An Investigation into the accuracy of follow-on GPRS/Mobile Data CDRs

Abstract

This study investigated the accuracy of 3G and 4G follow-on GPRS (General Packet Radio Service)/mobile data CDRs (Call Detail Records) from three UK mobile network operators (EE, Vodafone and Three). Follow-on GPRS/mobile data CDRs are currently considered to be more open to misinterpretation than voice/SMS CDRs as uncertainties exist regarding the correspondence between the timestamp and the Cell ID presented within the CDRs. Consequently, follow-on GPRS/mobile CDRs may be disregarded during criminal investigations, potentially losing valuable intelligence and evidence. To assess the accuracy of follow-on GPRS/mobile data CDRs, connected mode RF (Radio Frequency) surveys were conducted while simultaneously producing follow-on GPRS/mobile data CDRs in a travelling vehicle. This allowed a comparison of the start Cell ID presented in the CDR and the Cell ID that provided coverage to the device at the start time of the CDR to assess the correspondence between the timestamp and the Cell ID presented within the CDRs, and to consider the validity of the terminology used by experts. It was found that individual follow-on GPRS/mobile data CDRs cannot consistently place a device within the coverage area of the start Cell ID at the start time of the CDR. Instead, the results indicate that a terminology which places the device within the coverage area of the start Cell ID 'at or before' the start time of the CDR is appropriate. It is crucial that follow-on GPRS/mobile data CDRs are analysed with this consideration in mind so to interpret the evidence correctly.

Keywords: cell site analysis, CDR, GPRS, mobile data

1. Introduction

As societies increasingly rely on technology there has been a concomitant increase in the use of digital devices during criminal activities [1], and it has been estimated that over 80% of crimes have a digital evidence component [2]. Therefore, there are potentially large amounts of digital data from mobile phones which can aid the investigation of a wide range of different crimes. Law enforcement may utilise information presented within the CDRs (Call Detail Records) related to a specific mobile subscription to provide timestamped intelligence relating to the geographical area within which a device could have been during a crime-related event, known as cell site analysis. The Forensic Science Regulator has emphasised that "there are inherent uncertainties within cell site analysis no matter which methods have been applied" [3, p.19]. These uncertainties are heightened with follow-on GPRS (General Packet Radio Service)/mobile data CDRs, as uncertainties exist regarding the correspondence between the timestamp and the Cell ID presented within the CDRs [4]. The substantial increased use of IM (Instant Messaging) applications has meant cell site analysts are increasingly required to interpret and base conclusions on GPRS/mobile data CDRs [4]. Very little published research has been undertaken to test these uncertainties. This is an issue, as it may lead to incorrect interpretations being formed or important data being disregarded due to concerns about its reliability, and thus its admissibility, in court.

Cell site analysis is an investigative technique used to provide evidence regarding the movements of a specific mobile subscription [5]. It often involves the comparison of the data captured within CDRs generated by Mobile Network Operators (MNOs) and the results of Radio Frequency (RF) surveys, with subsequent presentation of the data in court [4]. Each time a device is used to make or receive a phone call, text message or uses mobile data, a CDR is created within the billing record for that mobile subscription [5]. It should be noted that the term 'CDR' not only refers to the individual lines of data produced when a device is used to make or receive a phone call, text message or uses mobile data, but also to the overall document containing individual CDR lines. CDRs capture a variety of information, including the date/time the call was made or received, the length of the call and the start Cell ID, relating to the cell site used to make or terminate the exchange. RF surveys are used to indicate which cells provide service at particular locations [6]. RF surveys can be conducted in idle mode (where the device is available but not being used to make calls, send text messages, or use mobile data) or connected mode (where the device forms a radio connection with the serving cell to exchange calls and text messages or to use mobile data) [5]. Comparison of start Cell IDs captured within CDRs and the results of RF surveys can allow cell site analysts to define an area within which a mobile phone may have been present whilst in use [5], which, for criminal investigations, could help support or undermine a suspect's alibi [7].

IM applications have increased in popularity in recent years, and at the same time there has been a decrease in the quantity of 'traditional' text messages being sent [8]. Therefore, analysis of GPRS/mobile data CDRs provides, in theory, a useful supplementary tool for criminal investigations and an additional source of intelligence/evidence. However, follow-on GPRS/mobile data CDRs are currently considered inferior to voice/SMS CDRs due to uncertainties regarding the correspondence between the timestamp and the Cell ID presented within the CDRs [4].

When a device uses internet services, a new 'data session' is established, triggering a CDR to open. It is not practical for the network to produce a single CDR during the course of the data session, instead, a series of 'follow-on' CDRs are produced, each being generated by certain triggers such as a prolonged period of user inactivity, a data volume limit being hit, or a certain time limit being reached [5]. Each of these follow-on CDRs contain a start Cell ID, which can, in theory, provide useful evidence regarding the movements of the device. However, the core network is not necessarily updated with the Cell ID which the device was using at that time. In reality, the network may regard the most recently reported Cell ID as the one it is currently using, which can be out of date by "minutes or even hours" [4, p.26]. This lag in reporting occurs as the start Cell ID may only be updated when the device moves from one Location Area Code (LAC) or Tracking Area Code (TAC)

to another or during a change of technology (i.e., from 3G to 4G) [5]. The start Cell ID documented within the first CDR of a data session can be seen as accurate [5], therefore this uncertainty only relates to the subsequent follow-on CDRs within the same data session.

Therefore, the terminology used by cell site analysts must reflect the complexities observed with follow-on GPRS/mobile data CDRs to ensure they are interpreted correctly. In the absence of a universal, clearly defined industry standard for reporting, the terminology recommended by Forensic Analytics Ltd will used for the purposes of this paper. The terminology recommended by Forensic Analytics Ltd [4, p.14] is:

"the subject device was in the coverage area of Cell ID 12345 AT OR BEFORE the start time of the CDR and after the start time of the last consecutive record that had a different Cell ID".

In the statement above, the term 'coverage area' is defined as the area to which the Cell ID documented in the CDR is the serving cell. The 'before' term must be included within the terminology due to the lag in reporting of the most recently used Cell ID. However, this 'before' time may be constrained to the start time of the last consecutive record which had a different Cell ID, as the core network must have been updated within this time period for a new Cell ID to be displayed in the most recent CDR [4]. It should be noted that for the MNOs O2 and Three, the terminology is slightly different. O2 record the end Cell ID rather than the start Cell ID in their CDRs, meaning that the terminology recommended by Forensic Analytics Ltd which will be used for the purposes of this paper [4, p.16] is:

"the subject device was in the coverage area of Cell ID 12345 AT OF BEFORE the END time of the CDR and after the start time of the last consecutive record that had the same Cell ID".

GPRS/mobile data CDRs for Three don't currently display an end time, meaning it is not currently possible to determine whether successive CDRs are truly follow-on CDRs. Therefore, the 'before' time cannot be constrained, meaning the terminology recommended by Forensic Analytics Ltd which will be used for the purposes of this paper [4, p. 16] is:

"the subject device was in the coverage area of Cell ID 12345 AT OR BEFORE the START time of the CDR".

