' HSP = 'ANGLISS'.
IF HospitalName = 'WILLIAM ANGLISS' HSP = 'ANGLISS'.
IF HospitalName = 'WILLIAMS ANGLIS' HSP = 'ANGLISS'.
IF HospitalName = 'WILLIAMSTOWN' HSP = 'WESTERN HEALTH'.
IF HospitalName = 'WILLIAN ANGLISS' HSP = 'ANGLISS'.
IF HospitalName = 'WILLOW ANGLISS' HSP = 'ANGLISS'.
IF HospitalName = 'WIMMERA BASE' HSP = 'HORSHAM'.
IF HospitalName = 'WIMMERA BASE HO' HSP = 'HORSHAM'.
IF HospitalName = 'WIMMERIA BASE' HSP = 'HORSHAM'.
IF HospitalName = 'WINCHELSEA' HSP = 'WINCHELSEA'.
IF HospitalName = 'WN' HSP = 'VAT_H Missing'.
IF HospitalName = 'WO' HSP = 'VAT_H Missing'.
IF HospitalName = 'WODONA' HSP = 'WODONGA'.
IF HospitalName = 'WODONGA' HSP = 'WODONGA'.
IF HospitalName = 'WODONGA BASE' HSP = 'WODONGA'.
IF HospitalName = 'WODONGA DISTRIC' HSP = 'WODONGA'.
IF HospitalName = 'WODONGA HOSPITA' HSP = 'WODONGA'.
IF HospitalName = 'WONGTHAGGI' HSP = 'VAT_H Missing'.
IF HospitalName = 'WONTHAGGI' HSP = 'VAT_H Missing'.
IF HospitalName = 'XEPW' HSP = 'VAT_H Missing'.
IF HospitalName = 'Y918HERN' HSP = 'VAT_H Missing'.
IF HospitalName = 'YARRAM' HSP = 'YARRAM'.
IF HospitalName = 'YARRAM DISTRICT' HSP = 'YARRAM'.
IF HospitalName = 'YARRAWONGA' HSP = 'YARRAWONGA'.
IF HospitalName = 'YARRAWONGA HOSP' HSP = 'YARRAWONGA'.
IF HospitalName = 'YEA' HSP = 'YEA'.
IF HospitalName = 'YEA HOSPITAL' HSP = 'YEA'.
EXECUTE.

```

SORT CASES BY ADMDATE HSP PSTCODE MODAGE PERSEX.

SAVE OUTFILE 'VAT_H04Sort1.sav'.

DHS04Sort1.sps

```

GET FILE 'DHS04.sav'
  /RENAME admdate = Hadmdate modage = Hmodage pstcode = Hpstcode sex = Hsex.

NUMERIC ADMDATE(F6.0) ADMDATE1(F6.0) NEWADMDATE(F6.0).

COMPUTE ADMDATE = YRMODA(ADM_YEAR, ADM_MONTH, ADM_DAY) - YRMODA(1960,1,1).
COMPUTE ADMDATE1 = ADMDATE.
COMPUTE NEWADMDATE = (ADMDATE + YRMODY(1960,1,1)) * 86400.
EXECUTE.

FORMAT NEWADMDATE(DATE9).
EXECUTE.

NUMERIC MODAGE(F2.0).

