Lighting Res. Technol. 2024; **56**: 87–101

Incorrect categorisation of ambient light level at the time of a road traffic collision

S Fotios PhD and **CJ Robbins** PhD
School of Architecture, University of Sheffield, Sheffield, UK

Received 13 April 2021; Revised 24 September 2021; Accepted 21 November 2021

Ambient light level, one of the items recorded by police investigating a road traffic collision (RTC) was previously suggested to be incorrectly recorded in 5%–15% of cases. Significant erroneous categorisation of ambient light level, as suggested by the latter estimate, may lead to incorrect conclusions being drawn about the impact of light on RTCs which is critical where such data inform transport policy decisions. This study investigated the accuracy with which ambient light level was recorded in comparison to that determined using solar altitude at the time, date and location of the RTC. Data were drawn from the STATS19 database of RTCs in the UK for the period 2005 to 2015. Ambient light level was incorrectly reported in 103 021 (5.79%) of the 1 779 903 RTCs in that period. The percentage of errors was greater for RTCs occurring after dark than in daylight, and for RTCs where the scene was not attended by a police officer than those that were attended: ambient light level was incorrectly categorized in 8.72% unattended RTCs in darkness. The highest percentage of errors (57%) occurred within civil twilight; if these are omitted the overall percentage reduces to 2.81%, a similar error rate to that available for the determination of RTC location.

1. Introduction

One purpose of road lighting is to provide visual cues and reveal obstacles so that safe vehicular operation is possible.[1] A measure of the success of road lighting is, therefore, the degree to which the probability and/or severity of road traffic collisions (RTCs) after dark are reduced by road lighting, and this assumed benefit is used in the determination of whether or not to light a road.[2,3] The mitigating effect of road lighting is therefore one basis for the design criteria recommended in guidance documents and standards.

One approach for establishing the potential for road lighting to be of benefit is to investigate the relationship between ambient light level and RTCs. Many studies have used ambient light level as an independent variable or to categorise cases when investigating RTCs involving a range of roads users, vehicle types and situational contexts [e.g.,[4-9]].

A source for such analyses are national databases of RTCs as might be compiled from police reports. In the UK, this database is compiled from STATS19, the police report form for RTCs.[10] The variables recorded in STATS19 include general information such as the time, date and location (longitude and latitude) of the RTC, casualty severity, location information such as whether the RTC occurred at a junction and the speed limit of a given road, and environmental information such as ambient light level and weather.

According to the apparent movement of the sun across the sky, the diurnal cycle of ambient light tends to flow from daylight, through twilight (dusk) to dark, and through a second twilight stage (dawn) back to daylight. Daylight is defined by a solar altitude of greater than 0° - i.e., the sun

Address for correspondence: S Fotios, School of Architecture, University of Sheffield, Arts Tower, Western Bank, Sheffield S10 2TN, UK. E-mail: steve.fotios@sheffield.ac.uk

10.1177/14771535211069028

is above the horizon.[11] For solar altitudes of less than 0° daylight persists to some degree due to the reflection and scattering of sunlight towards the horizon of a terrestrial observer.[11] The twilight periods are thus not fully daylit nor dark but a gradual transition between the two as solar altitude decreases. There are three defined stages of twilight; civil twilight, nautical twilight and astronomical twilight. In civil twilight, where solar altitude is between 0° and −6°, in clear weather there is sufficient daylight illuminance to enable outdoor civil activity to continue unhindered without resorting to the use of electric road lighting.[11] Nautical twilight is the period where the solar altitude is between −6° and −12°. During this period, sailors are able to take reliable readings via well-known stars because the horizon is still visible, and outlines of terrestrial objects may still be visible but detailed outdoor activities are restricted without artificial light.[12] Finally, astronomical twilight is the period where the solar altitude is between −12° and −18°. In this period, the sky illumination is so faint that observers would regard the sky as fully dark, and the horizon is not discernible.[13]

In some RTC databases ambient light level is categorised as either daylight or dark. In others, the civil twilight period is identified separately, and either included with one or other of daylight or dark, considered separately, or omitted to provide a clear distinction between daylight and dark. Some databases record whether the road was lit or unlit.

It is common for researchers to use in their investigation the ambient light levels as categorised by the police rather than to independently establish ambient light level according to solar altitude for the time, date and location of the RTC, even when these data are available in the database. There are only a few examples in previous research where ambient light levels were independently established [e.g.,[14-17]]. The precision of the categorisation of ambient light level, and hence conclusions drawn from the investigation, therefore depends on the accuracy with which ambient light level was recorded. Police RTC reports are the main source of data for road safety research, despite that such use may not have been their intended purpose, and despite that they are likely to contain missing or inaccurate information.[18]

Shinar *et al.*[19] investigated the validity of police-reported RTC data by comparing a sample of reports against the data and conclusions of multi-disciplinary accident investigation teams who investigated the same RTCs. The expert investigators were assumed to provide the correct RTC data because the team comprised of four or more people (and these being specialists in vehicular, environmental and human factors) rather than, typically, a single police officer. The experts were also able to allocate more time to study each RTC than were the police. The data were 124 RTCs which had occurred in Monroe County, Indiana USA, between 1971 and 1975. Variables with the least reliable reporting included accident severity and speed limit, each misidentified by the police in about 30% of cases. Driver age was misclassified in about 12% of the cases. The variables most reliably reported were accident location, date, and the numbers of drivers, passengers and vehicles.

To give consideration to the accuracy of police-reported data such as RTC records requires a secondary source of data. One such source is hospital records, and these enable a comparison of injury severity following an RTC. Brubacher *et al.*[20] found poor agreement between police-reported RTC injury severity and hospital admission records, the focus being serious-injury RTCs defined as those requiring a night in hospital. Only about one-third (36%) of drivers who required hospital admission after the RTC had a police report which recorded major injury. In other words, for the remaining two-thirds of hospital admissions for a serious injury, the police record did not indicate a serious injury. In the reverse, only about half (56%) of drivers for whom the police indicated major injury actually required admission to hospital. Similarly, Tsui *et al.*[21] concluded that police tend to overestimate injury severity-casualties classified as seriously injured by the police but who were not really seriously injured.