It is clear that there is a requirement to rigorously test the reliability of follow-on GPRS/mobile data CDRs so that assessments can be made regarding the validity of the terminology used by cell site experts to ensure correct conclusions are made based on the evidence. A study was designed to address the following question: To what extent is the start Cell ID documented in follow-on GPRS/mobile data CDRs an accurate record of device activity for each MNO and technology?

2. Materials and Methods

2.1. Materials

In this study, two Samsung Galaxy S9 phones equipped with the RF surveying software 'Nemo Handy' were used. Two phones were necessary so that 3G and 4G technologies for each network could be tested simultaneously. Two 'pay-as-you-go' SIM cards for three MNOs (EE, Three and Vodafone) were acquired. The data was processed and analysed using CSAS (Cell Site Analysis Suite) (v2.6.2) and CSAS RF Survey (v1.4.1.14).

2.2. Selected MNOs and Technologies

This research considered three MNOs (EE, Vodafone and Three), and two technologies (3G and 4G), resulting in six MNO and technology combinations (Table 1). However, a number of issues were discovered on initial analysis of the data, meaning no analysis could take place for Three 3G and Vodafone 3G (the cause of these issues is discussed later in this section). This research did not investigate 2G as it has been shown that 4G and 3G carry the majority of data traffic, at approximately 90% and 10% respectively [9]. In order to allow identical data processing and analysis procedures to take place for each MNO in the study, O2 was not surveyed due to the differences that exist in the way the CDR data is presented by O2 compared to the other MNOs.

Table 1: Summary table showing the technologies and networks surveyed in this study.

Technology/Network Combinations
3G/EE
4G/EE
3G/Three
4G/Three
3G/Vodafone
4G/Vodafone

2.3. Data Collection Method

Connected mode RF surveys were conducted using the RF surveying software 'Nemo Handy' along a predetermined route to indicate which cells would, hypothetically, be selected by a phone to use mobile data. Simultaneously, each phone was used to stream live TV in order to generate a series of follow-on GPRS/mobile data CDRs. Live TV, opposed to live radio, was streamed as it uses higher volumes of data, thus increasing the likelihood of hitting data volume limits to trigger the generation of follow-on CDRs.

A circular route of approximately 110 miles was chosen, which took approximately two hours to complete by car. This route was chosen as it took place on motorways/A roads, instead of B roads, as it has been identified that 4G coverage within vehicles in the UK is available on 62% of motorways/A roads, and 46% of B roads [9]. A constant speed of 50-60 mph was maintained throughout the route, as driving at the slowest speed that is practically possible and safe means the RF surveying equipment captures sufficient data regarding the Cell ID at certain locations [5].

Data collection took place on weekdays as opposed to weekends as it was assumed that there would be similar levels of traffic and cell usage on weekdays, which may potentially fluctuate on weekends. The two devices were used to conduct 3G and 4G RF surveys simultaneously, and three repeats were carried out for each technology and MNO combination (see Table 1). The CDRs for the phone numbers of each testing device were obtained from the MNOs, and the RF survey results were extracted directly from the devices and uploaded into CSAS for analysis.

2.4. Data Processing and Analysis

Two data sets were obtained for each technology and MNO combination; the RF survey results and the CDRs. The CDRs were processed using CSAS and exported into Excel, then the data was cleaned using the following cleaning criteria:

- All CDR lines with start times before (with the exception of the last consecutive CDR line prior to the start) and after the specific data collection period for that repeat were deleted. The 'data collection period' is defined as any time outside of the bounds of the times recorded within the RF Survey for that specific repeat.
- All CDR lines associated with events other than GPRS/mobile data events were deleted, such as incoming/outgoing text messages.

The RF surveys were uploaded into CSAS RF Survey, which allowed the data to be visualised as a map or a dataset, showing the date, time, latitude, longitude, network, technology, and the Cell ID measured at one second intervals.

Each CDR list was analysed to determine whether they were truly follow-on records from the same data session. Vodafone provides an additional field in their CDRs called the 'Charging ID', where all records produced within the same data session will contain the same Charging ID [4]. For the other MNOs who do not provide a Charging ID in their CDRs, the CDR lists were analysed to determine whether they were follow-on records within the same data session through subtracting the start time of a CDR with the end time of the last consecutive CDR. A suggested threshold for the maximum interval between successive CDRs is 10 seconds, with any interval above this representing the start of a new data session [4]. The MNO 'Three' doesn't display the end time, therefore it couldn't be ascertained whether successive CDRs for Three are follow-on records within the same data session.

Next, each of the three repeats for each network/technology were analysed individually to determine the extent to which the information displayed in the CDRs were an accurate record of device activity, which was done via the procedure shown in Figure 1. A tolerance of ± 5 seconds was used either side of the CDR start time, which was used as it was unknown whether there may be a slight lag between the time recorded by the device and the time used by the MNOs.

