Garda Síochána Analysis Service

Measuring the recording error and over-recording of breath tests on MAT/MIT checkpoints on PULSE

Sampling methodology and findings

GSAS_V4_FINAL

Produced by: Garda Síochána Analysis Service
Date: 27/07/2017

Contents

1. Introduction

As part of the MAT/MIT examination by An Garda Síochána, exploratory analysis of MAT/MIT incidents on PULSE was conducted. This revealed recording errors for some of the MAT/MIT checkpoint records reviewed. For example, breath test figures mentioned in the incident narrative did not match those recorded under the "MIT Statistics" tab, or the number of vehicles stopped and controlled was higher than the total number of vehicles recorded as having passed through a checkpoint. In addition, the number of breath tests recorded on some checkpoints appeared too high when taking into account the duration and the number of members at the checkpoint. To estimate the scale of the recording error and over-recording of breath tests across the entire number of MAT/MIT checkpoints on PULSE, we drew a random sample of 2,136 checkpoint records and based our conclusions on the sample results, applying a 95% confidence interval and 3% margin of error.

The rest of the report outlines the sampling methodology and findings.

2. Sampling methodology

The examination period for the enquiry was between 7 June 2009[1] and 10 April 2017; however, the sampling frame included MAT/MIT checkpoints on PULSE which occurred between 1 July 2010 and 10 April 2017 (correct as at 12 April 2017). The later start date for the sampling frame was selected to only include MAT/MIT checkpoints recorded on PULSE within the last 7 years, as this is how long the Garda Information Services Centre (GISC) keeps

[1] MAT (Mandatory Alcohol Testing) checkpoints officially started being recorded on PULSE from 7 June 2009. However, a number of checkpoints have an "occurred" date prior to this date. This is because they were entered on PULSE retrospectively. MAT checkpoints were replaced with MIT (Mandatory Intoxication Testing) checkpoints on 4 December 2016 and now include testing for consumption of drugs as well as alcohol.

incident voice recordings. The ability to listen to the recordings for a number of the sampled checkpoints assisted in the verification of errors and provided additional insight into the possible recording issues. As the planned fieldwork visit to GISC was scheduled for the end of June/start of July, we limited the sample frame to the last 7 years from the planned visit.

The sampling frame consisted of **502,730** MAT/MIT checkpoints, supplied by the Information Analysis Service (IAS).

We split the sampling frame into two groups or strata:

1) checkpoints that appeared correctly recorded, on the face of it (**77%** of the total); and
2) checkpoints that appeared to have recording errors (**23%** of the total).

The initial call on whether a checkpoint was recorded correctly or incorrectly was made by conducting logic checks in Excel. The tests checked that the sum of negative and positive breath tests did not exceed or equal the number of vehicles passing through a checkpoint. On the majority of checkpoints, the pre-determined order of selecting motorists, which is specified prior to the operation of checkpoints, is to stop vehicles at random[2]. By this logic, there should be more vehicles passing through a checkpoint than vehicles stopped and controlled/people breath tested. Therefore, checkpoints where this was not the case were flagged as potentially containing recording errors.

[2] Based on the information supplied by the examination team and exploratory analysis.

The following formula was used to determine the required sample size for each group (appeared correct or appeared incorrect), using a 95% confidence interval and 3% margin of error:

$$ME = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where

- *ME* is the desired margin of error[3]; 3% in our case;
- *z* is the z-score. As our selected confidence interval[4] was 95%, the *z* value for this was 1.96;
- $\hat{p}$ is our judgement of the correct value of the proportion in the population. As we didn't know what proportion of the overall MAT/MIT checkpoint records on PULSE were recorded incorrectly or potentially over-recorded the number of breath tests, we took $\hat{p}$ to equal 0.5 to give the most conservative sample;
- *n i*s the sample size (to be found).

Based on this formula, the required sample size for each group was **1,068**. We chose to draw an equal number of checkpoint records from both groups, even though the group with potentially erroneous recordings based on the logic checks constituted only 23% of the overall number of MAT/MIT checkpoints in the sampling frame. This was to ensure that this group, with potentially more recording errors compared to the other group, was adequately represented. The results were weighted at the end to ensure that the findings based on the sample were representative of the overall number of MAT/MIT checkpoints in our sampling frame.

---

[3] The margin of error measures the maximum amount by which the sample results are expected to differ from those of the actual population. In our case, it measures the difference between the level of recording error and over-recording of breath tests on MAT/MIT checkpoints in the sample versus MAT/MIT checkpoints in the sampling frame.

[4] A 95% confidence interval is the most commonly used interval. It tells how confident one can be about their sample estimates. That is, if we used the same method to select different samples of MAT/MIT checkpoints and then computed a proportion of incorrectly recorded checkpoints for each sample, we would expect the true proportion to fall within the interval estimates 95% of the time.