A comparison of injury severity between police and hospital records may be unfair on the police. Injury severity is an expert judgement. The police officer at the scene of an RTC needs to attend to many issues and may make mistakes when recording injury severity. Furthermore, where hospital records define injury severity by length of stay, the reporting police officer will not know, a priori, how long the injured will remain in the hospital.[22]

Consider the recording of ambient light level. Shinar *et al.*[19] found that this was incorrectly categorised in about 5% of their cases. The cause of error was suggested to be "*probably due to misunderstanding of the coding procedures*" but that proposal was not substantiated. However, this estimate was drawn from only a small sample of RTCs (n = 124) and they did not state the ambient light level recorded by the police nor describe how ambient light level at the time of the RTC was established by their expert investigators. A more recent study[15] found that RTCs occurring after dark were incorrectly categorised as being in daylight in 15% of 4677 cases. The primary focus of that work was the influence of ambient light level on motorcycle and four-wheel motor vehicle RTCs as recorded in STATS19 for the period 2005 to 2015. For Shinar *et al.*[19] it is not known whether their incorrectly categorised cases were in darkness or daylight. There is quite a difference between these estimates: an error in 15% of cases gives more cause for concern than an error in 5% of cases.

It may not always be straightforward to categorise the ambient light level as being daylight, twilight or dark. When reporting ambient light level in STATS19, the available categories are daylight or darkness, with darkness having four options for describing the presence of road lighting (lit, unlit, no lighting and unknown). The instructions state that darkness is the period which lies between half an hour after sunset to half an hour before sunrise: all other times are classified as daylight.[23] The inclusion of a 30-minute period after sunset and before sunrise means that civil twilight should be included within daylight. This approach could lead to accurate reporting if the reporting police officer knew the precise time of sunset (or sunrise) at that location and at that time of year, and also if they know the exact time at which the RTC occurred. On the other hand, a police officer may instead use judgement informed by observation: unlike injury severity, light level is not an expert judgement in most cases. Such judgement may be accurate in periods of definite daylight and darkness, but is more difficult at twilight. If the categorisation of ambient light level is made by on-site observation this may be more error-prone during civil twilight than during periods further away from the daylight and darkness thresholds.

If there is a tendency for ambient light level to be incorrectly categorised in police reports it would also be useful to consider when and why. Brubacher *et al.*[20] suggested that non-attendance of a police officer at the RTC is a reason for the under-reporting of major injuries, although they did not substantiate that proposal. As with any variable, ambient light level may be misreported due to human error (the wrong box was ticked) or because the available response categories did not offer the required option.[24]

In summary, many studies investigating the effect of ambient light level on the frequency and/or severity of RTCs use the ambient light level as reported by the police rather than that determined independently according to solar altitude. Errors in the recorded ambient light category can be suspected because errors are found in other recorded variables. Incorrect categorisation of ambient light level may lead to incorrect conclusions being drawn about the effectiveness of lighting: this is critical where such data inform decisions. For example, the decision to install road lighting may be informed by a cost-benefit analysis[2,3] where the benefit is the reduction in RTCs associated with installing (or improving) lighting.

The current study investigates the accuracy of the ambient light level recorded in STATS19 by comparing this with the light level determined using the time, date and location of the RTC. These data are first used to establish the

prevalence of error, thus to arbitrate between previous findings that ambient light level was incorrectly categorised in 5%,[19] or 15%,[15] of cases. To extend previous work[15] the analysis includes RTC data across the whole day and the whole year to provide a better estimate of the prevalence for incorrectly reporting ambient light level. The analysis includes comparison of the tendency to make this error in attended and absent RTC scenes, and at different periods of the day.

2. Method

Data were drawn from the STATS19 database of UK RTCs recorded by the police between the years 2005 and 2015. These data are openly accessible via the UK Government Web site.[10] This resulted in 1 780 515 individual RTCs. For each RTC, solar altitude was calculated using the National Oceanic and Atmospheric Administration (NOAA) method.[25] This method calculates the precise solar altitude for each RTC using the date and time of the RTC, and the longitude and latitude at the location of the RTC, variables which are recorded in STATS19.

The data were filtered to categorise RTCs by ambient light according to solar altitude. RTCs that occurred when the solar altitude was less than −6° were categorised as occurring in dark and those when the solar altitude was greater than 0° were categorised as occurring in daylight. Of the 1 780 515 RTCs, 1 682 292 (94.5%) occurred in dark or daylight, the remaining 98 223 RTCs (5.5%) were those occurring in civil twilight.

Police-reported categorisation of ambient light level at the time of the RTC is that given in the STATS19 variable 'light condition'. The five categories are daylight (1 304 387 RTCs, 73.26%), darkness-lights lit (349 697 RTCs, 19.64%), darkness-lights unlit (8182 RTCs, 0.46%), darkness-no lighting (98 937 RTCs, 1.08%) and darkness-lighting unknown (19 312 RTCs, 1.08%). There were no cases where the ambient light level was missing or out of range. For the current analysis, four categories were collated into one labelled darkness (darkness-lights lit, darkness-lights unlit, darkness-no light and darkness-lighting unknown), this representing 476 128 (26.74%) cases.