COMPUTE MODAGE = Hmodage.
EXECUTE.

```

NUMERIC PSTCODE(F4.0).

COMPUTE PSTCODE = NUMBER(Hpstcode,F4.0).
EXECUTE.

STRING PERSEX(A1).
EXECUTE.

IF Hsex = 1 PERSEX = 'M'.
IF Hsex = 2 PERSEX = 'F'.
EXECUTE.

STRING HSP(A25).
EXECUTE.

COMPUTE HSP = 'DHS Missing'.
EXECUTE.

IF campname = 'Alexandra District Hospital' HSP = 'ALEXANDRA'.
IF campname = 'Alfred The [Pahran]' HSP = 'ALFRED'.
IF campname = 'Alpine Health [Bright]' HSP = 'BRIGHT'.
IF campname = 'Alpine Health [Mount Beauty]' HSP = 'MOUNT BEAUTY'.
IF campname = 'Alpine Health [Myrtleford]' HSP = 'MYRTLEFORD'.
IF campname = 'Angliss Hospital' HSP = 'ANGLISS'.
IF campname = 'Austin Hospital' HSP = 'AUSTIN'.
IF campname = 'Bairnsdale Regional Health Service' HSP = 'BAIRNSDALE'.
IF campname = 'Ballarat Health Services [Base Campus]' HSP = 'BALLARAT'.
IF campname = 'Ballarat Health Services [Queen Elizabeth Campus]' HSP = 'BALLARAT'.
IF campname = 'Bass Coast Regional Health' HSP = 'BASS'.
IF campname = 'Benalla & District Memorial Hospital' HSP = 'BENALLA'.
IF campname = 'Bendigo Health Care Group - Anne Caudle' HSP = 'BENDIGO'.
IF campname = 'Bendigo Hospital The' HSP = 'BENDIGO'.
IF campname = 'Boort District Hospital' HSP = 'BOORT'.
IF campname = 'Box Hill Hospital' HSP = 'BOX HILL'.
IF campname = 'Casterton Memorial Hospital' HSP = 'CASTERTON'.
IF campname = 'Central Gippsland Health Service [Sale]' HSP = 'SALE'.
IF campname = 'Cobram District Hospital' HSP = 'COBRAM'.
IF campname = 'Cohuna District Hospital' HSP = 'COHUNA'.
IF campname = 'Colac Area Health' HSP = 'COLAC'.
IF campname = 'Coleraine District Health Services' HSP = 'COLERAINE'.
IF campname = 'Dandenong Campus' HSP = 'DANDENONG'.
IF campname = 'Daylesford District Hospital' HSP = 'DAYLESFORD'.
IF campname = 'Dimboola District Hospital' HSP = 'DIMBOOLA'.
IF campname = 'Djerriwarrh Health Service [Bacchus Marsh]' HSP = 'BACCHUS MARSH'.
IF campname = 'East Grampians Health Service [Ararat]' HSP = 'ARARAT'.
IF campname = 'East Wimmera Health Service [Birchip]' HSP = 'BIRCHIP'.
IF campname = 'East Wimmera Health Service [Charlton]' HSP = 'CHARLTON'.
IF campname = 'East Wimmera Health Service [Donald]' HSP = 'DONALD'.
IF campname = 'East Wimmera Health Service [St Arnaud]' HSP = 'ST ARNAUD'.
IF campname = 'Echuca Regional Health' HSP = 'ECHUCA'.
IF campname = 'Edenhope & District Hospital' HSP = 'EDENHOPE'.
IF campname = 'Frankston Hospital' HSP = 'FRANKSTON'.
IF campname = 'Geelong Hospital' HSP = 'GEELONG'.
IF campname = 'Goulburn Valley Health [Shepparton]' HSP = 'GOULBURN VALLEY'.
IF campname = 'Goulburn Valley Health [Tatura]' HSP = 'GOULBURN VALLEY'.
IF campname = 'Goulburn Valley Health [Waranga]' HSP = 'GOULBURN VALLEY'.
IF campname = 'Hamilton Base Hospital' HSP = 'HAMILTON'.
IF campname = 'Hesse Rural Health Service [Winchelsea]' HSP = 'WINCHELSEA'.
IF campname = 'Inglewood & District Health Service' HSP = 'INGLEWOOD'.

IF campname = 'Kerang District Health' HSP = 'KERANG'.
 IF campname = 'Kilmore & District Hospital The' HSP = 'KILMORE'.
 IF campname = 'Korumburra Hospital' HSP = 'KORUMBURRA'.
 IF campname = 'Kyabram & District Health Service' HSP = 'KYABRAM'.
 IF campname = 'Kyneton District Health Service' HSP = 'KYNETON'.
 IF campname = 'Latrobe Regional Hospital [Traralgon]' HSP = 'TRARALGON'.
 IF campname = 'Leongatha Memorial Hospital' HSP = 'LEONGATHA'.
 IF campname = 'Lorne Community Hospital' HSP = 'LORNE'.
 IF campname = 'Mallee Track Health & Community Service' HSP = 'MALLEE TRACK'.
 IF campname = 'Mansfield District Hospital' HSP = 'MANSFIELD'.