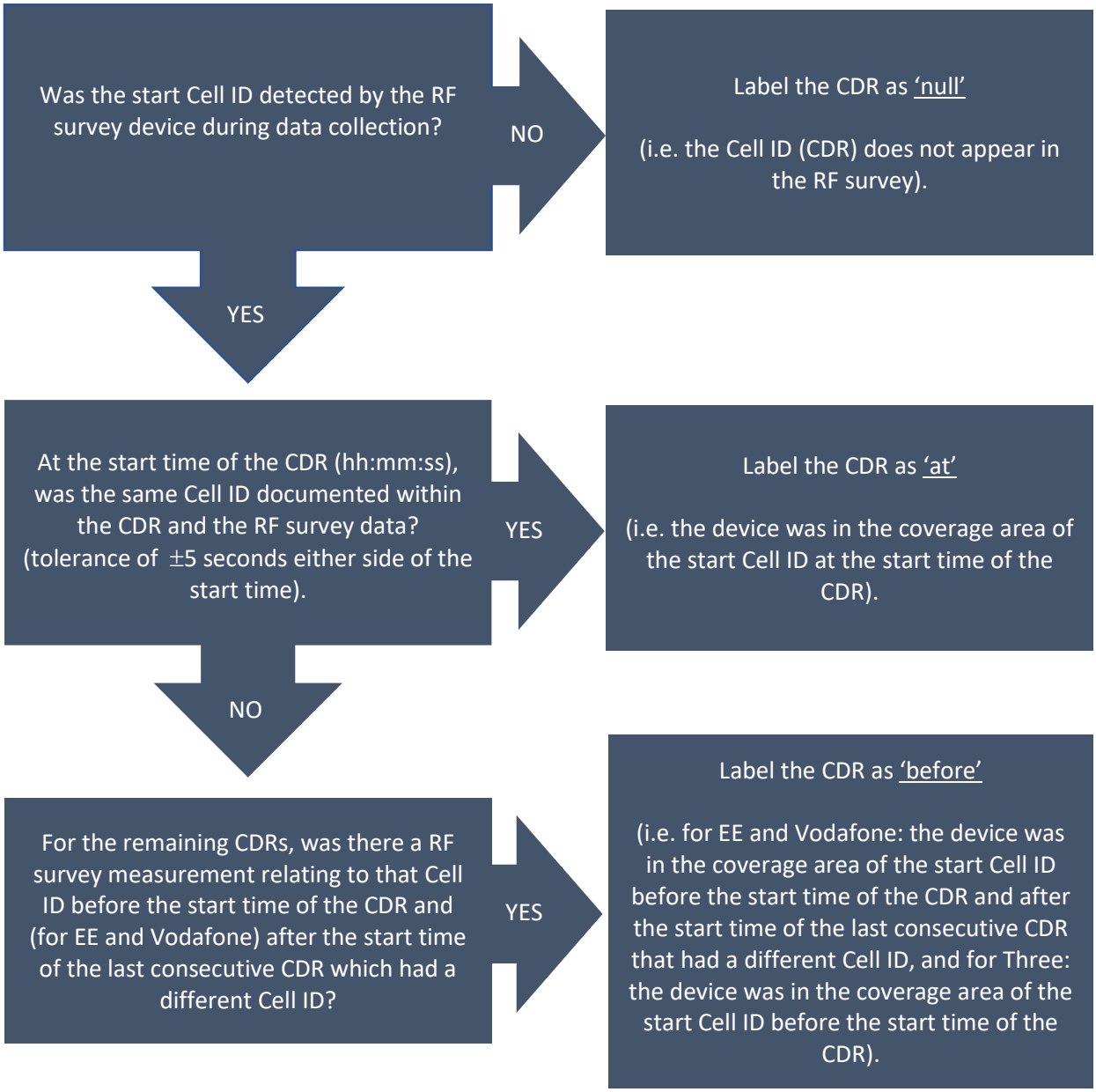

Figure 1: Data analysis scheme for the comparison of the CDR the RF survey data.

The extent to which the start Cell ID for each CDR was an accurate record of device activity are presented as bar charts, using the three categories; 'before', 'at' and 'null' (defined in Table 2). For the CDRs labelled as 'before', further analysis took place to determine the time elapsed between the start time of each CDR labelled as 'before' and the most recent RF survey data point associated with the Cell ID recorded in the CDR, which signified the last time the device was in the coverage area of the Cell ID recorded in the CDR. Finally, to investigate inconsistencies within the EE 3G/4G data, LAC/TAC maps were produced within CSAS to show the LAC/TAC boundaries along the route.

Table 2: Coded labels and meanings for the data presented within the bar graphs for this study.

Coded label		Meaning
At		Device was in the coverage area of the start Cell ID at the start time of the CDR (±5 s).
Before	EE and Vodafone	Device was in the coverage area of the start Cell ID before the start time of the CDR and after the time of the last consecutive record had a different Cell ID.
	Three	Device was in the coverage area of the start Cell ID before the start time of the CDR.
Null		Cell ID (CDR) does not appear in RF survey.

2.5. Data Issues

A number of issues were discovered upon initial analysis of the data. First, the 3G CDRs for Three returned with vacant Cell ID columns. The cause of this was unknown, so for this study no analysis could take place for 3G Three.

Second, the RF survey device did not record a serving cell for sections of the Vodafone 3G route, which meant an incomplete data set was produced. However, the device appeared to continually stream live TV during the route though retaining a 3G connection (as the device was locked onto the 3G network). To understand the cause of the inconsistent RF survey data further investigation is required, which is beyond the scope of this paper. Therefore, no analysis could take place for 3G Vodafone.

3. Results

CDRs and RF survey data obtained for each network and technology combination were analysed to determine the extent to which the start Cell ID documented in follow-on GPRS/mobile data CDRs were an accurate record of device activity. As explained previously, no analysis could take place for the 3G Vodafone or 3G Three data. The results are displayed as bar charts using the terminology outlined in Table 2.

3.1. EE Results

The number of CDRs produced for the EE 3G and 4G repeats and the maximum time interval between successive CDRs is presented in Table 3. The maximum time interval did not exceed 10 seconds for all repeats, therefore using the 10 second tolerance as discussed in section 2.4., all the CDRs were follow-on records within the same data session.

Table 3: Table displaying the number of CDRs and maximum time interval between successive CDRs produced for each repeat for the different technologies for EE.

Technology	Repeat No.	Number of CDRs	Maximum time interval between successive CDRs
3G	1	19	00:00:10
	2	20	00:00:08
	3	20	00:00:02
4G	1	31	00:00:08
	2	30	00:00:02
	3	33	00:00:08

3.1.1. EE 3G Results

The results for the 3G EE data collection are shown in Figure 2, using the coded labels in Table 2.

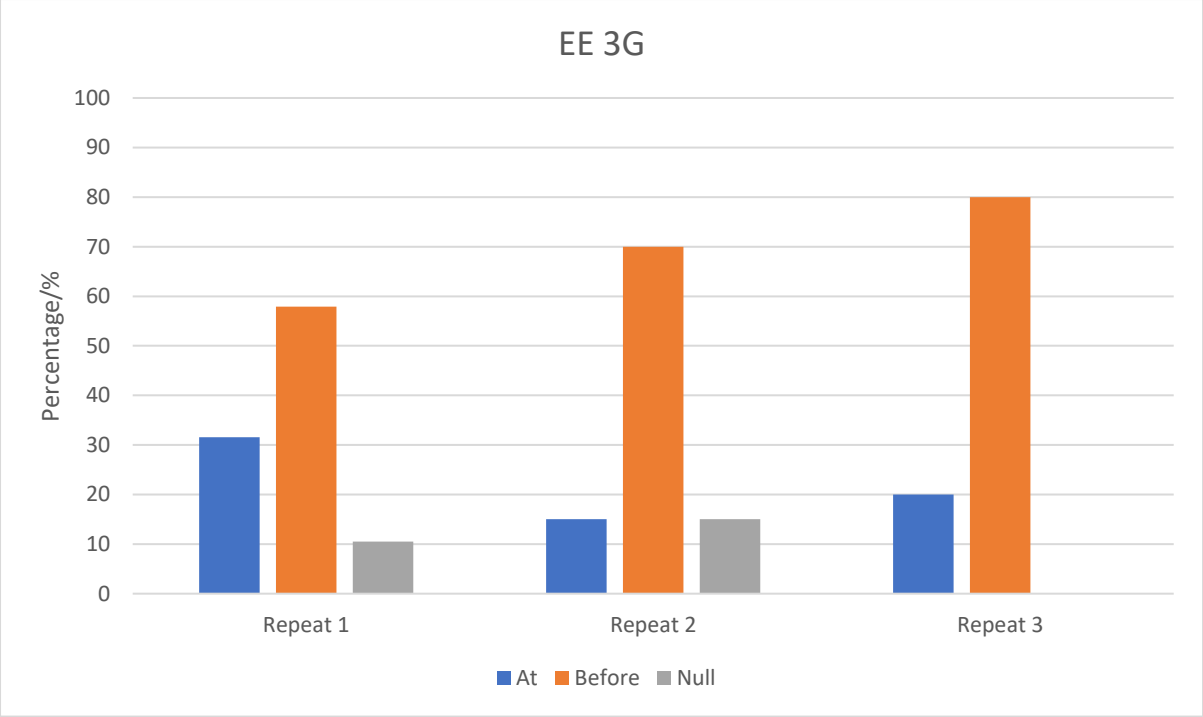

Figure 2: Graph showing the extent to which the start Cell ID displayed within the 3G EE CDRs were an accurate record of device activity over three repeats (repeat 1, n(CDR) = 19, repeat 2, n(CDR) = 20, repeat 3, n(CDR) = 20).