To reflect the variability of potentially erroneous/valid recordings over time, the sample was drawn proportionally to the number of such recordings each year. As MAT/MIT checkpoints on PULSE are generally recorded by GISC, there was no need to stratify the sample by division.

Table 1 shows a breakdown of potentially erroneous/valid recordings in the sampling frame over the years. It indicates that the number of erroneous recordings based on the logic checks reduced between 2010 and 2017, thus justifying proportional selection by year. **Table 2** shows potentially erroneous/valid recordings in the sampling frame as a proportion of the total number of checkpoints by year, while **Table 3** shows the number of checkpoints sampled from each year.

Each checkpoint was assigned a random number using Excel's RAND() function. The incidents were then sorted in an ascending order by group, that is, by potentially erroneous/valid recordings (based on the logic checks), and random number. Finally, the number of checkpoint records was selected for each group as per Table 3[5].

[5] The sample for each year was drawn from the individual yearly files.

Table 1: Potentially erroneous/valid recordings in the sampling frame by year, 2010 - 2017

	Total 2010-2017*		2010*		2011		2012		2013		2014		2015		2016		2017*	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
No. of MAT/MIT checkpoints	**502,730**	**100%**	**31,927**	**100%**	**70,211**	**100%**	**71,024**	**100%**	**77,637**	**100%**	**76,673**	**100%**	**75,138**	**100%**	**76,551**	**100%**	**23,569**	**100%**
Potentially erroneous recordings	**116,749**	**23%**	10,191	32%	21,772	31%	19,368	27%	17,392	22%	15,818	21%	14,479	19%	13,520	18%	4,209	18%
Potentially valid recordings	**385,981**	**77%**	21,736	68%	48,439	69%	51,656	73%	60,245	78%	60,855	79%	60,659	81%	63,031	82%	19,360	82%

**2010 starts from 1 July; 2017 includes data up to 10 April.*

Table 2: Potentially erroneous/valid recordings in the sampling frame – proportion of the total, 2010 - 2017

	Total 2010-2017*	2010*	2011	2012	2013	2014	2015	2016	2017*	Total
Potentially erroneous recordings - % of the total	116,749	9%	19%	17%	15%	14%	12%	12%	4%	**100%****
Potentially valid recordings - % of the total	385,981	6%	13%	13%	16%	16%	16%	16%	5%	**100%****

**2010 starts from 1 July; 2017 includes data up to 10 April.*
*** The figures do not add up to 100% due to rounding.*

Table 3: Sample breakdown by year

Year	Potentially erroneous recordings	Potentially valid recordings	Total
Sample drawn	**1,068**	**1,068**	**2,136**
2010	93	60	153
2011	199	134	333
2012	177	143	320
2013	159	167	326
2014	145	168	313
2015	132	168	300
2016	124	174	298
2017	39	54	93

3. Checkpoint coding

The checkpoints sampled were reviewed and coded manually. Each checkpoint was coded as correctly recorded or not, based on the information on PULSE. For any checkpoint coded as incorrect, the reason for it being incorrect was provided. The following were the reasons why a checkpoint was treated as incorrectly recorded:

- The number of vehicles through a checkpoint was equal to the number of vehicles stopped and controlled, when random selection was specified at the outset[6]. This rule was only applied to the checkpoints with 4 or more vehicles passing through. Where the number of vehicles through a checkpoint was 3 or fewer, the incident was coded as correct, as the number of vehicles was perhaps too small to be stopped at random.
- The pre-determined order of vehicle selection was to stop every vehicle. However, the number of vehicles through a checkpoint was different from the number of vehicles stopped and controlled.
- The number of vehicles stopped and controlled was higher than the number through a checkpoint.
- Typographical error in the negative/positive/failed/refused breath test fields - extra zero. By default, fields for roadside breath tests under the

[6] The authorising officer specifies the method of selecting motorists to be stopped prior to the operation of checkpoints.

"MIT Statistics" tab on PULSE have zeros pre-filled. When entering the figures in these fields, a member or GISC call-taker has to delete the zero. However, there are instances where zeros are not deleted causing inflation of the number of breath tests recorded. Generally, this error is easy to spot based on other figures entered under the "MIT Statistics" tab. For example, the number of vehicles stopped and controlled entered is 2 and the number of negative breath tests is 20.

- The checkpoint should have been invalidated on PULSE as, for various reasons, it hadn't been conducted, but this invalidation did not occur[7]. For example, members were called to assist with other matters or there was no Dräger device available.
- A checkpoint was conducted but no figures were entered on PULSE under the "MIT Statistics" tab.
- Figures in the narrative did not match those recorded under the "MIT Statistics" tab.
- Other reasons (vehicles through a checkpoint or vehicles stopped and controlled box was not filled in, vehicle selection method not stated, authorised by tab details not filled in, or Dräger count not recorded in the designated location).