STATS19 offers three categories to record attendance of a police officer at the scene of an RTC: a police officer attended the scene of an RTC, a police officer did not attend the scene of RTC but filled in the corresponding STATS19 form, and a police officer did not attend scene of RTC and the STATS19 form was completed by the driver or other member of the public involved in the RTC. Of the 1 780 515 RTCs in the 2005 to 2015 period, 1 438 782 (80.81%) were attended by a police officer who completed the STATS19 record. In the current analysis, these are labelled as *attended* RTCs. In 341 121 cases (19.16%) the RTC was not attended by the police but a police officer filled in the corresponding STATS19 form; these are here labelled as *absent* RTCs.

Of the remaining cases, 335 (0.02%) were self-reported, where the RTC was not attended by the police and the STATS19 form was completed by a member of the public involved in the RTC. In a further 277 cases (0.01%) RTC attendance was recorded as unknown. These two categories were not included in the current analysis.

Comparisons of errors for attended and absent RTCs, and also in reported ambient light level in daylight and darkness, were investigated using Odds Ratios (OR).[26,27] The OR and associated 95% confidence intervals (95%CI) were established using equations (1) and (2) following Johansson *et al.*[28] To determine the significance of departure from 1.0, the *p*-value for each OR was determined using a Chi-square test. Table 1 defines the data used when calculating ORs for each comparison. For the dark errors versus daylight errors comparison, an OR significantly greater than 1.0 indicates greater risk of incorrectly reporting ambient light level for RTCs occurring in darkness than daylight.

$$\mathrm{OR} = A/B/C/D \tag{1}$$

Table 1 Description of terms for calculating OR (equation (1)) and 95%CI (equation (2)) to compare incorrect categorisation of ambient light level according to (1) police attendance or Absence at the RTC scene and (2) whether the RTC occurred in daylight or darkness as established using solar altitude

Effect measured by the odds ratio	Terms of equation			
	A	B	C	D
1 Police attendance or absence at the RTC scene	RTCs where police were absent and light condition is incorrect	RTCs where police attended and light condition is incorrect	RTCs where police were absent and light condition is correct	RTCs where police attended and light condition is correct
2 RTCs occurring in daylight or darkness	RTCs occurring in darkness and light condition is incorrect	RTCs occurring in daylight and light condition is incorrect	RTCs occurring in darkness and light condition is correct	RTCs occurring in daylight and light condition is correct

$$95\%\mathrm{CI} = \exp\left(\ln(OR) \pm 1.96 \times \sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}}\right) \quad (2)$$

3. Results

Table 2 shows the total number of RTCs grouped by ambient light level as defined by solar altitude at the time of the RTC. These numbers are reported for RTCs occurring in five periods as defined by solar altitude: daylight (≥0°); darkness (≤-6°); daylight and darkness combined; civil twilight; and finally for daylight, darkness and civil twilight combined (i.e. the whole 24-hour period). The data were further categorised by police attendance, and for each category are shown the numbers of cases where ambient light level was incorrectly reported. RTCs occurring in civil twilight were considered as correct when categorised as daylight and incorrect when categorised as darkness, following the instructions provided in STATS19.

Ambient light level was incorrectly categorised in 5.79% of all cases. The greatest proportion of errors (56.96%) is found in civil twilight, where cases that should have been recorded as daylight were instead recorded as darkness. The percentage of incorrect categorisations is slightly higher for attended than absent cases. Outside of civil twilight, the greatest percentage (8.72%) of incorrectly classified ambient light levels occurs with unattended RTCs in darkness.

Table 3 shows the ORs comparing ambient light level errors between attended and absent cases for the same five ambient light periods as shown in Table 2. For daylight and darkness there is a significantly greater risk of incorrectly reporting the ambient light level for absent cases than for attended cases. For civil twilight, however, the reverse situation is found: there is a greater risk of incorrect reporting of ambient light level for attended RTCs than for absent.

Figure 1 shows the percentages of incorrect ambient light level categories for each half-hour period of the day. These data are for the four-week periods centred on the 22nd day of December, March and June, marking the solstices and equinox. The autumn equinox in September, not shown in Figure 1, displays a very similar trend to that of the spring equinox in March. The greater percentages are found at around the times of civil twilight, with these percentages being greater for the evening than the morning civil twilight periods.

While Figure 1 and Table 2 show that the greater percentages of ambient light category errors are found in civil twilight, there are still errors at other times of the day. These data also show that the percentages of incorrect categorisations are

higher for RTCs occurring in darkness than in daylight. Analysis using an OR of the cases occurring in daylight and dark (Table 4) suggests that the police are three times more likely to incorrectly categorise ambient light level for RTCs occurring in darkness than for RTCs in daylight.

Darkness, the time between evening and morning civil twilight, can be further sub-divided into three phases: nautical twilight, astronomical twilight and night. Table 5 shows the numbers and percentages of RTCs occurring in each of these phases. The percentage of errors is greatest in nautical twilight and then progressively decreases for astronomical twilight and night, yet even at night this is incorrectly recorded as daylight in nearly 6% of cases. For all phases of darkness, the percentages are greater for absent than attended cases.

Table 2 Numbers of RTCs grouped by ambient light level as defined by solar altitude, and the numbers (and percentages) of these where STATS19 incorrectly reports light condition, as a function of police attendance at the RTC (absent or attended)

Ambient light as defined by solar altitude	Numbers of RTCs according to police attendance		Cases where reported ambient light category is incorrect	
			Number	%
Daylight (0° and above)	Absent	259 377	4953	1.91
	Attended	1 001 030	17 998	1.80
	Both	1 260 407	22 951	1.82
Darkness (−6° and below)	Absent	63 450	5530	8.72
	Attended	358 279	18 854	5.26
	Both	421 729	24 384	5.78
Civil Twilight	Absent	18 294	9516	52.02
	Attended	79 473	46 170	58.10
	Both	97 767	55 686	56.96
Darkness and daylight (civil twilight excluded)	Absent	322 827	10 483	3.25
	Attended	1 359 309	36 852	2.71
	Both	1 682 136	47 335	2.81
Darkness, daylight and civil twilight	Absent	341 121	19 999	5.86
	Attended	1 438 782	83 022	5.77
	Both	1 779 903	103 021	5.79