The results were further examined to determine the mean and standard deviation for the three criteria ('before', 'at' or 'null') over the three repeats (Table 4).

Table 4: The mean (rounded to nearest integer) and standard deviations of the 3G EE results.

Terminology	Mean/%	Standard deviation/%
Device was in the coverage area of the start Cell ID at the start time of the CDR (±5 s). ('At')	22	8.5
Device was in the coverage area of the start Cell ID before the start time of the CDR and after the time of the last consecutive record had a different Cell ID. ('Before')	69	11.07
Cell ID (CDR) does not appear in RF survey. ('Null')	9	7.7

A higher proportion of CDRs fell into the 'before' category than the 'at' category (mean values of 69% and 22% respectively) (Table 4). Relatively large standard deviations were observed for each category, highlighting the variation between the repeats. For repeats 1 and 2, some of the Cell IDs documented in the CDRs were not detected by the RF survey device, meaning no analysis could be completed for these CDRs.

Further analysis took place to investigate the CDRs labelled using the 'before' terminology. Table 5 show the difference in time elapsed between the start time of each CDR labelled as 'before' and the most recent RF survey data point associated with the Cell ID recorded in the CDR, which signifies the last time the device was in the coverage area of the Cell ID recorded in the CDR.

Table 5: Further analysis for the CDRs labelled with the 'before' criteria for EE 3G (a) repeat 1, (b) repeat 2, (c) repeat 3.

(a) Repeat 1		
CDR No.	**Cell ID**	**Δt**
1	57272	00:01:08
3	32935	00:05:11
7	11514	00:12:48
8	40009	00:00:12
11	4982	00:11:12
13	48471	00:10:09
14	48471	00:19:30
15	48471	00:25:51
16	48471	00:32:26
18	30312	00:06:23
19	57265	00:00:57

(b) Repeat 2		
CDR No.	**Cell ID**	**Δt**
1	57272	00:02:04
2	32935	00:02:31
5	11514	00:06:01
6	11514	00:14:20
7	11514	00:21:47
8	11564	00:01:24
9	11564	00:08:16
10	11564	00:14:21
11	11564	00:21:17
13	48471	00:04:39
14	48471	00:11:03
15	48471	00:17:49
16	48471	00:24:54
17	48471	00:31:27

(c) Repeat 3		
CDR No.	**Cell ID**	**Δt**
1	57272	00:03:47
2	57272	00:11:31
6	11514	00:05:13
7	11514	00:13:24
8	11514	00:21:47
9	5538	00:02:00
10	5538	00:09:18
11	5538	00:16:08
12	5538	00:22:44
14	48471	00:05:08
15	48471	00:12:02
16	48471	00:19:40
17	48471	00:27:22
18	48471	00:34:07
19	63182	00:02:09
20	48011	00:03:30

For many of the CDRs, the time elapsed between the moment the device was last in the coverage area of the Cell ID and the start time of the CDR was highly variable (Table 5). In some cases, the same Cell ID was documented over a series of successive CDRs, such as the CDRs which document 48471 as their start Cell IDs. These CDRs are shown to extend to over 30 minutes after the device was last in the coverage area of cell 48471.

To explore the cause of these inconsistencies, a map showing the different LACs encountered during the route was created (Figure 3) which shows that the device encountered five different LACs. Within individual LACs it seems that it is common for the same Cell ID to be reported in successive CDRs, which appears to update once the device had moved into a new LAC. Looking in particular at the CDRs which state cell 48471 as its start cell (CDR numbers (a) 13-16, (b) 13-17, (c) 14-18 in Table 5), through inspection of the CDR and RF survey results, it was discovered that each of these CDRs were generated whilst the device was within LAC 1134 (Figure 3). However, before the generation of the next consecutive CDR (i.e., CDR numbers (a) 17, (b) 18, (c) 19), the device had travelled into another LAC (1146), consequently updating the core network of the last used Cell ID. This is repeated within other successive CDRs (i.e. CDRs for (b) 5-7 and 8-11, (c) CDRs 6-8 and 9-12). The causes and implications of this lack of continual update of Cell ID is explored within section 4.2.

Figure 3: Map showing different LACs for the RF survey data collected for 3G EE repeat 1.

3.1.2. EE 4G Results

The results for the 4G EE data collection are shown in Figure 4, using the coded labels in Table 2.

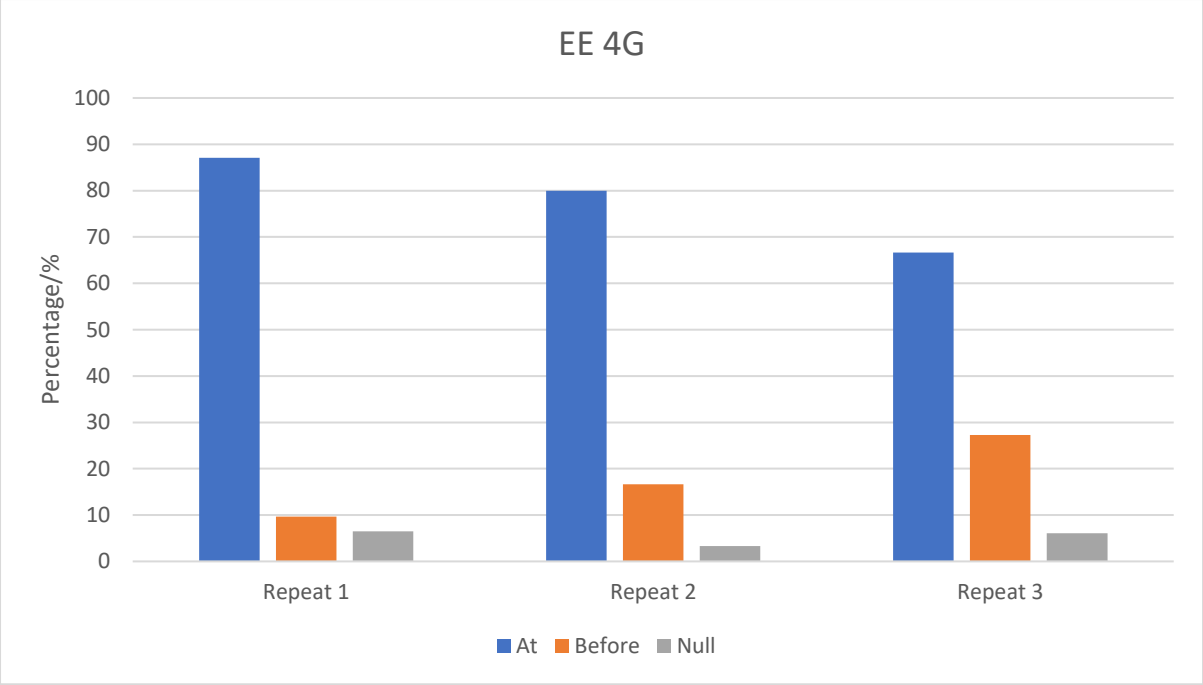

Figure 4: Graph showing the extent to which the start Cell ID displayed within the 4G EE CDRs were an accurate record of device activity over three repeats (repeat 1, n(CDR) = 31, repeat 2, n(CDR) = 30, repeat 3, n(CDR) = 33).