 IF campname = 'Maroondah Hospital [East Ringwood]' HSP = 'MAROONDAH'.
 IF campname = 'Maryborough District Health Service [Dunolly]' HSP = 'MARYBOROUGH'.
 IF campname = 'Maryborough District Health Service [Maryborough]' HSP = 'MARYBOROUGH'.
 IF campname = 'McIvor Health & Community Services [Heathcote]' HSP = 'HEATHCOTE'.
 IF campname = 'Mercy Hospital for Women' HSP = 'MERCY'.
 IF campname = 'Mercy Public Hospitals Inc [East Melbourne]' HSP = 'MERCY'.
 IF campname = 'Mercy Public Hospitals Inc [Werribee]' HSP = 'MERCY'.
 IF campname = 'Mildura Base Hospital' HSP = 'MILDURA'.
 IF campname = 'Monash Medical Centre [Clayton]' HSP = 'MONASH'.
 IF campname = 'Monash Medical Centre [Moorabbin]' HSP = 'MONASH'.
 IF campname = 'Mount Alexander Hospital [Castlemaine]' HSP = 'CASTLEMAINE'.
 IF campname = 'Northeast Health Wangaratta' HSP = 'WANGARATTA'.
 IF campname = 'Northern Hospital The [Epping]' HSP = 'EPPING'.
 IF campname = 'Omeo District Health' HSP = 'OMEO'.
 IF campname = 'Orbost Regional Health' HSP = 'ORBOST'.
 IF campname = 'Otway Health & Community Services [Apollo Bay]' HSP = 'APOLLO BAY'.
 IF campname = 'Portland District Health' HSP = 'PORTLAND'.
 IF campname = 'Private' HSP = 'Private'.
 IF campname = 'Robinvale District Health Services' HSP = 'ROBINVALE'.
 IF campname = 'Rochester & Elmore District Health Service' HSP = 'ROCHESTER'.
 IF campname = 'Rosebud Hospital' HSP = 'ROSEBUD'.
 IF campname = 'Royal Childrens Hospital [Parkville]' HSP = 'ROYAL CHILDRENS'.
 IF campname = 'Royal Childrens Hospital [Travancore Psych]' HSP = 'ROYAL CHILDRENS'.
 IF campname = 'Royal Melbourne Hospital - City Campus' HSP = 'ROYAL MELBOURNE'.
 IF campname = 'Royal Melbourne Hospital - Royal Park Campus' HSP = 'ROYAL MELBOURNE'.
 IF campname = 'Royal Melbourne Hospital' HSP = 'ROYAL MELBOURNE'.
 IF campname = 'Royal Melbourne Hospital Mental Health' HSP = 'ROYAL MELBOURNE'.
 IF campname = 'Royal Womens Hospital [Carlton]' HSP = 'ROYAL WOMENS'.
 IF campname = 'Rural Northwest Health [Hopetoun]' HSP = 'RURAL NORTHWEST'.
 IF campname = 'Rural Northwest Health [Warracknabeal]' HSP = 'RURAL NORTHWEST'.
 IF campname = 'Sandringham & District Memorial Hospital' HSP = 'SANDRINGHAM'.
 IF campname = 'Seymour District Memorial Hospital' HSP = 'SEYMOUR'.
 IF campname = 'South Gippsland Hospital [Foster]' HSP = 'FOSTER'.
 IF campname = 'South West Healthcare [Camperdown]' HSP = 'SOUTH WEST'.
 IF campname = 'South West Healthcare [Warrnambool]' HSP = 'SOUTH WEST'.
 IF campname = 'St Vincents Hospital' HSP = 'ST VINCENTS'.
 IF campname = 'Stawell Regional Health' HSP = 'STAWELL'.
 IF campname = 'Sunshine Hospital' HSP = 'WESTERN HEALTH'.
 IF campname = 'Swan Hill District Hospital [Nyah]' HSP = 'SWAN HILL'.
 IF campname = 'Swan Hill District Hospital [Swan Hill]' HSP = 'SWAN HILL'.
 IF campname = 'Terang & Mortlake Health Service [Terang]' HSP = 'TERANG'.
 IF campname = 'Timboon & District Healthcare Service' HSP = 'TIMBOON'.
 IF campname = 'Upper Murray Health & Community Services [Corryong]' HSP = 'Corryong'.
 IF campname = 'West Gippsland Healthcare Group [Warragul]' HSP = 'WARRAGUL'.
 IF campname = 'West Wimmera Health Service [Jeparit]' HSP = 'WEST WIMMERA'.
 IF campname = 'West Wimmera Health Service [Kaniva]' HSP = 'WEST WIMMERA'.
 IF campname = 'West Wimmera Health Service [Nhill]' HSP = 'WEST WIMMERA'.
 IF campname = 'West Wimmera Health Service [Rainbow]' HSP = 'WEST WIMMERA'.
 IF campname = 'Western Hospital [Footscray]' HSP = 'WESTERN HEALTH'.
 IF campname = 'Williamstown Hospital' HSP = 'WESTERN HEALTH'.