Table 3 ORs, 95% CIs and associated *p*-values comparing accuracy of reported ambient light level according to police attendance and absence at RTCs

Ambient light as defined by solar altitude	Accuracy of police-reported ambient light level	Number of RTCs		OR	95% CI	Sig
		Absent	Attended			
Daylight (0° or above)	Incorrect	4953	17 998	1.06	1.03-1.10	$p < 0.001$
	Correct	254 424	983 032			
Darkness (−6° or below)	Incorrect	5530	18 854	1.72	1.67–1.78	$p < 0.001$
	Correct	57 920	339 425			
Darkness and daylight (civil twilight excluded)	Incorrect	10 483	36 852	1.20	1.18–1.23	$p < 0.001$
	Correct	312 344	132 2457			
Civil Twilight	Incorrect	9516	46 170	0.78	0.76–0.81	$p < 0.001$
	Correct	8778	33 303			
Darkness, daylight and civil twilight	Incorrect	19 999	83 022	1.02	1.00–1.03	$p < 0.05$
	Correct	321 122	135 5760			

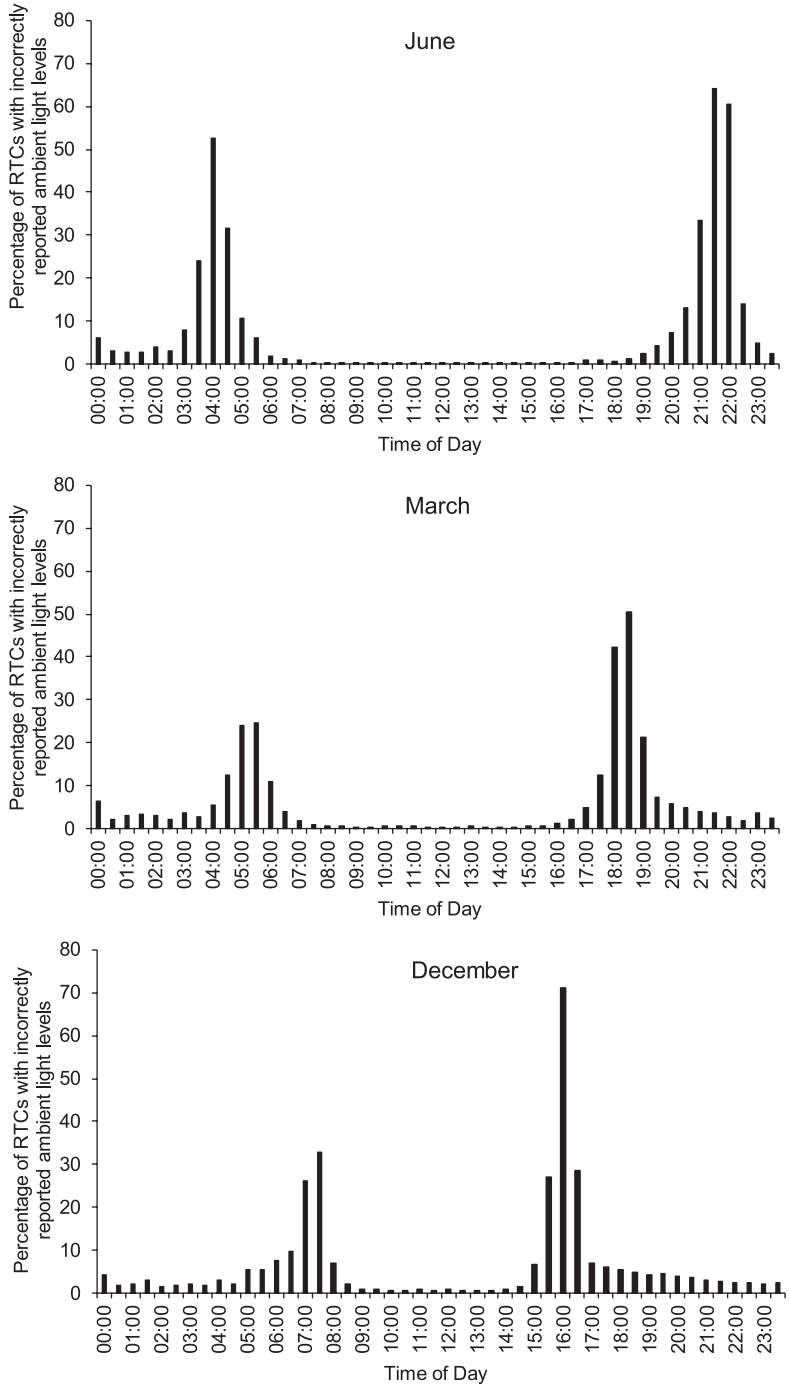

Figure 1 Percentage of RTCs in each half-hour interval for which ambient light level was incorrectly recorded in STATS19. These data are for the four-week periods centred on the 22nd day of the stated month

Table 4 ORs, 95% CIs and associated *p*-values for ambient light level as determined using solar altitude, compared to categories reported by police reports

Accuracy of police-reported ambient light level	Number of RTCs		OR	95% CI	Sig
	Darkness	Daylight			
Incorrect	24 384	22 951	3.31	3.25–3.37	$p < 0.001$
Correct	397 435	1 237 456			

Table 5 The total number of RTCs that occurred in the three stages of darkness (nautical twilight, astronomical twilight and night) and the number (and percentage) of these RTCs that were recorded as daylight

