LGRAS: 2012-05-18 meeting
Note on STATS 19 concerns

Introduction.

Pat Kilbey of the UK Department of Transport and Matt Perkins of the Scottish Government’s agency Transport Scotland had expressed interest in the results of the collation of a complete sample of the queries made by Falkirk Council of STATS 19 data collected by Central Scotland Police.

Falkirk Council’s road safety engineer (and CoSLA rep. at SCRAS) had, for many years expressed concern about data quality and had decided that it might help to attempt to identify and quantify the more obvious problems. A main motivation was that it had proved impossible to report quickly and conclusively to superiors on priorities for action at junctions or at sites with a “darkness” problem.

Obviously, this has negative consequences for the rate at which we accidents can be reduced through roads engineering.

So, around 50 days of the year’s 200 working days were spent in validating accident data to a greater degree than previously, so that it might be possible to:

1. Compare the results with similar attempts made by commentators, earlier, and
2. Identify, more clearly, those variables causing difficulty.

It is important to note that, contrary to the claim made at the 2011 meeting of SCRAS, validating the STATS 19 data for one accident is NOT a 5 minute job. The whole cycle of reading, comprehending, querying, amplifying, and agreeing and, ultimately, updating the local database seems to take around one day per 5 accidents. Assuming a 7.5hr day, then the whole process takes around 1.5hrs per accident – at the local authority end. To that figure, however, the time spent by the data collectors in responding the users’ queries should be added.

Comparison with earlier work.

It had been hoped that Hutchinson’s 1987 Road Accident Statistics would have provided a basis for comparison, but Hutchison provides no suitable UK figures. The best that could be found at the time of writing were those of Professor Chris Wright. His findings of 1997-1999 for 1990-1996 data relating to the Isle of Thanet were compared with 2010 data relating to the Falkirk Council area.

A further misfortune is that Professor Wright’s final report on: The Development of Improved Methods for Representing Road Accident Data refers specifically to only a very few of the STATS19 variables. However, it is possible that a comparison of those with current ones may be taken as indicative of the overall situation. The results are shown below:

<table>
<thead>
<tr>
<th>STATS19 variable</th>
<th>1990-1996 Thanet area</th>
<th>2010 Falkirk area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.14 Road Type</td>
<td>2.51</td>
<td>9.09</td>
</tr>
<tr>
<td>1.15 Speed Limit</td>
<td>9.68</td>
<td>4.13</td>
</tr>
<tr>
<td>1.16 Junction Detail</td>
<td>12.92</td>
<td>24.79</td>
</tr>
<tr>
<td>1.17 Junction Control</td>
<td>14.96</td>
<td>11.57</td>
</tr>
<tr>
<td>1.20b Ped. Xg physical</td>
<td>1.31</td>
<td>5.37</td>
</tr>
<tr>
<td>Number in sample</td>
<td>1912</td>
<td>242</td>
</tr>
</tbody>
</table>

This suggests that matters are no better. Some of the large differences in percentage between areas (in either direction) may be related to the different natures of the areas involved. If, for example, Falkirk Council area is markedly more urban than Thanet, for instance, one might expect fewer speed limit miscodings and more pedestrian crossing miscodings.
Clearly a comparison based on the same area and similar time periods would be better, but no-one appears to have attempted this. Nonetheless, it seems quite safe to say that this comparison suggests that there has been no marked improvement in STATS19 data quality despite the reviews conducted since 1996.

(One exception is in the allocation of OS grid co-ordinates, but this may be a result of using GPS.)

**Variables causing most difficulty.**

A minimal number of 1,030 queries (the Contributory Factor data having been treated somewhat cursorily) arose from the STATS19 data for 242 personal injury accidents in the Falkirk Council area in 2010. Those variables most frequently queried were distributed thus:

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Number of Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Movement</td>
<td>140</td>
</tr>
<tr>
<td>Compass</td>
<td>120</td>
</tr>
<tr>
<td>Accident Location</td>
<td>100</td>
</tr>
<tr>
<td>Junction</td>
<td>80</td>
</tr>
<tr>
<td>Movement</td>
<td>60</td>
</tr>
<tr>
<td>Accident Description</td>
<td>40</td>
</tr>
<tr>
<td>Junction Detail</td>
<td>20</td>
</tr>
<tr>
<td>Contributory Factor</td>
<td>10</td>
</tr>
</tbody>
</table>

It should be noted that nearly every query resulted in a change to the data and that, as responses were received, more questions arose. The validating process requires practice. The additional queries related to lighting and weather, so they are under-reported in the above chart. So also, by a considerable margin, are the CF codes (represented by a single column above). These were not consistently queried as their subjectivity and lack of mutual exclusivity makes them barely worth considering. A proportion seemed to demand querying, and were queried, but if all suspect CF selections had been queried the chart would have required re-scaling.