```
IF campname = 'Wimmera Base Hospital [Horsham]' HSP = 'HORSHAM'.
IF campname = 'Wodonga Regional Health Service' HSP = 'WODONGA'.
IF campname = 'Yarram & District Health Service' HSP = 'YARRAM'.
IF campname = 'Yarrawonga District Health Service' HSP = 'YARRAWONGA'.
IF campname = 'Yea & District Memorial Hospital' HSP = 'YEA'.
EXECUTE.
```

```
SORT CASES BY ADMDATE HSP PSTCODE MODAGE PERSEX.
```

```
SAVE OUTFILE 'DHS04Sort1.sav'.
```

HMerge04_1.sps

```
GET FILE 'DHS04Sort1.sav'
  /RENAME septype = Hseptype injcode = Hinjcode age = Hage lenstay = Hlenstay.
```

```
MATCH FILES FILE = 'VAT_H04Sort1.sav' /IN=INVAT
  /FILE = * /IN=INDHS
  /BY ADMDATE HSP PSTCODE MODAGE PERSEX
  /MAP.
```

```
EXECUTE.
```

```
SAVE OUTFILE = 'HMERGE04_1.sav'.
```

HSplit04_1.sps

```
GET FILE='HMERGE04_1.SAV'.
```

```
TEMP.
```

```
SELECT IF INVAT=1 AND INDHS=0.
SAVE OUTFILE='VAT_H04Left1.sav'
  /KEEP pers_id TO HSP.
```

```
TEMP.
```

```
SELECT IF INVAT=0 AND INDHS=1.
SAVE OUTFILE='DHS04Left1.sav'
  /KEEP acctype1 TO NEWADMDATE ADMDATE ADMDATE1 MODAGE PSTCODE PERSEX
HSP.
```

```
SELECT IF INVAT=1 AND INDHS=1.
SAVE OUTFILE='DMATCH04_1.sav'
  /DROP ADMDATE1 INVAT INDHS.
```

VAT_H04Sort2.sps

```
GET FILE 'VAT_H04Left1.sav'.
```

```
SORT CASES BY ADMDATE PSTCODE MODAGE PERSEX.
```

```
SAVE OUTFILE 'VAT_H04Sort2.sav'.
```

DHS04Sort2.sps

```
GET FILE 'DHS04Left1.sav'.
```

```
SORT CASES BY ADMDATE PSTCODE MODAGE PERSEX.
```

```
SAVE OUTFILE 'DHS04Sort2.sav'.
```

HMerge04_2.sps

```
GET FILE 'DHS04Sort2.sav'.

MATCH FILES FILE = 'VAT_H04Sort2.sav' /IN=INVAT
      /FILE = * /IN=INDHS
      /BY ADMDATE PSTCODE MODAGE PERSEX
      /MAP.
EXECUTE.

SAVE OUTFILE = 'HMERGE04_2.sav'.
```

HSplit04_2.sps

```
GET FILE='HMERGE04_2.SAV'.

TEMP.
SELECT IF INVAT=1 AND INDHS=0.
SAVE OUTFILE='VAT_H04Left2.sav'
      /KEEP pers_id TO HSP.