Phase of darkness	Accuracy of police-reported ambient light level	Numbers (%) of RTCs		OR	95% CI	Sig
		Absent	Attended			
Nautical Twilight	Incorrect	2658 (15.6%)	8408 (10.3%)	1.51	1.44–1.59	$p < 0.0001$
	Correct	14 415 (84.4%)	73 287 (89.7%)			
	Total	17 073	81 695			
Astronomical Twilight	Incorrect	1117 (8.3%)	3546 (5.0%)	1.66	1.55–1.78	$p < 0.001$
	Correct	12 284 (91.7%)	67 202 (95.0%)			
	Total	13 401	70 748			
Night	Incorrect	1755 (5.6%)	6900 (3.4%)	1.62	1.54–1.71	$p < 0.001$
	Correct	31 221 (97.7%)	198 936 (96.6%)			
	Total	32 279	205 836			

4. Discussion

4.1 Percentage error

Of the 1 779 903 RTCs recorded in STATS19 in the period 2005 to 2015 for which absent or attended status is known, ambient light level at the time of the RTC was incorrectly categorised in 103 021 cases (5.79%). We did not encounter any reason why this error would be specific to STATS19 and would hence differ significantly for ambient light levels reported by the police for other national records.

The greatest proportion of errors (56.96%) is found for RTCs occurring in civil twilight, cases which should have been recorded as daylight according to STATS19 guidance but which were erroneously recorded as darkness. Outside of civil twilight, the tendency to incorrectly categorise ambient light was greater for RTCs that occurred in darkness (5.78%) than in daylight (1.82%).

The overall percentage of errors (5.79%) is similar to that (5%) reported by Shinar *et al.*[19] The percentage of errors in darkness (5.78%) is, however, smaller than that (15%) reported previously[15] despite use of the same data set (STATS19) for the same period of time (2005–2015). One reason for this is that the previous analysis included only those data within time windows (ranging in duration from 1 minute to near 60 minutes) chosen to ensure a definite daylight-to-darkness (or vice versa) difference in the weeks before and after clock change: it was therefore a smaller sample, located near, but not within, civil twilight, and omitted those cases within astronomical twilight and night which suggest a lower degree of error (Table 5). The incorrect cases in nautical twilight, for absent and attended RTC scenes combined, represent 11.2% of the total nautical twilight cases, which more closely matches the previous finding.

For RTCs in daylight or darkness, incorrect categorisation of ambient light level is more likely in absent (3.25%) than attended (2.71%) cases (Tables 2 and 3). While this trend is retained for the three phases of darkness (Table 5), the reverse situation is found in civil twilight, with a greater tendency of incorrectly reporting ambient light level for attended RTCs than for absent.

4.2 Alternative analysis

The results shown in Tables 2–5 considered RTCs occurring across the whole year, which is the data set used in some investigations.[28,29] Other studies, however, have considered only the one,[14,15] two,[17,30] or 5 weeks[31,32] immediately before and after the biannual clock changes. In these studies, a time window is chosen that is in darkness before clock change and daylight after clock change (or vice versa) and RTCs in these before and after periods are compared against changes in RTCs for control periods. Doing so aims to isolate the effect of ambient light from other factors such as seasonal change in weather and driver demography.

The current analysis was thus repeated but including only those RTCs occurring 1 week immediately before and 1 week immediately after clock changes in Spring and Autumn for the 11-year period 2005 to 2015 (see Supplemental file 1). This revealed that the police are more than twice as likely to incorrectly categorise ambient light level for RTCs occurring in darkness than RTCs in daylight. It also showed that for RTCs occurring in darkness (but not in daylight), there was a significantly greater risk of incorrect reporting of ambient light level for absent than attended RTCs. For RTCs in daylight, the whole year method suggests a significant effect of police attendance but the clock change method does not: this difference in conclusions may be due to the larger sample of RTCs in the whole year analysis. The clock change data have smaller sample sizes than do the whole year data which leads to larger confidence intervals: for daylight cases, the ORs are closer to 1.0 than found for darkness or darkness and daylight combined, and hence for the clock change daylight data the 95%CI now crosses OR = 1.0. The data associated with a clock-change analysis reveals similar percentages of incorrectly reported cases as does analysis of the whole-year data (Table 6).

4.3 Limitations

For each RTC, STATS19 records the ambient light level, the date, time and location. In the current study, ambient light level was also independently determined according to solar altitude, as established from the date, time and location reported in STATS19. Where the recorded and determined light categories disagree, the current analysis assumes the determined categories are correct. This assumption itself assumes that the date, time and location of the RTC were accurately reported in STATS19, despite assumed error in the ambient light category. This is consistent with Shinar *et al.*[19] who found in their comparison of police and expert reports that the most reliably reported police data were those concerned with accident location, time, date and number of drivers.

Others have raised the potential for errors in the recording of time and location. A study of RTCs in Croatia found that the location was incorrect in 33.5% of the 8550 cases.[33] The main error was that police officers incorrectly recorded the street name: assuming that the metropolitan area (city, town or village) was correctly recorded then this would be unlikely to influence significantly the determined ambient light level.

Regarding time, RTC records may round this (in particular where determined from witness statements) rather than time being reported at the minute-level of accuracy. Imprialou and Quddus[18] found that crash times recorded in STATS19 tend to be rounded to the nearest 5-minute point. They analysed data for 2014 only and found the minute indication ended with a 0 or 5 in 66.4% of cases, with further bias to the 0- and 30-minute points. Checking this using STATS19 data for the period

Table 6 The overall percentage of RTCs were the ambient light level was incorrectly categorised, according to light condition calculated by solar altitude, for daylight RTCs, civil twilight and darkness RTCs, for both the whole year analysis and the clock change weeks

Parameter	Whole-year analysis, %	Clock-change analysis, %
Errors in daylight (i.e., recorded as dark)	1.82	2.44
Errors in civil twilight (i.e., recorded as dark)	43.2	43.0
Errors in darkness (i.e., recorded as daylight)	5.78	6.08