The chart appears to suggest an “order” of difficulty, but it should be borne in mind that, whereas 2.8 Vehicle Movement Compass seems worst, all the accidents involved had at least one set of compass points assigned to a vehicle. Not all accidents would (e.g.) have a junction associated with them. It is probably easier to make a “null” choice than to choose between junction types. So the number of junction code mistakes may tend to be lower. Direction codes are also fairly easy to correct, although there are many of them. A variable such as 1.20b Pedestrian Crossing - Physical Facilities, although relevant to a small number of accidents, causes much more trouble proportionally, perhaps because reference to, and comprehension of, STATS 20 is required.

A particular concern is 2.1 Record Type, as this represents a query regarding the number of vehicles reported. In some cases no record had been made of a vehicle which had clearly contributed to the accident, but which was not damaged. A suspicion is that the current fashion for referring to accidents as collisions contributes to these failures. Clearly, therefore, this chart is only a starting point.
Now, some may say that querying a mere 1,030 of the 242x76 (excluding up to 18 CFs and assuming only one vehicle and one casualty) = 18,392 (i.e. a maximum of 5.6%) data items collected is not significant. But that is not the point. There are at least three, more relevant, ones.

First, as can be seen from the 40 variables (excluding CFs) featured in the chart, these are vital to the discharge of local road (or traffic) authorities' statutory obligations to study road accidents and to act upon their results. In some cases, a miscoding in one data item can change, for example, the significance of the item in a cluster, or the priority a blacksite or area will receive. So, every one of these data items must be correct.

Second, the amount of time being spent on validating accident data is unacceptable to local authorities, as it must be to the data collectors.

Third, the current situation should not exist. Little appears to have changed since T. P. Hutchinson’s “Road Accident Statistics” was published in 1987. Both Professor Chris Wright’s work and the 2003 presentation to LGRAS of Doctor Brendan Wallace’s review of a set of contributory factors (kindly facilitated by Strathclyde Police), seem to have been ignored – and no-one has yet demonstrated how CRASH will improve matters. Indeed, we seem to have regressed – as the inclusion of the unreliable “contributory factors” - seems to indicate.

With a new review cycle possibly about to start, every road safety engineer in Scotland will be hoping that the approach will be one which takes into account the need to apply up-to-date methods in the design and testing of the system. A system, which is mutually exclusive and exhaustive, and which has been subjected to the reliability testing (all described by Wallace & Ross in 2006) in amplification of Wright’s work will, in turn, “reduce the burden” on all involved and make a real, measurable, improvement in data quality.

Some thoughts on the current problems.

It may be helpful to look at some of the variables that are causing problems a little more closely and consider what the problem might be. Taking them on the order in which they appear on the chart above, and assuming that it is the reporting officer who completes the form (which may not be the case universally - another problem):

2.8 Vehicle Movement Compass Point

There is no doubt that a large proportion of the population has difficulty in distinguishing north from south, east from west and many also confuse north-south with east-west. It appears that some associate up with north and down with south, although this effect has not been noticed with east or west. STATS 19, however, requires the use not only of the cardinal points of the compass, but also the ordinal ones. There is a suspicion that the latter are simply ignored by some reporting officers, but querying them, over a year, seems to have made a difference. There is always scope for disagreement when a direction is very close to a cardinal or ordinal one, and so one wonders whether the use of compass points should be discontinued and compass bearings substituted.

A frequent failure is to record the intended direction at junctions and roundabouts. Some further guidance may be required on how to deal with gyratory systems and reversing while parking.

2.10 Junction Location of Vehicle

There seems to be a lack of awareness of the existence of code 8 - Mid junction - on roundabout or on main road, although this may be where a large proportion of vehicles actually collide. A possible reason is the position of the code on the list. A more logical list might be:
0 - Not within 20m of junction
1 - Within 20m on approach to junction
8 - Mid junction
2 - Within 20m on exit from junction

The above omits reference to entering roundabouts/leaving roundabouts and main roads, as these codes are not mutually exclusive of those in the list.