TEMP.
SELECT IF INVAT=0 AND INDHS=1.
SAVE OUTFILE='DHS04Left2.sav'
      /KEEP acctype1 TO NEWADMDATE ADMDATE ADMDATE1 MODAGE PSTCODE PERSEX
HSP.

SELECT IF INVAT=1 AND INDHS=1.
SAVE OUTFILE='DMATCH04_2.sav'
      /DROP ADMDATE1 INVAT INDHS.
```

VAT_H04Sort3.sps

```
GET FILE 'VAT_H04Left2.sav'.

SORT CASES BY ADMDATE1 HSP PSTCODE MODAGE PERSEX.

SAVE OUTFILE 'VAT_H04Sort3.sav'.
```

DHS04Sort3.sps

```
GET FILE 'DHS04Left2.sav'.

SORT CASES BY ADMDATE1 HSP PSTCODE MODAGE PERSEX.

SAVE OUTFILE 'DHS04Sort3.sav'.
```

HMerge04_3.sps

```
GET FILE 'DHS04Sort3.sav'.

MATCH FILES FILE = 'VAT_H04Sort3.sav' /IN=INVAT
      /FILE = * /IN=INDHS
      /BY ADMDATE1 HSP PSTCODE MODAGE PERSEX
      /MAP.
EXECUTE.
```

SAVE OUTFILE = 'HMERGE04_3.sav'.

HSplit04_3.sps

GET FILE='HMERGE04_3.sav'.

TEMP.

SELECT IF INVAT=1 AND INDHS=0.
SAVE OUTFILE='VAT_H04Left3.sav'
/KEEP pers_id TO HSP.

TEMP.

SELECT IF INVAT=0 AND INDHS=1.
SAVE OUTFILE='DHS04Left3.sav'
/KEEP acctype1 TO NEWADMDATE ADMDATE ADMDATE1 MODAGE PSTCODE PERSEX
HSP.

SELECT IF INVAT=1 AND INDHS=1.
SAVE OUTFILE='DMATCH04_3.sav'
/DROP ADMDATE INVAT INDHS
/RENAME ADMDATE1 = ADMDATE.

VAT_H04Sort4.sps

GET FILE 'VAT_H04Left3.sav'.

SORT CASES BY ADMDATE1 PSTCODE MODAGE PERSEX.

SAVE OUTFILE 'VAT_H04Sort4.sav'.

DHS04Sort4.sps

GET FILE 'DHS04Left3.sav'.

SORT CASES BY ADMDATE1 PSTCODE MODAGE PERSEX.

SAVE OUTFILE 'DHS04Sort4.sav'.

HMerge04_4.sps

GET FILE 'DHS04Sort4.sav'.

MATCH FILES FILE = 'VAT_H04Sort4.sav' /IN=INVAT
/FILE = * /IN=INDHS
/BY ADMDATE1 PSTCODE MODAGE PERSEX
/MAP.

EXECUTE.

SAVE OUTFILE = 'HMERGE04_4.sav'.

HSplit04_4.sps

GET FILE='HMERGE04_4.sav'.

TEMP.

SELECT IF INVAT=1 AND INDHS=0.

```
SAVE OUTFILE='VAT_H04Left4.sav'  
  /KEEP pers_id TO HSP.
```

```
TEMP.
```

```
SELECT IF INVAT=0 AND INDHS=1.  
SAVE OUTFILE='DHS04Left4.sav'  
  /KEEP acctype1 TO NEWADMDATE ADMDATE ADMDATE1 MODAGE PSTCODE PERSEX  
HSP.
```

```
SELECT IF INVAT=1 AND INDHS=1.  
SAVE OUTFILE='DMATCH04_4.sav'  
  /DROP ADMDATE INVAT INDHS  
  /RENAME ADMDATE1 = ADMDATE.
```

VAT_H04Sort5.sps