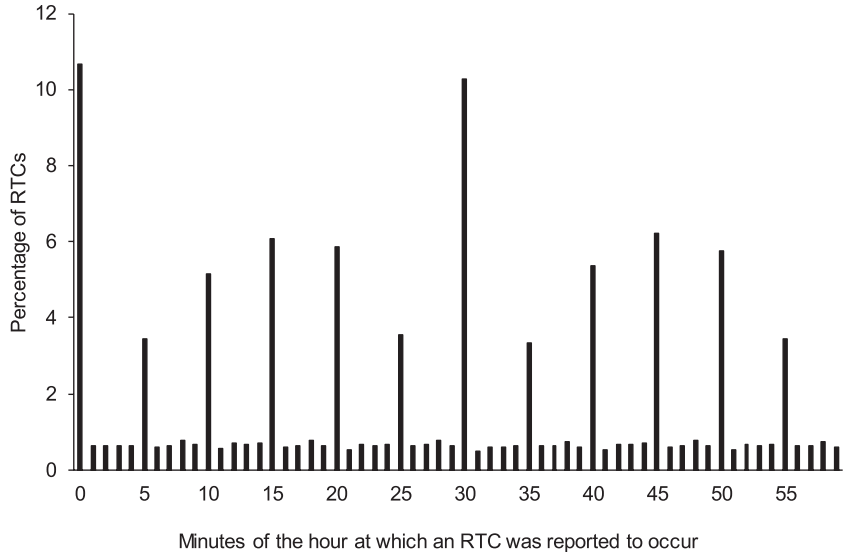

Figure 2 Minute of the hour at which RTCs were reported to occur as recorded in STATS19 for the period 2005 to 2015

2005 to 2015 reveals a near-identical pattern (Figure 2). For most of the day, this rounding would not be of relevance.

Where it becomes significant is at the thresholds of civil twilight (0° and −6°): for the solstices, solar altitude changes by about one half of a degree in 5 minutes (estimated for London, latitude 51.5°) which would lead to some RTCs occurring in daylight being categorised as darkness (and vice versa). What is not known, however, is whether there is a tendency to round up or to round down to the nearest preferred five-minute interval: if this is assumed to be equally distributed then time rounding would not influence the numbers of RTC cases in each ambient light category.

Previous studies have determined ambient light level according to the solar altitude for a specific RTC using the date, time and location of the RTC rather than using the reported categories.[15-17] Doing so does not, however, ensure perfect categorisation of ambient light level. Following previous work,[28] one study[17] used a binary day/dark categorisation, the distinction being a solar altitude of 0°, and thus included civil twilight in the dark period. Subsequent analysis demonstrated that this underestimates the effect of ambient light on RTC risk.[14,34] Twisk and Reurings[16] established solar altitude for the time of an RTC but they did so only for one location, the centre of their study region, The Netherlands. For a northerly (Leeuwarden) and Southerly

(Maastricht) city in the Netherlands, differences in the times of evening civil twilight (for example) range from about 2 to 19 minutes, depending on the time of year. While this is sufficient to incorrectly categorise the ambient light level for a specific RTC, it may be self-counterbalanced for analyses of data across the year and across the nation. Furthermore, categorisation of ambient light level by solar altitude alone may be misleading; for example, obstructions such as high-rise buildings might reduce ambient light levels to darkness that might otherwise be defined by solar altitude as daylight. In further work this might be investigated at a macro level by comparing data from urban and suburban areas, or at a micro level by checking the environment at each RTC location.

There may also be an influence of weather conditions. For example, mist and fog reduce visibility[35,36] and cloud cover will reduce the amount of daylight[37] which may lead to an evaluation that the ambient light level category is darkness when solar altitude would otherwise indicate it to be daylight.

An additional analysis was conducted to investigate whether the police were more likely to incorrectly categorise RTCs in adverse weather compared to clear weather (Supplemental file 2). STATS19 includes five categories of adverse weather: raining, no high winds; snowing, no high winds; raining, high winds; snowing, high winds; and fog or mist. Clear weather includes two categories: fine, no high winds; and fine, high winds. Weather conditions labelled as other or unknown were omitted from this analysis. For RTCs occurring in darkness, an incorrect categorisation is to label this as daylight: this error is less likely to be made in adverse weather. For RTCs occurring in daylight, an incorrect categorisation is to label this as darkness: this error is three times more likely to be made in adverse weather. RTCs were more likely to be incorrectly categorised when the police officer was absent than when they attended the scene with this difference being significant ($p < 0.001$) in clear weather in daylight and darkness and adverse weather in darkness, but was not suggested to be significant ($p = 0.68$) for RTCs occurring in adverse weather in daylight.

A police officer is more likely to attend a fatal RTC than an RTC with only slight injury.[38] Therefore, incorrect categorisation of ambient light level could change as a function of accident severity. However, such an analysis was not possible with the current data as there were insufficient sample sizes for some cases (e.g., unattended fatal RTCs).

Finally, we raise the possibility that the reporting form, STATS19, might itself prompt errors. The available categories are daylight, darkness-lights lit, darkness-lights unlit, darkness-no lighting and darkness-lighting unknown. There are no options for road lighting to be lit or unlit during daylight which may lead to the attending police officer to assume that if the road lighting is lit it must be darkness. While road lighting should not be lit in daylight, faults or incorrect settings in the control gear may mean that it is, leading to an incorrect conclusion that the ambient light level is darkness.

4.4 What might be done

We consider here whether the error is of significance and what to do about it.