The results were further examined to determine the mean and standard deviation for the three criteria ('before', 'at' or 'null') over the three repeats (Table 6).

Table 6: The mean (rounded to nearest integer) and standard deviations of the 4G EE results.

Terminology	Mean/%	Standard deviation/%
Device was in the coverage area of the start Cell ID at the start time of the CDR (±5 s). ('At')	78	10.37
Device was in the coverage area of the start Cell ID before the start time of the CDR and after the time of the last consecutive record had a different Cell ID. ('Before')	18	8.86
Cell ID (CDR) does not appear in RF survey. ('Null')	4	1.61

Table 6 shows that the majority of the CDRs fell into the 'at' category compared to the 'before' category (mean values of 78% and 18% respectively). Relatively high standard deviations were observed for the 'at' and 'before' categories, highlighting variations between repeats. Similar to the 3G EE results, some of the Cell IDs documented in the 4G CDRs were not detected by the RF survey device, meaning no comparison could be made.

Further analysis took place to investigate the CDRs labelled using the 'before' terminology. Table 7 shows the difference in time elapsed between the start time of each CDR labelled as 'before' and the most recent RF survey data point associated with the Cell ID recorded in the CDR, which signifies the last time the device was in the coverage area of the Cell ID recorded in the CDR.

Table 7: Further analysis for the CDRs labelled with the 'before' criteria for EE 4G (a) repeat 1, (b) repeat 2, (c) repeat 3.

(a) Repeat 1		
CDR No.	**Cell ID**	**Δt**
2	4707328	00:00:30
18	7083520	00:00:25
21	8290562	00:01:24

(b) Repeat 2		
CDR No.	**Cell ID**	**Δt**
5	6741248	00:00:55
20	7198210	00:01:05
22	7974914	00:00:08
25	7553026	00:00:25
28	7243520	00:00:56

(c) Repeat 3		
CDR No.	**Cell ID**	**Δt**
9	8649473	00:00:09
14	6696194	00:00:14
15	5494529	00:00:13
16	4352256	00:00:18
17	4786434	00:00:22
19	5316354	00:00:38
21	3682306	00:00:17
24	7974914	00:00:09
26	7151872	00:00:10

Table 7 shows that the time elapsed between the moment the device was last in the coverage area of the Cell ID and the start time of the CDR was generally low, with the majority being under one minute. As for the 3G EE analysis, a map displaying the different TACs encountered during the course of the route was created (Figure 5) to explore why smaller time gaps were observed between the moment the device was in the coverage area of the Cell ID and the time the CDR was generated for the 4G EE data opposed to the 3G EE data. Figure 5 shows that more TAC boundaries were crossed during the 4G data collection than LAC boundaries for the 3G EE data collection, and the possible reasons for this are explored within section 4.3.

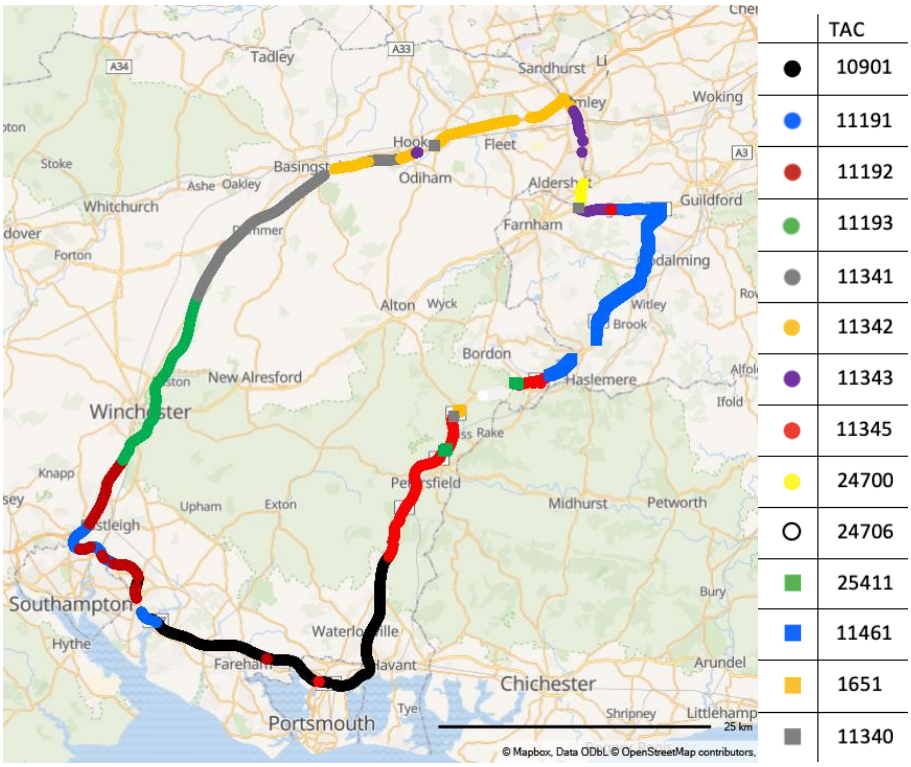

Figure 5: Map showing different TACs for the RF survey data collected for 4G EE repeat 1.

3.2. Vodafone Results

Table 8 shows the number of CDRs produced for the Vodafone 4G experiments, and the maximum time interval between successive CDRs. It is shown that the maximum time interval is 0 seconds, providing confidence that all the CDRs were truly follow-on records. Additionally, each CDR within each repeat contained the same Charging ID, which confirms that the CDRs were truly follow-on records.

Table 8: Table displaying the number of CDRs and maximum time interval between successive CDRs produced for each repeat for 4G Vodafone.

Technology	Repeat	Number of CDRs	Maximum time interval between successive CDRs
4G	1	34	00:00:00
	2	34	00:00:00
	3	34	00:00:00

3.2.1. Vodafone 4G Results

The results for the Vodafone 4G data collection are shown in Figure 6, using the coded labels in Table 2.