2.7 Manoeuvres

A difficulty with this code (particularly with major roads) is the inability to code, e.g. “Going ahead left hand bend” with “Overtaking … “. There is a lack of mutual exclusivity. Consequently, there is a wide variety of mistakes, many understandable. A vehicle pulls out to turn right from a side road into a main road is described as “moving off”, rather than “turning right”. On a well-known double bend, where one vehicle should be coded “Going ahead left hand bend” and the other “Going ahead right hand bend”, BOTH are coded “Going ahead other”.

A possibility is that the list is too long and tries to cope with too many different entities - the road nature and the driver’s action or intention. What if the driver is waiting to turn left on a right hand bend?

Accident Description

There are clear terminological difficulties. The differences between “lane” and “carriageway”, “offside” and “nearside”, “north”, “south”, “east” and “west” are frequently confused. The quality is highly variable, and some are very good indeed. All are essential to making sense of the codes.

1.16 Junction Detail

There is a basic difficulty here, in that private drives are treated differently from all other junctions. Perhaps a preliminary code is required to discriminate between accidents actually involving junction manoeuvres. This is the business of the taxonomist.

Another problem may be order, which might explain, e.g., the selection of roundabout instead of mini-roundabout.

An alternative might be:

00 - Not at or within 20m of junction
08 - At or within 20m of junction with private driver or entrance in use
03 - At or within 20m of T jn (3 legs)
10 - At or within 20m of staggered junction (4 legs)
06 - At or within 20m of cross roads (4 legs - no stagger)
07 - At or within 20m of complex junction (more than 4 legs - but not roundabout)
02 - At or within 20m of mini-roundabout
01 - At or within 20m of roundabout
05 - At or within 20m of slip road
09 - At or within 20m of other junction (provide description).

Accident Location (Description)

These were much more difficult previously, as reporting officers generally guessed distances rather than estimated them. With GPS, provided the co-ordinates are correct, there should be little need for a distance to be included. Occasionally, roads are incorrectly named, or there is confusion (at junctions) as to which road the accident actually occurred upon. There are mistakes over road identities, generally. There is a clear need for reporting officers and local authority engineers to be referring to the same data when describing locations as road names are not always clear from the maps generally available. Local authorities, however, have a responsibility for maintaining lists of roads and are, therefore, in a good position to assist.
**Contributory Factor 1**

The Contributory Factors (CF) record is a minefield, and it is detracting from the quality of the other STATS 19 records.

For example, a comparison of the use of 3.11 Pedestrian Movement (PM) codes 2, 4 and 6 were compared with the use of CF 801 during 2005 and 2006 was made in 2008. These codes all relate to masking. It was found that searches for accidents in which a pedestrian was “masked” in some way produced 51 when based on 3.11 Pedestrian Movement and 46 when based on CFs. In fact, detailed study showed that there were 69 personal injury accidents which involved “masking”, but in only 28 was there “agreement” between the codes selected at 3.11 PM and the selection of CF 801.

It seems that some reporting officers are recording “masking” in the Casualty Record, but not in the Contributory Factors record and that others are doing the opposite. In the sample referred to above, “agreement” was achieved in only 41% of “masking” accidents.

An earlier examination of the codes used in recording the presence of oil, diesel, mud, chippings, etc on the road surface, it was found, not only that the question “to how many accidents did oil or diesel [on the road surface] contribute?” not be answered promptly, it could not be answered at all. In the course of the examination, it was also found that, of the 28 accidents in which the textual data suggested that the presence of oil or diesel may have contributed, in only 2 were the data sufficiently consistent for a researcher to be reasonably sure that oil or diesel on the road surface did contribute.

Further complications are introduced by the presentation of CFs as an array and the requirements first to assign them to a participant and then to provide a level of confidence. Factors are frequently assigned to the wrong participant. Several factors appear twice in the matrix, virtually guaranteeing that mistakes will be made, as reporting officers (inevitably) forget to check the row headings.

The most useful of the CFs might usefully be included in the Circumstance, Vehicle or Casualty codes, once their taxonomy has been attended to.