As to whether the error in categorisation of ambient light level is of significance, consider it alongside the percentages of errors found in other factors recorded about RTCs. Errors have been found in a range of RTC data. Injury severity is underestimated by the police compared with hospital records in 15%[39] to 64%[20] of cases; similarly, reports of major injury are an overestimate of severity in about half of those cases.[20] There are also errors in location information. Loo[40] examined RTCs in Hong Kong between 1993 and 2004 and found that road names and districts were incorrect in 12.7% and 9.7% of cases respectively. Austin[41] found that carriageway type was incorrectly coded or located in 21% of RTCs examined in the Humberside region of the UK. Chung and Chang[42] considered the use of vehicle black box systems to investigate accuracy

of location information for RTCs in Korea and found that 68% of RTCs were within 50 m of the reported location and 88% were included within 150 m of the reported location. Imprialou *et al.*[43] examined 10 520 motorway and trunk road RTCs reported in STATS 19 for the year 2012 and found an error in 7.3%. They suggested using fuzzy logic to improve RTC location reporting, and this reduced the error to 1.1%. Similarly, Deka and Ouddus[44] used artificial neural networks to reduce RTC location errors to 2.9% of cases.

Regarding what to do about it, we propose two courses of action. For researchers, ambient light level should be determined using solar altitude rather than relying on reported categories. Where there is no option but to use reported categories, then those RTCs occurring in civil twilight should be omitted, although this assumes that civil twilight is included as a separate category to daylight and dark, which is not the case for STATS19. Civil twilight is the phase of ambient light where the percentage of errors is greatest, and removing these cases reduces the daily error rate from 5.79% to 2.81% (Table 2) which is a similar error rate to that obtained for location using fuzzy logic or neural networks. These data hence support the decision in some studies to omit RTCs occurring in civil twilight.[14,30]

For the process of recording RTC data at the scene of an RTC we suggest that the police continue to record both the ambient light level and the time of the crash. When recording ambient light level, we suggest that guidance related to the time of sunset and sunrise is omitted, this giving a potentially false sense of accuracy, and instead rely on the observation of the attending police officer. For absent (non-attended) cases, ambient light level should be either estimated from the time of day but with a caveat that the event was not attended, or should not be recorded at all. The time at which an RTC occurred should be established with the greatest accuracy possible, ideally to the nearest minute rather than 5-minute rounding. When police reports are entered into the database, ambient light level according to solar altitude should be determined according to location, time and date; where this computed light level disagrees with the police-observed light level, then this should be flagged to be checked by the reporting police officer.

To reduce the risk of assumed association between lit and unlit road lighting and daylight or darkness, the STATS19 report form should offer sub-categories of categories of daylight similar to those for darkness: lights lit, lights unlit, no lighting and lighting unknown.

5. Conclusion

This study investigated the tendency for ambient light level (daylight or darkness) to be incorrectly reported in police records of RTCs. An incorrect categorisation is one which disagrees with that determined according to solar altitude. Estimates in previous studies suggested an error in 5%–15% of RTCs.

Of the 1 779 903 RTCs reported in STATS19 in the period 2005 to 2015, ambient light level was incorrectly reported in 103 021 cases (5.79%). In that period 421 819 RTCs occurred after dark, of which 24 384 (5.78%) were incorrectly reported as being in daylight; 1 260 407 RTCs occurred in daylight of which 22 951 (1.82%) were incorrectly reported as being in darkness. Errors after dark are three times more likely than errors in daylight.

According to the instructions for STATS19, RTCs occurring in the 30-minute periods before sunrise and after sunset, approximately representing civil twilight, should be recorded as daylight. Over 50% of RTCs occurring in civil twilight were incorrectly recorded as occurring in darkness. This high degree of error supports the decision of some studies to omit RTCs occurring in civil twilight.

One explanation for incorrect categorisation is that the police did not attend the RTC. For RTCs in daylight or darkness (but not in civil twilight) the data suggest a significantly greater risk of incorrectly reporting ambient light level when a police officer did not attend the RTC (absent cases) than for attended cases. The greatest percentage (8.72%) of incorrectly categorised ambient light levels is those that occurred in darkness and which

were not attended by a police officer. However, even for attended RTC scenes, errors persist.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), grant reference EP/S004009/1.

ORCID iDs

S Fotios https://orcid.org/0000-0002-2410-7641
CJ Robbins https://orcid.org/0000-0002-6076-5983

Supplemental material

Supplemental material for this article is available online.

References

1 Commission Internationale de l'Éclairage. CIE-115:2010. *Lighting of Roads for Motor and Pedestrian Traffic*. Vienna: CIE, 2010.
2 Commission Internationale De l'Éclairage. CIE-093:1992. *Road Lighting as an Accident Countermeasure*. Vienna: CIE, 1992.
3 Highways Agency. Design manual for roads and bridges. Volume 8; Section 3. TA 49/07. In *Appraisal of New and Replacement Lighting on the Strategic Motorway and All Purpose Trunk Road Network*. London: Highways Agency. 2007.
4 Abdel-Aty M, Ekram A-A, Huang H, Choi K. A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis and Prevention* 2011; 43: 1730–1737.
5 Boufous S, de Rome L, Senserrick T, Ivers R. Risk factors for severe injury in cyclists involved in traffic crashes in Victoria, Australia. *Accident Analysis and Prevention* 2012; 49: 404–409.
6 Das S, Avelar R, Dixon K, Sun X. Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis and Prevention* 2018; 111: 43–55.
7 Islam S, Brown J. A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. *Accident Analysis and Prevention* 2017; 108: 163–171.
8 Li Y, Fan WD. Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: a case study of North Carolina. *Accident Analysis and Prevention* 2009; 131: 284–296.
9 Missikpode C, Peek-Asa C, Young T, Hamann C. Does crash risk increase when emergency vehicles are driving with lights and sirens? *Accident Analysis and Prevention* 2018; 113: 257–262.
10 Department for Transport. *Road Safety Data*, 2019. Retrieved from, https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.
11 Muneer T. *Solar Radiation and Daylight Models for Energy Efficient Design of Buildings*. Oxford: Architectural Press, 1997.
12 Time and Date. Nautical twilight – nautical dawn & dusk. Retrieved on 03/09/2020 from, https://www.timeanddate.com/astronomy/nautical-twilight.html 2020.
13 Time and Date. Astronomical twilight – astronomical dawn & dusk. Retrieved on 03/09/2020 from, https://www.timeanddate.com/astronomy/astronomical-twilight.html 2020.
14 Raynham P, Unwin J, Khazova M, Tolia S. The role of lighting in road traffic collisions. *Lighting Research and Technology* 2020; 52: 485–494.
15 Robbins CJ, Fotios S. Motorcycle safety after-dark: the factors associated with greater risk of road-traffic collisions. *Accident Analysis and Prevention* 2020; 146: 105731.