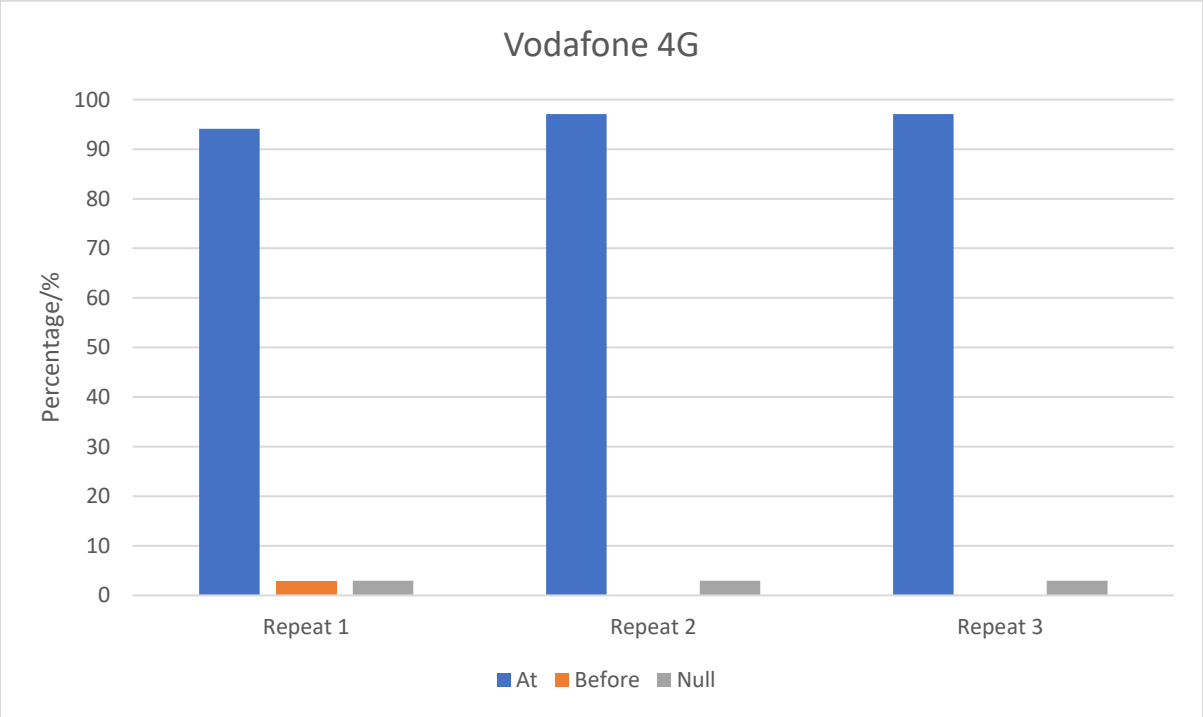

Figure 6: Graph showing the extent to which the start Cell ID displayed within the 4G Vodafone CDRs were an accurate record of device activity over three repeats (repeat 1, n(CDR) = 34, repeat 2, n(CDR) = 34, repeat 3, n(CDR) = 34).

The results were further examined to determine the mean and standard deviation for the three criteria ('before', 'at' or 'null') over the three repeats (Table 9).

Table 9: The mean (rounded to nearest integer) and standard deviations of the 4G Vodafone results.

Terminology	Mean/%	Standard deviation/%
Device was in the coverage area of the start Cell ID at the start time of the CDR (±5 s). ('At')	96	1.7
Device was in the coverage area of the start Cell ID before the start time of the CDR and after the time of the last consecutive record had a different Cell ID. ('Before')	1	1.7
Cell ID (CDR) does not appear in RF survey. ('Null')	3	0

Table 9 shows that the majority of the CDRs fell into the 'at' category (96%), meaning that the Cell ID documented in each CDR was generally an accurate record of device activity. The only exception to this is a CDR from repeat 1. This CDR had been labelled as 'before' as the device was within the coverage area of the Cell ID before the start time of the CDR. However, no RF survey measurements were made between 11:37:48 and 11:38:33, meaning no RF survey measurements were made within the ±5 second tolerance of the start time of the CDR (11:37:54). The implications relating to the absence of RF survey measurements within this time frame are explored in section 4.4. Small standard deviations were observed for all categories, showing repeatability between the repeats. Some of the Cell IDs documented within the CDRs were not detected by the RF survey device during data collection, meaning no comparison could be made for these CDRs.

3.3. Three Results

The differences in the way Three display information in their CDRs mean it is not possible to ascertain whether the successive CDRs are truly follow-on records. This has implications on the terminology used to discuss the CDRs, as the 'before' time cannot be constrained in the same way as with EE and Vodafone.

3.3.1. Three 4G Results

The results for the Three 4G data collection are shown in Figure 7, using the coded labels outlined in Table 2.

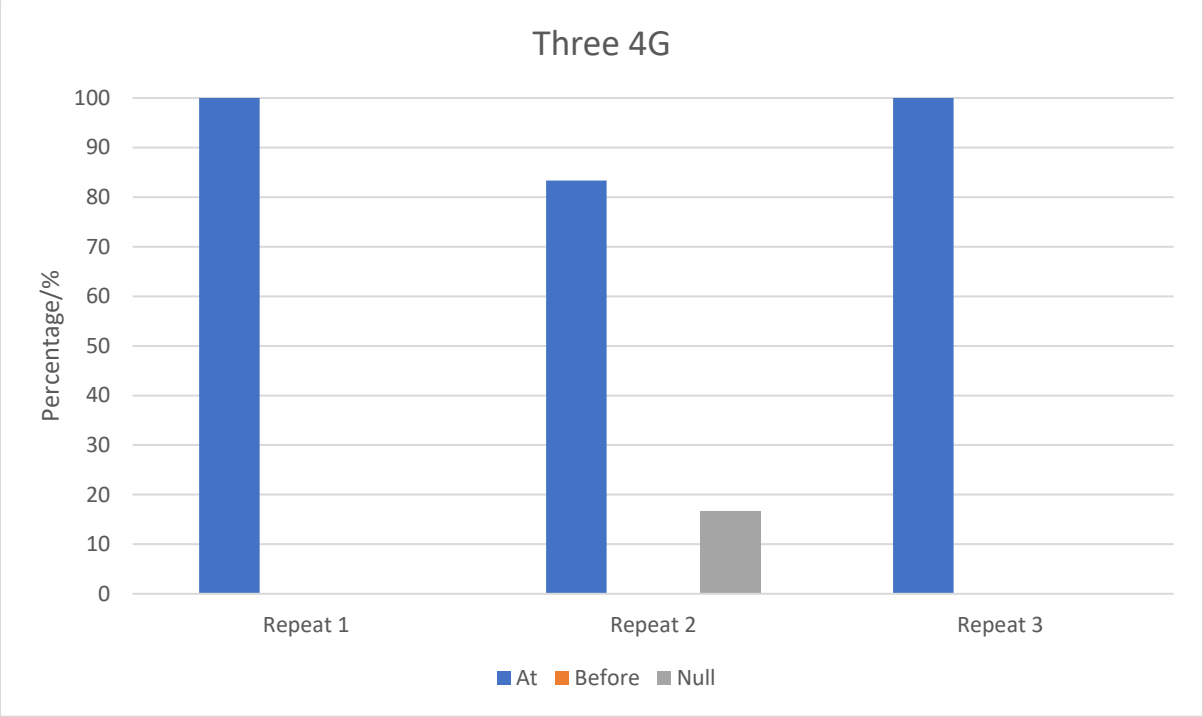

Figure 7: Graph showing the extent to which the start Cell ID displayed within the 4G Three CDRs were an accurate record of device activity over three repeats (repeat 1, n(CDR) = 8, repeat 2, n(CDR) = 6, repeat 3, n(CDR) = 9).