```
GET FILE 'VAT_H04Left4.sav'.
```

```
SORT CASES BY ADMDATE PSTCODE PERSEX.
```

```
SAVE OUTFILE 'VAT_H04Sort5.sav'.
```

DHS04Sort5.sps

```
GET FILE 'DHS04Left4.sav'.
```

```
SORT CASES BY ADMDATE PSTCODE PERSEX.
```

```
SAVE OUTFILE 'DHS04Sort5.sav'.
```

HMerge04_5.sps

```
GET FILE 'DHS04Sort5.sav'.
```

```
MATCH FILES FILE = 'VAT_H04Sort5.sav' /IN=INVAT  
  /FILE = * /IN=INDHS  
  /BY ADMDATE PSTCODE PERSEX  
  /MAP.
```

```
EXECUTE.
```

```
SAVE OUTFILE = 'HMERGE04_5.sav'.
```

HSplit04_5.sps

```
GET FILE 'HMERGE04_5.sav'.
```

```
NUMERIC AGEDIFF(F2.0).  
COMPUTE AGEDIFF=ABS(tage-Hmodage).  
EXECUTE.
```

```
RECODE AGEDIFF(2 THRU 9999=1)(ELSE=0).  
EXECUTE.
```

```
TEMP.
```

```
SELECT IF (AGEDIFF=1 OR (INVAT=1 AND INDHS=0)).  
SAVE OUTFILE='VAT_H04Left5.sav'  
  /KEEP pers_id TO HSP.
```

```
TEMP.  
SELECT IF (AGEDIFF=1 OR (INVAT=0 AND INDHS=1)).  
SAVE OUTFILE='DHS04Left5.sav'  
      /KEEP acctype1 TO NEWADMDATE ADMDATE ADMDATE1 MODAGE PSTCODE PERSEX  
HSP.
```

```
SELECT IF (AGEDIFF=0 AND INVAT=1 AND INDHS=1).  
SAVE OUTFILE='DMATCH04_5.sav'  
      /DROP ADMDATE1 INVAT INDHS AGEDIFF..
```

VAT_H04Sort6.sps

```
GET FILE 'VAT_H04Left5.sav'.  
  
SORT CASES BY ADMDATE1 PSTCODE PERSEX.  
  
SAVE OUTFILE 'VAT_H04Sort6.sav'.
```

DHS04Sort6.sps

```
GET FILE 'DHS04Left5.sav'.  
  
SORT CASES BY ADMDATE1 PSTCODE PERSEX.  
  
SAVE OUTFILE 'DHS04Sort6.sav'.
```

HMerge04_6.sps

```
GET FILE 'DHS04Sort6.sav'.  
  
MATCH FILES FILE = 'VAT_H04Sort6.sav' /IN=INVAT  
      /FILE = * /IN=INDHS  
      /BY ADMDATE1 PSTCODE PERSEX  
      /MAP.  
EXECUTE.  
  
SAVE OUTFILE = 'HMERGE04_6.sav'.
```

HSplit04_6.sps

```
GET FILE='HMERGE04_6.sav'.  
  
NUMERIC AGEDIFF(F2.0).  
COMPUTE AGEDIFF=ABS(tage-Hmodage).  
EXECUTE.  
  
RECODE AGEDIFF(2 THRU 9999=1)(ELSE=0).  
EXECUTE.  
  
TEMP.  
SELECT IF (AGEDIFF=1 OR (INVAT=1 AND INDHS=0)).  
SAVE OUTFILE='VAT_H04Left6.sav'  
      /KEEP pers_id TO HSP.  
  
TEMP.  
SELECT IF (AGEDIFF=1 OR (INVAT=0 AND INDHS=1)).  
SAVE OUTFILE='DHS04Left6.sav'
```

```
/KEEP acctype1 TO NEWADMDATE ADMDATE ADMDATE1 MODAGE PSTCODE PERSEX  
HSP.
```

```
SELECT IF (AGEDIFF=0 AND INVAT=1 AND INDHS=1).  
SAVE OUTFILE='DMATCH04_6.sav'  
/DROP ADMDATE1 INVAT INDHS AGEDIFF.
```

VAT_H04DHS04.sps

```
ADD FILES FILE='DMATCH04_1.sav'  
/FILE='DMATCH04_2.sav'  
/FILE='DMATCH04_3.sav'  
/FILE='DMATCH04_4.sav'  
/FILE='DMATCH04_5.sav'  
/FILE='DMATCH04_6.sav'.  
EXECUTE.
```

```
SAVE OUTFILE='VAT_H04DHS04.sav'.
```