**2.16 First Point of Impact**

It seems surprising that this code should create any difficulty at all. However, the terms “nearside” and “offside” are frequently confused. As alternatives, “left” and “right” are hardly likely to improve matters, and so the following is suggested:

0 - No impact
1 - Front
2 - Right hand drive vehicle - driver’s side
5 - Left hand drive vehicle - passenger side
7 - TWMV - accelerator side
9 - Pedal cycle - chain side
2 - Back
4 - Right hand drive vehicle - passenger side
6 - Left hand drive vehicle - driver’s side
8 - TWMV - clutch side
10 - Pedal cycle - rear brake side
11 - Roof
12 - Other (explain)

**2.1 Record Type and 1.5 Number of Vehicle Records**

The main problem with these is the regular failure to record details of “vehicles which did not suffer damage, nor caused not contained casualties, but which contributed to the accident (includes parked, stationary, temporarily held-up or moving vehicles).” In correspondence with the LGRAS, it has been suggested that there are now thousands of “missing vehicles”. Hence, the incidence of single vehicle accidents is likely to be highly exaggerated and possibly noticeable, even at the coarse, “high” level.
A memorable example involved an HGV parked on a mini-roundabout while its driver asked for directions. Two vehicles collided while trying to manoeuvre around it, but the set of accident records contained only two, not three, for vehicles. It is, therefore, very important for reporting officers to realise that their responsibility is to record details of the ACCIDENT, not merely of those vehicles which may have collided. In this respect, it would be helpful if current trend to substitute the term “collision” for “accident” was reversed. The assistance of the police is essential to this, and it is hoped that it will not be withheld.

Other problems

The foregoing is intended to give a flavour of the problems being encountered by data users within the local authorities, some of whom have been driven, possibly by the current financial constraints, to give up the effort involved in querying the data. This is a consequence of local government re-organisation, but much ground could be made up, if modern methods were applied.

Over the past year or so, data problems have been reported to DfT as they occurred. A sample is given below.

16 December 2010
2.9 Vehicle location at time of accident
To differentiate between types of lay-by, STATS 20 might better refer to “contiguous lay-by”. 2.7 Manoeuvres might perfectly well be coded 08 or 10 if the lay-by is of another type.

31 December 2010
STATS 20
3. Vehicles to be reported
Wording to 3.1 (e) might usefully include for untraced vehicles.

13 January 2011
STATS 20
Spelling mistake.
Previously existing code not removed.
Need to adopt a consistent use of “class” in respect of roads. Roads are classed by the Minister: trunk, principal other. They are divided: M, A, B, C and numbered. STATS 19 currently requires a road identity, consisting of division and number, NOT road class.

19 January 2011
STATS 21
Casually records coded both “front seat car passenger” AND seated PSV passenger. Consistency check required - or better planned form.

19 January 2011
STATS 19/20
**Intended** direction of vehicle in one direction. Accident happens during brief stop, as vehicle is reversed in lay-by, i.e. it is briefly travelling in the opposite direction. Further note (G) required at 2.8?

19 January 2011
STATS 19/20
Since the Highway Code states that drivers must park with flow, should the “parked” codes at 2.8 be removed and a reference to parked vehicles “facing” direction be added to Note E?

21 February 2011
STATS19/20
2.12 Hit Object in C’way
An initial misunderstanding led to the discovery of a few accidents in which a parked vehicle managed to hit an object in the carriageway.
18 May 2011
STATS 19/20
1.25 Carriageway Hazards
Discovered that, from 2005 onwards, there was an upsurge in the use of code 2, where the only "other object" was a pedestrian casualty.

15 June 2011
STATS 19/20
2.16 First points of contact
2.24 Hit and run
"Impossible" combinations discovered - first point of impact (V1) front + non-stop vehicle (V1), not hit and first point of impact (V1) none + hit and run yes.

11 July 2011
STATS 19/20
1.26 Did a Police Officer Attend?
It was discovered that (contrary to popular opinion) the %age of over the counter reports appears to have fallen, at least in the Falkirk Council area.

01 August 2011
STATS 19/20
2.12 Hit Object in C'way
Reporting officer chose code 11 "other object", despite having mentioned both “kerb” and “island” in the Accident Description. Consequently, 1.25 Carriageway Hazards could not be coded "0", although there was no “other unexpected object".
The problem here is hard to understand - we need to know why the reporting officer overlooked “kerb” AND “bollard/ refuge”. Consensus in the way the system is applied does not appear to have been achieved.