The results were further examined to determine the mean and standard deviation for the three criteria ('before', 'at' or 'null') over the three repeats (Table 10).

Table 10: The mean (rounded to nearest integer) and standard deviations of the 4G Three results.

Terminology	Mean/%	Standard deviation/%
Device was in the coverage area of the start Cell ID at the start time of the CDR (±5 s). ('At')	94	9.62
Device was in the coverage area of the start Cell ID before the start time of the CDR. ('Before')	0	0
Cell ID (CDR) does not appear in RF survey. ('Null')	6	9.62

All the CDRs which displayed start Cell IDs which were detected by the RF survey device were shown to be an accurate record of device activity (mean value of 94%) (Table 10). However, some of the Cell IDs documented in the CDRs were not detected by the RF survey device, meaning no comparison could be made for these CDRs.

4. Discussion

The aim of the study was to determine the extent to which the start Cell ID documented within follow-on GPRS/mobile data CDRs were an accurate record of device activity for the MNOs EE (3G and 4G), Vodafone (4G) and Three (4G). The results indicate that in all the tested scenarios it is not appropriate to solely use the 'at' terminology, as each network and technology combination presented instances where the device was in the coverage area of the start Cell ID before the start time of the CDR, or the analysis gave 'null' results.

4.1. Null Results

Each of the MNOs surveyed in this study gave 'null' results, meaning at least one of the Cell IDs documented within a CDR was not detected within the RF survey. This does not necessarily mean the device had not been within the coverage area of the Cell ID at or prior to its generation. A singular cell site doesn't usually provide the infrastructure for a singular cell; instead, it may hold multiple 'stacked' antennae relating to different Cell IDs [5]. Therefore, the RF survey may have failed to detect the Cell ID as an alternate cell may have been detected instead which was held on the same site. This reasoning is hypothetical; further research must take place to confirm whether the Cell IDs documented within the 'null' CDRs were located in proximity to the Cell ID identified within the RF survey. However, as this only accounts for 3-9% of CDRs within this study, conclusions may be drawn from the CDRs which documented a Cell ID which also appeared within the RF survey.

4.2. EE 3G

For EE 3G (n=3), there was a consistently higher proportion of instances where the device was in the coverage area of the start cell 'before' the start time of the CDR rather than 'at' the start time of the CDR (mean values of 69% and 22% respectively). Further inspection of the 'before' scenarios revealed that if the device was in the coverage area of the Cell ID before the start time of the CDR, the device had also been in the coverage area of the start cell after the time of the last consecutive record which had a different Cell ID. Constraining this 'before' time is beneficial as it allows placement of the device (and thus its user) within an area in proximity to the location the device was in at the start time of the CDR. However, as seen in Table 5(a-c), within certain scenarios in this study, extensive time periods were observed between the start time of the CDR and the moment the device was last in the coverage area of the associated Cell ID. When a device is travelling within a particular LAC, the core network is not necessarily updated with the cell which the device last used for mobile data, instead the core network is updated when the device moves from one LAC to another [5]. Therefore, these findings indicate that forming conclusions on the basis of multiple consecutive follow-on CDRs displaying the same Cell ID as the start Cell ID should be done with caution, as the device may not have been in the coverage area of that Cell ID for an extended period of time prior to the CDR being generated. However, in each of the 'before' instances, the device had been in the coverage area of the start cell after the time of the last consecutive record which recorded a different Cell ID, suggesting that the terminology*: the device was in the coverage area of the start Cell ID at or before the start time of the CDR and after the time of the last consecutive record which had a different Cell ID* is appropriate for use whilst discussing 3G EE follow-on CDRs.

4.3. EE 4G

For EE 4G (n=3), there was a consistently higher proportion of instances where the device was in the coverage area of the start cell 'at' the start time of the CDR rather than 'before' the start

time of the CDR (with mean values of 78% and 18% respectively). Further inspection of the 'before' scenarios revealed that, similarly to the 3G EE data, if the device was in the coverage area of the Cell ID before the start time of the CDR, the device had also been in the coverage area of the start cell after the time of the last consecutive record which had a different Cell ID. Therefore, the 'before' time could also be constrained. However, in contrast to the 3G EE results, the difference in time between the start time of the CDR and the moment the device was last in the coverage area of the associated Cell ID was less variable than with the 3G EE data, generally being less than one minute. This lower uncertainty may have resulted from the fact that more TAC boundaries were crossed for the 4G data collection than LAC boundaries for the 3G data collection (see Figures 3 (3G) and 5 (4G)), consequently meaning the core network would have been updated with the last used cell more frequently for the 4G EE experiments than the 3G EE experiments. However, it appears the last used cell was updated more frequently than just when TAC boundaries were crossed. This would be worthy of further research to determine the cause of the more frequent cell update. Therefore, it would appear that the terminology: *the device was in the coverage area of the start Cell ID at or before the start time of the CDR and after the time of the last consecutive record which had a different Cell ID* is appropriate for use whilst discussing 4G EE follow-on CDRs.

4.4. Vodafone 4G

For Vodafone 4G (n=3), the device was consistently in the coverage area of the start Cell ID 'at' the start time of the CDR in the majority of instances (mean value of 96%) compared to 'before' the start time of the CDR (mean value of 1% relating to one CDR over the three repeats). On further inspection, it was revealed that no RF survey measurements were taken at the start time of the CDR which was labelled as 'before'. However, the last documented Cell ID to appear within the RF survey results was the same as the start Cell ID documented within that CDR. Therefore, the start cell documented in the CDR may have been the last cell that the device was using prior to losing service, which would result in that cell being documented as the start cell in the CDR. Even though the majority of the CDRs were labelled using the 'at' terminology in this study, these findings cannot be used to recommend sole use of the 'at' terminology as there were occasions where the accuracy could not be ensured. Therefore, it is suggested that the terminology: *the device was in the coverage area of the start Cell ID at or before the start time of the CDR and after the time of the last consecutive record which had a different Cell ID* is appropriate when discussing 4G Vodafone follow-on CDRs.

4.5. Three 4G

For Three 4G (n=3), the device was consistently within the coverage area of the start Cell ID 'at' the start time of each CDR, with the only exceptions being the 'null' results. The results may indicate that the 'at' terminology could be used to discuss Three 4G CDRs, however the low number of CDRs to which this analysis was made must be noted. A total of 24 CDRs were generated over the repeats for the Three/3G study compared to a total of 59, 94 and 102 CDRs for the EE/3G, EE/4G and Vodafone/4G studies respectively. Therefore, it would not be possible to recommend sole use of the 'at' terminology on the basis of these findings and further research must be carried out to determine whether different results would be obtained. Therefore, on the basis of this study, it is suggested that the terminology: *the device was in the coverage area of the start Cell ID at or before the start time of the CDR* is appropriate when discussing successive 4G Three CDRs.