The sample was also checked for potential over-recording of breath tests. We assumed that it takes, on average, 4 minutes for one member to administer a breath test (based on the report from an experienced traffic sergeant). This is from the time a vehicle is stopped, spoken to by the member, documents inspected, the test conducted and the result determined. Checkpoint start and end times are entered on PULSE, which allowed us to determine the overall duration of each checkpoint. Thus any checkpoints that had more breath tests recorded than was feasible to do within the time period and based on the manpower available (that is, the number of members at the checkpoint), using the 4 minute per breath test assumption, were deemed as over-recorded.

[7] Checkpoints can be entered on PULSE once they have been authorised.

4. Findings

The checkpoints containing recording errors and over-recorded breath tests were weighted to ensure that the findings were representative of the MAT/MIT checkpoints in the sampling frame. **Table 4** shows the weights that were applied to the two groups.

Table 4: Sample weights

Sample	No. of incidents sampling frame	No. of incidents sample	Proportion sampling frame	Sample proportion	Weight applied
Potentially erroneous recordings	116,749	1,068	0.23	0.5	0.46
Potentially valid recordings	385,981	1,068	0.77	0.5	1.54
Total	**502,730**	**2,136**	**1**	**1**	

According to the sample, 10% of the MAT/MIT checkpoint records had recording errors (**Table 5**) and 6% over-recorded the number of breath tests conducted, based on the 4 minute per breath test assumption (**Table 6**). These figures are not mutually exclusive. Extrapolating these proportions to the overall number of MAT/MIT checkpoints in the sampling frame translates into between **7%** and **13%** of all checkpoints containing recording errors (between **35,191** and **65,355** checkpoints). Similarly, between **3%** and **9%** of all MAT/MIT checkpoints in the sampling frame (between **15,082** and **45,246** checkpoints) are estimated to over-record the number of breath tests conducted, based on the 4 minute assumption. This is based on the 3% margin of error and 95% confidence interval.

It is difficult to calculate the number of breath tests that were over-recorded. Based on the sample, the average number of over-recorded breath tests per

checkpoint was 7.04[8]. Therefore, we can estimate that between **106,177** and **318,530** breath tests in the sampling frame are potentially over-recorded.

Table 5: Incorrect recordings, weighted

Is the checkpoint correctly recorded?	Potentially erroneous recordings	Potentially valid recordings	Overall
Total	**100%**	**100%**	**100%**
No	31%	3%	10%
Yes	69%	97%	90%
Base (weighted)	**496**	**1,640**	**2,136**

Table 6: Over-recorded breath tests, weighted

Is the number of breath tests higher than expected?	Potentially erroneous recordings	Potentially valid recordings	Overall
Total	**100%**	**100%**	**100%**
No	91%	95%	94%
Yes	9%	5%	6%
Base (weighted)	**496**	**1,640**	**2,136**

In terms of the recording errors, the most common error (occurred on 68% of all checkpoints coded as incorrectly recorded) (**Table 7**) was related to the fact that a random method of stopping vehicles had been specified but all vehicles passing through a checkpoint were stopped.

[8] There were 150 checkpoints (unweighted) in the sample which over-recorded breath tests based on the 4 minute per breath test assumption. The total number of breath tests over the expected number was 1,056. This translates into an average of 7.04 breath tests per checkpoint.

Table 7: Reasons checkpoints were coded as incorrect, weighted

Reasons coded incorrect	Potentially erroneous recordings	Potentially valid recordings	Overall
Number of vehicles through a checkpoint equals vehicles stopped and controlled, when random design specified at the outset	75%	47%	68%
Checkpoint should've been invalidated	16%	3%	13%
Figures in the narrative don't match those under the "MIT Statistics" tab	2%	19%	7%
Checkpoint conducted - no figures entered under the “MIT Statistics” tab	2%	17%	6%
Number of vehicles stopped and controlled higher than vehicles through a checkpoint	2%	0%	2%
Random vehicles stopped when every vehicle to be stopped specified at the outset	0%	6%	1%
Typo in the breath test fields - extra zero	1%	0%	0%
Other*	1%	8%	3%
Base (weighted)	**152**	**55**	**208**

**Other category includes vehicles through a checkpoint or vehicles stopped and controlled box not filled in, vehicle selection method not stated, authorised by tab details not filled in and Dräger count not recorded in the designated location reasons.*

While not adhering to the pre-determined order of selecting motorists does not mean that the rest of the figures are incorrect, it nevertheless sheds doubt on their reliability. That is, it is not clear whether members really had stopped all the vehicles passing through the checkpoint or whether there was a misunderstanding of what information should have been recorded in the vehicles stopped and controlled and other fields under the “MIT Statistics” tab on PULSE. Based on the feedback received as part of this examination, there was confusion, at least among some of the members, in relation to these fields, which made us question the overall reliability and validity of the information recorded under the “MIT Statistics” tab on PULSE.