16 Twisk DAM, Reurings M. An epidemiological study of the risk of cycling in the dark: the role of visual perception, conspicuity and alcohol use. *Accident Analysis and Prevention* 2013; 60: 134–140.
17 Uttley J, Fotios S. The effect of ambient light condition on road traffic collisions involving pedestrians on pedestrian crossings. *Accident Analysis and Prevention* 2017; 108: 189–200.
18 Imprialou M, Quddus M. Crash data quality for road safety research: current state and future directions. *Accident Analysis and Prevention* 2019; 130: 84–90.
19 Shinar D, Treat JR, McDonald ST. The validity of police reported accident data. *Accident Analysis and Prevention* 1983; 15: 175–191.
20 Brubacher JR, Chan H, Erdelyi S. Injury severity in police collision reports correlates poorly with requirement for hospital admission. *Journal of Transport and Health* 2019; 14: 100606.
21 Tsui KL, So FL, Sze NN, Wong SC, Leung TF. Misclassification of injury severity among road casualties in police reports. *Accident Analysis and Prevention* 2009; 41: 84–89.
22 Couto A, Amorim M, Ferreira S. Reporting road victims: assessing and correcting data issues through distinct injury scales. *Journal of Safety Research* 2016; 57: 39–45.
23 Department for Transport. STATS 20. *Instructions for the Completion of Road Accident Reports* 2011, Retrieved from, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/230596/stats20-2011.pdf.
24 Rolison JJ. Identifying the causes of road traffic collisions: using police officers' expertise to improve the reporting of contributory factors data. *Accident Analysis and Prevention* 2020; 135: 105390.
25 National Oceanic and Atmospheric Administration (NOAA). *Earth System Research Laboratory NOAA Solar Calculator*. Boulder, CO: NOAA, 2005. Retrieved from, www.esrl.noaa.gov/gmd/grad/solcalc/with downloadable_spreadsheet_version_from www.esrl.noaa.gov/gmd/grad/solcalc/NOAA_Solar_Calculations_day.xls.
26 Nussbaum EM. *Categorical and Nonparametric Data Analysis: Choosing the Best Statistical Technique*. New York: Routledge, 2014.
27 Szumilas M. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent* 2010; 19(3): 227–229.
28 Johansson Ö, Wanvik PO, Elvik R. A new method for assessing the risk of accident associated with darkness. *Accident Analysis and Prevention* 2009; 41: 809–815.
29 Fotios S, Uttley J, Fox S. A whole-year approach showing that ambient light level influences walking and cycling. *Lighting Research and Technology* 2019; 51: 55–64.
30 Sullivan JM, Flannagan MJ. The role of ambient light level in fatal crashes: inferences from daylight saving time transitions. *Accident Analysis and Prevention* 2002; 34: 487–498.
31 Sullivan JM, Flannagan M J. Risk of fatal rear-end collisions: is there more to it than attention?: Proceedings of the 2nd International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design, Park City, UT, 21–24 July: 2003.
32 Sullivan JM, Flannagan MJ. Determining the potential safety benefit of improved lighting in three pedestrian crash scenarios. *Accident Analysis and Prevention* 2007; 39: 638–647.
33 Miler M, Todić F, Ševrović M. Extracting accurate location information from a highly inaccurate traffic accident dataset: a methodology based on a string matching technique. *Transportation Research Part C: Emerging Technologies* 2016; 68: 185–193.
34 Fotios S, Robbins CJ, Uttley J. A comparison of approaches for investigating the impact of ambient light on road traffic collisions. *Lighting Research and Technology* 2020; 53: 249–261.
35 Cavallo V, Colomb M, Doré J. Distance perception of vehicle rear lights in fog. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 2001; 43: 442–451.

36 Moore RL, Cooper L. *Fog and Road Traffic (TRRL Report LR 446)*. Crowthorne, UK: Transport and Road Research Laboratory, 1972.
37 Li DHW. A review of daylight illuminance determinations and energy implications. *Applied Energy* 2010; 87: 2109–2118.
38 Tay R, Kattan L, Bai Y. Factors contributing to police attendance at motor vehicle crash scenes. *Journal of the Transportation Research Forum* 2014; 53: 101–116.
39 McDonald G, Davie G, Langley J. Validity of police-reported information on injury severity for those hospitalized from motor vehicle traffic crashes. *Traffic Injury Prevention* 2009; 10: 184–190.
40 Loo BPY. Validating crash locations for quantitative spatial analysis: a GIS-based approach. *Accident Analysis and Prevention* 2006; 38: 879–886.
41 Austin K. The identification of mistakes in road accident records: part 1, locational variables. *Accident Analysis and Prevention* 1995; 27: 261–276.
42 Chung Y, Chang I. How accurate is accident data in road safety research? An application of vehicle black box data regarding pedestrian-to-taxi accidents in Korea. *Accident Analysis and Prevention* 2015; 84: 1–8.
43 Imprialou M-IM, Quddus M, Pitfield DE. High accuracy crash mapping using fuzzy logic. *Transportation Research Part C: Emerging Technologies* 2014; 42: 107–120.
44 Deka L, Quddus M. Network-level accident-mapping: distance based pattern matching using artificial neural network. *Accident Analysis and Prevention* 2014; 65: 105–113.