4.6. Implications

The uncertainties associated with follow-on GPRS/mobile data CDRs have meant that they are currently considered inferior to voice/SMS CDRs due to uncertainties regarding the correspondence

between the timestamp and the Cell ID presented within the follow-on records [4]. This issue impacts the potential to use mobile phone records going forward given that IM applications are being increasingly used [10]. Without establishing whether the current terminology for evaluating this form of evidence and presenting it in a court setting is appropriate, there is a possibility that useful intelligence or evidence may be omitted from consideration in an investigation [4]. Therefore, the findings from this study form part of an evidence base to underpin the interpretation of follow-on GPRS/mobile data CDRs and the conclusions that are reached from that data to contribute to establishing empirically supported opinions that can assist the court [1,2,11,12]. A balanced approach must be applied by the expert witness (for example, through following the Case Assessment and Interpretation (CAI) model [13]) which takes into consideration the propositions presented by both the prosecution and the defence [14]. Therefore, these conclusions are ultimately opinion evidence, with the expert forming an opinion on the likelihood of observing the data under competing propositions stated by the prosecution and/or the defence [14].

The results of this study indicate that it is appropriate to use the 'at or before' terminology when discussing follow-on GPRS/mobile data CDRs but suggests that it would not be appropriate to solely use the 'at' terminology. This is an important distinction to make, and this 'before' term must be included in order to account for a possible lag in reporting by the network for the start Cell ID used by the device at the start time of the record [4]. Additionally, the results show that, for 3G/4G EE and 4G Vodafone CDRs (but not 4G Three), the 'before' time can be constrained back to the start time of the last consecutive CDR which had a different Cell ID. This has substantial implications for criminal investigations. If the terminology used was solely 'at or before' without constraining the 'before' time, the 'before' time would extend back to the beginning of the data session. A single data session can remain open for an extended period of time, which could be two or even twelve hours [4]. Therefore, constraining the 'before' time is useful as it narrows the time frame to which the device was within the coverage area of the cell, which can prove beneficial in supporting or refuting an alibi. Understanding these complexities observed with follow-on GPRS/mobile data CDRs, cell site analysts are better able to present the evidence in a way which is aligned with the requirements of the Forensic Science Regulator's Codes of Practice and Conduct FSR-C-118 [14], where the evidence is presented "in a way that conveys its strengths and limitations in a clear, complete, correct and consistent manner" [14, p.6].

5. Conclusion

This study aimed to determine the extent to which the start Cell ID documented within follow-on GPRS/mobile data CDRs were an accurate record of device activity. The results indicate that the accuracy was variable, meaning it is not appropriate to use the terminology: the device was in the coverage area of the start Cell ID at the start time of the CDR, as for each of the networks/technologies tested, there were instances where the start Cell ID documented in the CDR was used by the device prior to (not at) the time the CDR was generated, along with instances where the Cell ID documented within the CDR was not detected by the RF survey device. These inconsistencies mean the terminology 'at or before' should be used when discussing follow-on GPRS/mobile data CDRs.

For the 3G/4G EE and 4G Vodafone CDRs, it was established that the 'before' time could be constrained, therefore the results suggest the terminology: the device was in the coverage area of the start Cell ID at or before the start time of the CDR and after the time of the last consecutive record which had a different Cell ID is appropriate when discussing 3G/4G EE and 4G Vodafone CDRs. A slightly different terminology must be used when discussing 4G Three CDRs, as it cannot be ascertained whether the succession of CDRs are truly follow-on records, and therefore the results suggest that the terminology: the device was in the coverage area of the start Cell ID at or before the start time of the CDR is appropriate when discussing 4G Three CDRs. As discussed in section 2.5, a number of issues were discovered upon analysis of the 3G Vodafone and Three data, meaning this research cannot comment on the correspondence between the timestamp and the Cell ID documented in the follow-on GPRS/mobile data CDRs. This is an acknowledged gap in the findings, and further work to determine the cause of these errors is recommended.

This study demonstrates the value of empirical research that offers data and insights that can support cell site analysts in forming robust, reproducible, and transparent conclusions based on follow-on GPRS/mobile data CDRs. Agreeing the terminology used in court is vital to ensure the interpretation and presentation of evidence is as clear as possible to assist the court with establishing its probative value.

References

[1] McMillan, J., Glisson, W. and Bromby, M. (2013). 'Investigating the Increase in Mobile Phone Evidence in Criminal Activities', *46th Hawaii International Conference on System Sciences,* Maui, Hawaii, USA, 7-10 January.

[2] House of Lords Science and Technology Select Committee. (2019*). Forensic Science and the criminal justice system: a blueprint for change.* 3rd Report of session 2017-2019 HL Paper 333. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/333.pdf (Accessed 13th May 2021)

[3] Forensic Science Regulator. (2020). *Codes of Practice and Conduct, Appendix: Digital Forensics – Cell Site Analysis FSR-C-135,* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/918946/135_FSR-C-135_Cell_Site_Analysis_Issue_2.pdf (Accessed 8 August 2021).

[4] Forensic Analytics Ltd. (2020). *GPRS Billing: Using CDR Data Evidentially, Briefing Paper Version 3.1*, Available at: https://www.forensicanalytics.co.uk/wp-content/uploads/2020/01/0058-BRF_Briefing_Paper-GPRS_Billing_v3.1.pdf (Accessed 21 May 2020).

[5] Hoy, J. (2015). *Forensic Radio Techniques for Cell Site Analysis*. Chichester: John Wiley & Sons, Ltd.

[6] Tart, M., Brodie, I., Gleed, N. and Matthews, J. (2012). Historic cell site analysis – Overview of principles and survey methodologies. *Digital Investigation*, 8(3-4), pp.185-193.

[7] Blank, A. (2011). The Limitations and Admissibility of Using Historical Cellular Site Data to Track the Location of a Cellular Phone. *Richmond Journal of Law & Technology*, 18(1), pp.1-43.

[8] Ofcom. (2019). *Communications Market Report 2019*, Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/117065/communications-market-report-2019.pdf (Accessed 21 May 2020).

[9] Ofcom. (2019). *Connected Nations 2019 UK Report,* Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/186413/Connected-Nations-2019-UK-final.pdf (Accessed 2 July 2020).

[10] Anglano, C. (2014). Forensic Analysis of WhatsApp Messenger on Android Smartphones. *Digital Investigation,* 11(3), pp.201-213.

[11] Mnookin, J. L.,Cole, S.A., Dror, I.E., Fisher, B., Houck, M.M., Inman, K., Kaye, D. H., Koehler, J. J., Langenburg, G., Risinger, D. M., Rudin, N., Siegel, J. and Stoney, D. A. (2011). The need for a research culture in the forensic sciences. *UCLA Law Review*, 58(3), pp.725-780.

[12] Morgan, R. M. (2017). Conceptualising forensic science and forensic reconstruction; Part I: a conceptual model. *Science and Justice,* 57(6), pp.455-459.

[13] Tart, M. (2020). Opinion evidence in cell site analysis. *Science & Justice*, 60(4), pp.363-374.

[14] Forensic Science Regulator. (2021). *Codes of Practice and Conduct, Development of Evaluative Opinions FSR-C-118,* Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960051/FSR-C-118_Interpretation_Appendix_Issue_1__002_.pdf (Accessed 10 October 2021).