01 August 2011
2.13 Vehicle leaving carriageway
In the same accident as that referred to above, the officer could not sensibly code the situation where a vehicle ended up at a lamp column on a “splitter” island on the approach to a roundabout. He chose “left carriageway straight ahead at junction”, but this is excluded by NOTE 3. “Left carriageway offside onto central reservation” seems more appropriate, but it cannot be used, other than with a road type “dual carriageway”. This demonstrates a lack of exhaustiveness in the system.

03 August 2011
STATS 19/20
2.8 Vehicle movement compass points
Concern expressed about how these codes are being applied at very large roundabouts.

10 October 2011
STATS 19/20
1.20a and 1.20b Pedestrian Crossing - Human Control and Physical Facilities
The choice of code associated with these variables is not intuitive. Many reporting officers record the presence of crossing facilities whether or not they were actually in use, or contributed to the accident. Recording their presence is the intuitive thing; whether or not they were in use ought, perhaps, to be a separate question.
In the course of checking data relevant to a similar situation to that in the last EXAMPLE on p 32 of STATS 20, it was noted that the choices 1.20a: 2 - control by an authorised person and 1.20b: 5 - pedestrian phase at traffic signal junction should only be chosen IF a pedestrian had been involved (within 50m of the crossing) or if the authorised person or the crossing facility itself was considered a factor. This is not clear from the example.
23 November 2011
STATS 19/20
2.10 Junction Location of Vehicle

Some difficulty is being experienced in deciding when to use the entering/leaving codes. An explanatory note to the 1990 edition had been lost, apparently around 2005, and it is doubtful whether these codes are reliable.

30 November 2011
STATS 19/20
1.21 Light Conditions

The data suggest that a large number of accidents have occurred in 30 miles/hr areas where there are no street lights. This is nonsense, but may have resulted from the removal of any reference to street lighting columns. Such a reference (to their height) may previously have triggered the right response. It may, of course, be that the current legends are being misunderstood. Does the phrase, "street lights present" suggest that they ought to be LIT? The choice of a code further down the list, "no street lighting" suggests that officers, who may well be aware that the columns are there, simply choose "no street lighting" because street lights are not normally lit in daylight.

This seems to emphasise the need for "consensus" - i.e. every reporting officer must be in agreement with every other reporting officer as to what the legends mean.

There are many more such problems - and some have been addressed.

Most could be avoided if, as has been said many times before, the STATS19 system was MEE, mutually exclusive and exhaustive. But what is also clear is that there must be agreement as to what the legends associated with the codes mean.

It seems extraordinary that in other arenas such as that of the air, rail and nuclear industries - none of which kill and injure anything like as many people as the road one does, safety systems have been developed which are MEE and have been subjected to the consensus or reliability testing which is now a fundamental requirement of any such system.

It is encouraging to those who have been concerned about the state of STATS 19 since the 1990s, to discover that they may not be alone, after all. The following abstract confirms that the problem is recognised elsewhere:

"Many different classification schemes have been used in the analysis of road traffic accidents but the agreement between coders using the same classification scheme is rarely tested and/or reported. As a high intercoder agreement is a prerequisite for a study's validity, this is a serious shortcoming. The aim of the present study was, therefore, to test the intercoder agreement of the Driving Reliability and Error Analysis Method (DREAM) version 3.0 by letting seven coders from different European countries analyse and classify the causes of the same four accident scenarios. The results showed that the intercoder agreement for genotypes (contributing factors) ranges from 74% to 94% with an average of 83%, while for phenotypes (observable effects) it ranges from 57% to 100% with an average of 78%. The results also showed that weaknesses in classification schemes, methods, training of coders as well as in presentation of accident information can be identified by testing the intercoder agreement."

It is to be hoped that the UK STASTS 19 system will at last begin to catch up.

2012-05-09

Andrew Fraser
CoSLA representative at SCRAS/Senior Accident Investigation Officer
Falkirk Council
FALKIRK FK2 7YZ
E: andrew.fraser@falkirk.gov.uk
T: 01324 50 4931

2012-05-09

Andrew Fraser
CoSLA representative at SCRAS/Senior Accident Investigation Officer
Falkirk Council
FALKIRK FK2 7YZ
E: andrew.fraser@falkirk.gov.uk
T: 01324 50 4931